

Design and analysis of experiments with high throughput biological assay data

David M. Rocke*

Division of Biostatistics, University of California, Davis, CA 95616, USA

Abstract

The design and analysis of experiments using gene expression microarrays is a topic of considerable current research, and work is beginning to appear on the analysis of proteomics and metabolomics data by mass spectrometry and NMR spectroscopy. The literature in this area is evolving rapidly, and commercial software for analysis of array or proteomics data is rarely up to date, and is essentially nonexistent for metabolomics data. In this paper, I review some of the issues that should concern any biologists planning to use such high-throughput biological assay data in an experimental investigation. Technical details are kept to a minimum, and may be found in the referenced literature, as well as in the many excellent papers which space limitations prevent my describing. There are usually a number of viable options for design and analysis of such experiments, but unfortunately, there are even more non-viable ones that have been used even in the published literature. This is an area in which up-to-date knowledge of the literature is indispensable for efficient and effective design and analysis of these experiments.

In general, we concentrate on relatively simple analyses, often focusing on identifying differentially expressed genes and the comparable issues in mass spectrometry and NMR spectroscopy (consistent differences in peak heights or areas for example). Complex multivariate and pattern recognition methods also need much attention, but the issues we describe in this paper must be dealt with first.

The literature on analysis of proteomics and metabolomics data is as yet sparse, so the main focus of this paper will be on methods devised for analysis of gene expression data that generalize to proteomics and metabolomics, with some specific comments near the end on analysis of metabolomics data by mass spectrometry and NMR spectroscopy.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: Gene expression; Mass spectrometry; Metabolomics; Microarray; NMR spectroscopy

1. Principles of experimental design

There was at one time the widespread notion that the large number of measurements obtained in a gene expression array would somehow make up for small sample sizes. In fact, experiments with many responses in most cases need to be as large (and well designed) as if there were only one response. An experiment that is inadequate for understanding a single response is most likely inadequate for understanding 20,000 responses. The multiple measurements obtained from a gene expression, proteomics, or metabolomics experiment do not make the experiment more effective in such a way that less replication is needed.

1.1. Biological replicates are more important than technical replicates

Most of the variation in any biological experiment or clinical trial is biological. Usually, the largest differences are between organisms, though there are cases in which such factors as diurnal variation can make the within-organism variability larger than the between-organism variability. In almost all cases, the analytical variability, sometimes called technical variability is usually small by comparison with the biological variability. Given the same biological sample, the standard deviation of replicate data from genes that are highly expressed is usually no more than 5–10% (for example, see [1]). The percentage difference for levels of mRNA or other biologically active compounds that are present at low levels can be much smaller, but this is actually a consequence

* Tel.: +1 530 752 0510; fax: +1 530 752 8894.

E-mail address: dmrocke@ucdavis.edu.

of using the wrong measure of change rather than increased variability at low levels. This is explained in more detail below.

This has several consequences for the design of experiments and studies. For data with relatively small amounts of variability, such as cell lines, only a few replicates may be needed, but even for inbred strains of mice or rats, at least a few replicates per condition are likely to be required. In many cases, use of several experimental factors can reduce the need for replication within each treatment combination, but does not decrease the total number of animals required (usually in the tens at least). For human studies and trials, hundreds of patients and controls may be needed before accurate answers may emerge, just as is the case in drug and surgical clinical trials.

1.2. *To pool or not to pool?*

Pooling samples can reduce the cost. Given a measure in which the variation between subjects is σ_1^2 and in which the variation due to the analysis is σ_2^2 , a study with n arrays or other analyses with k subjects pooled on each array has variability of the mean of all the measurements of $(\sigma_1^2/kn) + (\sigma_2^2/k)$. Suppose that sampling a subject costs c_1 and running an array (or mass spec, NMR) costs c_2 , so that the total cost of this experiment is $nkc_1 + kc_2$. Using, for example, Excel's solver routine, we can solve for the design that has the minimum variance subject to keeping the total cost below some threshold C . As an example, if the cost per sample is US\$ 100 and the cost per array is US\$ 500 and if the standard between samples is 0.3 and the analytical standard deviation is 0.1, then the best design by this criterion that keeps the cost below \$10,000 is eight arrays with seven samples pooled on each array for a total cost of US\$ 9600. To analyze the 56 samples with 56 arrays would cost a much greater sum of US\$ 33,600. In general, the largest gains from pooling are when the cost per sample is low and the cost per analysis is high.

There are several factors that are left out in this analysis, however. One is that a single bad or unusual sample will spoil an entire array, which can lead to loss of power. In the case of human samples, it also does not lead to an ability to detect differences after the fact between patients that lead to differences in gene expression or metabolic profiles. It is not clear how these disadvantages balance against reduced cost, but it may be useful in some cases to consider pooling as a design factor [2].

1.3. *Use multiple experimental factors when appropriate*

It is often the case that the investigation of multiple factors at the same time may be more efficient and effective than a series of experiments aimed at each factor alone. This may seem counter-intuitive, but it is a well known to students of the statistical design of experiments [3]. Consider a study of

the effect of administration of a toxic substance on gene expression or on metabolite profiles in the mouse lung. The primary factor is treatment or control, but one may also consider mice of, say, four developmental ages and both sexes. This gives a total of 16 treatment groups. With two mice/arrays per group, the total experimental effort is 32 runs (arrays, mass spec runs, etc.).

First consider the issue of efficiency. One can compare treatment and control with the precision of 16 versus 16, and one can also compare male vs. female at the same time, as well as the effect of developmental age on the response. A series of experiments on male/female, age comparison, and treatment/control would take far more experimental runs.

Even more important, this design allows more complex questions to be asked. Do males and females respond differently to the administration of the toxic substance? Does the response vary by age? Is the age variation in response to the toxic substance itself dependent on sex? These interactions may be the most important findings of the experiment.

Given a single response, a standard statistical model is:

$$\begin{aligned} \text{Response} = & \text{mean} + \text{dose} + \text{sex} + \text{age} + \text{dose} \times \text{sex} \\ & + \text{dose} \times \text{age} + \text{sex} \times \text{age} \\ & + \text{dose} \times \text{sex} \times \text{age} + \text{error} \end{aligned}$$

All of these effects can be estimated and their statistical significance explored by a standard analysis of variance (ANOVA) or linear model statistical program such as those found in SAS, S-Plus, R, or Stata. In many cases, consultation with a biostatistician on issues of design and analysis may be helpful. Software (LMGene) for fitting such models for each gene in an array is available in the form of R code on the website <http://www.cipic.ucdavis.edu/~dmrocke>.

1.4. *Two color arrays can make designs more complex, but may reduce variability*

In two-color arrays, two samples using different colored dyes (conventionally called red and green) are hybridized to the same array. The advantage of this is that a direct comparison of these two samples may be more accurate than if they are hybridized to two different arrays, since such sources of error as array-to-array variability and differential spot size will cancel out. Of course, there may be dye bias that needs to be accounted for [4].

Frequently, the data from a two-color array experiment is immediately reduced to log ratios, but this is probably a mistake for reasons explained below in the section on putting the data on the right measurement scale. Briefly, log ratios work well when both conditions have the gene expressed well above background, but work badly when the gene is essentially off in one condition or the other.

A different issue is the effect on experimental design of two-color arrays. Given the 32-array experiment described in the last section, design issues with one-color arrays are relatively straightforward. The 32 arrays should be run in a

random order in the same lab by the same operator with the same batch of reagents (the random order is to avoid biases due to time order). With two-color arrays, we have further design choices. We can provide a universal control (such as by pooling a large number of similar samples), and always use the control on red and the sample on green or vice versa. We can make half the arrays with the control on red and half with the control on green. Or we can use loop designs [5]. In the latter, we use no control, but always hybridize one experimental sample labeled with red with another labeled in green. This is more efficient, but creates some complexities. We need to make sure that comparisons are balanced between red and green. Direct comparisons of two conditions are the most accurate, but usually we cannot accommodate all of these. For example, with 16 treatment combinations, there are 120 possible pairs (or 240 if we do one each of red/green and green/red). With only 32 arrays this cannot be accommodated. Efficient design in this case requires use of optimal design software such as that found in SAS.

At this stage of our understanding, it may be that loop designs and the like should be used only for quite simple experiments, and that the universal control, though statistically perhaps less efficient, should be used for more complex designs.

2. Principles of analysis

2.1. Analyze the experiment in accord with the design

Suppose we have an experiment with four types of cell lines, with two samples per type, and one array or other analysis per sample. With one response, we would typically analyze this by the following one-way ANOVA model:

$$\text{Response} = \text{mean} + \text{type} + \text{error},$$

which is in accord with the design. It may be tempting to use some canned analysis tools to, for example, cluster genes based on eight expression measurements. Since this ignores the treatment (cell-line type) it is unlikely to lead to a good analysis. The simplest first step is to test each gene (MS peak, etc.) as if it were the one response and determine if that gene had changed in a manner that could not be easily explained by chance. Some care is required in this process, but the basic idea is sound.

The entire experiment for all genes (replace gene by peak or bin for MS and NMR) and all samples can also be framed as a single ANOVA model in the form

$$\text{Response} = \text{mean} + \text{gene} + \text{array} + \text{gene} \times \text{type} + \text{error}.$$

Here the array effect is a normalization on the transformed scale (comparable to a multiplicative normalization on the original scale), and the effects of main interest are the gene \times type terms. Note, however, that the overall mean square error from the model pooling all the genes should probably

not be used for testing, even after a variance stabilizing transformation (see below).

2.2. Put the data on the right scale for analysis

In general, the intuitively most appealing methods of comparison for the expression level of a gene or concentration of a protein or metabolite in two groups are multiplicative. We might like to say that a gene is two-fold over expressed in one group compared to another. It turns out that it is often better in such cases to log transform the data and conduct an analysis of differences. This is essentially equivalent since $\log(x/y) = \log(x) - \log(y)$. In addition, more complex comparisons such as the one-way ANOVA above can be conducted on the transformed data with greater validity.

However, one problem with this strategy is when one or more of the conditions has the gene essentially unexpressed (protein essentially absent, etc.), an extremely common phenomenon. This is shown by very low measured expression values, near, or even below zero.¹ The answer to this problem is to use a slight variant of the logarithmic transform, called the generalized logarithm or glog transformation [6–9]. Much additional work has been devoted to methods of using this transformation to improve analyses of microarray data [1,10–12]. This work was motivated by the desire to stabilize the variance of microarray data, and the glog transformation is the natural choice when the variability of microarray data is modeled as in [13].

Fig. 1 shows the glog and log transforms over the range of a microarray data set. Note that the difference is difficult to discern except at the very lowest level. Fig. 2 shows the two transformations near zero, showing that the glog transformation deals more gracefully with low-level data, but is otherwise similar to the log transform.

2.3. Account for multiplicity of tests

Given that there are many probes or probe sets on a microarray, there is a strong possibility of false positives in tests for differentially expressed genes. For example, there are approximately 40,000 probe sets on the latest Affymetrix human genome array. If each gene is tested for differential expression, and if the chance of a false positive is 1% for each test, then the expected number of false positives is 400. If there are 1000 probe sets that are declared to be differentially expressed, then probably 40% of them are *false discoveries*.

¹ Of course the actual concentration cannot be negative. However, if the true concentration is zero, and the measurement process is unbiased, then negative measurements will always occur. The interpretation of a negative measurement is in terms of the confidence interval. If the standard deviation near 0 is, say, 10, then a measured value of 10 suggests a concentration between 0 and 30 (plus or minus two standard deviations, but truncated at 0). A measured concentration of -10 suggests a true concentration between 0 and 10, thus is indicative of a lower true concentration than a measured value of 10 or 0. This implies that negative measurements should not be censored or reduced to 0 or half the detection limit to avoid low-level bias.

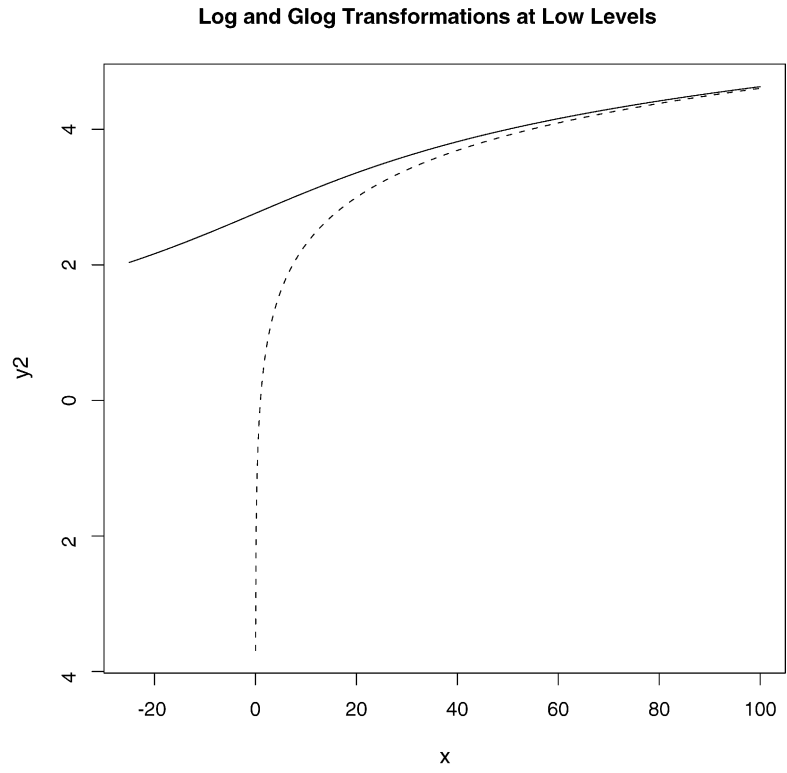


Fig. 1. Logarithm (log) and generalized log (glog) transforms over a range consistent with microarray data.

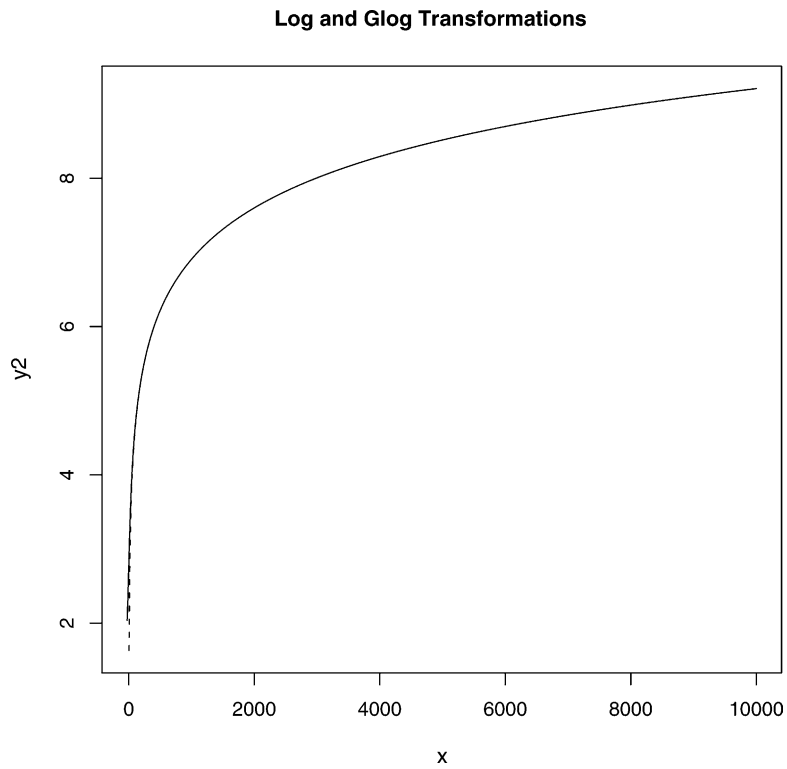


Fig. 2. Logarithm (log) and generalized log (glog) transforms over a range near zero.

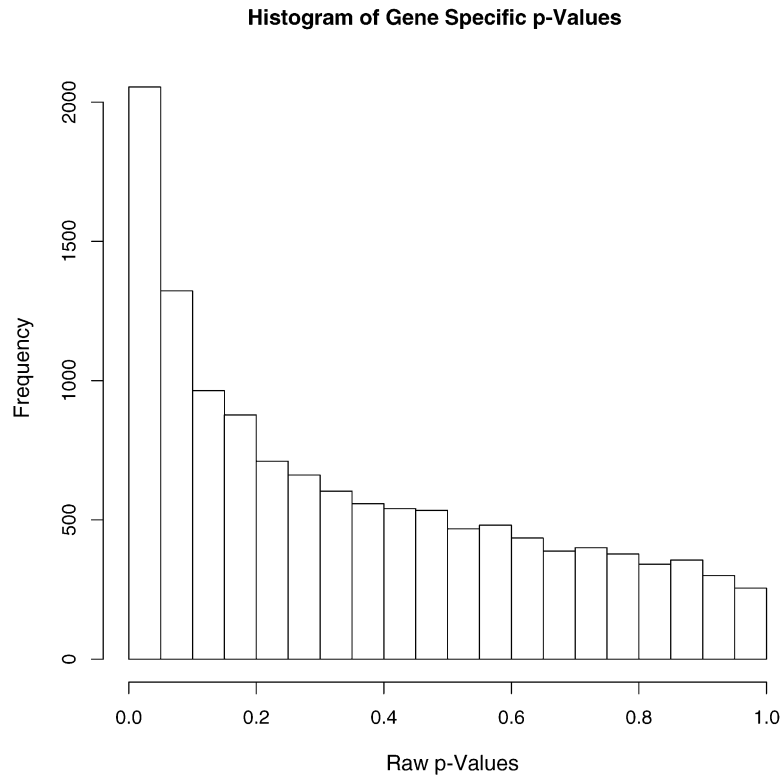


Fig. 3. Histogram of all 12,625 p -values from a microarray experiment, in which each gene is analyzed separately.

Generally, this is too high a fraction. False discovery rates can be controlled by various procedures, but perhaps the most popular are based on the methods of Benjamini and Hochberg ([14–18]). These are easy to perform using standard software and can give at least some assurance that the number of false positives is approximately at or below some threshold. The same argument holds if there are 1000 peaks per sample in a mass spectrogram or 256 binned spectral values in an NMR study.

The alternative of avoiding statistical tests at all, and screening by some fold-change threshold is a poor substitute. In the first place, fold change can be inflated if the expression is low in one of the conditions, though this can be solved by using the glog ratio instead of the log ratio. Second, fold change is only easily applicable to comparison of two alternatives and is difficult to use with more complex designs, even a one-way design with more than two levels. Third, the natural variability of measurements by a given probe is

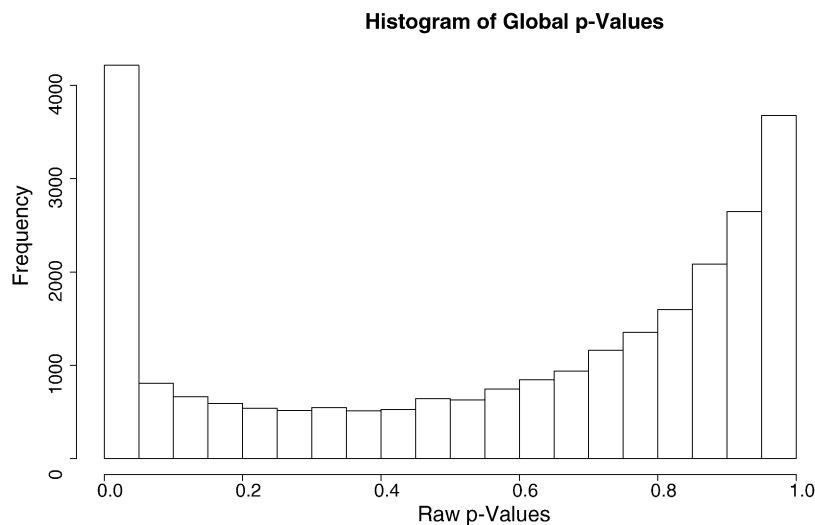


Fig. 4. Histogram of all 12,625 p -values from a microarray experiment, in which the error term for the analysis of each gene comes from a single pooled value across all genes.

not the same. In addition to the well-known dependence of the standard deviation on the mean [13], there are large differences in natural variability between measurements from probe sets even though the mean expression measurements may be very similar (see Section 2.4). This implies that any fold-change cutoff will tag too many probe sets that happen to have large natural variability, and too few that happen to have small natural variability. What counts is whether the differences between conditions are larger than can be explained by the natural variability attendant on that probe set, and that requires statistically valid analyses.

2.4. What is the right denominator for *t*-tests and *F*-test?

If tests are conducted one gene or probe set (or MS peak, etc.) at a time, then each such test should be approximately valid. In many cases, however, the test will not be sensitive due to relatively poor estimates of the error term. Consider, for example, the experiment with four types of cell lines and two arrays per cell line. The ordinary ANOVA test of differential expression for a single probe set uses an error estimate with only four degrees of freedom. Fig. 3 shows the collection of 12,625 *p*-values produced by analyzing each probe set of the Affymetrix U95A chips that were used for this experiment. The probe sets on each array were summarized using the mean log transformed PM probe values (results are similar with, for example, the RMA summary method [19]), and each probe set was tested for significance using a one-way

ANOVA as described above. Clearly, there are many probe sets that are significantly different (if there were none, the *p*-value histogram should be flat). However, none of the *p*-values passes the 5% FDR threshold, meaning that we are sure that there are probe sets that are differentially expressed, but not sure which ones exactly that they are. This is primarily due to the fact that the MSE is poorly estimated in a one gene at a time analysis.

An alternative is to pool all the probe sets together and use the MSE with more than 50,000 degrees of freedom. Fig. 4 shows the *p*-values from this procedure. The peak at the right end of the plot corresponds to *F*-ratios that are extremely small, which should not occur under the model. The reason that they do occur is the aforementioned differences in natural variability in probe set measurements even after variance stabilization. If the natural variability of a gene is low, then the pooled MSE will be much too large, and the *F*-ratio will be much too small. Correspondingly, many of the apparently significant probe sets with very low *p*-values will only be there because the MSE is much smaller than the natural variability of that probe set. This method is therefore not recommended since it does not lead to statistically valid results.

A middle ground is to use a hierarchical model and estimate it using an empirical Bayes methodology. This method is fully described in a manuscript available on the website <http://www.cipic.ucdavis.edu/~dmrocke> (Rocke, Heterogeneity of variance in gene expression microarray data;

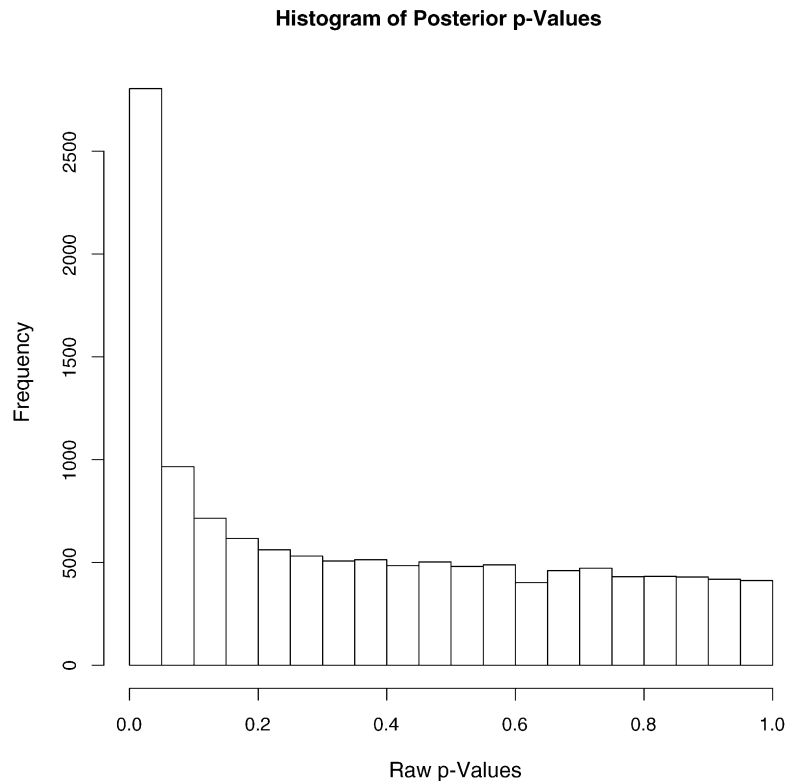


Fig. 5. Histogram of all 12,625 *p*-values from a microarray experiment, in which the error term for the analysis of each gene comes from an empirical Bayes analysis.

2003), and implemented in an R package (LMGene) available on the same site. This is essentially the same method is used in [20]. The outcome of this methodology is that small gene-specific MSE's are "shrunk" towards a middle value, and therefore increased, whereas large gene-specific MSE's shrink towards the middle value and decrease. The information content increases from 4 d.f. to about 8.6 d.f., which makes the procedure much more powerful. Fig. 5 shows all the p -values, and the pathology of Fig. 4 is not present. The number of genes identified as differentially expressed using a 5% FDR procedure increases from 0 to 866, showing the greatly improved power of this procedure.

3. Evaluation of alternative analysis tools

3.1. Look for evaluations in the open literature

In all too many cases, procedures for analysis of array data are not validated by those who propose them. At a minimum, a newly proposed procedure should be compared to existing procedures and shown to be at least as effective. Large comparative evaluations of alternative methodologies are especially helpful (e.g. [19] and [21]).

3.2. Know what the method does and why it is appropriate

Always, the first issue should be to answer the specific scientific question behind the experiment. Tools of analysis should be chosen to address these specific questions. For example, if one conducts a dose–response study, then the primary questions are things like (1) Which genes change measurably at all over the different doses? (2) Which genes change in patterns thought to be biologically relevant, such as increasing with the dose, decreasing with the dose, or rising and then falling? Commonly, the first resort is to conduct a cluster analysis of the genes, but this is not necessarily aimed at answering the scientific questions at hand. Clustering can be useful, but should generally not be the first resort.

4. Proteomics and metabolomics by mass spectrometry and NMR spectroscopy

The data processing and statistical analysis problems with proteomics and metabolomics data by mass spectrometry and NMR spectroscopy resemble those presented by gene expression data. Instead of background correction of arrays, we have baseline estimation. Instead of spot intensities, we have peak

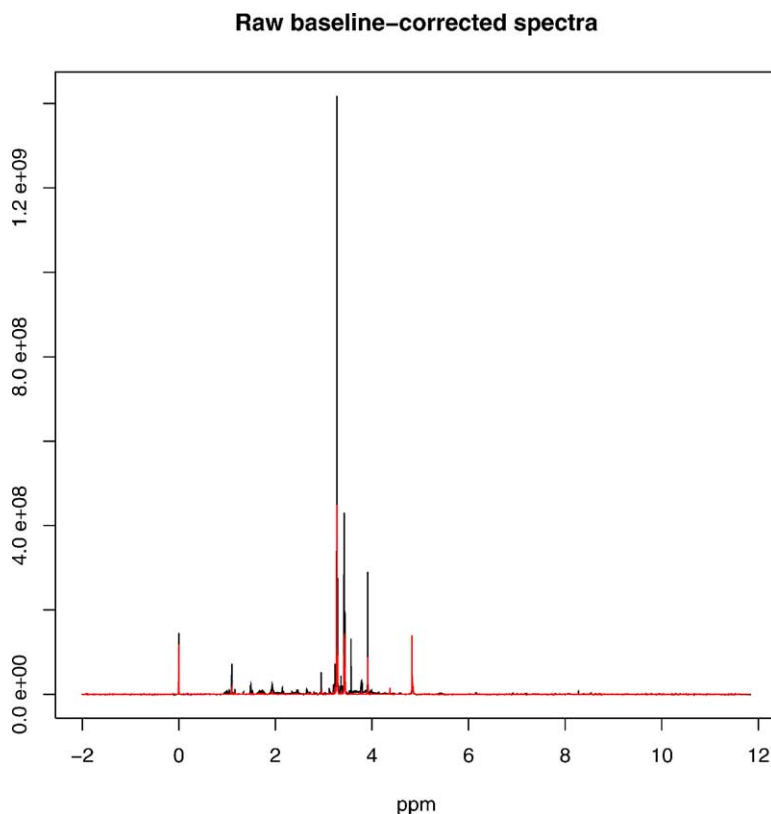


Fig. 6. NMR spectra from two abalone, one healthy and one diseased.

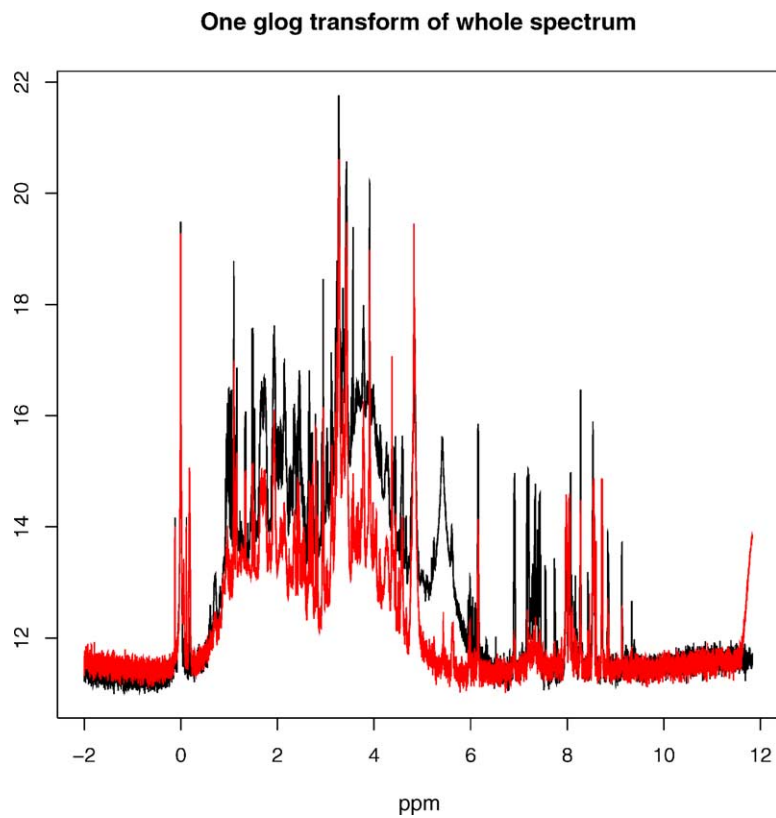


Fig. 7. Two NMR spectra after glog transformation.

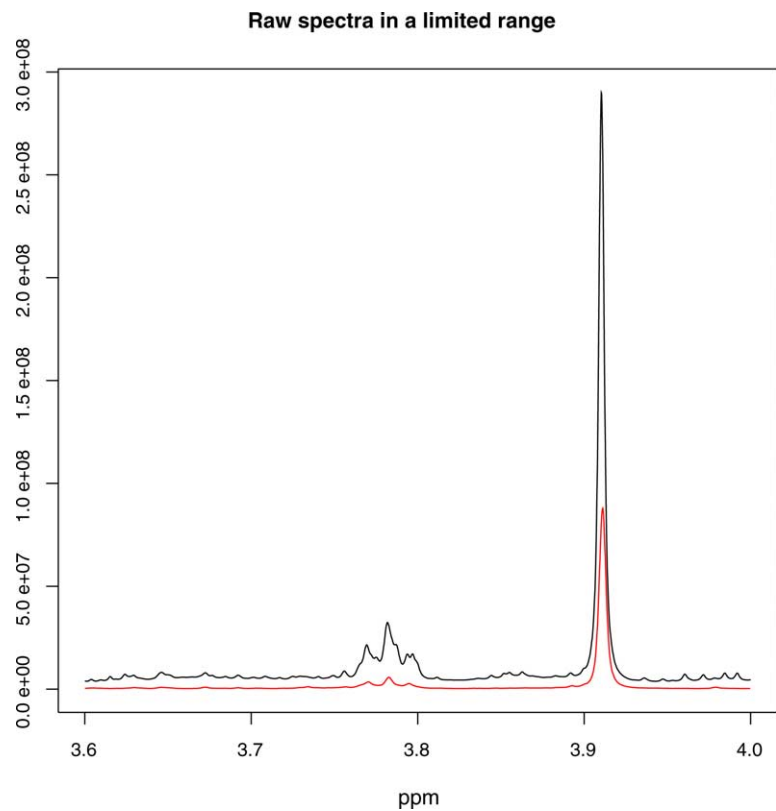


Fig. 8. NMR spectra in a limited range.

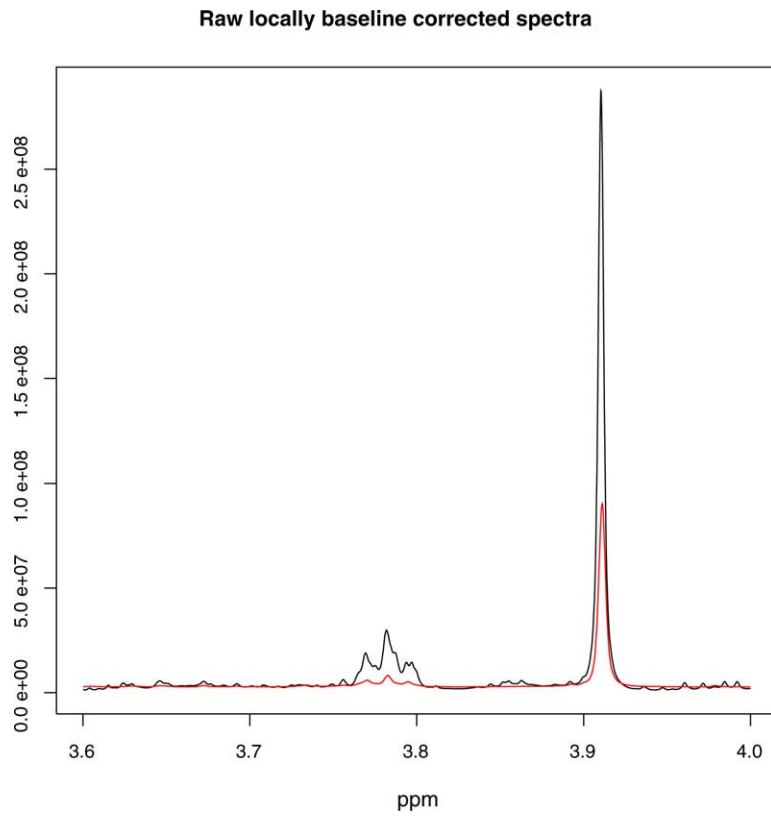


Fig. 9. NMR spectra in a limited range after local baseline correction.

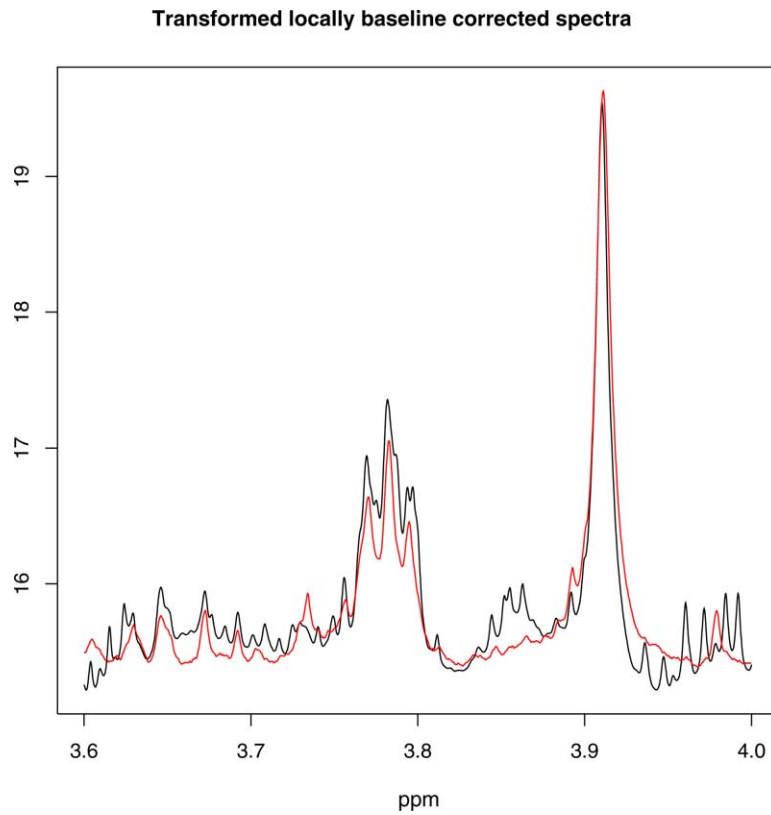


Fig. 10. NMR spectra in a limited range after local baseline correction, glog transformation, and additive normalization.

heights, peak areas, or binned spectral values. This approach is illustrated for proteomics by mass spectrometry data in [22] and for NMR metabolomics data in [23].

We illustrate the issues involved by a look at the NMR spectra for one healthy and one diseased abalone chosen from a recent study [24]. Obviously, there are not enough data for a statistical analysis with only one sample from each group (a more complete analysis is in [23]), but it will serve to illustrate the commonality of issues with gene expression data analysis.

Fig. 6 shows the two spectra, which have been globally baseline corrected by a standard quadratic fit algorithm [25]. On this scale, differences are difficult to see, since the entire picture is dominated by a few extremely high peaks. Fig. 7 shows the same data after a possible glog transformation. Note that it is now apparent that there are problems in the baseline correction, since the noise regions at either end are not aligned. These problems are invisible on the original scale. Fig. 8 shows a small segment of the spectrum, in which we can easily see the inadequate baseline correction. This must be repaired, since otherwise differences in baselines may translate into apparent differences in the spectrum that are entirely artifactual. Fig. 9 shows the same segment after local baseline adjustment. Finally, Fig. 10 shows the spectra after local baseline adjustment, glog transformation, and additive normalization on the glog scale. After these adjustments, we are much more easily able to see differences in the spectrum than we could before. If these differences are consistent across several samples of healthy and sick abalone, we have potentially both the ability to discriminate and the possibility to identify a specific compound that is a biomarker for health/sickness [23].

There are many open questions to the use of NMR spectroscopic data for metabolomics (and similar questions for the use of mass spectrometry). Here are some questions that require further research:

1. How should the baseline be estimated so as not to distort small but significant peaks?
2. How should peaks be aligned, given that their chemical shift locations will vary from run to run depending on pH and other factors?
3. Should we quantify the spectrum via peak heights or areas, or should the spectrum be binned as in [26–29]?

In addition, there is much work needed in improved pattern recognition, once the data have been baseline corrected, transformed, and normalized.

5. Conclusions

Analysis of high-throughput biological assay data—including gene expression microarray data and proteomics and metabolomics data by mass spectrometry and NMR spectroscopy—is neither mysterious on the one hand nor trivial on the other. Long standing principles of statistical design and analysis of experiments still apply, but there are

many complexities that require consideration by the experimenter. Since the literature is evolving rapidly, and since commercial software is usually at least several years behind the literature, the best answer is to have someone on the research team who is expert in the area of analysis of these data.

Acknowledgements

The research reported in this paper was supported by grants from the National Science Foundation (ACI 96-19020, and DMS 98-70172) and the National Institute of Environmental Health Sciences, National Institutes of Health (P43-ES04699 and P01-ES11269).

References

- [1] Geller S, Gregg J, Hagerman PJ, Rocke DM. Transformation and normalization of oligonucleotide microarray data. *Bioinformatics* 2003;19:1817–23.
- [2] Kendziorski CM. Pooling biological samples in microarray studies in DNA microarrays and statistical genomics techniques: design, analysis, and interpretation of experiments. New York: Marcel Dekker; 2003.
- [3] Box GEP, Hunter WG, Hunter JS. Statistics for experimenters: an introduction to design, data analysis, and model building. New York: Wiley; 1978.
- [4] Kerr K, Martin M, Churchill G. Analysis of variance for gene expression microarray data. *J Comput Biol* 2000;7:819–37.
- [5] Churchill GA. Fundamentals of experimental design for cDNA microarrays. *Nat Genet* 2002;32(Suppl S):490–5.
- [6] Durbin BP, Hardin JS, Hawkins DM, Rocke DM. A variance-stabilizing transformation for gene-expression microarray data. *Bioinformatics* 2002;18:S105–10.
- [7] Hawkins DM. Diagnostics for conformity of paired quantitative measurements. *Stat Med* 2002;21:1913–35.
- [8] Huber W, von Heydebreck A, Sültmann H, Poustka A, Vingron M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* 2002;18:S96–104.
- [9] Munson P. A ‘consistency’ test for determining the significance of gene expression changes on replicate samples and two convenient variance-stabilizing transformations. GeneLogic workshop on low level analysis of Affymetrix GeneChip data; 2001.
- [10] Durbin B, Rocke DM. Estimation of transformation parameters for microarray data. *Bioinformatics* 2003;19:1360–7.
- [11] Durbin BP, Rocke DM. Variance-stabilizing transformations for two-color microarrays. *Bioinformatics* 2004;20:660–7.
- [12] Rocke DM, Durbin BP. Approximate variance-stabilizing transformations for gene-expression microarray data. *Bioinformatics* 2003;19:966–72.
- [13] Rocke DM, Durbin BP. A model for measurement errors for gene expression arrays. *J Comput Biol* 2001;8:557–69.
- [14] Benjamini Y, Hochberg Y. Controlling the false discovery rate. *J R Stat Soc Ser B* 1995;57:289–300.
- [15] Dudoit S, Shaffer JP, Boldrick JC. Multiple hypothesis testing in microarray experiments. *Stat Sci* 2003;18(1 February):71–103.
- [16] Reiner A, Yekutieli D, Benjamini Y. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* 2003;19:368–75.

- [17] Storey JD, Taylor JE, Siegmund D. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *J R Stat Soc Ser B-Stat Methodol* 2004;66(Part 1):187–205.
- [18] Tsai CA, Hsueh HM, Chen JJ. Estimation of false discovery rates in multiple testing: application to gene microarray data. *Biometrics* 2003;59(4 December):1071–81.
- [19] Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res Methods Online* 2003;31:e15.
- [20] Wright GW, Simon RM. A random variance model for detection of differential gene expression in small microarray experiments. *Bioinformatics* 2003;19:2448–55.
- [21] Dudoit S, Fridlyand J, Speed TP. Comparison of discrimination methods for the classification of tumors using gene expression data [article]. *J Am Stat Assoc* 2002;97(457 March):77–87.
- [22] Purohit PV, Rocke DM. Discriminant models for high-throughput proteomics mass spectrometer data. *Proteomics* 2003;3:1699–703.
- [23] Purohit PV, Rocke DM, Viant MR, Woodruff DL. Discrimination models using variance-stabilizing transformation of metabolomic NMR data. *Omics* 2004;8:118–30.
- [24] Viant MR, Rosenblum ES, Tjeerdema RS. A powerful approach for characterizing the effects of environmental stressors on organism health. *Environ Sci Technol* 2003;37(21):4982–9.
- [25] Hoch JC, Stern AS. *NMR Data Processing*. New York: Wiley-Liss; 1996.
- [26] Nicholson JK, Connelley J, Lindon JC, Holmes E. *Nat Rev Drug Discov* 2002;1:153.
- [27] Homes E, Nicholson JK, Tranter G. *Chem Res Toxicol* 2001;14:182.
- [28] Azmi J, Griffin JL, Antti A, Shore RF, Johansson E, Nicholson JK, Holmes E. *Analyst* 2002;127:271.
- [29] Beckwith-Hall BM, Brindle T, Barton RH, Coen M, Homes E, Nicholson JK, Antti H. *Analyst* 2002;127:1283.