Sample Size Estimation for Microarray Experiments

Gregory R. Warnes*

Department of Biostatistics and Computational Biology

Univeristy of Rochester

Rochester, NY 14620

and

Peng Liu

Department of Biological Statistics and Computational Biology

Cornell University

Ithaca, NY 14853

July 22, 2006

SUMMARY

RNA Expression Microarray technology is widely applied in biomedical and pharmaceutical research. The huge number of RNA concentrations estimated for each sample make it difficult to apply traditional sample size calculation techniques and has left most practitioners to rely on rule-of-thumb techniques. In this paper, we describe and demonstrate a simple method for performing and visualizing sample size calculations for microarray experiments. We then summarize simulation

* *email:* warnes@bst.rochester.edu

results showing that this method performs well, even under a variety of departures from our simplifying assumptions, provided that the underlying formula for computing sample size has been appropriately selected. Although we demonstrate sample size calculation only for the two-sample pooled t-test, it is trivial to substitute an alternative sample size formula appropriate to the problem at hand. The described method has been implemented in the `ssize` R package, which is available from the Bioconductor project (`http://www.bioconductor.org`) web site.

KEY WORDS: Sample Size, Gene Expression, Microarray, t-test

1. Introduction

High-throughput microarray experiments allow the measurement of expression levels for tens of thousands of genes simultaneously. These experiments have been used in many disciplines of biological research, including as neuroscience (Mandel *et al*., 2003), pharmacogenomic research, genetic disease and cancer diagnosis (Heller, 2002). As a tool for estimating gene expression and single nucleotide polymorphism (SNP) genotyping, microarrays produce huge amounts of data which are providing important new insights.

Microarray experiments are rather costly in terms of materials (RNA sample, reagents, chip, etc), laboratory manpower, and data analysis effort. As a consequence, such experiments often employ only a small number of replicates (2 to 8) (Yang and Speed, 2003). In many cases, the selected sample size may not be adequate to provide reliable statistical answers to the scientific question of inter-

est. Conversely, a given sample size may be larger than necessary for answering the question at hand. Either case can result in a waste of resources. It is important, therefore, to perform proper experimental design, including sample size estimation, before carrying out microarray experiments. Since tens of thousands of variables (gene expressions) may be measured on each individual chip, it is essential to take into account multiple testing and dependency among variables when calculating sample size. Estimating sample size for large number of statistical tests adds new questions to those traditionally posed by sample size calculations, including: Should one use one or multiple values for power? Should a single value for minimum effect size be used? How should error rates be adjusted for multiple-testing? How can we account for dependency among the variables? Further, traditional methods for sample size calculation apply only to a single hypothesis test and cannot be directly applied to the microarray problem where tens of thousands of hypothesis tests are preformed.

Possible approaches to estimating sample size for microarray experiments range between two extremes. At one extreme, sample size estimation can be performed by constructing a model for the entire system, including the realistic error structures and interdependencies among variables. Provided that the model is appropriate, this choice should generate a highly accurate answer. However, it can take tremendous effort to find an appropriate and general model, not to mention the complexity necessary for fitting such a model. Since the behavior of microarray data from different microarray technologies and for different biological systems is extremely varied, the applicability of such model and its heuristic

parameters to a particular problem is uncertain.

At the other extreme, standard sample size calculation methods can be applied individually to each gene, and a summary can be constructed to provide an overall estimate of the required sample size. With this choice, the calculations are simple to set up, the results can be obtained quickly, and it is easy to incorporate available data. Further, such methods are easy to understand and to explain to applied scientists. However, this approach might not capture enough of the structure of the system, leading to incorrect estimates for sample size.

Several papers have addressed the sample size calculation. For example, (Zien *et al.*, 2003) applied the first approach and proposed a hierarchical model including several different sources of error and suggest heuristic choices for the key parameters in the model. Unfortunately, the model itself is non-identifiable, preventing the use of historical data to directly estimate the necessary parameters. Some proposed methods have set the parameter being the same and address the sample size calculation for different microarray designs (Lee and Whitmore, 2002), (Dobbin and Simon, 2005) or specific purposes such as classification (Hua *et al.*, 2005). Several recent papers investigated the problem of multiple testing such as controlling the false discovery rate (FDR) (Liu and Hwang, 2005), (Yang *et al.*, 2003), (Pawitan *et al.*, 2005), (**?**),(**?**). The sample size calculation has been an applied topic that attracts many statisticians.

We have applied this simple approach, i.e., to calculate sample size individually for each gene. We propose a new visualization method that can inform biologists about the tradeoff between sample size and power. This has been applied in

our consultation work and proved to be very helpful in deciding sample size with consideration of experimental costs and statistical power. The proposed method in this paper can be combined with other methods mentioned above as long as the calculation of sample size per gene is available. Since the method is based on simplified assumptions, we also describe a simulation study demonstrating that our method functions well despite its simplicity.

The paper is organized as follows. Section 2 describes our proposed method in detail. Simulation studies to check the performance of the method under a variety of conditions are outlined in Section 3. Section 4 presents results and observations from the simulation study. Section 5 provides discussion of the simulation results and comparison of our method with other proposed methods.

2. Method

To illustrate our method, we assume that a microarray experiment is set up to compare gene expressions between one treatment group and one control group. The scanned intensity data for microarray experiments usually go through quality control, transformation and normalization, as reviewed in (Smyth *et al*., 2003; Quackenbush, 2002). We assume that data first go through those steps before applying statistical tests. Before the experiment, we have no observations to check the distribution. It seems reasonable to make a convenient assumption that the distribution of the pre-processed data is normal and hence a standard 2-sample pooled-variance t-test is applicable detect differentially expressed genes. The same assumption has been made by other methods to calculation sample size for

microarray experiments (Yang *et al.*, 2003), (Liu and Hwang, 2005), (Dobbin and Simon, 2005), (Pawitan *et al.*, 2005). The tested hypothesis for each gene is:

$$H_0 : \mu_T = \mu_C$$

versus

$$H_1 : \mu_T \neq \mu_C$$

where μ_T and μ_C are means of gene expressions for treatment and control group respectively.

The proposed procedure to estimate sample size is:

1. Estimate standard deviation (σ) for each gene based on *control samples* from existing studies performed on the same biological system.

2. Specify values for

   (a) minimum effect size, Δ, (log of fold-change for log-transformed data)

   (b) maximum family-wise type I error rate, α

   (c) desired power, $1 - \beta$.

3. Calculate the per-test Type I error rate necessary to control the family-wise error rate (FWER) using the Bonferroni correction:

$$\alpha_G = \frac{\alpha}{G} \tag{1}$$

where G is the number of genes on the microarray chip.

4. Compute sample size separately for each gene according to the standard formula for the two-sample t-test with pooled variance:

$$\begin{aligned} & 1-\beta \\ & = 1 - T_{n_1+n_2-2}\left(t_{\alpha_G/2,n_1+n_2-2}|\frac{\Delta}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right) \\ & \quad + \; T_{n_1+n_2-2}\left(-t_{\alpha_G/2,n_1+n_2-2}|\frac{\Delta}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right) \end{aligned} \tag{2}$$

   where $\mathrm{T}_d(\bullet|\theta)$ is the cumulative distribution function for non-central t-distribution with d degree of freedom and the non-centrality parameter θ.

5. Summarize the necessary sample size across all genes using a cumulative plot of required sample size verses power. An example of such a plot is given in Figure 1 for which we assume equal sample size for the two groups, $n = n_1 = n_2$.

On the cumulative plot (Figure 1), for a point with x coordinate n, the y coordinate is the proportion of genes which require a sample size smaller than or equal to n, or equivalently the proportion of genes with power greater than or equal to the specified power $(1-\beta)$ at sample size n. This plot allows users to visualize the relationship between power for all genes and required sample size in a single display. A sample size can thus be selected for a proposed microarray experiment based on user-defined criterion. This will provide not just a single number but the

Figure 1. After estimation for each gene separately, all estimated sample sizes are summarized in this cumulative plot. Sample size can then be selected by balancing the number of genes achieving the desired power against the required sample size.

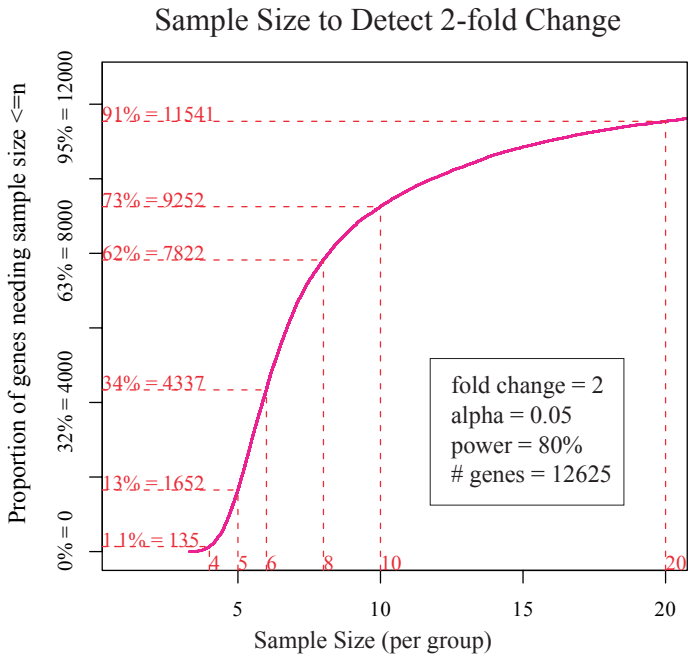


Figure 2. Given sample size, this plot allows visualization of the fraction of genes achieving a specified power.

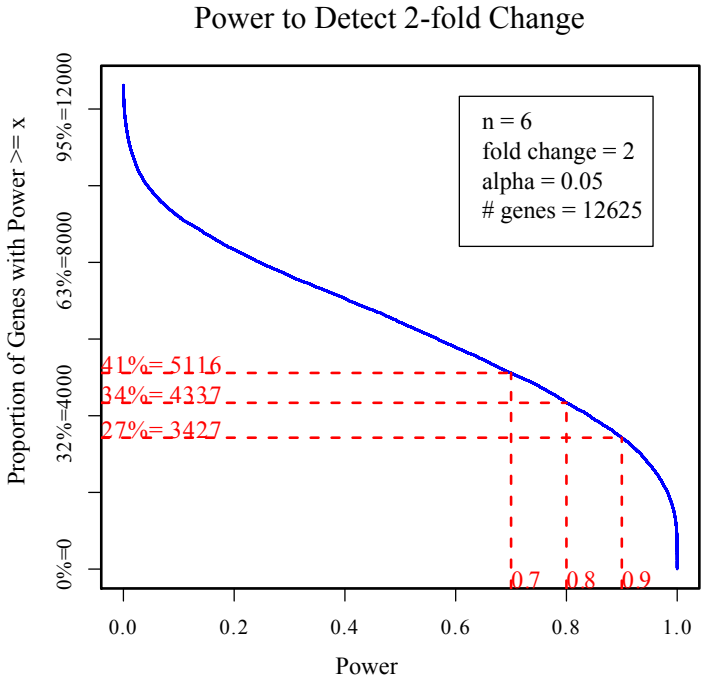
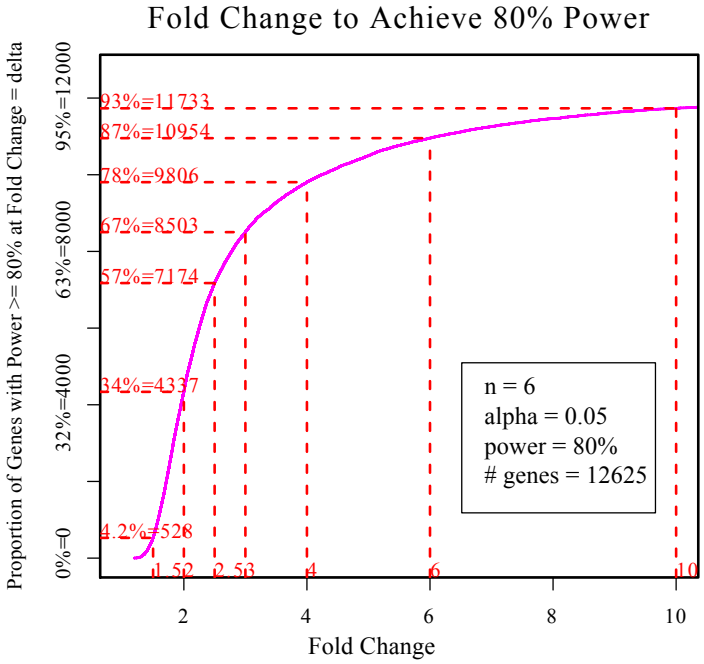


Figure 3. Given sample size, this plot allows visualization of the fraction of genes achieving the specified power for different fold changes.



tradeoff in selecting sample size with respect to statistical power. For the plot in Figure 1, for example, requiring 70% of genes to achieve the 80% power yields a sample size of 10.

Similar plots can be generated by fixing the sample size and varying one of the other parameters, namely, significance level (α), power ($1 - \beta$), or minimum effect size (Δ). Two such plots are shown in Figures 2 and 3.

Our method for computing sample size is easily adapted to other statistical tests for detecting differentially expressed genes. For example, if several groups of treatments are to be compared using ANOVA, our method can be modified by replacing the pooled t-test sample size formula (2) with the corresponding formula for the F-test. Similarly, formula (2) can be replaced with a paired t-test sample size formula to allow the use of a paired t-test.

3. Simulation Study

Given the simplifying assumptions used by our proposed method, we chose to perform a simulation study in order to check its performance under violations of our assumptions and to determine the influences of additional factors that could affect the accuracy of the results in the context of 2-sample t test. Similar studies can be performed to other tests that will be applied for microarray study.

Many variables could affect the accuracy of our sample size estimation method.For example, our method assumes independence between tests. However, biology predicts–and measured gene expressions demonstrate–strong dependency patterns among genes (e.g. among co-regulated genes). Unequal variance between the

Table 1

Tested Variables in Simulation Study. FDR indicates the Benjamini and Hochberg (1995) procedure to control false discovery rate

Description	Variable	Tested Levels
Proportion of genes that are dependent	γ	100%, 80%, 50%, 0%
Proportion of genes with true differential expression	a	100%, 50%, 10%, 5%
Variance ratio between two groups	r	1, 3, 10
Minimum effect size for scientific relevance	Δ	2-fold 4-fold
Choice of multiple comparison method	m.method	Bonferroni, FDR

control and treatment groups is another potential violation of our method's assumptions. Table 1 summarizes each of the variables that we included in our simulation study.

For different combinations of the parameters in Table 1, we simulated hundreds of micrarray data sets with sample size determined by the calculation under the simplified assumptions as described in Method. We then obtained the real power after performing hypothesis testing. Then the comparison between the simulation and calculation would show whether the proposed method provide a reasonable guide in practice.

The simulation passes for each combination of the simulation parameters follow the steps outlined in Figure 4.

Each pass starts by generating a "true" underlying distribution by sampling

standard deviations (for independent genes) and covariance matrices (for groups of interdependent genes) from a reference data set. Our reference data set contains gene expression values for smooth muscle cells from a control group of untreated healthy volunteers processed using Affymetrix U133 chips and normalized per the Robust Multi-array Average (RMA) method of Irizarry *et al*. (2003). The covariance matrices were calculated by selecting for each gene the most correlated 49 genes and calculating the covariance matrix for the 50 genes as a group. Then we have 12625 covariance matrices to sample from and each matrix has the dimension of 50x50. A specified fraction a of the genes is randomly selected as exhibiting a true change in expression consequent to the applied treatment. For these genes, the mean is set to the selected effect size Δ. For all other genes, the mean is set to 0.

We then mimic the process of using our method in practice. This can be viewed in two branches. For initial sample size estimation (left branch in Figure 4), we generate a set of samples based on the generated "true" underlying distribution for use as the historical reference population in our sample size estimation method. This serves the role of pilot study which is often available before designing experiment in real problem. We then compute a standard deviation for each gene based on this data set. These standard deviations are then used as to estimate the cumulative sample size curve for 80% power.

Once the sample size estimation has been completed, 985 separate experimental populations are generated (right branch of Figure 4). We elected to use 985 samples sets after calculating that 983 samples are required to provide 95% confi-

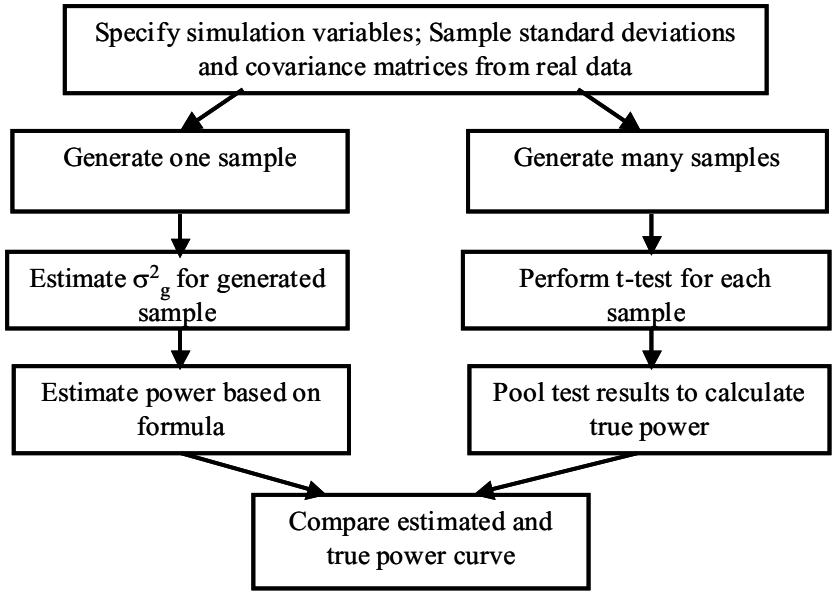
dence that the power obtained from simulation result is within a margin of 2.5% of true power. The standard two-sample t-test with pooled variance is then applied to each individual population using a range of sample sizes. Error rate within populations is controlled using one of two different methods, either the Bonferroni method for family-wise Type I error rate or the Benjamini and Hochberg procedure (1995) for False Discovery Rate (FDR). The tests from all populations are then pooled to determine the true power for each gene at each sample size. This allows the generation of the true cumulative sample size by summarizing across genes, which is then compared with the estimated curve generated by our proposed method. If the violations to the assumptions does not affect the accuracy of the sample size calculation, we would expect to see the overlap of the two curves obtained from simulation and from calculation.

4. Results

4.1 *Dependency*

Independence of the genes is a convenient statistical assumption, allowing application of statistical tests separately to each gene, which is applied in the proposed method. However, there is considerable biological and statistical evidence that the expression levels of genes are not independent (e.g. many transcription factors collaborate to regulate the expression of any one gene). To check the performance of the proposed method in estimating sample size when gene expressions are dependent in groups, we compared estimation of sample size based on proposed method with different proportions (γ) of genes that are in fact interdepen-

Figure 4. Simulation study performed to check the performance of proposed method. One set of gene expression sample is generated and calculated for sample size following the steps descried in Method. Many sets of gene expression data are generated and tested with standard pooled-variance t-test. Test results are pooled to get the true power for each sample size. Estimated sample size and simulation results are compared at the end.



dent in groups containing up to 50 genes. The covariance structure are based on our reference data set as described in Simulation Study. Figure 5 presents the result for the two most extreme cases: all genes are independent of each other (panel (a)) and 100% of genes come from interdependent groups (panel (b)). The other parameters for the two displayed panels are identical: $\alpha = 0.05$, $1 - \beta = 80\%$, $a = 1$, $r = 1$, $\Delta = 1$ (2-fold change) and Bonferroni adjustment is applied to controlling family wise type I error (FWER). In both cases, as well as at intermediate levels of dependency, the estimated sample size curve overlaps the results from simulation. This demonstrates that interdependency among genes has little or no effect on sample size estimation using our method. This result is reassuring, given the high degree of interdependence known to exist among genes.

4.2 *Proportions of Genes with True Differential Expression*

The proportion of genes with true differential expression will vary considerably among treatments. We suggest using all available genes for calculating cumulative plot unless information is available indicating that a specific subset of genes is more likely to be truly differentially expressed. In this latter case, this subset or "focus" list of genes can be used for sample size estimation instead of or in addition to performing the calculations for all genes.

We simulated four proportions of genes with true differential expression, $a = 100\%, 50\%, 10\%$ and 5%. None of the tested levels showed any effect on the accuracy of sample size estimation, i.e., the estimated sample size curve overlapped with true result from simulation for all tested proportions of differentially

Figure 5. Effect of interdependency of genes on sample size estimation. Cumulative plots for sample sizes from both estimation and simulation result are generated for the following variable values: $a = 1, r = 1, \Delta = 1$ (2-fold change) and Bonferroni adjustment is applied for controlling family-wise type I error. Dashed green lines are estimated sample sizes while solid black lines are sample sizes obtained from simulation result.

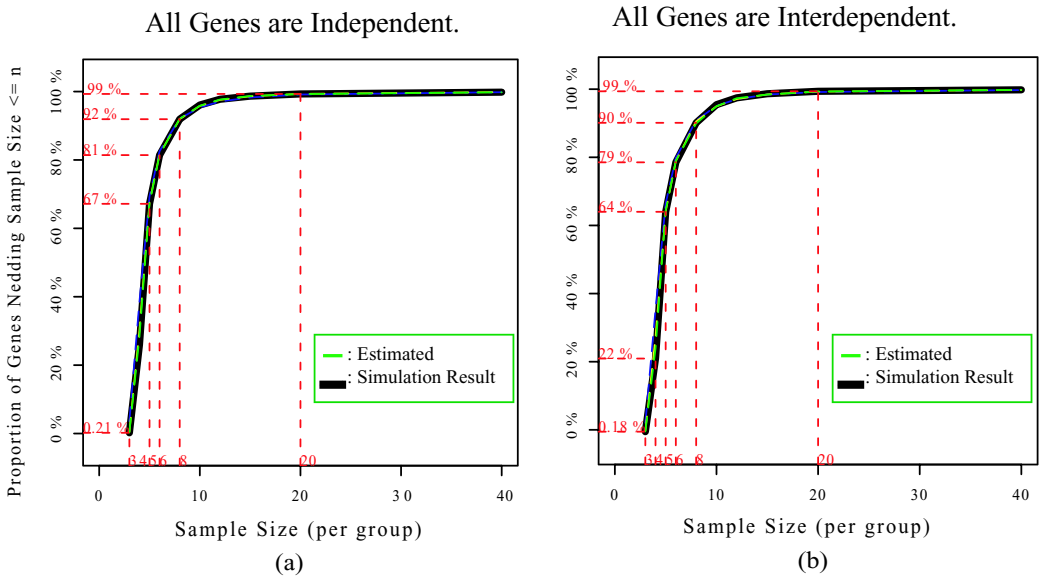
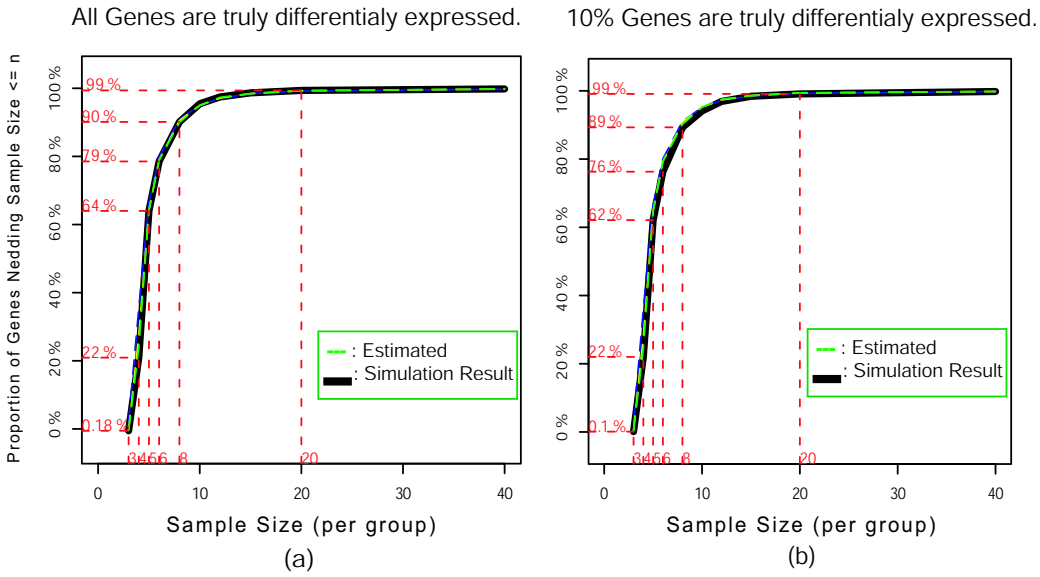


Figure 6. Effect of the proportion of genes with true differential expression on sample size estimation. Cumulative plots for sample sizes from both estimation and simulation result are generated for the following variable values: $\gamma = 1, r = 1, \Delta = 1$ (2-fold change) and Bonferroni adjustment is applied for controlling family-wise Type I error. Dashed green lines are estimated sample sizes while solid black lines are sample sizes obtained from simulation result.



expressed genes (Figure 6).

4.3 *Variance Ratio*

As described here, our method employs the sample size formula for the two-sample pooled-variance t-test. This test is only valid when the comparison groups have approximately equal variance. To determine the sensitivity of our method to the equal variance assumption, we simulated variance ratio between two groups of 1, 3 or 10. Since the problem is symmetric, it was unnecessary to test the ratios 1/3 and 1/10.

As expected, when the equal variance assumption is satisfied (variance ratio

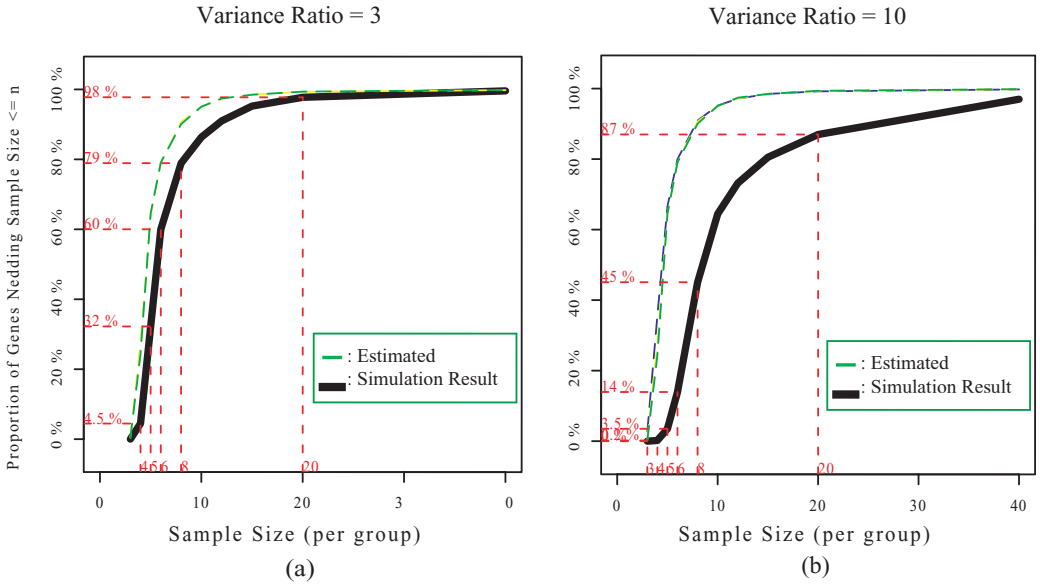
is 1) the estimated sample size agrees exactly with the true sample size (Figure 5 and 6). When the variance ratio is larger than one, the simulations show that the true sample size increases while the estimated sample size remains constant (Figure 7). This is as expected and simply confirms that the two-sample pooled t-test (2) is not robust against unequal variance.

Thus, when the treatment and control groups are expected to have a variance ratio greater than 3 or smaller than 1/3, we recommend replacing the pooled t-test sample size formula (2) with an appropriate unequal variance t-test sample size formula, and, of course, using the unequal-variance t-test for detecting differentially expressed genes.

4.4 *Minimum effect Size*

As the minimum effect size (Δ) increases, genes that are truly differentially expressed are easier to detect and the required sample size is correspondingly smaller. In our simulation, when we require at least 80% genes to achieve 80% power, a per-group sample size of 6 is required to detect 2-fold changes ($\Delta = 1$), while only 4 samples per group is needed to detect 4-fold changes ($\Delta = 2$) when $\gamma = 1$, $a = 1$, variance ratio $= 1$ and using the Bonferroni correction to control family-wise type I error. Our simulations demonstrate that, as expected, change in Δ has no effect on the accuracy of our sample size estimation methods since the value of Δ is explicitly included in the sample size formula 2.

Figure 7. Effect of variance ratio between treatment groups on sample size estimation. Cumulative plots for sample sizes from both estimation and simulation results are generated for the following variable values: $\gamma = 1, a = 1, \Delta = 1$ (2-fold change) and Bonferroni adjustment is applied for controlling family-wise Type I error. Dashed green lines are estimated sample sizes while solid black lines are sample sizes obtained from simulation result.



4.5 *Multiple Comparison Method*

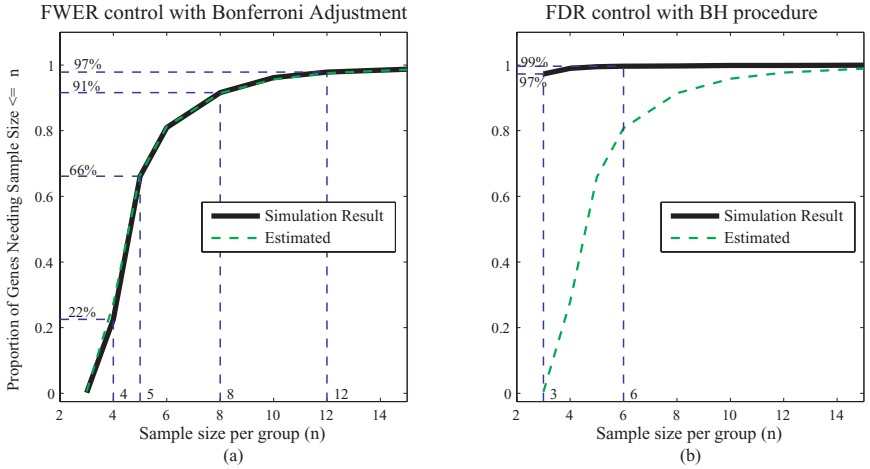
Many different methods have been proposed to control error rate for multiple testing. These include the Bonferroni correction for strong control of family-wise Type I error (FWER), the Benjamini and Hochberg (1995) method for controlling false discovery rate (FDR) and Storey's (2002) q-value procedure to control FDR or positive FDR (pFDR) . Both the FDR and pFDR methods control the proportions of false positives among all positive findings. These allow a certain proportion of Type I errors within the list of positive calls, usually resulting in higher power than Bonferroni adjustment which attempts to control the probability of any Type I error.

Although the FDR methods are applied more often in practice, we chose to utilize the Bonferroni method in sample size estimation for two reasons: First, the Bonferroni method is simple to apply. Second, it provides a conservative estimate of power and hence sample size. To determine the extent of the conservatism, we compared the estimated sample size when controlling FWER using the Bonferroni correction with the true sample size needed when controlling FDR via the Benjamini and Hochberg (1995) procedure (BH procedure). As expected, Figure 8 shows that the true sample sizes needed for the FDR method is considerably smaller than the sample size estimated using the Bonferroni adjustment.

5. Discussions

The number of chips included in microarray experiments directly affects the reliability of any conclusions from data analysis. Thus, it is important to have an

Figure 8. Effect of multiple comparison method on sample size estimation. Cumulative plots for sample sizes from both estimation and simulation result are generated for the following variable values: $\gamma = 1, a = 1, r = 1, \Delta = 1$ (2-fold change). Dashed green lines are estimated sample sizes while solid black lines are for sample sizes obtained from simulation result.



appropriate method for selecting the number of chips required to obtain reliable data in order to avoid wasting effort and resources. The huge number of correlated outcomes prevents traditional methods of estimating sample size from being directly applied to microarray experiments, at the same time that the increasing frequency of microarray experiments demands that appropriate methods be developed.

The literature proposes several methods of sample size calculation for microarray experiments. None are as straightforward and flexible as the method proposed here. Hwang *et al*. (2002) proposed a method that first identifies differentially expressed genes and then calculates power and sample size on a reduced parameter space. Dow studied relationship between minimum detection size and sample size

for a specific experiment (2003). Both methods cannot be applied to cases where preliminary results, including a knowledge of differentially expressed genes, are not available. Zien *et al.* (2003) proposed a hierarchical model which includes several different sources of error and recommend heuristic choices for the key parameters in the model. After attempting unsuccessfully to fit this model to real data, we discovered that that the model itself is non-identifiable. This makes it inappropriate to employ across the variety of sample types we see in practice.

We have instead proposed a very straightforward method for estimating required sample size that is easy to apply and is simple to adapt or extend. The key component of our method is the generation of cumulative plot of the proportion of genes achieving a desired power as a function of sample size, based on simple gene-by-gene calculations. While this mechanism can be used to select a sample size numerically based on pre-specified conditions, its real utility is as a visual tool for helping clients to understand the trade off between sample size and power. In our consulting work, this latter use as a visual tool has been exceptionally valuable in helping scientific clients to make the difficult trade offs between experiment cost and statistical power.

In order to check the performance of our proposed method, we performed an extensive simulation study. Of the variables tested in the simulation (Table 1), only variance structure and the multiple testing method had some substantial impact on the accuracy of the sample size estimation. Both of these were expected *a-priori* due to the specific assumptions employed and are easily corrected by appropriate modifications to the method (i.e. use of the appropriate sample-size formula

and use of the appropriate multiplicity correction during sample size estimation, respectively).

Neither the proportion of genes that are interdependent (γ) nor the proportion of genes with true differential expression (a) have any meaningful effect on the accuracy of our sample size estimates. This is fortunate, since a problem in either area would be difficult to correct based on information available before the experiment is run. That is, while we expect that genes have high correlation within regulatory and functional groups, it is currently impossible to determine the level of correlation among sets of genes for a given experiment before it is run. Likewise it is difficult to predict the fraction of genes with true differential expression before running the experiment.

In all of the cumulative plots of sample size versus number of genes achieving 80% power we found that there is a steep increase at small sample numbers. For example, Figure 5 shows that an increase of sample size from 4 to 5 assures 40% more genes with desired power (from 22% to 64%). Another increase of one chip (sample size = 6) results in about 80% of genes with the desired power. Taking this into account, an ideal sample size might be the leftmost number at the top of the steep portion of the curve.

Our simulation has also confirmed that the Benjamini and Hochberg procedure for controlling FDR has much more power than Bonferroni adjustment (Figure 8). A useful extension of proposed method would be the use of a FDR control instead of the conservative Bonferroni adjustment. One approach, suggested by Yang *et al.* (2003) computes an individual type I error to control FDR based on an initial

guess of how many genes will be differentially expressed (n_0). This could be used to modify our method to provide a better estimate when a FDR adjustment will be employed. Unfortunately, the proposed transformation of FDR to type I error rate is conservative and may control FDR at a level more strict than desired. We have a method that applies FDR-control during the sample size estimation process that can has been shown to provide more accurate estimation and can be combined to the tool that is presented here.

Acknowledgment

This work was supported by Pfizer Global Research and Development.

References

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing, *Journal of Royal Statistical Society B*, **57:1**, 289-300.

Dobbin, K., Simon, R. (2005) Sample size determination in microarray experiments for class comparison and prognostic classification, *Biostatistics*, **6:1**, 27-38.

Dow,G.S. (2003) Effect of sample size and p-value filtering techniques on the detection of transcriptional changes induced in rat neuroblastoma (NG108) cells by mefloquine, *Malaria Journal*, **2**, 4.

Heller, M. J. (2002) DNA microarray technology: devices, systems, and applications, *Annual Review in Biomedical Engineering*, **4**, 129-153.

Hua, J., Xiong, Z., Lowey, J., Suh, E., Dougherty, E. R. (2005) Optimal number of features

as a function of sample size for various classification rules, *Bioinformatics*, **21:8**, 1509-1515.

Hwang,D., Schmitt, W. A., Stephanopoulos, G., Stephanopoulos, G. (2002) Determination of minimum sample size and discriminatory expression patterns in microarray data, *Bioinformatics*, **18:9**, 1184-1193.

Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., Speed, T.P. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data, *Biostatistics*, **4:2**, 249-264.

Lee, M. T., Whitmore, G. A. (2002) Power and sample size for DNA microarray studies, *Statistics in Medicine*, **21**, 3543-3570.

Liu, P., Hwang, J. T. G. (2005) Quick Calculation for Sample Size while Controlling False Discovery Rate and Application to Microarray Analysis, *working paper*.

Mandel, S., Weinreb, O., Youdim, M. B. H. (2003) Using cDNA microarray to assess Parkinson's disease models and the effects of neuroprotective drugs, *TRENDS in Pharmacological Sciences*, **24:4**, 184-191.

Pawitan, Y., Michiels, S., Koscielny, S., Gusnanto, A., Ploner, A. (2005) False discovery rate, sensitivity and sample size for microarray studies, *Bioinformatics*, **21:13**, 3017-3024.

Quackenbush, J. (2002) Microarray data normalization and transformation, *Nature Genetics Supplement*, **32**, 496-501.

Smyth, G. K. and Yang, Y. H. and Speed, T. (2003) Statistical issues in microarray data analysis, *Methods in Molecular Biology*, **224**, 111-136.

Storey, J., (2002) A direct approach to false discovery rates, *Journal of Royal Statistical Society B*, **64:3**, 479-498.

Yang, M. C. K., Yang, J. J., McIndoe, R. A., She, J. X. (2003) Microarray experimental design: power and sample size considerations, *Physiological Genomics*, **16**, 24-28.

Yang, Y. H., Speed, T. Design and analysis of comparative microarray experiments *Statistical analysis of gene expression microarray data*, Chapman and Hall, 51.

Zien, A., Fluck, J., Zimmer, R., Lengauer, T. (2003) Microarrays: how many do you need?, *Journal of Computational Biology*, **10:3-4**, 653-667.