PowerAssist™: A Microarray QC Tool

Charles D. Johnson, Yuanhong Li, Mike Wilson, and Timothy S. Davison Asuragen, Inc. – A Spin-Off of Ambion, Austin, Texas 78744.

ABSTRACT

As part of the FDA MicroArray Quality Control project (MAQC), Asuragen, Inc., has developed a novel method to assess differences in inter-laboratory reproducibility, intrinsic platform-dependent factors, and multiple data processing methods. This graphical method is based on a modified two-sided t-test power analysis that evaluates both within-group variability and effect-size between experimental groups for each probe set on the array. Using the two-sample t-test power analysis in this way makes it a more useful quality assessment tool compared to conventional power analysis that accounts only for changes in standard deviation. Our method takes a real-world approach by calculating the average fold change between the two groups for each gene on the array and applies a pooled, standard deviation to calculate an empirically measured power for every gene. Results are then displayed on a cumulative plot that captures the contributions of all probes on the array, and more importantly, multiple plots can be overlaid to highlight the effects of changing laboratory or data processing methods. Application of this tool to the MAQC datasets from each of the major commercial array platforms demonstrates the general utility and sensitivity of this tool

INTRODUCTION

A graphical method was created as part of the FDA MicroArray Quality Control project (MAQC), to evaluate both the within group noise, and effect size between two groups for every probe set on the array as a quality assessment tool based on a two-sample t-test power analysis (Fig. 1). For a detailed description of the MAQC project see the September 2006 issue of Nature Biotechnology (Vol 24, No 9). With one graphic we can show the effect of the within-group variation (site effects for example), intrinsic platform-dependent noise, effects of sample titration, and normalization method. This analysis uses the standard formula for the power of two-sample t-test using a pooled estimate of standard deviation (s_{pooled}). The power of a statistical test is the probably of rejecting the null hypothesis given that the alternative hypothesis is true (1-β). Power depends on the type of statistical test, sample size, effect size, α -level, and sample standard

Fig.1

Two group pooled SD

$$\sigma_{pooled} = \sqrt{\frac{s_1^2(n_1-1)+s_2^2(n_2-1)}{n_1+n_2-2}}$$

Calculated Effect Size

Power Equation

$$1-\beta = 1-T_{n1+n2-2}\left(t_{\alpha/2,\,n1+n2-2}\,\middle|\,\frac{\Delta}{\sigma_{pooled}\sqrt{\frac{1}{n1}+\frac{1}{n2}}}\right)+T_{n1+n2-2}\left(-t_{\alpha/2,\,n1+n2-2}\,\middle|\,\frac{\Delta}{\sigma_{pooled}\sqrt{\frac{1}{n1}+\frac{1}{n2}}}\right)$$

deviation (s). Traditionally power analysis is conducted prospectively, solving for one of three variables; sample size, effect size, and power for a given a and s. The results provide an estimate of the necessary sample size, and effect size with the given level of confidence. Because we want to examine the effect of changing both effect size and sample noise we have elected to use a non-traditional approach where we vary two parameters within the power analysis simultaneously. These results are then plotted for each gene on the array using a cumulative frequency plot. The fact that we have solved for power in this way makes it useful as a quality assessment tool, but the results should not be confused with the previous microarray power analysis methods (Hwang et al., 2002; Seo et al., 2006; Tibshirani, 2006; Page et al., 2006) where all the parameters are imputed with the exception of stand deviation. Our analysis was developed to assay the performance of microarray platforms and data processing methods by both estimating standard deviation within the sample group and the measured distance between groups. Traditional power analysis methods failed to show this dynamic. This analysis is suited to examine mixture studies such as the MAQC project. Unlike a spike-study where only a relative few genes are examined, but the truth is known, our study describes the measurement characteristics of 12,091 common genes, but must rely on probabilistic inferences. Therefore, we have elected to use the term assisted or apparent power to distinguish our findings from those studies that use titrations of known quantities of a given gene. We used power analysis to show the effects of normalization (Fig. 2) and site-to-site variability (Fig. 3) across platforms

Figures and some text were first published in:

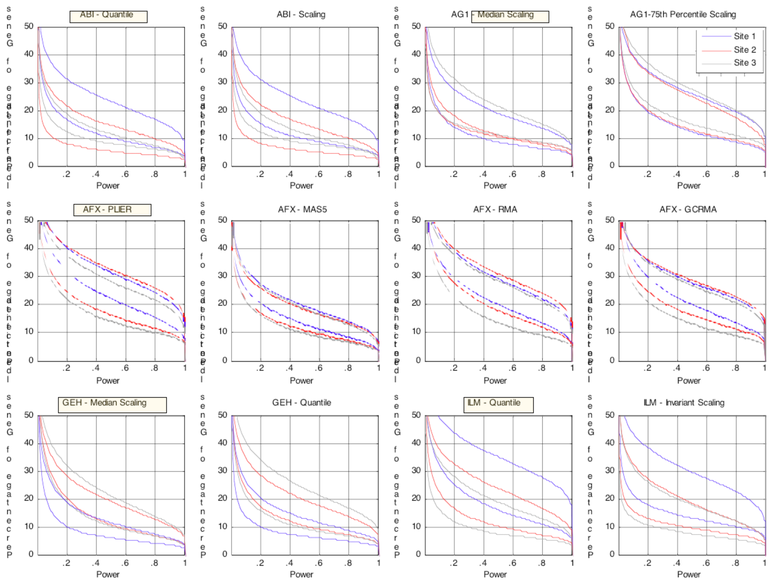
1. Shi L, et al. (2006). The MicroArray Quality Control (MAQC) project shows inter- and intra-platform reproducibility of gene expression measurements. Nat Biotechnol. Sep 8;24(9):1151-1161.

2. Shippy R, et al. (2006). Using RNA sample titrations to assess microarray platform performance and normalization techniques. Nat Biotechnol. Sep 8;24(9):1123-31

Asuragen, Inc. 2150 Woodward St. Austin, TX 78744
T: 1.512.651.0191 | 1.877.777.1874 | F: 1.512.651.0601
support@asuragen.com | www.asuragen.com

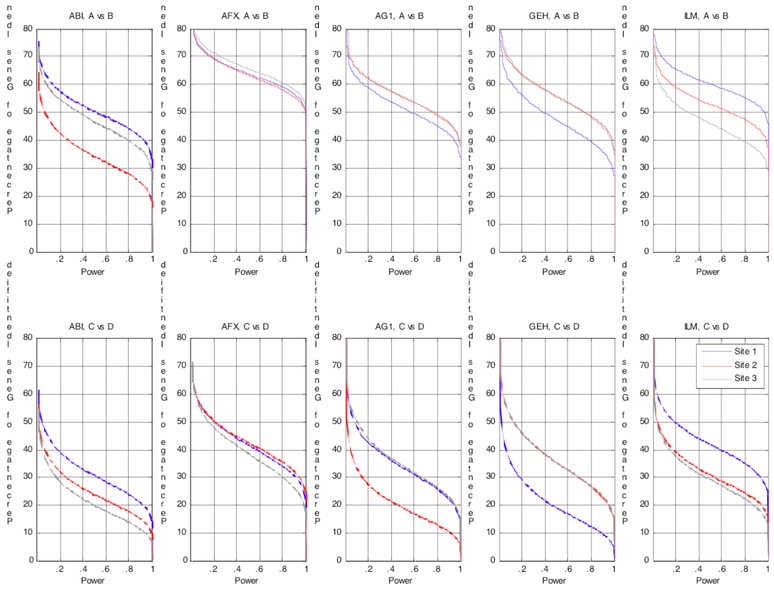
RESULTS

Fig. 2



PowerAssist used for comparison of normalization methods.
In this figure each microarray platform and data preprocessing method is represented in a separate panel. The data preprocessing methods highlighted in yellow for each platform represent the manufacturer's recommended method used in the MAQC study and represented in the main MAQC manuscript. The solid lines in each graph illustrate the discrimination power between the A and C samples while the dashed lines illustrate the discrimination power between the B and D samples. The effect size was calculated from the data, sample size = 5, α = 0.0001. The x-axis displays the calculated power and the y-axis displays the percentage of genes that have that power or greater. Figures from Shippy et al. 2006

Fig. 3

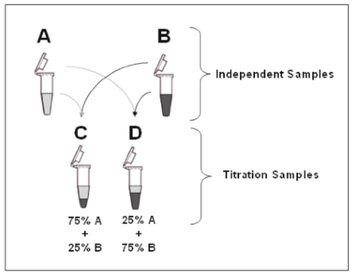


Laboratory-to-laboratory variability by platform.
Comparing groups A vs. B, and C vs. D cumulative plot of the proportion of genes achieving a desired power for a given sample size (n = 5), multiple test corrected a (0.05/number of tests) using the Bonferonni method, and a probe-by-probe S_{pooled} for each site and measure difference between groups at each site. The x-axis displays the calculated power and the y-axis displays the percentage of genes that have that power or greater. Figures from Shi et al. 2006

METHODOLOGY

The utility of PowerAssist was implemented for demonstrating differences in platforms, normalization and laboratories. The MAQC PowerAssist analysis is based on data from the 12,091 common genes set from across five commercially available microarray platforms: Applied Biosystems (ABI, Foster City, CA); Affymetrix (AFX, Santa Clara, CA); Agilent Technologies (AGL and AG1, Santa Clara, CA); GE Healthcare (GEH, Tempe, AZ); Illumina (ILM, San Diego, CA) . There were three laboratory sites per platform. Asuragen was selected as one of the Affymetrix sites. The study design was a mixture of two reference RNAs: Universal Human Reference RNA (UHRR) from Stratagene (La Jolla, CA) and Human Brain Reference RNA (HBRR) from Ambion (Austin, TX). The four pools included the two RNA samples as well as two mixtures of the original samples: Sample A) 100% UHRR; Sample B) 100% HBRR; Sample C) 75% UHRR: 25% HBRR; and Sample D) 25% UHRR: 75% HBRR (Fig. 4). Microarray data were normalized as specified in Shippy et al. (2006) and log2 transformed. After log2 transformation, the signal for all microarrays approximated a normal distribution (data not shown). We implemented a novel power analysis based on Warnes & Liu's method (http://www.bioconductor.org/repository/devel/vignette/ssize.pdf) with four key modifications. The average difference between groups was explicitly calculated for each probe, a pooled estimate of σ (s_{pooled}) was used, and plotting of experimentally derived power. In addition, the method was generalized so it could be used for all microarray platforms. The key component of this analysis is the generation of a cumulative plot of the proportion of genes achieving a desired power for a given sample size (n = 5), α (0.0001), and a probe-by-probe spooled for each site and measure difference between groups at each site. The results are expressed as the number of genes on the y-axis with a calculated power equal to or greater than a given power on the x-axis (Fig. 2 and 3). For each comparison, the power analyses for all test sites using the same microarray platform are grouped to display the extremes of test site performance.

MAQC Study Design Fig. 4



CONCLUSION

The Effect of Normalization as Measured by PowerAssist (Fig. 2)
We calculated cumulative power for A vs. C and B vs. D comparisons at each site, using the standard and alternative normalization methods. Cumulatively, each platform has similar power across the 12,091 genes, but for each of the platforms, at least one site shows a significant loss of apparent power due to increased technical noise. Figure 2 also illustrates the impact of normalization or data preprocessing methods on the ability to discriminate between the samples statistically. Interestingly, for all platforms with the exception of Illumina, the MAQC "standard" normalization or data preprocessing method performance was slightly inferior to the secondary method, especially in the power analysis. This result highlights the observation noted throughout this study that data processing methods determined to be optimal under one set of circumstances may not always prove appropriate under all conditions, particularly if primary assumptions underlying those data processing methods are violated.

The Effect of Laboratory Selection as Measured by PowerAssist (Fig. 3)
Power was calculated for two separate group comparisons at each site: sample A replicates vs. sample B replicates as well as sample C replicates vs. sample D replicates. For each comparison, the power analyses for all test sites using the same microarray platform are grouped to display the extremes of test site performance. As expected, the comparisons of the A vs. B replicates demonstrated greater average power than the comparisons of the C vs. D replicates, because the titrated samples can show at most a 3-fold change in gene abundance. Cumulatively each platform has similar power across the 12,091 set of common genes, but for each platform there was at least one site that showed a significant loss of power due to increased technical noise. For example, Applied Biosystems test site 2 had a lower labeling efficiency in sample A28 which impacted its performance in the power analysis relative to the other platforms for the A vs. B comparison. An increase in power was observed at Illumina site 1 compared to sites 2 and 3. These relative differences illustrate the importance of a detailed review of laboratory performance in microarray facilities

KEY FINDINGS

- Those laboratories with the greater average power for a platform are able to distinguish a greater number of genes that are truly different between groups. The key is consistency.
- Difference between samples has a considerable effect on power
- The selection of array platform impacts statistical power
- Normalized method can have a significant impact on power