Author's response to reviews

Title: Application of Affymetrix Array and Massively Parallel Signature Sequencing for Identification of Genes Involved in Prostate Cancer Progression

Authors:

Asa J Oudes (aoudes@systemsbiology.org)
Jared C Roach (jroach@systemsbiology.org)
Laura S Walashek (lwalashek@systemsbiology.org)
Lillian J Eichner (leichner@systemsbiology.org)
Lawrence D True (ltrue@u.washington.edu)
Robert L Vessella (vessella@u.washington.edu)
Alvin Y Liu (aliu@u.washington.edu)

Version: 2 Date: 9 June 2005

Author's response to reviews: see over
Dr.
1441 NE 34th Street
Seattle, WA 98103

June 6, 2005

Dear Editor,

We have revised our manuscript according to the reviewer’s suggestions. We have also addressed the reviewer’s concerns regarding technical aspects of the paper. For revisions that we felt were beyond the scope of the paper we have explained our position.

The manuscript has also been formatted according to the BMC cancer article template and references have been updated to the Endnote style “biomedcentral.ens” downloaded from BMC on 06/06/05.

With regard to classification of the article as a “Technical Advance”, the BMC cancer description of the article type is: “Technical advances should present a new experimental method, test or protocol for a laboratory experiment, a surgical intervention or a diagnostic procedure. The method described may either be completely new, or may offer a better version of an existing method. The article must describe a demonstrable advance on what is currently available. The method needs to have been well tested and ideally, but not necessarily, used in a way that proves its value.”, which we interpret to imply that the methods used were developed by the authors. The technology used in the reported work was not developed in our laboratory; therefore, we do not believe that it fits the criteria for a Technical Advance article. The definition of “Technical advance” notes that the method may be completely new or be a better version of an existing method. Again, our work does not conform to this classification either, because we did not modify the Affymetrix or MPSS protocols in any way. We believe that the manuscript would be most appropriately categorized as a Research article.

RE: Application of Affymetrix array and massively parallel signature sequencing for identification of genes involved in prostate cancer progression
Response to reviewers:

Response to Dr. Gerald’s review of our manuscript:

Major revisions

1. Regarding a quantitative analysis approach comparing Affymetrix signal with TPM counts. Affymetrix does not recommend the use of signal intensity for quantitation. Our comparison of Affymetrix signal intensity and the number of transcripts present (TPM) as counted by MPSS (fig. 2) indicates that there is too much variability in the relationship between signal intensity and TPM to assess a
correlation. Therefore, we did not pursue comparison of signal intensity vs. TPM further.

2. To address the reviewer’s feeling that we painted the Affymetrix platform in a negative light we have removed text which is skewed toward MPSS technology: “Additionally, there is visual support for a tpm = 0 intercept with a ProbeSet signal around 100, suggesting that signals much below 100 may not guarantee transcript presence.” We believe our comparison of the technologies was balanced and pointed out problems with both MPSS (zeros) and Affymetrix (signal saturation, probeset effects) platforms. To address questions about the accuracy of the MPSS protocol we reference Stolovitzky et al. [1], which provides a thorough analysis of the statistics (and thus limitations) of MPSS. The Stolovitzky et al. paper addresses each step in the MPSS process. We also added text in the results section: “Furthermore (5), measurement of outliers for either or both technologies can randomly obscure the correlation between TPM and Affymetrix signal.” and to account for outliers “…..and (4) used established statistical methods [1, 2] to limit spurious data contributed by outliers.”

3. We have revised the description of the methodology used to compile the unique gene lists with respect to clarity and rational. The new text regarding Affymetrix reads:
“We chose the cutoff points of ≥50 raw-signal and a call of present in both replicates for Affymetrix data because it was the most liberal method of determining if a gene was expressed by a cell-line. In establishing the cutoff points it was our aim to limit bias in our initial data analysis so that further analyses would have the maximum potential data to work with.”
The new text regarding MPSS reads:
“Like the Affymetrix data, the ≥1 tpm cutoff was chosen because it was the most liberal definition of gene expression for the MPSS data.”

Minor revisions

1. The RT-PCR experiments conducted with human tissue (cancer progression panel) were representative expression profiles derived from tissue samples of different individuals. We have added the text: “The expression of each gene was determined in 3 different biological replicates and was consistent among the replicates; figure 6 data is representative of the replicates.” to emphasize that the expression profiles are representative of 3 biological replicates.

Response to Dr. Li’s review of our manuscript:

Minor revisions

1. We feel that extensive protein level work is beyond the scope of the paper as the study was primarily an investigation of the ability of the transcriptome profiling platforms to identify all the genes expressed in an RNA sample. Our application of the data to human tissue demonstrates that the approach is able to identify genes that may be involved in prostate cancer progression from a pair of cell lines that have been extensively investigated.
2. Regarding RT-PCR internal controls. Since internal controls are necessary for quantitative comparison we’ve modified the language of the text to only indicate qualitative comparisons of RT-PCR experiments. In figures 4 and 5 each set of primers was screened against the same cDNA sample and each cDNA sample was prepared from the same amount of RNA. Therefore, given the high sensitivity of 35 cycles of PCR, if a gene is detected in both cDNA samples in any of the paired samples (for example fig. 5 column 1 and 2, row B shows expression in both cell types which is essentially acting as a control for cDNA quality) it is reasonable to make the qualitative assertion that the gene was detected or not among samples where only one of the paired samples shows signal. For figure 6, similar logic is applicable given that the genes PRG-3, ACA2SL, and PAK1 were detected in all tissue samples.

3. We have clarified the number of replicates performed for each transcriptome profiling experiment in the text. The reviewer is interested in the number of technical replicates that were performed for each sample. For Affymetrix data, technical replicates are not necessary as the correlation of genes expressed between two hybridizations of labeled target made from the same RNA is extremely high (>99.5% according to Affymetrix’s literature). Due to the cost of MPSS ($30,000 per sample) we were able to obtain data for only one sample of LNCaP and one sample of C4-2 RNA. Although the use of multiple tissue samples (biological replicates) was prohibitively expensive, the MPSS process samples millions of sequences (a form of technical replicate) We have chosen to analyze our MPSS data with the method of Stolovitzky et al. [1] which takes into account the number of sequences analyzed.


2. **Statistical Algorithms Description Document**
   [http://www.affymetrix.com/support/technical/whitepapers'affx]