Author's response to reviews

Title: A Systematic Comparison and Evaluation of High Density Exon Arrays and RNA-seq Technology used to Unravel the Peripheral Blood Transcriptome of Sickle Cell Disease.

Authors:

Nalini Raghavachari (nraghavachari@nhlbi.nih.gov)
Jennifer Barb (barbj@mail.nih.gov)
Yanqin Yang (Yanqin.yang@mail.nih.gov)
Poching Liu (pcliu@nhlbi.nih.gov)
Kimberly Woodhouse (kwoodhous@nhlbi.nih.gov)
Daniel Levy (levyd@nhlbi.nih.gov)
Chris O'donnell (odonnellc@nhlbi.nih.gov)
Peter J Munson (munson@mail.nih.gov)
Gregory Kato (gkato@nhlbi.nih.gov)

Version: 9 Date: 30 April 2012

Author's response to reviews: see over
We thank the reviewer again for the careful review of our revised manuscript. We have addressed and corrected the manuscript as per the reviewer’s valuable suggestions.

**Minor Essential Revisions**

1) On page 9, paragraph 2, line 4. The word "epsilon" should be replace by the corresponding greek letter.

   *As per the reviewer’s suggestion, we have replaced the word epsilon by the greek letter in the revised manuscript (page 9)*

2) On page 9, paragraph 2, last line. "(" -> "(".

   *As per the reviewer’s suggestion, we have corrected the last line in paragraph 2.

3) On page 9, paragraph 2, last line. The reference to the FDR method should be the following paper:

   *4. As suggested by the reviewer, we have added the Benjamini Yoav, Hochberg, Hosef reference for the FDR method in the revised manuscript.*

5. As requested by the Editor, We give below a short version of our response to the question from the last review on why you selected different cutoff values on fold changes for microarray and RNA-seq.

   *It is well known that microarray tends to underestimate the true fold change of expression. We observed this fact with RNA-seq reporting twice the log fold change compared with Exon array (our manuscript figures 4 and 6 clearly depict this compression of fold changes in array). In order to perform an unbiased comparative analysis of the two platforms and knowing that the microarrays have a compressed fold change, we chose two different cutoff values (requiring at least a 2-fold change in microarray data or a greater than 4-fold change in RNA-seq data) to select differentially expressed transcripts, guided by the observed compression in fold-change in the Exon array data.*
We used a very conservative and reasonable background cutoff level for microarray, on the RMA scale, corresponding to the level at which 50% of the probesets show the Affymetrix DABG (Detection Above Background) value of $p \leq 0.01$. For RNAseq we used a background cutoff of 1.0 RPKM, a level used as a cutoff by one of the pioneers of this method (Mortazavi et al. Nature Methods 2008), where they imply that at a sequencing capacity of 40 M mapped reads, transcript detection was robust at 1.0 RPKM and above for a typical 2-kilobase (kb) mRNA with a $P$ value $<10^{-16}$.