*Microarray Data Analysis*

Blood samples (1-3 mL) were collected in Tempus tubes (Applied Biosystems, CA, USA) and stored at -20°C. Whole blood RNA was processed and hybridized into Illumina Human WG-6 V4 beadchips (47,323 probes) and scanned on the Illumina Beadstation 500 as described ([46](#_ENREF_46), [47](#_ENREF_47)). The data is deposited in the NCBI Gene Expression Omnibus

Illumina GenomeStudio Version 4 software was used to subtract background and scale average signal intensity for each sample to the global average signal intensity of all samples. To perform further normalization and analyses it was used GeneSpring™ GX 7.3 software (Agilent Technologies) as previously described ([48-50](#_ENREF_48)). Transcripts were first selected if they were present in greater than 10% of all samples and then filtered to select the most variable probes (those that had a minimum of two fold expression change compared with the median intensity across all samples, in greater than 10% of all samples) ([49](#_ENREF_49)).

For the analyses we used several analytical tools and followed different approaches to obtain robust and validated data. First we used GeneSpring to perform the unsupervised and supervised analysis. Unsupervised clustering was performed on 140 samples that passed quality control to assess whether samples are grouping according to known factors and identify whether there are unknown subclasses within the data set. Supervised analysis was performed using statistical filtering and class comparisons to identify transcripts differentially expressed between study groups as previously described ([49](#_ENREF_49)). Non-parametric Wilcoxon-Mann-Whitney test for comparisons across the studied groups was applied with a p value of 0.05.

Second, we performed functional analyses of differentially expressed genes using modular analysis ([20](#_ENREF_20)). A set of 62 transcriptional modules was used as pre-existing framework. Gene expression levels were compared between children vaccinated and healthy controls on module-by-module basis. Module activity was measured by the percentage of transcripts derived from comparison and showing significant differences. Spot intensity indicates the percentage of differentially expressed transcripts for each module. Modules containing transcripts showing over expression or under expression when compared to healthy controls are shown in red and blue, respectively. Modular transcript content and annotations are publicly available online at <http://www.biir.net/public_wikis/module_annotation/V2_Trial_8_Modules>.

Third, to confirm and validate our module-based analyses findings, an alternative analytical approach was carried out by applying a linear mixed model. Briefly, gene expression values were log2 transformed and analyzed using a one-way ANOVA test with the software JMP Genomics 6.0 (SAS Institute, Cary, NC) using a linear mixed model. Genes with a p value < 0.01 were considered differentially expressed ([51](#_ENREF_51))