

## PhalanxBio TechNotes

---

# Comparison of Phalanx Biotech Human OneArray™ to MicroArray Quality Control Project Data

**Large-scale investigations, such as the FDA-led MicroArray Quality Control (MAQC) project, have clearly demonstrated the repeatability of various microarray platforms and the compatibilities between the different platforms. Additionally, the MAQC study also brought to light the importance of applying technical repeats to the quality of the microarray data, and the value of validation of data from one microarray platform with another microarray platform. Unfortunately, many researchers are still unable to follow the thorough MAQC-like experimentation due to the high costs traditionally associated with using commercially available microarrays. However, recent breakthroughs in high volume manufacture of the Phalanx Biotech OneArray™ now enable researchers to test more samples and important controls at costs much lower than other commercially available arrays - without sacrificing quality. Phalanx Biotech Human OneArray microarray was tested using the same target samples and experimental design as the MicroArray Quality Control project. Testing and data analysis shows Phalanx Biotech Human OneArray yields high quality data that is comparable to the commercial platforms included in the original MAQC study.**

## The MicroArray Quality Control Project

The explosion of genomic information in recent years has led to the development of numerous microarray platforms for gene expression identification. While research using microarrays has had a profound impact on clinical and academic research, no consensus has emerged to compare data from one platform to the next. A number of articles have been critical of the validity of microarray data in identifying differentially expressed genes<sup>1</sup>. They point out low levels of gene overlap and poor repeatability and reproducibility of different microarray platforms. Although there has been data generated to refute these claims<sup>2</sup>, for some time the microarray community lacked a set of reference data to evaluate the quality of products and experiments. The MicroArray Quality Control (MAQC) Consortium<sup>3</sup>, a community-wide effort led by FDA scientists with the cooperation and support of academic, government, and commercial researchers, was established to address these issues and validate the microarray as a research and diagnostic tool.

The MAQC project<sup>4</sup> has tested whole genome microarrays and alternative gene expression platforms at independent sites using a set of commercially available high-quality RNA samples. The RNA samples, Universal Human

Reference RNA provided by Stratagene and Human Brain Reference RNA from Ambion, were produced in batch specifically for the MAQC researchers. The project was reported in a series of articles in a dedicated issue of Nature Biotechnology in September 2006. The team gathered data from more than one thousand microarrays to show how standardized experimentation, reduction, and analysis can generate reproducible and comparable results among different platforms and laboratories.

## MAQC Methodology

The MAQC project compared the performance of seven DNA microarray platforms: Applied Biosystems, Affymetrix, Agilent, Eppendorf, GE Healthcare, Illumina,

and the National Cancer Institute<sup>5</sup>. In addition, differential gene expression was measured in alternative platforms including TaqMan Gene Expression Assays (Roche), StaRT-PCR® (Gene Express), and QuantiGene® (Panomics) and compared to microarray platforms. Four samples were tested; Sample A containing 100% Universal Human Reference RNA (UHRR from Stratagene), Sample B containing 100% Human Brain Reference RNA (HBRR from Ambion), Sample C containing 75% UHRR / 25% HBRR, and Sample D containing 25% UHRR / 75% HBRR. Because these RNAs are commercially available in the same batches as used by MAQC, experimenters and microarray producers now have an accessible standard to test platforms and protocols.

Each microarray platform in the project was tested at three different sites using five replicates of the four target

---

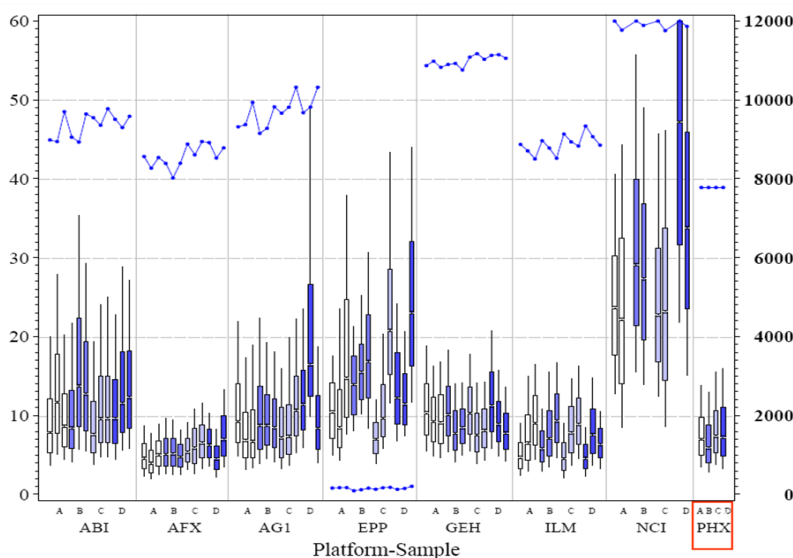
<sup>1</sup> See Tan, *et al.* (2003) *Nucleic acids Res.* **31**: 5676-5684; Miklos *et al.* (2004) *Nat. Biotechnol.* **22**:615-621; Frantz, S (2005) *Nat Rev Drug Discov.* **4**:362-363.

<sup>2</sup> See Petersen *et al.* (2005) *BMC Genomics* **6**:63; Dobbin *et al.* (2005) *Clin. Cancer Res.* **11**:565-572; Irizarry *et al.* (2005) *Nat. Methods* **2**:345-350; Larkin *et al.* (2005) *Nat. Methods* **2**:337-344; Kuo *et al.* (2006) *Nat. Biotechnol.* **24**: 832-840.

<sup>3</sup> <http://www.fda.gov/nctr/science/centers/toxicoinformatics/maqc/>

<sup>4</sup> MAQC Consortium (2006) *Nat Biotech* **24**(9): 1151-1161.

<sup>5</sup> NCI arrays were produced internally; probes designed and supplied by Operon.



**Figure 1:** Reproducibility of microarray platforms.

samples (A, B, C, and D)<sup>6</sup>. The platform supplier provided hybridization protocols and software for preliminary data reduction, using its own criteria for detection calls and signal magnitude. In an effort to further eliminate any interplatform differences due to protocols, software, or hardware, subsequent analysis compared magnitudes within each platform as ratios to normalize the data prior to comparison. Only genes that were detected in three of the five replicates were included in the calculations. All data analysis was performed by Expression Analysis Inc., a member of the MAQC consortium.

Much of the difficulty in comparing data from different microarray platforms has been attributed to differences in annotation. These differences were minimized by the MAQC project by first mapping the probe sequences from all the platforms to the RefSeq human mRNA database<sup>7</sup> and the AceView database<sup>8</sup>. This generated a common set of 15,615 RefSeq entries, which contained genes assayed by probes present on all of the high-density platforms. Furthermore, only probes that assayed a single, unique Entrez gene were analyzed in the study to adhere to a one-

probe-to-one-gene scheme<sup>9</sup>. This limited the final analyses to a common set of 12,091 probes mapping to 12,091 unique genes.

## Probe Mapping of Phalanx Biotech Human OneArray (HOA) With the MAQC Common Set

In order to test the compatibility of Human OneArray (HOA), the MAQC project experiments were reproduced using the Phalanx Biotech platform. The RNA samples used were from the same batch of RNA used in the MAQC project. Hybridizations were

performed using an internally developed agitation-based hybridization system. The samples were tested at a single site using five replicates of the four RNA samples. After hybridization, fluorescence was detected using an Axon GenePix microarray scanner and signals were measured using GenePix Pro software. Detection calls were assigned to features that had a signal-to-noise ratio > 5; each feature that met this criterion underwent regional background subtraction to calculate the signal value. Only features that were detected in 4 of 5 replicates and 3 of 4 samples were retained. Data were then normalized using the quantile method and submitted to Expression Analysis for processing with the existing MAQC project data. Data for 15,116 genes was submitted for analysis; of these genes 11,061 were common to the shared set of 12,091 in the original MAQC publications.

In the figures that follow, it is this set of 11,061 genes that are included in the analysis for HOA. With the exception of Figure 4, all graphics were provided by Expression Analysis. Figure 4 was generated by the authors using processed data from Expression Analysis and additional in-house data. Abbreviations used in the original publication are retained for ease of comparison: PHX = Phalanx Human OneArray, ABI = Applied Biosystems, AFX = Affymetrix, AG1 = Agilent Technologies, EPP = Eppendorf, GEH = GE Healthcare, ILM = Illumina, NCI = National Cancer Institute, TAQ = Roche, GEX = Gene

<sup>6</sup> The experimental design varied for platforms using two fluorophores or containing more than one array per slide. Alternative platforms tested each sample in triplicate or quadruplicate at a single test site.

<sup>7</sup> <http://www.ncbi.nlm.nih.gov/RefSeq>

<sup>8</sup> <http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly>

<sup>9</sup> Only the probe closest to the 3' end was used when a platform had multiple probes mapping to a single RefSeq entry.

Express, QGN = Panomics. Underscores indicate different test sites.

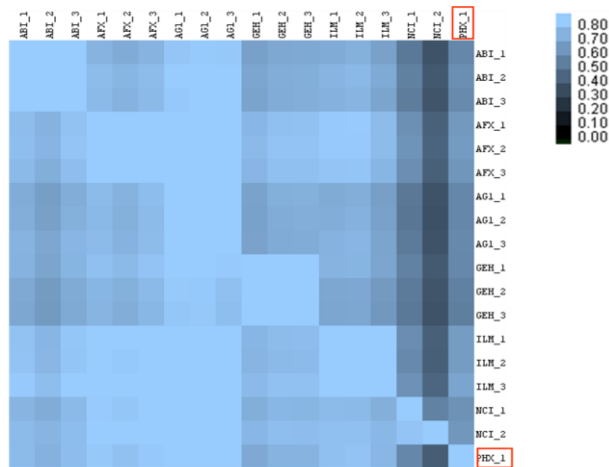
## HOA Data Repeatability and Reproducibility

An important characteristic of any microarray is array-to-array reproducibility. HOA meets strict quality control standards, starting with total spot number (TSN). Every HOA microarray is scanned to ensure features are present and aligned properly. In addition, samples are set aside each day of production and undergo probe immobilization and hybridization quality control. All of this is done to ensure your HOA microarray will provide the highest quality data possible.

The MAQC project evaluated repeatability and reproducibility by measuring the coefficient of variation (CV) of the signal magnitude values for intra-site replicates (Figure 1). The red box (PHX) indicates the addition of the HOA data. Whisker plots of the CV for the various platforms indicate median, interquartile, and 10<sup>th</sup> and 90<sup>th</sup> percentile (left ordinate). Coloring is consistent with the original publication: Sample A = white, Sample B = light blue, Sample C = light purple, Sample D = dark blue. The data were calculated for genes detected in at least four of five HOA replicates, and for three of five replicates for the remaining platforms. The graph shows comparable CV values across all commercially available platforms, including HOA. In addition, all HOA CV values were  $\leq 8\%$ , showing excellent array-to-array consistency. The number of genes detected by HOA is taken from a reduced common set (right ordinate). The experiments have also been reproduced at a second site (Yang Ming Medical College; data not shown), and the results show similar intersite reproducibility and good gene detection overlap.<sup>10</sup>

## Data Compatibility of HOA with MAQC Participants

One of the goals of the MAQC project was to evaluate interplatform data concordance. This can be assessed by determining whether differences between two samples (for example, an experimental sample and a control) can be detected on each platform. Figure 2 shows the concordance of differentially expressed genes between MAQC Sample A and Sample B.



**Figure 2:** Differential gene list overlap

Genes were identified as differentially expressed with  $\geq$  twofold change and  $P < 0.001$ . Lists of differentially expressed genes were generated for each test site and compared pair wise. The coloring in the matrix of Figure 2 indicates the percent overlap in differentially expressed genes between platforms. HOA (see red box) shows good concordance with other commercially available microarrays, indicated by the light colors in the PHX\_1 horizontal. (The colors in the PHX\_1 vertical indicate low overlap due to the smaller common set of genes available for HOA analysis.) The data provides assurance Phalanx Biotech production technology generates reliable, sensitive arrays for gene expression detection.

## Assessing Relative Accuracy Between HOA and Other MAQC Platforms

Instead of simply comparing lists of differentially expressed genes, agreement across platforms can also be measured by evaluating the magnitude of the ratios. This comparison is a good indication of whether a differential expression will yield similar ratios from one platform to the next. Phalanx Biotech's thermo jet spotting technology and rigorous quality control ensure every microarray will detect differential gene expression with the same certainty as the industry's leading platforms. The MAQC project measured the correlation of differential gene expression by analysis of log ratio ranking. The log ratio is calculated from the ratio of signal intensities for a probe between two samples

<sup>10</sup> Data is available upon request.

(in this analysis, Sample A and Sample B). In order to make this comparison, the log ratios for each platform are first ranked from most positive to most negative. The log ratio rank list is then compared between pairs of sites to measure the correlation. The matrix in Figure 3 shows the pair wise correlation of these rankings between sites. Lighter colors indicate good rank correlation, whereas darker colors show larger differences.

HOA shows excellent rank correlation with commercially available microarray platforms as well as the alternative gene detection methods (TAQ, GEX, and QGN). This comparison is also shown in Figure 4 by calculating the compression or expansion difference between site pairs. Perfect correlation of two arrays or test sites would result in identical numerical values for the log ratio (again, between Sample A and Sample B). A plot of these values would, ideally, result in a slope value of 1; compression or expansion occurs when the slope deviates from this ideal. The deviation is calculated as a percent difference between the two datasets. (For a more detailed explanation, please refer to the methods section of reference 4.) Darker values in Figure 4 indicate good agreement between sites or platforms. Lighter colors, either magenta (a negative difference in slope) or cyan (a positive difference in slope) indicate greater deviations from ideal correlation. Again, HOA performs comparatively to other platforms, showing little expansion or compression in comparison with other commercially available microarrays.

## Correlation of HOA with TaqMan®

The gold standard for sensitive gene identification is the use of real-time PCR products, such as TaqMan® (Roche Molecular Systems, Inc). A high correlation between these methods and microarrays, as was determined by the MAQC project, emphasizes the comparability of data gathered by different platforms. The MAQC project examined correlation between TaqMan data and various arrays, showing good correlations ( $r > 0.90$ ) for most commercially available arrays. Consistent with the MAQC data for other platforms (data not shown), Phalanx HOA shows good correlation ( $r = 0.89$ ) with TaqMan in identifying gene targets (Figure 5).

A qualitative analysis reveals most deviation from an ideal correlation occurs for small Sample B : Sample A ratios. Note that the other region of deviation is at large B : A ratios. The systematic compression pattern is present when *PhalanxBio TechNotes*:

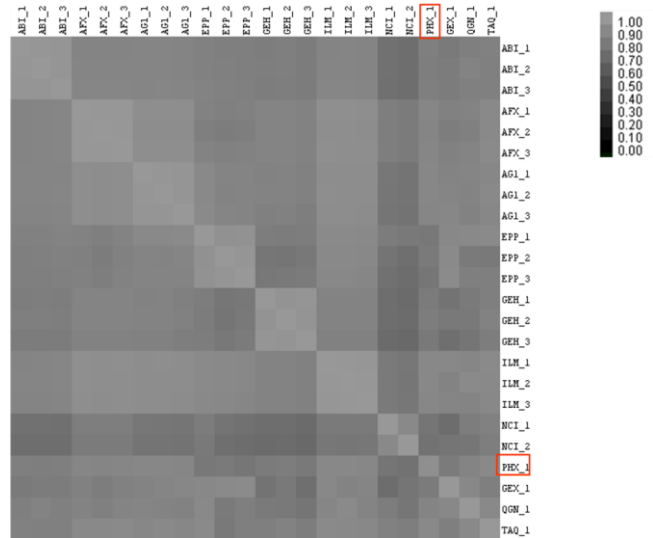


Figure 3: Log ratio rank correlation

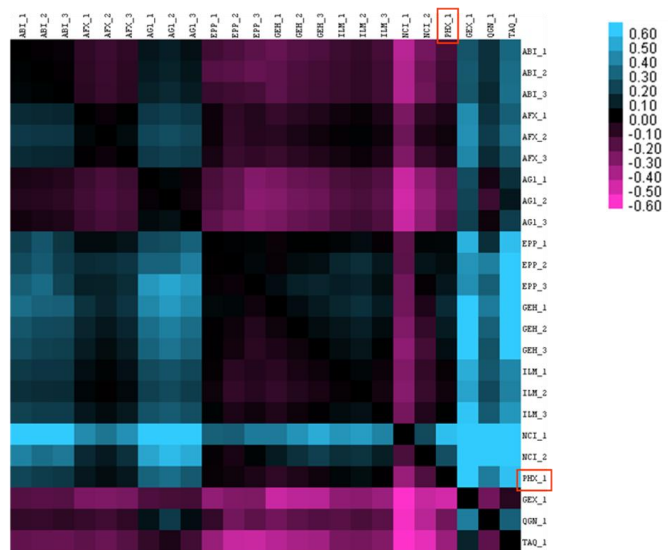
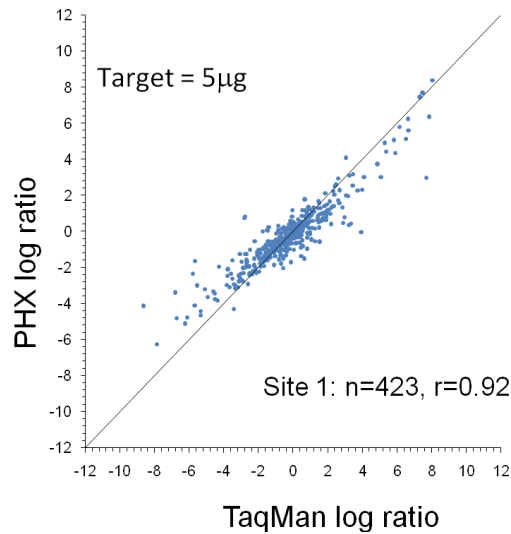
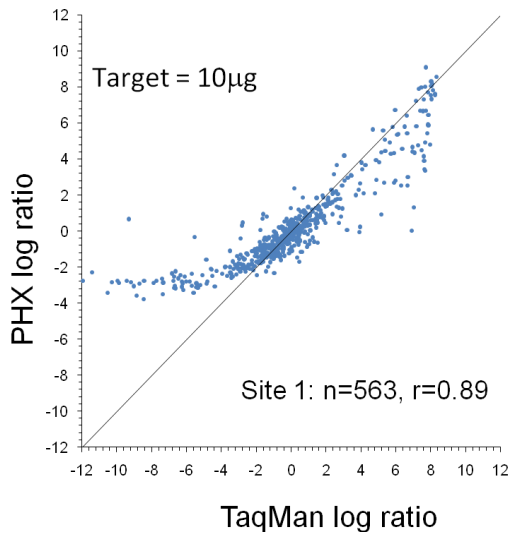


Figure 4: Site pair compression or expansion

comparing all microarray platforms with TaqMan. The compression can be explained by the inherent sensitivity of real-time PCR experiments. This apparent sensitivity issue can be resolved by lowering the mass of target sample in the experiment from 10  $\mu\text{g}$ , as established in the original Phalanx Biotech MAQC protocol to 5  $\mu\text{g}$ ; the correlation increases ( $r = 0.92$ ) and the data shows a qualitative improvement in the region of lower B:A ratios (Figure 5).

## A New Era in Microarray Research

The MAQC Consortium has succeeded in providing benchmark data by which microarrays can be evaluated. In the first phase of study, the MAQC project has shown that



**Figure 5:** Correlation with TaqMan

differential gene expression is comparable across platforms. Furthermore, Phalanx Biotech Human OneArray, although not originally included in the MAQC study, has been tested using the same experiments as the MAQC project. The results of these tests indicate that HOA shows excellent repeatability, with coefficients of variation consistently  $\leq 8\%$ . HOA's ability to recognize differentially expressed genes is comparable to other platforms included in the MAQC project. Finally, HOA shows excellent correlation ( $r = 0.92$ ) with real-time PCR gene expression identification. Taken as a whole, HOA exhibits excellent performance when compared to the industry's leaders in microarray research.

Breakthroughs end eras and begin new ones. Phalanx Biotech's innovative manufacturing process provides scientists with the same high-quality microarrays they have come to expect, but at a much more affordable price. In addition to standing up well to the standards of the MAQC project microarrays, HOA probe design also shows comparable interrogation power when aligned to the human genome (See PhalanxBio TechNotes: Whole Genome Microarray Probe Content Interoperability). The thermo jet technology generates unparalleled spot uniformity; the printing mechanism also ensures these spots are virtually free of cross-contamination. The array production facility couples exceptional array-to-array consistency with unprecedented volume and capacity capabilities, producing as many as 2,000 high quality arrays per day. Each microarray is designed and manufactured using a standard and familiar format for easy adaptation to your existing instrumentation and workflow. These technological breakthroughs make Phalanx Biotech OneArray the best

*PhalanxBio TechNotes:*

*Comparison of OneArray™ to MAQC Project Data*

value in microarrays, enabling you to do more experiments, more repeats, and more time points that allow you to understand more in the process.

## Learn More

For more information about Human OneArray and other whole genome microarrays and services, visit [OneArray.com](http://OneArray.com), email us at [info@phalanxbiotech.com](mailto:info@phalanxbiotech.com), or contact Phalanx Biotech directly at 650-320-8669.

*Human OneArray™ and OneArray™ are trademarks of PhalanxBiotech Group, Inc. in the United States and in other countries.*

*All trademarks and copyrights used in this report belong to their respective owners and are the sole property of their respective owners.*

*Last updated October 2007*

*© 2005 - 2007 Phalanx Biotech Group. All rights reserved.*

**Phalanx Biotech Group, Inc.**

1301 Shoreway Road, Suite 160  
Belmont, CA 94002 USA  
<http://www.OneArray.com>  
[info@phalanxbiotech.com](mailto:info@phalanxbiotech.com)  
PH: 650-320-8669  
FAX: 650-508-9889