# An Analysis of Matched mRNA Measurements from the High-Density HU6800 and cDNA Microarrays

**Winston Patrick Kuo, [1,2,3] Tor-Kristian Jenssen,[4] Ju Han Kim, [1,5]**
**Isaac S. Kohane, [1,3] Lucila Ohno-Machado [2,3]**

**Keywords**: cDNA, oligonucleotide, DNA microarray, cross-platform comparison, cluster

## 1 Introduction.

Microarray technology has enabled researchers to measure expression of thousands of genes simultaneously, but there are still important challenges in the analysis of results. Recent publications indicate researchers are working more or less independently with the microarray technology of their choice, using a variety of different formats, procedures, data collection and organizational strategies in their studies. Can the scientific community share their data for collaborative research and analyze data from different technologies? We recently reported mRNA measurements from two large-scale independent studies measuring gene expression in cancer cell lines, one from spotted cDNA microarrays and the other from low-density HU6800 microarrays[1]. Affymetrix recently introduced high-density microarrays, as well as new algorithms for their analysis. We investigated whether the high-density oligonucleotide arrays produce measurements that are more correlated with those from cDNA microarrays than do the low-density oligonucleotide arrays.

## 2 Methods and Results.

In this study, we refine our previous observations by an analysis of high-density HU6800 microarray with a cDNA microarray measuring gene expression in cancer cell lines. mRNA expression was measured in the NCI60 cancer cell lines. The details of the gene expression data sets have been described and are publicly available[2,3]. 53 of the 60 cancer cell lines from the standard panel of cancer cell lines from the National Cancer Institute as well as the genes were matched for both high-density HU6800 and cDNA microarrays.

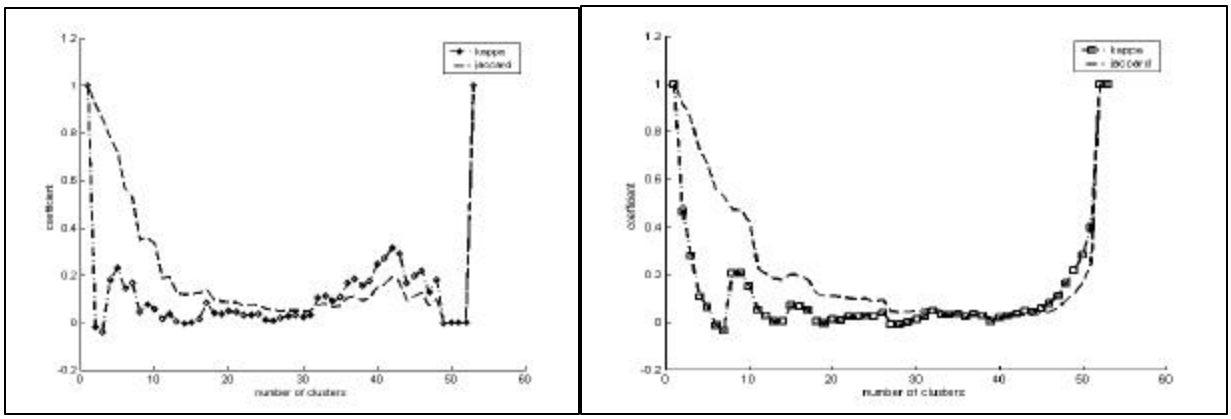
The Pearson and Spearman rank-order correlation coefficients were calculated for genes, cell lines, and across all 153,435 matched pairs of measurements. Table 1 illustrates the correlation between RAT2N and oligonucleotide measurements, as well as CH2D and oligonucleotide measurements. We also compared the two generations of oligonucleotide arrays and the overall average Pearson and Spearman correlation of 0.8902 and 0.7818, respectively.

Hierarchical clustering was performed using Euclidean distance as a similarity measurement with average linkage heuristic. We computed Jaccard and Kappa coefficients as a measure of the similarity of clusters. Figure 1 shows that there was some concordance between the clusters resulting from the oligonucleotide data and those resulting from the cDNA data.

[1] Children's Hospital Informatics Program, Harvard Medical School, Boston, MA 02115. E-mail: **wpkuo@mit.edu**
[2] Decision System's Group, Harvard Medical School, Boston, MA 02115.
[3] Division of Health Science and Technology, Harvard University and MIT, Cambridge, MA 02139
[4] Department of Computer and Informational Science, Norwegian University of Science and Technology, Trondheim, Norway.
[5] Seoul Nat'l Univ. Biomedical Informatics, Seoul Nat'l Univ. School of Medicine, Seoul 110-799, Korea.

Table 1. Correlation and standard Deviation of high-density oligonucleotide and cDNA microarray measurements

| | Pearson | | | Spearman | | |
|---|---|---|---|---|---|---|
| cDNA | Overall | Cell lines | Genes | Overall | Cell lines | Genes |
| RAT2N | 0.0373 | 0.0569 (0.045) | 0.3356(0.3364) | 0.2454 | 0.2463(0.0586) | 0.2944(0.2991) |
| CH2D | 0.3197 | 0.3339(0.0475) | 0.2551(0.2869) | 0.4551 | 0.4675(0.0361) | 0.2209(0.2477) |




(a) (b)

Figure 1. Concordance of clusters from hierarchal clustering measured by Jaccard and Kappa coefficient between (a) CH2D and Average Difference, (b) RAT2N and Average Difference

## 3 Discussion.

The results indicate sub-optimal correlation between the results obtained from high-density oligonucleotide arrays and those from spotted cDNA microarrays, though the results were slightly better than the low-density comparison. Improving consistency and reproducibility across technologies are very important pre-requisites for enabling effective sharing of gene expression data.

## Acknowledgment

WPK supported in part by the grant "Research Training in Health Informatics" by the National Library of Medicine, 5T15 LM07092-07.

## References

[1] Kuo WP, Jenssen TK, Butte AJ, Ohno-Machado L, Kohane IS. Analysis of matched mRNA measurements from two different microarray technologies. *Bioinformatics.* 2002:In press.

[2] Staunton JE, Slonim DK, Coller HA, et al. Chemosensitivity prediction by transcriptional profiling. *Proc Natl Acad Sci U S A.* Sep 11 2001;98(19):10787-10792.

[3] Ross DT, Scherf U, Eisen MB, et al. Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet.* Mar 2000;24(3):227-235.