

The Now and Later of Oligo GEArray® HybPlate Express Hybridization Timing

Introduction

Multi-tasking research scientists have often inquired how changing the length of the hybridization time would affect microarray results. With SuperArray's Oligo GEArray® HybPlate Express protocol and proper background correction and normalization, hybridization length does not influence larger fold changes in gene expression. However, overnight or longer hybridization times increase the statistical significance of smaller fold-changes in gene expression. Typical concerns with the overnight hybridization option on the HybPlate include increased background signal, decreased call present percentages, and altered fold-change results. On the other hand, the 3-hour HybPlate hybridization time raises questions about the consistency of the results. This Technical Note demonstrates that these Oligo GEArray® performance parameters are not affected by the length of hybridization.

Percent Present Call

RNA samples from biological triplicates of MCF7 cells treated with or without 5-fluorouracil (5-FU) were characterized with the Human p53 Signaling Pathway Oligo GEArray® HybPlate Format (EHS-027) using various hybridization times (3, 6, 18, or 42 hours) in the GEArray® Express ThermoShaker. Visual examination of the microarray photographs shows a major increase in overall signal intensity between 3 and 6 hours of hybridization time with a corresponding increase in the non-specific background. Because the present/absent threshold is calculated from the absolute, non-specific background intensity, the signal intensity increase does not produce an increase in the number of genes called present at each hybridization time (Table 1). Therefore, hybridization time does not influence the number of detectable genes.

Table 1: Increasing Hybridization Time Does Not Increase Percent Present Call. The average percent present calls from replication experiments and their standard deviations at different hybridization times are listed, as determined by the GEArray® Expression Analysis Suite.

	Hybridization Time (h)			
	3	6	18	42
Percent Present Call	58	56	61	61
Standard Deviation	1.4	2.3	2.0	1.9

Signal Intensity: The Raw and the Cooked

The median intensity value (the actual data point directly in the middle of the data set) is a better metric for overall intensity on a microarray than the mean (or average). Unlike the mean, the magnitude of the data points falling outside the linear dynamic range of the system (high or low) does not affect the median value. In this experiment, the average median raw intensity value of the replicate arrays and the variance across those replicates both increase between 3 and 6 hours and then plateau at longer hybridization times (Figure 1).

To correct the raw data and account for variation between replicate arrays, we used our standard minimum value background subtraction and interquartile normalization recommendations. Once adjusted, there is no significant difference between the hybridization times in their intensity value medians across the replicates or their standard deviations (Figure 1). Therefore, after correcting for differences in absolute, non-specific background intensities and systemic variation between the replicates, it can be concluded that hybridization time does not affect the overall intensity on the microarrays.

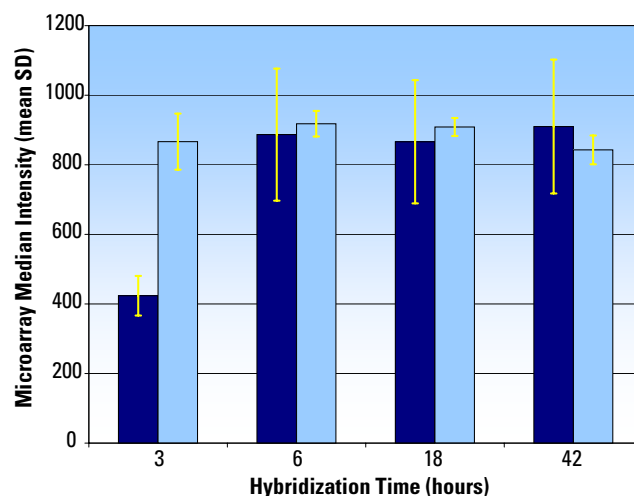


Figure 1: Adjusted Signal Intensities Plateau Before the 3-Hour Hybridization Time Point. Shown are the average median intensities for the raw (dark blue) and background-subtracted, normalized (light blue) data from each hybridization time. The error bars represent one standard deviation. The adjusted intensity values are multiplied by 1000 to place them on the same scale as the raw data.

Fold Changes in Gene Expression

Changes in gene expression, reported as adjusted data ratios, are the ultimate result of a microarray experiment. The fold-changes in the expression of p53 pathway genes induced by 5-FU were

calculated for each hybridization time point and compared. Excellent agreement is observed between the average fold-changes for each gene across the triplicate arrays determined at the 3-hour, and both the 18- and 42-hour hybridization time points, with correlation coefficients of 0.945 and 0.937, respectively (Figure 2, blue points and black trend line). These results indicate that the hybridization time also does not affect the gene expression profile on the microarray.

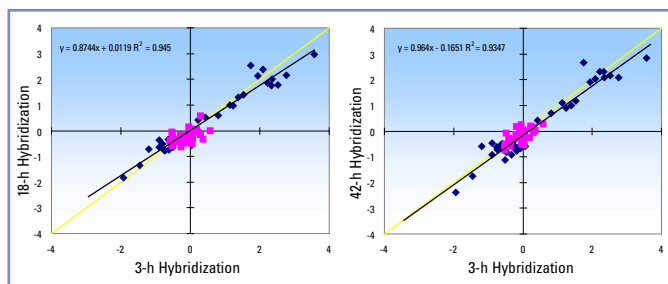


Figure 2: Fold-Changes in Gene Expression Determined at Different Hybridization Times Are Comparable. Fold-change ratios (\log_2 transformed) from the 3-h and either the 18-h (A) or 42-h (B) hybridization times are compared in scatter plots. Only data for “present” genes ($n=70, 68$) are shown. Blue symbols are statistically significant ($p < 0.05$) fold-changes for at least one hybridization time ($n=34, 40$), while pink ones are not. The trend line, equation and correlation coefficient are shown for the statistically significantly data (blue points) while the yellow line indicates the position of a perfect 1:1 agreement.

Volcano Plot

The use of replicate samples permits the application of a t-test to determine the statistical significance for every fold-change ratio. Plotting the average fold-changes against their p-value generates a “volcano plot” that easily visualizes the most significant gene expression changes (Figure 3). Zone A delineates genes with large fold-changes (>2 -fold) and statistical significance ($p < 0.05$). The genes in Zone B are also characterized as statistically significant ($p < 0.05$), but exhibit smaller expression changes (≤ 2 -fold). The genes in these zones are worthy of real-time PCR validation and further characterization. Zone C contains genes with unchanged expression levels, while zone D identifies genes with large average fold-changes but unusually high levels of variance in expression. Both of these groups are generally of minor interest to investigators.

The number of genes in Zone A shows little change with increasing hybridization time (Table 2). Indeed, most of the genes (12) are identified at all 3 hybridization times. In contrast, the length of the microarray hybridization substantially increases the number of genes in Zone B. Collectively examining all of the genes that show statistically significant fold-changes

($p < 0.05$, Zones A and B) indicates that longer hybridization times improve the overall reliability of detecting small changes in gene expression.

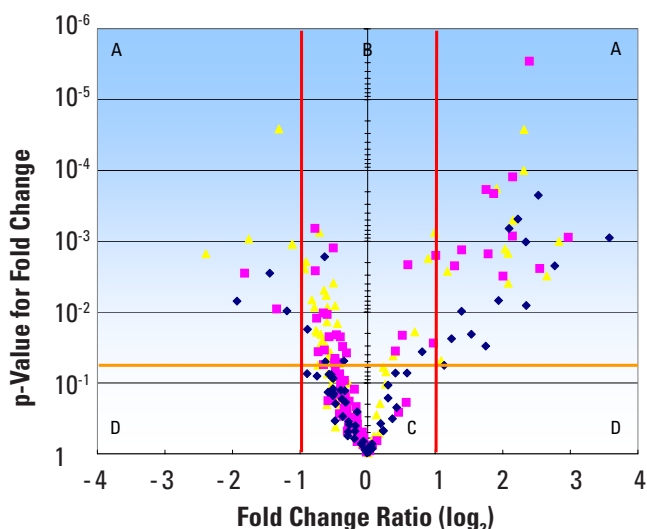


Figure 3: Volcano Plots Readily Define the Statistically Significant Fold-Changes in Gene Expression. For the 3-, 18- and 42-h (blue, pink, yellow symbols, respectively) hybridization times, the \log_2 fold-change is plotted against its p-value to produce a “volcano plot”. The higher the position, the more significant the gene’s fold-change while genes plotted farther from the central axis have larger changes. Fold-change (2-fold = red lines) and significance ($p < 0.05$ = gold line) thresholds define four zones on this graph (A, B, C and D).

Conclusion

Hybridization time has little effect on the results from the Oligo GEArray® HybPlate Express protocol for most experimental applications. We recommend that investigators interested in studying small changes in gene expression use an overnight hybridization, while those only concerned with larger changes may use the 3-hour hybridization time that permits a quicker execution of the experiment. In order to make accurate comparisons across arrays, it is important to remember that the same hybridization time must be used for all microarrays within a single experiment.

Table 2: Increasing Hybridization Times Improves the Statistical Significance of Fold-Change Results. The number of genes in zones A and/or B of Figure 4 for each hybridization time are counted and reported here.

	Hybridization Time (h)		
	3	18	42
Median p-value	0.147	0.078	0.042
# Zones A & B Genes, $p < 0.05$	19	32	41
# Zone A Genes, $p < 0.05$ & > 2 -fold	15	14	16
# Zone B Genes, $p < 0.05$ & ≤ 2 -fold	4	18	25