Research

Assessment of the relationship between signal intensities and transcript concentration for Affymetrix GeneChip® arrays
Eugene Chudin*†, Randal Walker†‡, Alan Kosaka†‡, Sue X Wu†‡, Douglas Rabert‡, Thomas K Chang‡ and Dirk E Kreder§

Addresses: *Illumina, San Diego, CA 92129, USA. ‡Roche Bioscience, Palo Alto, CA 94304, USA. §Abgenix Inc., Fremont, CA 94555, USA.
†These authors contributed equally to this work
Correspondence: Eugene Chudin. E-mail: echudin@illumina.com

Abstract

Background: Affymetrix microarrays have become increasingly popular in gene-expression studies; however, limitations of the technology have not been well established for commercially available arrays. The hybridization signal has been shown to be proportional to actual transcript concentration for specialized arrays containing hundreds of distinct probe pairs per gene. Additionally, the technology has been described as capable of distinguishing concentration levels within a factor of 2, and of detecting transcript frequencies as low as 1 in 2,000,000. Using commercially available arrays, we assessed these representations directly through a series of ‘spike-in’ hybridizations involving four prokaryotic transcripts in the absence and presence of fixed eukaryotic background. The contribution of probe-target interactions to the mismatch signal was quantified under various analyte concentrations.

Results: A linear relationship between transcript abundance and signal was consistently observed between 1 pM and 10 pM transcripts. The signal ceased to be linear above the 10 pM level and commenced saturating around the 100 pM level. The 0.1 pM transcripts were virtually undetectable in the presence of eukaryotic background. Our measurements show that preponderance of the signal for mismatch probes derives from interactions with the target transcripts.

Conclusions: Landmark studies outlining an observed linear relationship between signal and transcript concentration were carried out under highly specialized conditions and may not extend to commercially available arrays under routine operating conditions. Additionally, alternative metrics that are not based on the difference in the signal of members of a probe pair may further improve the quantitative utility of the Affymetrix GeneChip® array.

Background

Even though the DNA microarray is still an emerging technology, its usefulness as a profiling tool is well established. Affymetrix GeneChip® arrays enable the concurrent assessment of expression levels for thousands of genes in a single experiment. At the molecular level, however, the microarray experiment is a challenging biophysical problem that is extremely dependent on probe-target kinetics, specificity and design. Among the principal sources of variability are the nonspecific interactions due to combinatorial complexity of the genome, the thermodynamic equivalence of probes, the accuracy and spatial uniformity of probe synthesis and the preparation, amplification and fractionation of cDNA and cRNA.
Results

Biotinylated target cRNA from four prokaryotic genes - *lys*, *dap*, *phe* and *thr* from *Bacillus subtilis* - was collectively hybridized to Test2 arrays with and without 0.05 g/l of complex cRNA background generated from human brain total RNA. This background concentration roughly corresponds to 150 nM (0.05 g/l / (330 g/(mol x nucleotide) x 1,000 nucleotides)). The target transcript frequency can then be simply computed by dividing transcript concentration by total RNA concentration. *B. subtilis* genes were selected on the basis of their designation as standardization controls. In accordance with the probe array design, genes encoded on the Test2 array contain three probe sets corresponding to the 5', middle and 3' regions. Thus, a total of 12 distinct controlled target-transcript measurements were obtained per hybridization. Precision in target-transcript concentration was increased through a series of tenfold dilutions from a presynthesized in vitro transcribed mixture. Each combination of dilution and background was replicated four times (array availability) to account for and assess variability.

Figures 1 and 2 show that the plot of intensity versus target concentration has the typical sigmoidal shape encountered in chemical kinetics. The error bars reflect the range of observed values, which increases for measurements obtained subsequent to the lag phase and is greatest for the 100 pM level.

Table 1 supplements these graphs by distinguishing values at low transcript concentrations. Table 2 depicts averaged, pairwise ADI ratios across consecutive dilution levels (four replicates per dilution; 16 pairs in total) after normalization. Given the sensitivity of the ratio to small absolute values, the variance of the ADI ratios in the 1-0.1 pM range was comparable to that of the mean, and was highly influenced by the outlier-removal algorithm. It is worth noting that normalization did not have a significant effect owing to the extremely uniform conditions of this study.

A cursory review of Table 2 will establish that under both hybridization conditions (target transcripts versus target transcripts + cRNA background), a tenfold reduction in the quantity of target transcript was consistently reflected by the ADI in the 1-10 pM range. Additionally, the window of target-transcript concentration under which proportionality was preserved varied in accordance with background cRNA content. In the cRNA sample comprised solely of target transcripts and hybridization controls (*bioB*, *bioC*, *bioD*, *cre* and *dap*), approximate ratio equality was witnessed between 0.1-10 pM. Under conditions more akin to a standard assay (brain cRNA background), the range of proportionality shifted to 1-100 pM. In both cases, the ADI peaks and subsequently plateaus above 100 pM (Figures 1, 2 and Table 1 for PheX_3) as the rate of increase in the MM signal becomes equivalent to or greater than that of the PM signal. In the veritable absence of nonspecific interactions, background noise is bounded by the absolute value of the 0.1 pM MM signal. Analogously, the nonspecific interaction component of the MM signal is bounded by the observed increase in the 0.1 pM signal in the presence of complex cRNA. Consequently, by comparing both sides of Table 1, we conclude...
Figure 1
PM, MM, and ADI signals without cRNA background. PM, red; MM, blue; ADI, green. (a) PheX₃; (b) ThrX₅; (c) DapX₅; (d) LysX₅.
Figure 2
PM, MM, and ADI signals with cRNA background. PM, red; MM, blue; ADI, green. (a) PheX_3; (b) ThrX_5; (c) DapX_M; (d) LysX_5.
that less than 20% of the MM signal at the 100 pM level reflects nonspecific binding. The major component of the MM signal (66-80%) therefore constitutes interaction with the intended PM target transcript. Absolute values for the same transcript concentration differ significantly, despite averaging over 20 probe pairs, which shows the influence of selected probe sequences (Table 1).

Addition of a complex cRNA background elicited a profound effect on observed intensities, especially at the lower range of concentration. The absolute values of PM and MM increased for the 0.1 pM level whereas they decreased for the 10 pM and 100 pM levels (Table 1). Surprisingly, the decrease in absolute values for 100 pM resulted in the improved sensitivity of the 100-10 pM average ADI ratio (factor of 6 versus factor of 3). In addition, the signal-to-noise ratio (when defined as ADI/MM) decreased dramatically in the 0.1-10 pM range whereas the standard error of the 10-1 pM ADI ratios increased (Table 1). For the 0.1 pM target concentration, all 12 target transcripts were detectable in the absence of brain cRNA background. However, the presence of complex cRNA essentially rendered them undetectable. It is evident that the complex cRNA background had a nonlinear effect on the ADI, with up to a fourfold difference in the average ADIs observed for the same quantity of transcript (Figure 3).

To study the effect of hybridization duration, the 0.1, 1 and 10 pM hybridizations were repeated in the presence of fixed eukaryotic background with the hybridization time doubled from 16 to 32 hours. A marginal increase in the average ADI was observed (by around 10%) although such increase was within the observed 16-hour range. Additionally, the increase in signal for the 10 pM level remained consistently less than half of that obtained in the absence of complex cRNA. The increased hybridization time did result in partial detection of prokaryotic controls at 0.1 pM (5 out of 12 transcripts).

Finally, to examine reproducibility, a scatter plot of the log of gene ADIs for two distinct hybridizations (10 pM transcript concentration in the presence of fixed eukaryotic background) was produced (Figure 4a). About 10% of the

| Table 1 |

| Prokaryotic targets only | Fixed eukaryotic background |
|--------------------------|-----------------------------|
| ADI | PM | MM | ADI/MM | ADI | PM | MM | ADI/MM |
| DapX_M | 10 nM | 19919.75 | 35677.5 | 15757.8 | 1.264104 | 19202 | 33809.95 | 14608.25 | 1.314463 |
| | 1 nM | 19021 | 28886.78 | 9865.59 | 1.927983 | 18687.25 | 26908.9 | 8221.955 | 2.272848 |
| | 100 pM | 15123.25 | 19903.23 | 4786.035 | 3.16386 | 8037 | 11747.13 | 2860.618 | 2.809533 |
| | 10 pM | 4450.5 | 5747.473 | 1296.938 | 3.43138 | 1307.25 | 2457.7 | 1256.453 | 1.040429 |
| | 1 pM | 475.25 | 902.4025 | 528.7788 | 0.898818 | 152.75 | 908.6613 | 960.415 | 0.159046 |
| | 0.1 pM | 53 | 555.3025 | 500.4675 | 0.105894 | 61.75 | 789.3088 | 963.1375 | 0.064113 |
| LysX_5 | 10 nM | 22589.25 | 32681.15 | 10091.97 | 2.238332 | 22421.25 | 31153.3 | 8732.253 | 2.56736 |
| | 1 nM | 18822.5 | 25673.68 | 6851.14 | 2.747309 | 16307.25 | 21212.53 | 4905.173 | 3.3245 |
| | 100 pM | 14646.25 | 18662.85 | 4016.605 | 3.646521 | 6483.75 | 1887.703 | 816.94 | 1.87664 |
| | 10 pM | 4277 | 5444.548 | 1167.745 | 3.663383 | 970.25 | 1887.703 | 816.94 | 1.87664 |
| | 1 pM | 393.75 | 942.0425 | 512.4125 | 0.786868 | 102.25 | 764.4975 | 662.2375 | 0.154401 |
| | 0.1 pM | 47 | 550.8263 | 500.82 | 0.093859 | 11.5 | 683.4175 | 681.955 | 0.016863 |
| PheX_3 | 10 nM | 16341 | 39198.25 | 22857.5 | 0.714908 | 16647.5 | 38609.2 | 21961.6 | 0.750828 |
| | 1 nM | 17515.3 | 35193.4 | 17677.73 | 0.990822 | 19560.25 | 32962.05 | 13401.65 | 1.45954 |
| | 100 pM | 19859.25 | 31405.65 | 12039.89 | 1.649474 | 13262.75 | 18867.58 | 5088.875 | 2.602624 |
| | 10 pM | 9842 | 14063.93 | 3625.85 | 2.714473 | 2404.75 | 4105.543 | 1517.398 | 1.584785 |
| | 1 pM | 1242.25 | 2125.22 | 789.3375 | 1.573464 | 230.75 | 1162.27 | 884.2738 | 0.260994 |
| | 0.1 pM | 156.25 | 683.4025 | 527.1663 | 0.29649 | 2.25 | 881.4363 | 859.7225 | 0.002617 |
| ThrX_5 | 10 nM | 21092.25 | 38711.13 | 17618.83 | 1.197148 | 21321.75 | 37907.63 | 16585.88 | 1.285536 |
| | 1 nM | 20334.75 | 32740.03 | 12405.41 | 1.639205 | 20344.25 | 30814.2 | 10469.95 | 1.943109 |
| | 100 pM | 16992.5 | 24897.65 | 7905.4 | 2.149453 | 10744.75 | 14774.15 | 4029.59 | 2.666462 |
| | 10 pM | 5183.75 | 7391.858 | 2208.143 | 2.347713 | 1716 | 3056.805 | 1341.128 | 1.27952 |
| | 1 pM | 567.5 | 1181.16 | 613.6475 | 0.92502 | 211.5 | 1026.143 | 840.3225 | 0.251689 |
| | 0.1 pM | 61.5 | 577.555 | 517.4338 | 0.118841 | 29.5 | 847.0175 | 822.4388 | 0.035869 |
The observed linear range of the ADI is narrower than that previously reported [2], and potentially inflated from what might typically be encountered given the reduction in procedural variability achieved through the comparatively uniform conditions of the experiments presented here. It is worth noting, however, that the linear range initially reported [2] was obtained using a custom array containing probe sets with more than 500 PM/MM probe pairs per gene. Consequently, it is not surprising that even with technological advances in chip manufacturing over the past four years, we are still unable to reproduce fully the linear range of the ADI using commercial chips with 20 or fewer PM/MM probe pairs per probe set.

The premise that subtraction of the MM from PM essentially serves to extract constructively the nonspecific components (cross-hybridization, noise, and so on) of the signal common to both should be reconsidered in the light of the results obtained in this study. It appears to perform inadequately for low and high transcript concentrations alike, although the basis for the failure differs. At high concentrations (above 100 pM level), the rate of increase of the MM signal can eclipse that of the PM signal (Figures 1d, 2d, Table 1 for PheX_3) resulting in an eventual decline in the ADI (Figures 1d, 2d for PheX_3). For example, the average ADI of the ThrX_3 probe set at 1 nM was approximately 47%
higher than at 10 nM (Table 1). Probe-target binding can be characterized by sigmoidal dose-response curves with different parameters. While both PM and MM seem to have comparable maximum slopes, due to lower target affinity, the linear response phase of the MM signal is shifted from that of the PM signal (it occurs at higher concentrations). Consequently, the ADI decreases for high target concentrations where the slope of the PM sigmoidal has tapered. Reports that the ADI saturates simply as a result of saturation of both the MM and PM signals [2] do not seem to be supported by our results. It should be noted that part of the observed saturation effect is attributable to limited scanner sensitivity, which might reflect a limitation of our installation. Scanner-related saturation is functionally related to absolute intensity, which, in turn, is dependent on underlying probe kinetics. Examination of individual probe signals suggests that scanner effects are likely to be more pronounced at higher target concentrations (1-10 nM), as many of the probes have reached the maximum detectable intensity (around 46,000). However, saturation in the 10-100 pM range is most likely to be dominated by the kinetic properties of the PM and MM probes.

At the lower end of transcript concentration, it seems that eukaryotic background affects the PM and MM probes nonsymmetrically, generally resulting in a decrease in the ADI. Notably, the DapX_M probe set in Table 1 has a positive ADI for 0.1 pM level even though the MM signal is greater than the PM signal before outlier removal. Additionally, the same outlier-removal algorithm rendered the small 0.1 pM ADI for the PheX_3 nonexistent, suggesting that simple heuristics can fail in a low signal-to-noise environment. Outlier classification is perhaps best addressed using a functional (kinetic), as opposed to a purely inferential, profile constructed from a large repository of experimental data. Given the large number of experiments used to derive the ADI, such an approach would seem eminently feasible. In lieu of an ideal functional profile, the use of alternative heuristics may be promising. For example, PM + MM - background (PM + MM also mentioned in [5]), while being less sensitive in the 1-10 pM range, was more sensitive for higher ranges and more robust for 0.1-1 pM.

Given that a significant number of genes of biological interest have transcript frequencies at or below 1 pM [7], the commercial usefulness of the system is constrained by the minimum abundance level that is reliably detectable. A current limit of 1 in 2 x 10^6 transcripts, that is, around 0.075 pM or 1 in 7 cells) has been reported [4]. Indeed, the Affymetrix GeneChip® array was able to detect this very low transcript level in the absence of eukaryotic background. However, after addition of cRNA background, transcripts at the 0.1 pM concentration became essentially undetectable for all 12 independent transcripts, whereas 1 pM transcripts remained robustly detected. Our results are consistent with a recent study [8] showing a range of detection between 1 in

![Figure 4](http://genomebiology.com/2001/3/1/research/0005.7)

Correlation between hybridization results obtained for the same RNA sample. (a) Log space plot; (b) linear space plot. Uniform ‘factor of 2’ region (see text) in the log space plot does not have an intuitive physical counterpart in linear space. It underestimates variability at the low end of ADIs.
It is possible to argue that post-hybridization amplification would improve detection, but obviously at the expense of potentially saturating expression levels of more abundant genes. Perhaps scanning images before and after amplification could maximize detection without suffering saturation penalties. Longer hybridization cycles seem to be a viable alternative, as these enabled partial detection of transcripts (about 5 out of 12) at the 0.1 pM level without significantly affecting high-end intensities. Given the current inability to localize transcripts with similar abundances on different arrays, care should be taken to ensure a sufficient overall number of replicates to obtain a small standard error for 0.1 pM level measurements. Also, new methods of mRNA amplification [9] that robustly increase the hybridization concentration of rare transcripts might improve detectability but could simultaneously distort the underlying RNA expression profile. The addition of eukaryotic background had a profound effect on the properties of thermodynamic equilibrium of probe-target binding. The decrease in 10 pM and 100 pM intensities suggests the presence of complex, stable interactions (that is, cross-target binding) which persist subsequent to the 32-hour hybridization. A similar hypothesis was previously proposed [4] where probe-probe interactions were implicated. We believe that cross-target interactions are more important in our case as the saturation level was not affected by cRNA background. The marginal increase in the signal after 32 hours of hybridization, though, suggests that

300,000 and 1 in 50,000 (0.5-1.5 pM). It is possible to argue that post-hybridization amplification would improve detection, but obviously at the expense of potentially saturating expression levels of more abundant genes. Perhaps scanning images before and after amplification could maximize detection without suffering saturation penalties. Longer hybridization cycles seem to be a viable alternative, as these enabled partial detection of transcripts (about 5 out of 12) at the 0.1 pM level without significantly affecting high-end intensities. Given the current inability to localize transcripts with similar abundances on different arrays, care should be taken to ensure a sufficient overall number of replicates to obtain a small standard error for 0.1 pM level measurements. Also, new methods of mRNA amplification [9] that robustly increase the hybridization concentration of rare transcripts might improve detectability but could simultaneously distort the underlying RNA expression profile. The addition of eukaryotic background had a profound effect on the properties of thermodynamic equilibrium of probe-target binding. The decrease in 10 pM and 100 pM intensities suggests the presence of complex, stable interactions (that is, cross-target binding) which persist subsequent to the 32-hour hybridization. A similar hypothesis was previously proposed [4] where probe-probe interactions were implicated. We believe that cross-target interactions are more important in our case as the saturation level was not affected by cRNA background. The marginal increase in the signal after 32 hours of hybridization, though, suggests that
thermodynamic equilibrium was not quite reached in 16 hours. The increase in signal for 0.1 pM is obviously due to increased background noise. The net result of cRNA background is that it has a nonlinear effect on the ADIs (Figure 3), which is impossible to compensate for using linear methods.

**Conclusions**

Our findings suggest that while high-density microarrays are a convenient way of monitoring thousands of genes simultaneously, increased care is needed in the design of experiments and scrutiny of the predictive utility of the numerical measure used to represent gene expression. An increase in the number of replicates is preferable to reliance upon magnitudes of fold changes, as the latter is not always linearly related to target concentration and is extremely variable for low transcript concentrations. In light of the fact that the MM signal predominantly characterizes interactions with the target transcript, a different heuristic was used to weight the MM signal accordingly for each probe pair might further improve the quality of Affymetrix GeneChip® array data. Perhaps such an approach underlies the announced supplanting of historical algorithms in the impending GeneChip 5.0 release. Finally, longer hybridization times can improve partial detection of transcripts expressed at very low levels.

**Materials and methods**

**Preparation of labeled targets for hybridization**

The Test2 array (Affymetrix, Santa Clara, CA) contains probes corresponding to commonly expressed genes from the human, mouse, rat and yeast genomes, along with several prokaryotic control genes. For each of these genes, probes derived from the 5′, middle and 3′ portions of the genes are arrayed. The prokaryotic controls used in this study contain engineered poly(A)+ tails and are available through the American Type Culture Collection [10] (dapB, ATCC 87486; lysA, ATCC 87482; pheB, ATCC 87483; and thrC, ATCC 87484). Methods for preparing cRNA and subsequent steps leading to hybridization and scanning of the Test2 arrays were provided by the manufacturer. Briefly, amplified and purified prokaryotic control vectors were linearized at the 5′ end using XhoI and purified by gel electrophoresis. Poly(A)+ cDNA was transcribed in vitro by incorporation of biotinylated CTP and UTP (Enzo Diagnostics, Farmingdale, NY) using a BioArray High Yield RNA Transcript Labeling kit according to the manufacturer’s instructions. The labeled cRNA was purified using RNeasy spin columns (Qiagen, Chatsworth, CA), followed by DNase I treatment and a second round of RNeasy spin-column purification. The integrity of all labeled and purified transcript was checked by denaturing gel electrophoresis. Each of the four transcripts, dap, lys, phe and thr, were pooled, and fragmented in 60°C total human brain RNA (Clonetech, Palo Alto, CA) and 10 μg/ml streptavidin R-phycocerythrin (Molecular Probes, Portland, OR) for 30 min at 25°C. After staining, the arrays were rinsed with non-stringent wash buffer at 25°C and scanned using a GeneArray confocal scanner (Agilent, Palo Alto, CA).

To evaluate the influence of hybridization duration on intensity values, the 0.1 pM, 1 pM and 10 pM prokaryotic target concentrations were repeated in experiments where the hybridization time was extended to 32 h with all other conditions left unchanged.

**Data analysis**

Data analysis was performed using the Affymetrix GeneChip array 4.0 software. Four chips representing one target concentration level (that is, 10 pM) were paired with four chips from the following target concentration level (that is, 1 pM) resulting in 16 total pairs. For each pair, the data was multiplied by a normalization factor (calculated with a mask excluding prokaryotic target transcripts) to make the average signal for both arrays equivalent. Fold changes were subsequently averaged across different pairs, excluding fold changes involving negative ADIs. All normalization factors were within 10% of 1. Statistical curve fitting was carried out using GraphPad Prism 2.01. Sigmoidal dose-response curves were fitted to PM and
MM data, allowing all four parameters to be variable. The ADI was fitted with a 100 segment cubic spline.

**Acknowledgements**

We thank Michael Liebman for his support of this project and Paul Cannon, Al Baecker and Debra Cockayne for their helpful suggestions.

**References**

1. Forman JE, Walton ID, Stern D, Rava RP, Trulson MO: Thermodynamics of duplex formation and mismatch discrimination on photolithographically synthesized oligonucleotide arrays. ACS Symp Ser 1998, 682:206-228.

2. Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL: Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat Biotechnol 1996, 14:1675-1680.

3. Wodicka L, Dong H, Mittmann M, Ho MH, Lockhart DJ: Genomewide expression monitoring in *Saccharomyces cerevisiae*. Nat Biotechnol 1997, 15:1359-1367.

4. Lipshutz RJ, Fodor SPA, Gingeras TR, Lockhart DJ: High density synthetic oligonucleotide arrays. Nat Genet 1999, 21:20-24.

5. Schadt EE, Cheng L, Cheng S, Wong WH: Analyzing high-density oligonucleotide gene expression array data. J Cell Biochem 1999, 80:192-202.

6. Mazzola LT, Frank CW, Fordor SPA, Mosher C, Lar tus R, Henderson E: Discrimination of DNA hybridization using chemical force microscopy. Biophys J 1999, 76:2922-2933.

7. Mahadevappa M, Warrington JA: A high-density probe array sample preparation method using 10- to 100-fold fewer cells. Nat Biotechnol 1999, 17:1134-1136.

8. Hill AA, Hunter CP, Tsung BT, Tucker-Kellogg G, Brown EL: Genomic analysis of gene expression in *C. elegans*. Science 2000, 290:809-812.

9. Baugh LR, Hill AA, Brown EL, Huter CP: Quantitative analysis of mRNA amplification by *in vitro* transcription. Nucleic Acids Res 2001, 29:1-2.

10. American Type Culture Collection [http://www.atcc.org]