

Microarray Platform Comparisons – Commentary on Barnes et al. (2005)

April 19, 2006

Abstract	2
Introduction.....	3
Experimental Design.....	3
Conclusions Drawn by Barnes et al.	4
Re-Analysis of Barnes Data.....	5
Exploratory Data Analysis	5
Statistical Data Analysis	6
Classification Analysis.....	9
Pathway Analysis.....	11
Considerations When Performing Platform Comparisons.....	13
Conclusions.....	14
References.....	16

Abstract

An extended analysis of the tissue dilution data generated by Barnes et al (1) on the Affymetrix GeneChip® and Illumina® BeadArray platforms is presented in this white paper. Although the study is statistically underpowered, the authors, Barnes et al., conclude that the two platforms show comparable performance. However, their data, displayed in Figure 1, demonstrates that the Affymetrix platform shows significantly greater reproducibility than the Illumina platform. A simple exploratory approach to the data highlights a clear difference in the ability of the two platforms to capture the underlying biology with the Affymetrix platform being significantly more effective. Statistical modeling of the data reveals clear differences in the signal/noise ratios for each platform; the Affymetrix and Illumina platforms exhibit signal/noise ratios of 8.19:1 and 1.20:1 respectively. As a direct result, the Affymetrix platform yielded 10 times more genes showing differential expression at a statistical significance of <0.05 when comparing the 5% Placenta (the lowest dilution) to 100% PBMC. Pathway analysis confirmed that the detected genes reflected the underlying biology and were not an artifact of the analysis approach. As a direct measure of the utility of each platform as a classification tool, the ability to reproducibly build classifiers with the data sets was tested. When presented with the most stringent sample set, the Affymetrix platform displayed a classification accuracy of 77% while the Illumina platform performed no better than random. With the understood statistical power limitation of this study design, these analyses suggest significant performance advantages to using the Affymetrix GeneChip system when studying real-life biological systems.

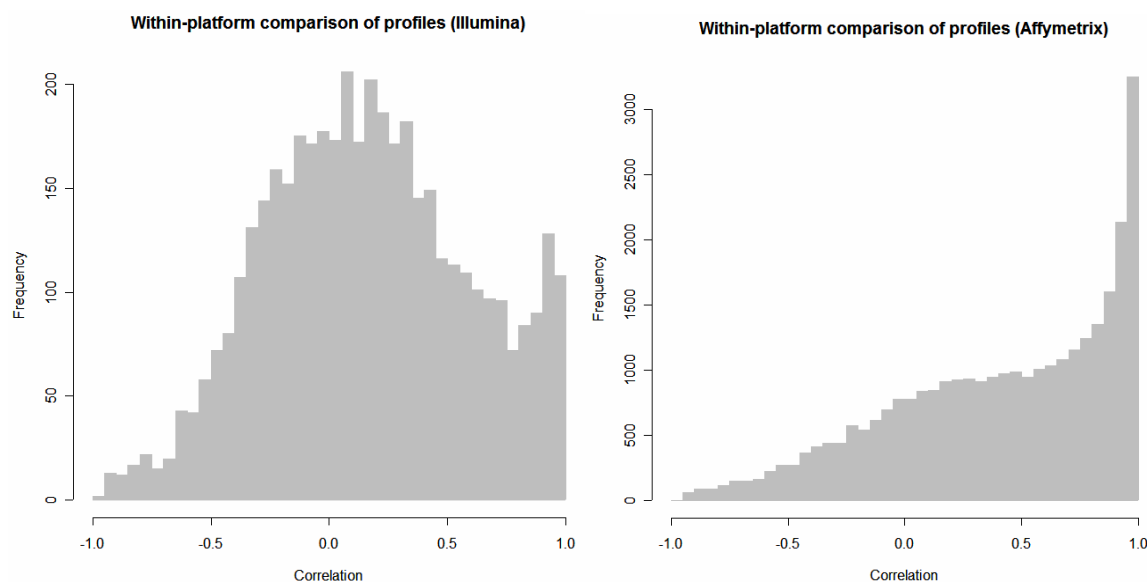


Figure 1. Within-platform correlation of signal estimates derived from two technical replicates. Bar graphs displaying the distribution of correlation coefficients for all probes on the Illumina and Affymetrix platforms respectively. A correlation coefficient closer to a value of one indicates higher agreement of expression values in the two technical replicates run. Barnes et al, supplemental data (<http://microarray.cu-genome.org/platformCompare/supplementary-figures.pdf>).

Introduction

With continued advances in microarray technologies, whole genome expression profiling has become a powerful research tool for elucidating genetic interactions, understanding gene pathway changes and identifying genomic biomarkers correlated with a broad assortment of biological outcomes; such as patient response to drug treatments, disease classification and toxicological liabilities.

For new users, there are a number of microarray platforms available to choose from, both commercial and home-brew. The primary advantage of selecting a commercial platform is the elimination of the manufacturing expertise required, quality uncertainties among a variety of other challenges associated with home-brew arrays thus permitting the user to obtain robust, reproducible results and focus solely on the biology of interest.

Once a decision has been made to purchase a commercial system, it is commonly of interest to understand which commercial system best meets the needs of the planned study. These needs will include but are not limited to the total content of the array, the availability of gene annotation for the content, the availability of standardized protocols and reagents, the maturity of data analysis options and support, platform performance, cost, overall ease-of-use and general acceptance. Although we will discuss many of these considerations, platform performance will be the primary focus of this white paper.

There have been a handful of platform comparison studies published in peer reviewed journals (1-4) specifically intended to evaluate the performance of the major commercial microarray platforms available. These can be helpful to the uninitiated user in understanding the performance expectation with the various platforms under consideration but can also be misleading dependent on how the studies are designed and how the data is processed and analyzed.

We will discuss the challenges involved in performing these types of comparisons and the pitfalls awaiting the unsuspecting user using as an example a recent publication by Barnes et al (1) who report a head-to-head evaluation of the Affymetrix GeneChip[®] and Illumina[®] BeadArray platforms. As the choice of the platform can be extremely critical in the long-term success of a research program, we also provide some suggestions for users to consider when making these decisions.

Experimental Design.

The up-front design of a study is critical to a researcher's ability to draw sound conclusions and will be guided by the types of questions being asked. There are a variety of approaches that can be employed when cross-platform comparisons are being performed, each having advantages and disadvantages in terms of the amount of information provided versus overall cost of the study.

As a basis for comparison, Barnes et al. chose a simple dilution design to assess platform variability and reproducibility. For this approach, the group started with peripheral blood mononuclear cell (PBMC) total RNA and titrated in predetermined amounts of placenta total RNA to create a total of six samples representing a dilution series from 0% through to 100% placental derived messages. Two technical replicates from each of the six dilutions were run on each platform for a total of 24 samples (12 on each platform). The resulting dataset was used to evaluate the ability of each platform to identify genes that are differentially expressed between the two tissues, PBMC versus placenta. For the purposes of their study, Barnes et al used the Human Genome U133 Plus 2.0 Array from Affymetrix (Affymetrix, Inc., Santa

Clara, CA) and the Sentrix HumanRef-8 Expression BeadChip from Illumina (Illumina, Inc., San Diego, CA).

This type of approach is useful in understanding the general performance of each platform; while theoretically possible with this type of dataset, not knowing *a priori* the concentration of any one transcript within the starting sample, an accurate assessment of platform sensitivity and specificity becomes challenging.

Of greater importance is the necessity of ensuring that a study is appropriately powered. That is, is it designed with sufficient numbers of replicates to provide enough statistical power to support the conclusions made? As a general rule, a minimum of three replicates is necessary to effectively capture system variability with a requirement for larger numbers of replicates as the system becomes noisier. Similarly, an increased number of replicates are required to detect smaller gene expression changes. With only two replicates for each dilution, the Barnes et al study is not ideal from this respect.

Conclusions Drawn by Barnes et al.

While at first glance, the presentation of data within the Barnes et al. publication suggests comparable performance between the Affymetrix and Illumina platforms, there are some key concepts to understanding the shortcomings to this conclusion. Consideration of the Barnes et al data in light of these key concepts will lead to key distinctions in the conclusions made by the authors.

- 1) With the experimental design chosen by Barnes et al, efforts must be taken to focus the analysis on probes and probe sets that are expected to detect the same transcript or pool of transcripts. This concept is echoed by Barnes et al. when they comment “the precise location of the probe on the genome affects the measurements to a substantial degree, such that two probes which do not map to the same location cannot be assumed to be measuring the same thing”.

Barnes et al did take steps to map probes and focus the analysis on those that map to identical regions of the genome as defined by ‘refGene’ and ‘knownGene’ annotation. This is an essential step necessary due to the experimental design chosen, however, due to genome complexity it cannot be guaranteed that probes designed on each platform to detect the same gene are measuring the same transcript and as a result some discordance is likely to still exist.

- 2) Larger expression changes (as mimicked by the 100% PBMC compared to 100% Placenta samples) are easier to detect than smaller changes (as mimicked by the 100% PBMC compared to a 5% Placenta samples), therefore, important differences in platform performance are expected to be more apparent when focusing on small changes.

Barnes et al focused their analysis on genes that show large expression differences between the two sample sets, the result of which is that neither platform is substantially challenged. The conclusion that the platforms display comparable performance is, thus, based upon large changes that are relatively easily detected. Therefore, important differences between the two different platforms observable in a real-life biological application would not be expected to surface in their analysis.

- 3) Platform accuracy cannot be effectively assessed with the experimental design chosen by the researchers, as described above, which, along with precision and bias, is a critical specification necessary for understanding the overall performance of any measuring device including a microarray.

Because the experimental design does not allow for assessment of platform accuracy, the researchers focus their analysis on platform precision, or the ability of the platforms to reproducibly report the same value when challenged with identical samples.

This assessment is based upon the two replicates run on each platform and is unfortunately only presented in the supplemental figures available on the web (at NAR Online or at <http://microarray.cu-genome.org/platformCompare>). For readers who care to delve into this additional information, they will discover expression estimate correlation graphs that clearly differentiate the two platforms at the level of reproducibility as shown in Figure 1.

In these graphs, all data points would show a correlation of one in a perfect system with the Affymetrix platform more closely resembling this ideal outcome. In fact, Barnes et al do come to similar conclusions when they comment “We first found that within-platform ‘reproducibility’ was substantially lower on the Illumina array than for Affymetrix ...”.

Re-Analysis of Barnes Data

Exploratory Data Analysis

Due to the limitation of the type of analysis conducted on the data set by Barnes et al, it was necessary for Affymetrix to carry out additional, more stringent analysis on the same data set to better understand if there are any performance differences between the platforms. The data files for all microarrays are available from GEO under accession number GSE3077

While it is of general interest to compare platforms at the level of the expression estimates, it is perhaps more telling to compare the ability of each platform to provide information about the biological question at hand. In the case of Barnes et al study, this would constitute the ability of the two platforms to detect transcripts unique to the placental sample. For this analysis, we will focus on the 100% PBMC compared to 5% placenta samples as this comparison provides the greatest detection challenge.

A simple scatter plot in which the average expression estimate of 100% PBMC replicates is plotted on the y-axis and the average expression estimate of 5% placenta replicates are plotted on the x-axis provides a visual method of identifying placenta specific genes. In these plots, genes common to both samples and expressed at similar levels will fall on the central diagonal while genes uniquely expressed in placenta are expected to fall near the x-axis.

The sensitivity of the platform can be roughly estimated by asking the simple question “how many genes expressed solely in placenta are detected by each platform?” As a first pass towards answering this question, approximately equivalent regions of each graph were selected and the captured data points were counted. This estimated that the Affymetrix platform detected over twice the number of placental genes compared to Illumina (125 versus 260) as illustrated in Figure 2.

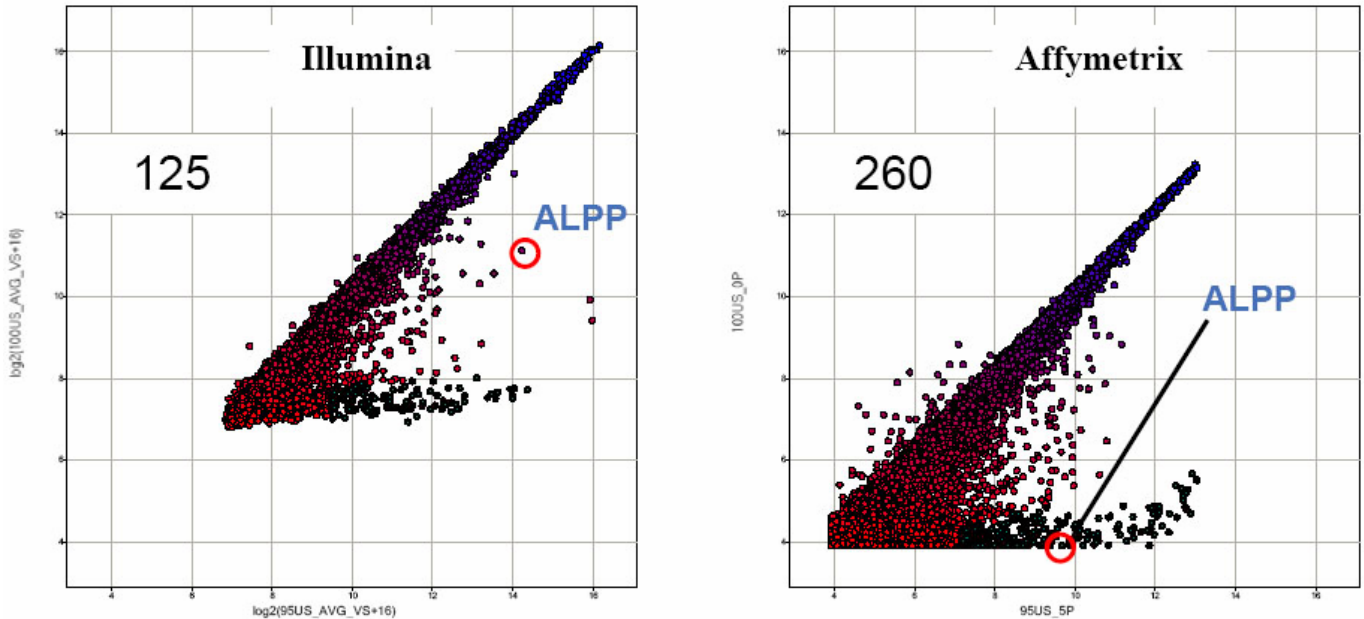


Figure 2. Scatter graphs of average expression estimates for the Illumina and Affymetrix datasets. Expression values from technical replicates were averaged and values from 100% PBMC and 5% placenta dilutions were plotted on the y-axis and x-axis respectively. Approximately equal regions of each graph were selected (captured points shown in grey) to estimate the numbers of placental genes detected on each platform. The position of placental alkaline phosphatase (ALPP), a placental marker, is highlighted in each figure.

It was of interest to confirm that genes assumed to be placenta specific from the analysis contained known placental markers. The position of placental alkaline phosphatase (ALPP), a key placenta marker, was identified in the respective scatter plots. As expected, this was contained within the cloud of genes originally selected as being placenta specific in the Affymetrix dataset. In contrast, in the Illumina dataset ALPP was not part of the chosen placenta specific gene set and was not clearly distinguished from genes common to each sample. This indicates that the Affymetrix platform is more sensitive at detecting a small change in expression values and that the small changes being detected are truly reflective of the underlying biology.

Statistical Data Analysis

Statistical methods are extremely powerful approaches for understanding the performance and quality of a microarray experiment. A common method chosen to both identify the experimental sources of variation and characterize the signal to noise ratio in a system is a model based approach such as Analysis of Variance (ANOVA).

The application of a Multi-Way ANOVA that includes in the same statistical model, the many factors that may contribute to total experimental variation, enables the variation in the observed expression values to be fully dissected and better understood. For example, in addition to the treatment (the tissue mixture in the case study described here) it is helpful to understand the contribution of Quality Control (QC) factors such as the date of hybridization, operator, scanner, etc. to the overall experimental variation observed.

Unfortunately, since Barnes et al. did not provide any of the relevant QC information, these important considerations could not be investigated individually.

However, the total technical variability can be separated from the biological variability to gain an understanding of the total noise within each platform. To do so, a One-Way ANOVA using Partek® Genomics Suite v6.2 (Partek, Inc. St. Louis, MO, USA) was applied to the two data sets and the biological effect (Mix) was separated from the noise (Error). Again, the analysis was focused on the 5% placenta versus 100% PBMC samples as this “treatment” is the only mixture included in the experimental design that reflects a typical biological microarray experiment, where the differential expression levels studied are likely subtle in the context of a biological treatment such as different doses of a compound or time points of biological treatment.

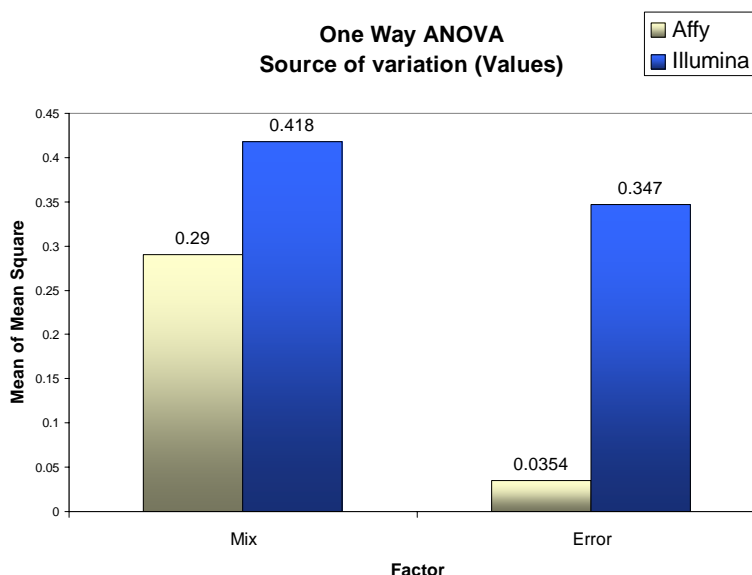
The output of an ANOVA includes two values for each probe or probe set; a p-value and a mean square for each factor tested. It is the mean of the mean square for each factor that is used to quantify the contribution of that factor to the total experimental variation. Any remaining variation in each data point (in this case the expression values) that is not attributable to one of the factors being tested in the statistical model is referred to as a residual. As a final step, the residuals are then summarized as the error (in our example this is variation in the expression values that are not due to the underlying biology) as illustrated in Figure 3.

For the model used in this example, the lower the error’s mean of mean square, the better the biology explains the variability in the system. Or in other words, the higher the observed error, the higher the noise in the system. Figure 3a shows a significantly lower (~10-fold) error value for the Affymetrix platform in comparison to the Illumina platform.

To identify the signal to noise ratio, the values of all factors were divided by the relevant error values so that the error of both platforms have a common value of 1. As a consequence, the biological variability effectively captured by each system can now be directly compared.

Figure 3b shows that, in the Affymetrix platform, the mix are more than 8-fold higher than the error while, in the Illumina platform, the biological effect is only 1.2-fold higher than the error. These findings indicate that Illumina data demonstrates a very low signal to noise ratio which is indirectly supported by its low reproducibility reported by Barnes et al. In summary, more biological differences between the samples are being captured on the Affymetrix platform. A brief analysis of top ranked differential expressed genes identified by the Affymetrix platform showed them to be of placental origin supporting this conclusion (data not shown).

a.



b.

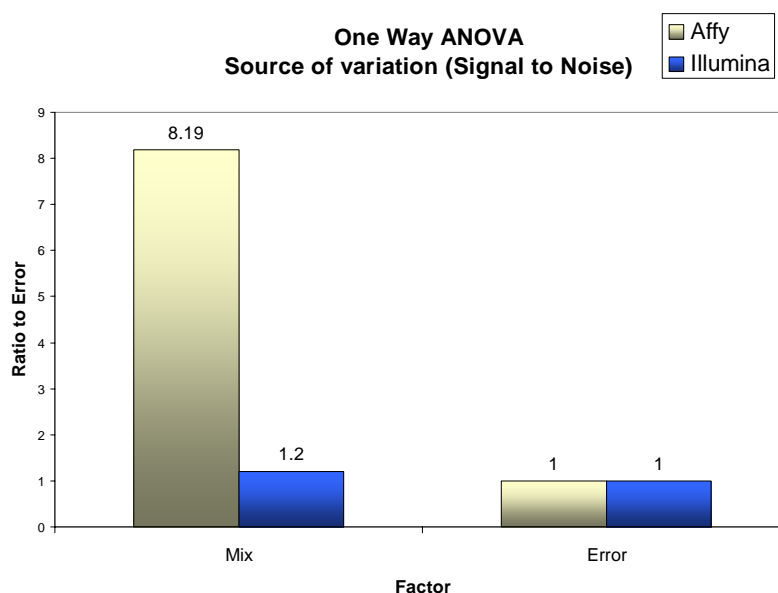


Figure 3. One-Way ANOVA to identify the noise (error) in each platform and the biological effect (Mix) as a source of variation. Panel a: The mean of mean square values shows a significantly lower (~10-fold) error value for the Affymetrix platform in comparison to the Illumina dataset. Panel b: Ratio to error analysis shows a high signal to noise ratio (8.19-fold) on the Affymetrix platform and a very low (1.2-fold) signal to noise ratio on the Illumina platform.

The second output values from an ANOVA analysis are the p-values reported for each probe or probe set. The p-values are important as they indicate the significance of an observed difference in expression between the treatment groups. The lower the p-value, the more confidence we have that the difference detected is not due to random chance.

An important consideration that must be taken into account when performing any microarray experiment is that multiple measurements are being taken in parallel; in this study the Illumina array assayed >24,000 genes and the Affymetrix array assayed greater than >54,500. As a consequence, the probability that any one gene is changing by random chance is additive across all genes tested resulting in a higher false positive rate than the individual p-value cut-off chosen would otherwise indicate. To adjust for p-value inflation, multiple testing corrections are typically applied.

The two most common methods for multiple test correction are the very conservative (i.e. results in a high rate of false negative) Bonferroni correction and the less stringent (i.e. allows for higher rate of false positives) False Discovery Rate (FDR) adjustment. Each method was applied to the data sets and the resulting number of significantly changed probes or probe sets are summarized in Table 1. After FDR adjustment, the Affymetrix platform detected ~10-fold more significant differences between the compared samples than the Illumina platform with a FDR p-value <0.05.

Due to the nature of the Bonferroni method, the resulting corrections are not expected to be the equivalent for multiple tests of different sizes. As a result, the adjusted p-value arrived at following Bonferroni correction is expected to be significantly more stringent for the Affymetrix dataset given that the array measures almost twice as many data points. Despite this increased challenge, the Affymetrix platform still showed 4-fold more significantly differentially expressed genes than Illumina.

Multiple test correction (Total probes or probe sets)	Affymetrix (54,675)	Illumina (24,114)
FDR <0.05	49	4
Bonferroni <0.05	4	1

Table 1. Number of probes or probe sets below an adjusted p-value threshold of <0.05 by FDR and Bonferroni, in the analysis of 5% placenta vs. 100% PBMC. On the Affymetrix platform more significantly differentially expressed genes are detected.

In Summary, Affymetrix detects more significantly differentially expressed genes in the 5% placenta “treatment”, which is in agreement with the exploratory analysis mentioned above. The downstream pathway analysis discussed below confirms that the genes determined to be significantly differentially expressed are biologically significant as the majority relate to placental specific networks.

Classification Analysis

Many microarray platform comparison studies focus on identifying agreement in the expression values of a common set of genes for all platforms that participated in the study. Such an evaluation approach is useful, especially for regulatory agencies such as the FDA (4), to assess agreement between platforms. However, these approaches are limiting as one is forced to evaluate only the performance of the common sequences between the different platforms. As the number of platforms compared in the study increases, the number of sequences that can be compared will decrease with the smallest microarray (i.e. the platform with the minimal content) being the limiting factor.

As a consequence, such approaches do not provide the information necessary to identify the platform that will provide the best overall performance for answering biological, medical research or diagnostic questions. The platform with the smallest content may have good technical performance; however, the available content may not be sufficient to answer these types of real-life biological questions. Instead, this is best achieved by allowing each platform to answer a common biological question, such as identification of biomarkers and classification accuracy, allowing all available content to be considered.

For such an analysis, a two-level nested cross validation is applied to provide an accuracy estimate for each microarray platform and assess their performance as a tool for identifying biomarkers useful for classifying or predicting biological or clinical samples.

One can apply hundreds of different classification models and provide an accuracy estimate based on the best performing classification model compared to the performance of all tested models. The process iteratively tests all selected models with different input lists.

To increase power and confidence in the accuracy estimate, each iteration of the process will leave one or more samples out of the analysis thereby eliminating the need to split the data set into training and test groups. This approach maintains a common ground for comparison allowing an assessment how each platform will perform in these real-life situations.

Towards this goal, a two-level nested cross validation with 24 classification models Using Partek Genomics Suite v6.2 (Partek, Inc. St. Louis, MO, USA) was applied. Different parameters settings in each classification method are considered as different classification models. The three parameters we used here are the distance measure (Euclidean or Pearson's Dissimilarity), the number of Nearest Neighbors in KNN (1, 3), and the number of top ranked probe sets (variables) from the ANOVA that are used as an input for classification (10, 50, 100, 250, 500, 1000).

The output of this classification approach includes an overall accuracy estimate for the best performing classifiers based on all models tested, and a list of probe sets that were selected as the best classifiers/biomarkers, ranked by the frequency of their detection as classifiers by the different classification models that were applied.

The concept behind this approach is that we increase confidence in the microarray platform as a tool to predict and classify if we show that many different classification models can be applied and the power to classify is not limited to one specific model. As a result the overall accuracy estimate is typically very conservative and, therefore increases confidence and sets the appropriate expectation in the likelihood for prediction of the classes studied in a new independent dataset from a similar experiment.

Table 3 shows the results of the two-level cross validation when applied to two treatment groups; 100% placenta versus 100% PBMC and 5% placenta versus 100% PBMC. While we expect very high classification accuracies when we classify between the two different tissues (100% PBMC versus 100% placenta), we see that using the multi-models classification approach, the conservative results are 85% for the Affymetrix platform and 76% for the Illumina platform.

Platform	100%PBMC vs. 100%Placenta	100%PBMC vs. 5%Placenta
Affymetrix	85%	77%
Illumina	76%	40%

Table 3. Accuracy estimate of classification. Two-level nested cross validation conservative results show 85% accuracy for Affymetrix and 76% for Illumina when 100% PBMC vs. 100% Placenta are tested, and 77% for Affymetrix and 40% (random) accuracy for Illumina when 5% placenta vs. 100% PBMC are tested.

This difference in classification accuracy is even more substantial (37% difference) when this approach is applied to the more biological reflecting “treatment” of 5% placenta versus 100% PBMC. Here Affymetrix shows a 77% classification accuracy versus the 40% achieved by Illumina. The Illumina result basically means that there is no ability to classify the 5% placental sample as a random dataset would be expected to have a higher classification accuracy. The bias from 50% may be explained by the under-powered limitation of this study with only two replicates run for each mixture.

Pathway Analysis.

The list of the top 100 classifiers/biomarkers identified by the ANOVA and the two-level nested cross validation classification approach were used as an input for pathway analysis to determine the biological relevance of the results.

It is important to note that the biomarkers contained in a classification signature will not always have a direct biological connection to the biology being investigated based on our current knowledge but may still serve as good biomarkers to predict classes.

The top network identified by Ingenuity® Pathways Analysis (Ingenuity Systems, Inc. Redwood City, CA, USA), as shown in Figure 4, includes the ALPP placental specific gene previously mentioned in the exploratory analysis section above. Ingenuity pathway analysis of the Affymetrix data resulted in 40 different networks while the Illumina data resulted in only 5 networks.

When applying a second pathway analysis software package, MetaCore™ (GeneGo, Inc., St. Joseph, MI, USA), 206 placental related networks and maps (Figure 5.) were identified with the Affymetrix dataset. These results confirm that the differentially expressed probe sets and the identified biomarkers are biologically relevant to the placental “treatment”.

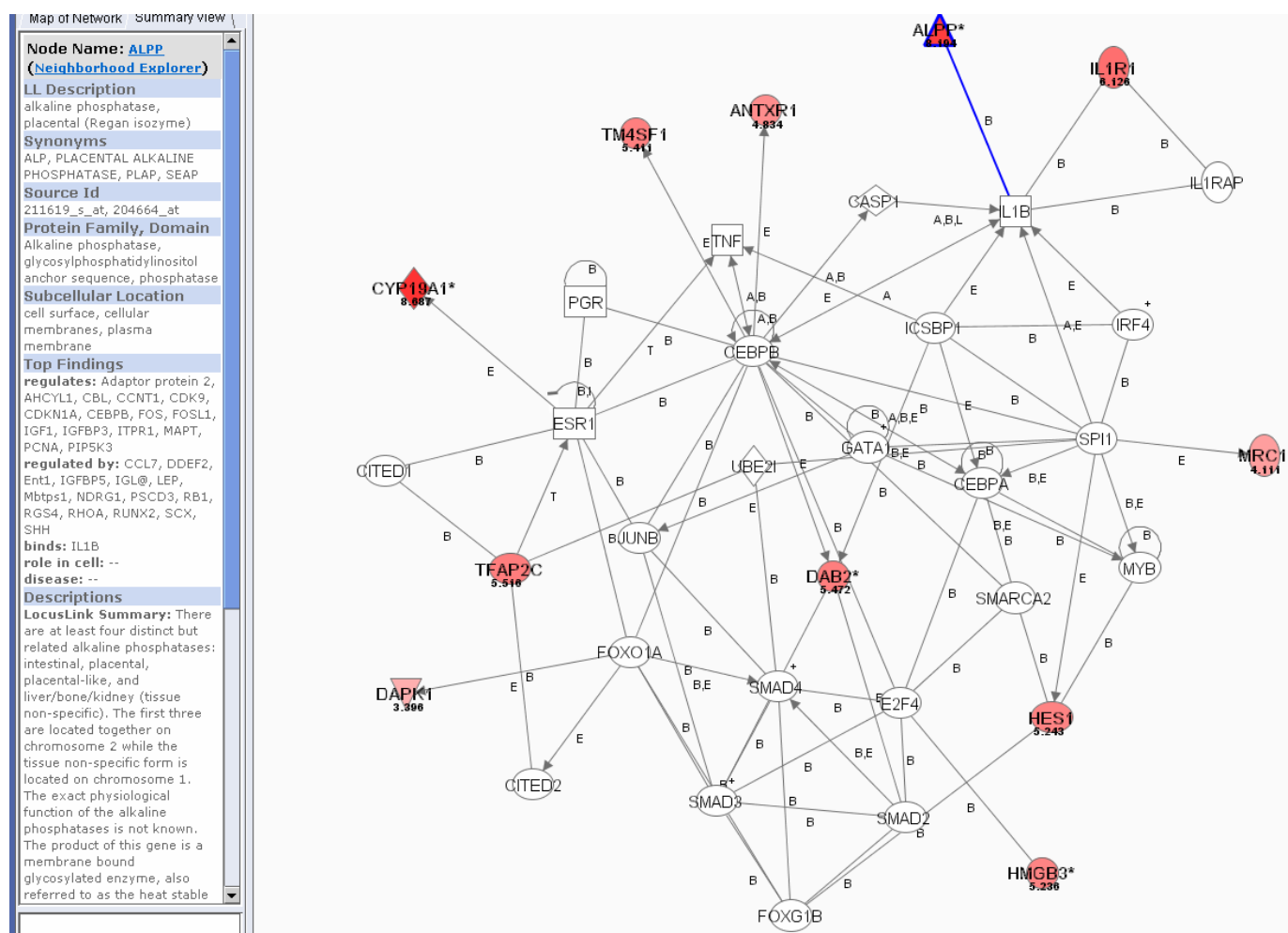


Figure 4. Top Ingenuity® network for the Affymetrix biomarkers includes the Placental Alkaline Phosphatase (ALPP). 40 Ingenuity networks were identified based on the Affymetrix dataset vs. 5 networks on Illumina.

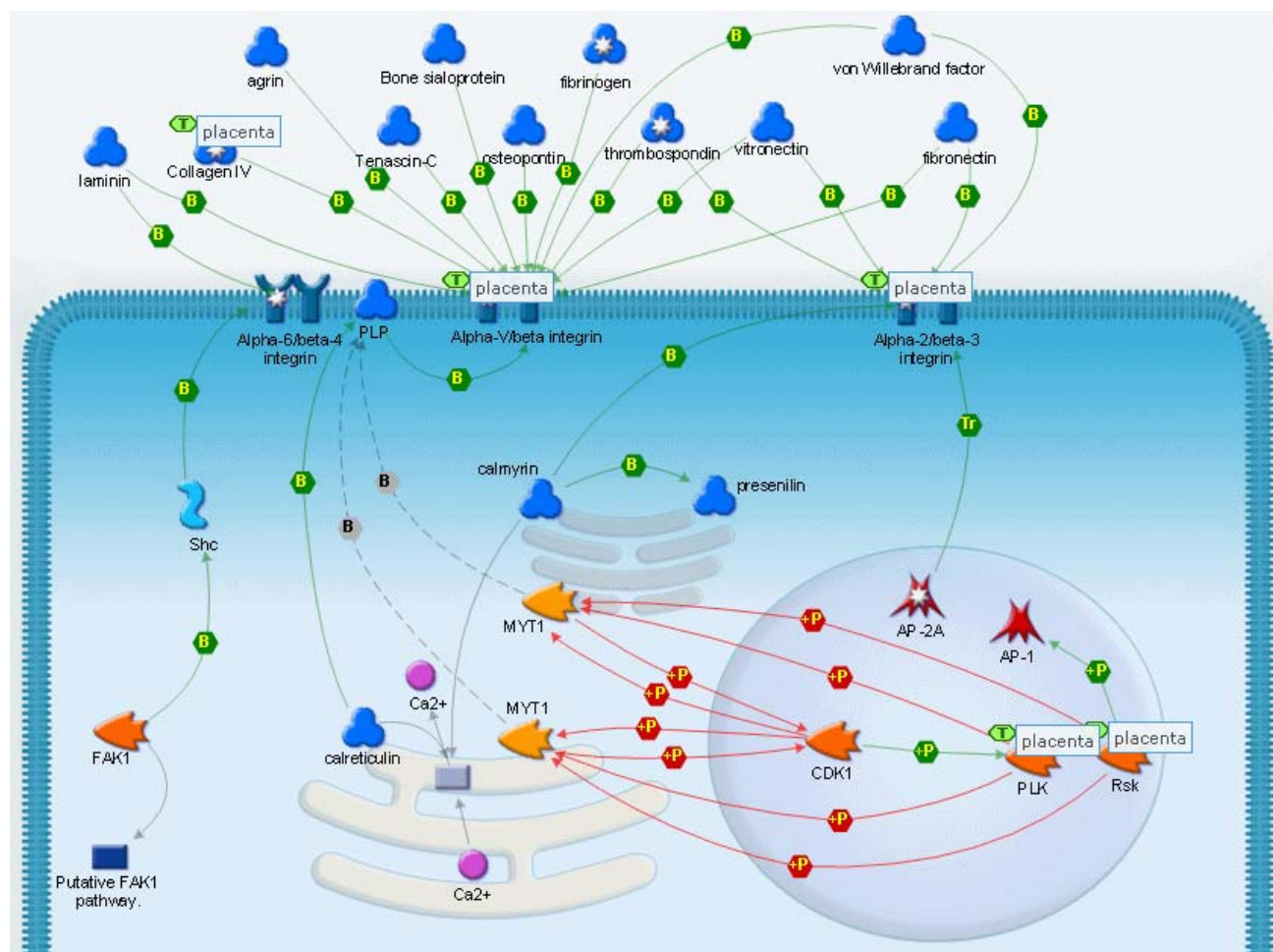


Figure 5. MetaCore™ Integrin Interactions with Extracellular Matrix map with placental related labeled genes. We identified 206 placental related maps based on the biomarkers identified on the Affymetrix platform.

Considerations When Performing Platform Comparisons

While the approaches taken by Barnes et al provide information useful to assessing the performance of the Affymetrix and Illumina platforms, there is information missing that is essential to fully understanding platform performance differences. To be able to fully describe a platform's performance in terms of accuracy (how closely a measurement is to its true value), precision (how reproducible a measurement is) and bias (systemic deviations from a true value that can potentially be mitigated if they are fully understood), the experimental design must include data points of known values.

This goal can be achieved by introducing mRNA spikes at predetermined concentrations; the most exhaustive approach would represent each transcript spike at each and every concentration tested as described by a Latin Square design (5). Such a microarray spike-in experiment is designed to provide a gold standard towards measuring the performance of a platform with respect to known transcript concentrations and differential changes.

Tissue mixture experimental designs can be useful in assessing the overall performance of the platform particularly in the sense of a classification/prediction tool (i.e. Responders and Non responders for a given drug treatment). To simulate subtle biological changes, low titration levels should be used such as: 1% Tissue 1 + 99% Tissue 2 up to 10% Tissue1 + 90% Tissue 2. A distinct drawback to using a mixture approach is that important biological changes that are invisible in the data are never seen as false-negatives.

Not only are limits set by the design and hybridization of samples, but careful consideration of the analysis methodologies employed must be given. For example, many normalization techniques designed to remove technical variation from the data can eliminate real biological variation if improperly applied. Again, the use of spike-in experiments provides the only true method for testing the resulting false-negative level.

The user should be extremely careful to base any conclusions on intersecting gene lists as these will accumulate false negatives. Instead, a statistical model, such as the ANOVA described above, can be extended to take into consideration the interactions between the different factors being studied. In this sense, the interactions between the biological and the platform effects can be quantified. That is, probes that respond differently on different platforms can be identified.

As a final thought, irrespective of the study design used or the questions being addressed, careful attention should be given to the number of sample and spike (if included) replicates to ensure that the study is sufficiently powered to support any conclusions made. Statistical approaches (power calculations) exist that allow the user to estimate the numbers of replicates required based on an estimate of the system variability and the required level differential detection.

Conclusions

Due to the large numbers of genes being measured in parallel in any microarray experiment, the challenge for the researcher is to identify those genes whose expression is changing due to direct effects of the biology under consideration (true signal) as opposed to random fluctuations in gene expression levels due to experimental variability, either technical or biological (noise). For this purpose, the application of standard statistical methods has become the preferred approach.

Taking this approach, a statistical analysis of the data generated in the Barnes et al comparison of the Affymetrix GeneChip and Illumina BeadArray platforms clearly highlights performance differences between the two platforms. The apparent signal detectable in the Illumina dataset is marginally higher than background while the Affymetrix dataset showed a signal-to-noise ratio of >8:1. The low signal-to-noise ratio observed for the Illumina data is supported by the higher variability highlighted by Barnes et al. This lower performance resulted in:

- Lower predictive ability to classify a sample of PBMC or placental origin.
- Fewer significant gene expression changes at a p-value <0.05 statistical cut-off.

Due to limitations in the Barnes et al study design, an accurate assessment of the balance between sensitivity (measured number of true positives) and specificity (measured number of false positives) on each platform could not be made. As previously discussed, this requires the inclusion of predefined mRNA spikes in the experimental design. However, it is of interest to note that the genes identified within the Affymetrix dataset as significantly differentially expressed between the different dilutions

compared were biologically relevant to the experiment, confirming that true biological differences were being detected.

Platform performance is of course of central importance to any microarray study, however, there are a number of other criteria that the user should keep in mind when making their final choice as data generation is only the first step in the experiment.

Having a comprehensive analysis workflow and access to tools that support that workflow is essential. This includes both statistical tools used to generate a “gene list” and annotation/pathway analysis tools that provide context to the gene list. Through the Affymetrix NetAffx™ Analysis Center and working openly with the 3rd party software community, a broad choice of tools are available to support what are now mature workflows greatly reducing the time and resources necessary to complete an Affymetrix GeneChip array experiment. Additionally, the open availability of Affymetrix probe sequences and array design parameters greatly simplify any follow-up validation experiments.

Comprehensive genomic content is an obvious requirement for any exploratory study. The Affymetrix whole genome 3' Expression Arrays are the most comprehensive designs available. Additionally, with ultra-high microarray densities a reality, new exon-level designs allow monitoring of a greater variety of transcript variants and alternative splicing to be investigated. The greater number of probes and equal distribution of those probes throughout a transcript provides an extremely robust design offering industry-leading sensitivity and specificity when used for gene level expression estimates. Such exon level designs (now available for human, rat and mouse) provide exciting new opportunities for understanding the link between genomics and proteomics.

Finally, platform, methodology and quality standardization is necessary to enabling multi-site studies, collaboration and data sharing between groups, and data acceptance within the research community. A high degree of Affymetrix data concordance even when run by different groups at geographically disparate sites has been carefully documented (6) emphasizing intra-platform reproducibility.

Supported by over 4,000 publications to date, a significant number of researchers are applying the GeneChip platform to answer key biological questions of interest. It is through such successes that the Affymetrix GeneChip System has become the platform standard of choice.

References

1. M Barnes, J Freudenberg, S Thompson, B Aronow and P Pavlidis: **Experimental comparison and cross-validation of the Affymetrix and Illumina gene expression analysis platforms.** *Nucleic Acids Research* 2005, Vol. 33, No. 18: 5914–5923.
2. Y Woo, J Affourtit, S Daigle, A Viale, K Johnson, J Naggert, G Churchill: **A comparison of cDNA, oligonucleotide, and Affymetrix GeneChip gene expression microarray platforms.** *J Biomol Tech* 2004, **15**:276-84.
3. A de Reyniès, D Geromin, J-M Cayuela, F Petel, Pe Dessen, F Sigaux and DS Rickman : **Comparison of the latest commercial short and long oligonucleotide microarray technologies.** *BMC Genomics* 2006, <http://www.biomedcentral.com/1471-2164/7/51>
4. Thompson, K.L., Rosenzweig, B.A., Pine, P.S., Retief, J., Turpaz, Y., Afshari, C.A., Hamadeh, H.K., Damone, M., Blomme, E., Ciurlionis, R., Waring, J., James C. Fuscoe, J.C., Paules, R., Tucker, J., Fare, T., Coffey, E.M., He, Y., Collins, J., Jarnagin, K., Fujimoto, S., Gander, B., Kiser, G., Kaysser-Kranich, T., Sina, J., Sistare, F.D. **Use of a mixed tissue RNA design for performance assessments on multiple microarray formats.** *NAR* 2005, 33: e187-199.
5. **New Statistical Algorithms for Monitoring Gene Expression on GeneChip® Probe Arrays.** Affymetrix Technical Note (<http://www.affymetrix.com>).
6. KK Dobbin, DG Beer, M Meyerson, TJ Yeatman, WL Gerald, JW Jacobson, B Conley, KH Buetow, M Heiskanen, RM Simon, et al: **Interlaboratory comparability study of cancer gene expression analysis using oligonucleotide microarrays.** *Clin Cancer Res* 2005, 11:565-72.