Statistical Algorithms Reference Guide

AFFYMETRIX®

Introduction

This guide is a reference tool for the Affymetrix Statistical Algorithms used in the expression analysis of GeneChip® probe arrays. It provides the user with a basic description of the mathematical concepts behind expression measurements for either single array or comparison analysis. Please refer to the References section of this guide for further information on published documentation.

The Statistical Algorithms were implemented in Affymetrix® Microarray Suite version 5.0. Previous versions of the GeneChip® Analysis Suite and Affymetrix® Microarray Suite used the Empirical Algorithms.

GeneChip® Technology

A GeneChip® probe array is a tool used to monitor gene expression for thousands of transcripts. A transcript is represented as a probe set. A probe set is made up of probe pairs comprised of Perfect Match (PM) and Mismatch (MM) probe cells. The intensities of each probe pair are the key ingredients used to make an expression measurement. This measurement is calculated for each probe set and is described in the form of qualitative and quantitative values. Additionally, the expression measurements of a baseline and experimental array can be compared to understand the relative change in abundance of a transcript.

Development of the Algorithms

The Statistical Algorithms were developed using standard statistical techniques. The performance was validated using an experimental design called the Latin Square. In this experimental design, transcripts, naturally absent in the complex background, were spiked in at known concentrations.

Single Array Analysis

Single array analyses can be used to build databases of gene expression profiles, facilitate sample classification and transcript clustering, and monitor gross expression characteristics. In addition, the analyses provide the initial data required to perform comparisons between experiment and baseline arrays.

This analysis generates a **Detection *p*-value** which is evaluated against user-definable cut-offs to determine the **Detection** call. This call indicates whether a transcript is reliably detected (Present) or not detected (Absent). Additionally, a **signal** value is calculated which assigns a relative measure of abundance to the transcript.

	Stat Pairs	Stat Pairs Used	Signal	Detection	Detection p-value
37984_s_at	16	16	92.2	P	0.000218
32102_at	16	16	59.5	P	0.000218
37900_at	16	16	72.6	P	0.000219
31697_s_at	16	16	664.2	P	0.000219
40567_at	16	16	502.3	P	0.000219
35808_at	16	16	212.6	P	0.000219
34819_at	16	16	143.0	P	0.000219
35787_at	16	16	295.7	P	0.000219
35758_at	16	16	301.0	P	0.000219
34817_s_at	16	16	339.6	P	0.000219
34644_at	16	16	723.9	P	0.000219
34608_at	16	16	3313.0	P	0.000219

Figure 1. Data analysis output (.CHP file) for a Single Array Analysis includes Stat Pairs, Stat Pairs Used, Signal, Detection, and the Detection *p*-value.

Detection Algorithm

The Detection algorithm uses probe pair intensities to generate a Detection *p*-value and assign a Present, Marginal or Absent call. Each probe pair in a probe set is considered as having a potential vote in determining whether the measured transcript is detected (Present) or not detected (Absent). The vote is described by a value called the Discrimination score [R]. The score is calculated for each probe pair and is compared to a predefined threshold Tau. Probe pairs with scores *higher* than Tau vote for the *presence* of the transcript. Probe pairs with scores *lower* than Tau vote for the *absence* of the transcript. The voting result is summarized as a *p*-value. The higher the discrimination scores are above Tau, the smaller the *p*-value and the more likely the transcript will be Present. The lower the discrimination score below Tau, the larger the *p*-value and the more likely the transcript will be Absent. The *p*-value associated with this test reflects the confidence of the Detection call.

— Detection *p*-value

A two-step procedure determines the Detection *p*-value for a given probe set.

1). Calculation of the Discrimination score [R] for each probe pair.

2). Test the Discrimination scores against the user-definable threshold Tau.

The Discrimination score is a basic property of a probe pair that describes its ability to detect its intended target. It measures the target-specific intensity difference of the probe pair (PM-MM) relative to its overall hybridization intensity (PM+MM):

$$R = (PM - MM) / (PM + MM)$$

For example, if the PM is much larger than the MM, the Discrimination score for that probe pair will be close to 1.0 (e.g., Probe pair 1 in Figure 2). If the Discrimination scores are close to 1.0 for the majority of the probe pairs, the calculated Detection *p*-value will be lower (more significant). A lower *p*-value is a reliable indicator that the result is valid and that the probability of error in the calculation is small. Conversely, if the MM is larger than or equal to the PM, then the Discrimination score for that probe pair will be negative or zero (e.g., probe pairs 8, 9 and 10 in Figure 2). If the Discrimination scores are low for the majority of the probe pairs, the calculated Detection *p*-value will be higher (less significant).

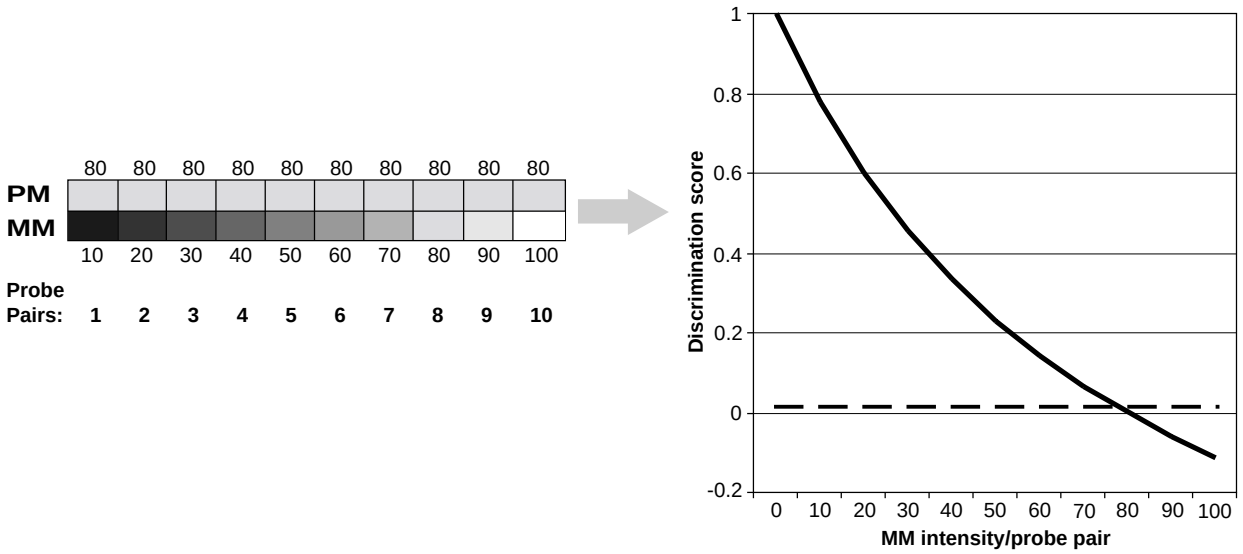


Figure 2. In this hypothetical probe set, the Perfect Match (PM) intensity is 80 and the Mismatch (MM) intensity for each probe pair increases from 10 to 100. The probe pairs are numbered from 1 to 10. As the Mismatch (MM) probe cell intensity, plotted on the x-axis, increases and becomes equal to or greater than the Perfect Match (PM) intensity, the Discrimination score decreases as plotted on the y-axis. More specifically, as the intensity of the Mismatch (MM) increases, our ability to discriminate between the PM and MM decreases. The dashed line is the user-definable parameter Tau (default = 0.015).

The next step toward the calculation of a Detection *p*-value is the comparison of each Discrimination score to the user-definable threshold Tau. Tau is a small positive number that can be adjusted to increase or decrease sensitivity and/or specificity of the analysis (default value = 0.015). The One-sided Wilcoxon's Signed Rank test is the statistical method employed to generate the Detection *p*-value. It assigns each probe pair a rank based on how far the probe pair Discrimination score is from Tau.

Tunable Parameter Tip:

Increasing the threshold Tau can reduce the number of false Present calls, but may also reduce the number of true Present calls. Note: Changing Tau directly influences the calculation of the Detection *p*-value.

— Detection Call

The user-modifiable Detection *p*-value cut-offs, Alpha 1 (a_1) and Alpha 2 (a_2) (See Figure 3), provide boundaries for defining Present, Marginal or Absent calls. At the default settings, determined for probe sets with 15-20 probe pairs (defaults $a_1 = 0.04$ and $a_2 = 0.06$), any *p*-value that falls below a_1 is assigned a Present call, and above a_2 is assigned an Absent call. Marginal calls are given to probe sets which have *p*-values between a_1 and a_2 (see Figure 3).
The *p*-value cut-offs can be adjusted to increase or decrease sensitivity and specificity.

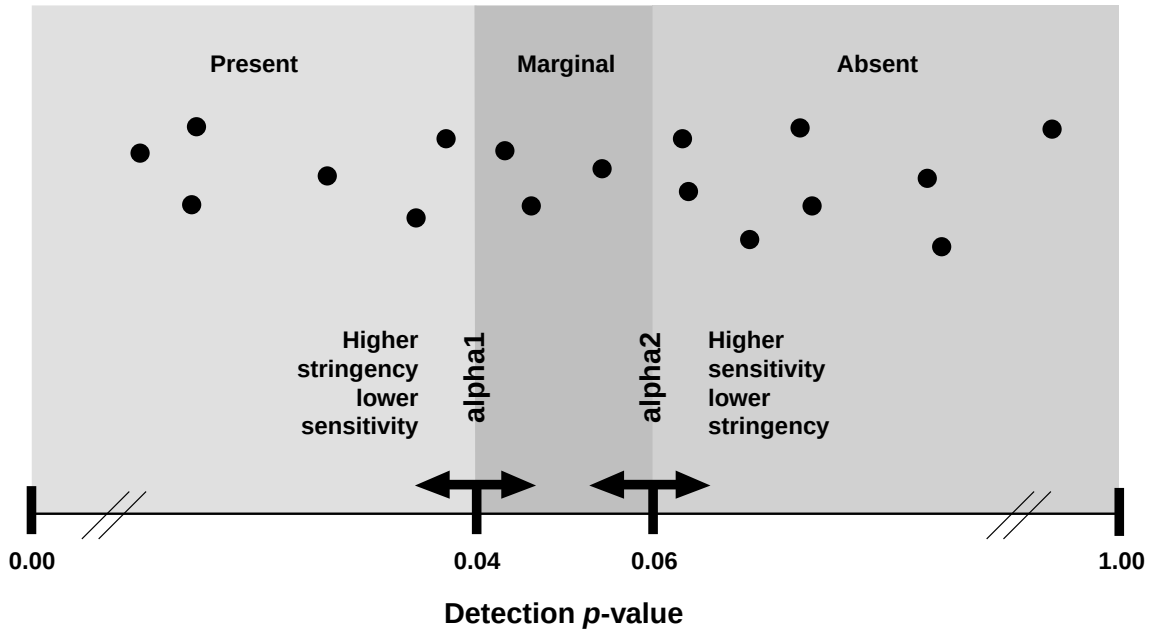


Figure 3. Significance levels a_1 and a_2 define cut-offs of *p*-values for Detection calls.

It is important to note that prior to the two-step Detection *p*-value calculation, the level of photomultiplier saturation for each probe pair is evaluated. If all probe pairs in a probe set are saturated, the probe set is immediately given a Present call. Note that when a probe pair is rejected from further analysis is when a Mismatch (MM) probe cell is saturated (MM = 46,000 for the 2500 GeneArray™ Scanner).

In summary, the Detection Algorithm assesses probe pair saturation, calculates a Detection *p*-value and assigns a Present, Marginal or Absent call.

Signal Algorithm

Signal is a quantitative metric calculated for each probe set, which represents the relative level of expression of a transcript. Signal is calculated using the One-Step Tukey's Biweight Estimate which yields a robust weighted mean that is relatively insensitive to outliers, even when extreme.

Similar to the Detection algorithm, each probe pair in a probe set is considered as having a potential vote in determining the Signal value. The vote, in this case, is defined as an estimate of the real signal due to hybridization of the target. The mismatch intensity is used to estimate stray signal. The real signal is estimated by taking the log of the Perfect Match intensity after subtracting the stray signal estimate (CT). The probe pair vote is weighted more strongly if this probe pair signal value is closer to the median value for a probe set. Once the weight of each probe pair is determined, the mean of the weighted intensity values for a probe set is identified. This mean value is the quantitative metric Signal.

When the Mismatch intensity is lower than the Perfect Match intensity, then the Mismatch is informative and provides an estimate of the stray signal. Rules are employed in the Signal algorithm to ensure that negative signal values are not calculated. Negative values do not make physiological sense and make further data processing, such as log transformations difficult. Mismatch values can be higher than Perfect Match values for a number of reasons such as cross hybridization. If the Mismatch is higher than the Perfect Match, the Mismatch provides no additional information about the estimate of stray signal. Therefore, an imputed value called Change Threshold (CT) is used instead of the uninformative Mismatch.

The following rules are applied:

Rule 1: If the Mismatch value is less than the Perfect Match value, then the Mismatch value is considered informative and the intensity value is used directly as an estimate of stray signal.

Rule 2: If the Mismatch probe cells are generally informative across the probe set except for a few Mismatches, an adjusted Mismatch value is used for uninformative Mismatches based on the bi-weight mean of the Perfect Match and Mismatch ratio.

Rule 3: If the Mismatch probe cells are generally uninformative, the uninformative Mismatches are replaced with a value that is slightly smaller than the Perfect Match. These probe sets are generally called Absent by the Detection algorithm.

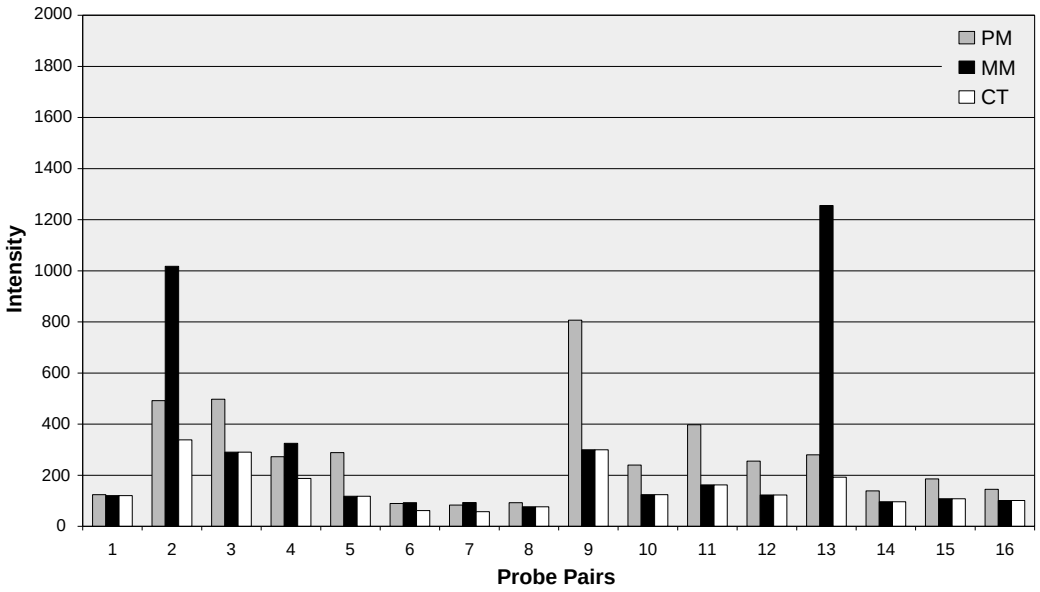


Figure 4. The grey bars illustrate the Perfect Match (PM) intensities and black bars the Mismatch (MM) intensities across a 16-probe pair probe set. The white bars, Change Threshold (CT), are the intensities of the Mismatch based on the Signal rules. In this example, most of the Perfect Match intensities are higher than the Mismatch intensities and therefore Mismatch values can be used directly (e.g., probe pair 9). When the Mismatch is larger than the Perfect Match (e.g., probe pairs 2, 4 and 13) the CT value is used instead of the Mismatch.

Comparison Analysis (Experiment versus Baseline arrays)

In a Comparison Analysis, two samples, hybridized to two GeneChip® probe arrays of the same type, are compared against each other in order to detect and quantify changes in gene expression. One array is designated as the baseline and the another as an experiment. The analysis compares the difference values (PM-MM) of each probe pair in the baseline array to its matching probe pair on the experiment array. Two sets of algorithms are used to generate change significance and change quantity metrics for every probe set. A change algorithm generates a **Change *p*-value** and an associated **Change**. A second algorithm produces a quantitative estimate of the change in gene expression in the form of **Signal Log Ratio**.

	Stat Common Pairs	Signal Log Ratio	Signal Log Ratio Low	Signal Log Ratio High	Change	Change p-value
35839_at	16	0.3	0.2	0.4	I	0.000014
1799_at	16	0.9	0.5	1.3	I	0.000015
35885_at	16	0.4	0.3	0.5	I	0.000015
34696_at	16	0.4	-0.1	0.9	I	0.000023
31356_at	16	1.8	0.8	2.8	I	0.000025
35202_at	16	0.4	0.2	0.6	I	0.000027
39651_at	16	0.4	0.3	0.5	I	0.000029
39777_at	16	0.4	0.1	0.6	I	0.000031
37610_at	16	0.4	0.2	0.5	I	0.000034
32070_at	16	0.3	0.2	0.4	I	0.000034
1561_s_at	16	0.7	0.1	1.3	I	0.000037
35283_at	16	0.5	0.3	0.6	I	0.000037

Figure 5. Data analysis output (.CHP file) for a Comparison Analysis includes Stat Common Pairs, Signal Log Ratio, Signal Log Ratio Low, Signal Log Ratio High, Change, and the Change *p*-value.

Before comparing two arrays, scaling or normalization methods must be applied. Scaling and normalization correct for variations between two arrays. Two primary sources of variation in array experiments are biological and technical differences. Biological differences may arise from many sources such as genetic background, growth conditions, dissection, time, weight, sex, age, and replication. Technical variation can be due to experimental variables such as quality and quantity of target hybridized, reagents, stain, and handling error. The minimization of variation is essential, but scaling and normalization techniques provide a means to remove differences and facilitate comparison analysis.

Normalization and scaling techniques can be applied by using data from a selected user-defined group of probe sets, or from all probe sets. When normalization is applied, the intensity of the probe sets (or selected probe sets) from the experiment array are normalized to the intensity of the probe sets (or selected probe sets) on the baseline array. When scaling is applied, the intensity of the probe sets (or selected probe sets) from the experimental array and the intensity of the probe sets (or selected probe sets) from the baseline array are scaled to a user-defined target intensity. In general, global scaling (scaling to all probe sets) is the preferred method when comparing two arrays.

An additional normalization factor is defined in the Robust Normalization section. This 'robust normalization,' which is not user modifiable, accounts for unique probe set characteristics due to sequence dependent factors such as affinity of the target to the probe and linearity of hybridization of each probe pair in the probe set.

Change Algorithm

As in the Single Array Analysis, the Wilcoxon's signed-rank test is used in Comparison Analysis to derive biologically meaningful results from the raw probe cell intensities on expression arrays. During a Comparison Analysis, each probe set on the experiment array is compared to its counterpart on the baseline array, and a Change *p*-value is calculated indicating an increase, decrease or no change in gene expression. User-defined cut-offs (gammas) are applied to generate discrete Change calls (Increase, Marginal Increase, No Change, Decrease, or Marginal Decrease).

— Robust Normalization

After scaling or normalization of the array (discussed in the Comparison Analysis overview), a further robust normalization of the probe set is calculated. Once the initial probe set normalization factor is determined, two additional normalization factors are calculated that are slightly higher and slightly lower than the original. The range by which the normalization factor is adjusted up and down is specified by a user-modified parameter called perturbation. This supplementary normalization accounts for unique probe set characteristics due to sequence dependent factors such as affinity and linearity. More specifically, this approach addresses the inevitable error of using an average intensity of the majority of probes (or selected probes) on the array as the normalization factor for every probe set on the array. The noise from this error, if unattenuated, would result in many false positives in expression level changes between the two arrays being compared. The perturbation value directly affects the subsequent *p*-value calculation. Of the *p*-values that result from applying the calculated normalization factor and its two perturbed variants, the one that is most conservative is used to estimate whether any change in level is justified by the data. The lowest value for perturbation is 1.00, indicating no perturbation. The highest perturbation value allowed is set at 1.49. Increasing the perturbation value widens the range allowed before a change is called. For example, changes that were called Increase with a smaller perturbation value, may be called No Change with a higher perturbation value. A default was established at 1.1 based on calls made from the Latin Square data set. The perturbation factor and the Latin Square data set are described in more detail in the Affymetrix Technical Notes referenced in the back of this guide.

— Change *p*-value

The Wilcoxon's signed rank test uses the differences between Perfect Match and Mismatch intensities, as well as the differences between Perfect Match intensities and background to compute each Change *p*-value.

From Wilcoxon's signed rank test, a total of three, one-sided *p*-values are computed for each probe set. These are combined to give one final *p*-value which is provided in the data analysis output (.CHP file). The *p*-value ranges in scale from 0.0 to 1.0, and provides a measure of the likelihood of change and direction. Values close to 0.0 indicate likelihood for an increase in transcript expression level in the experiment array compared to the baseline, whereas values close to 1.0 indicate likelihood for a decrease in transcript expression level. Values near 0.5 indicate a weak likelihood for change in either direction. Hence, the *p*-value scale is used to generate discrete change calls using thresholds. These thresholds will be described in Change Call section.

	Signal Log Ratio	Signal Log Ratio Low	Signal Log Ratio High	Change	Change p-value
l25069_s_at	-0.4	-0.5	-0.2	D	0.999979
m12303_s_at	-0.3	-0.4	-0.2	D	0.999999
m12347_s_at	-0.3	-0.3	-0.2	D	0.999997
m21495_f_at	-0.2	-0.3	-0.2	D	0.999763
m29793_s_at	-0.2	-0.2	-0.1	D	0.998697
m34173_at	-0.3	-0.4	-0.3	D	0.999505
m62867_s_at	-0.4	-0.5	-0.3	D	1.000000
n28127_rc_at	-0.4	-0.5	-0.3	D	0.997989
l07577_s_at	-0.4	-0.5	-0.4	D	1.000000
aa000380_s_at	0.4	0.2	0.6	I	0.002283
aa002704_at	0.7	0.5	0.9	I	0.000150
aa002761_s_at	0.4	0.2	0.6	I	0.000226
aa000148_s_at	0.3	0.2	0.4	I	0.001246
aa009154_s_at	0.4	0.3	0.6	I	0.000028
aa013647_s_at	0.3	0.2	0.4	I	0.002011
AA023300_at	0.2	-0.1	0.6	I	0.001140
aa023407_s_at	0.9	0.6	1.2	I	0.000089
AA408234_rc_g_at	0.5	0.3	0.6	I	0.000001

Figure 6. Data analysis output (.CHP file) for a Comparison Analysis illustrating Change *p*-values with the associated Increase (I) or Decrease (D) call. Increase calls have Change *p*-values closer to zero and Decrease calls have Change *p*-values closer to one.

Tunable Parameter Tip:

Increasing the perturbation value can reduce the number of false changes, but may also decrease the detection of true changes. Note: Changing perturbation factor affects the calculation of the *p*-value directly.

Change Call

The final Change *p*-value described above is categorized by cutoff values called gamma1(γ_1) and gamma2(γ_2). These cut-offs provide boundaries for the Change calls: Increase (I), Marginal Increase (MI), No Change (NC), Marginal Decrease (MD), or Decrease (D).

The user does not directly set γ_1 and γ_2; rather each is derived from two user-adjustable parameters, γL and γH. In the case of γ_1, the two user-adjustable parameters are called γ_1L and γ_1H (defaults for probe sets with 15-20 probe pairs: γ_1L= 0.0025 and γ_1H= 0.0025), which define the lower and upper boundaries for γ_1. Gamma2 (γ_2) is computed as a linear interpolation of γ_2L and γ_2H (defaults for probe sets with 15-20 probe pairs: γ_2L= 0.003 and γ_2H= 0.003) in an analogous fashion.

The ability to adjust the stringency of calls associated with high and low signal ranges independently makes it possible to compensate for effects that influence calls based on low and high signals. This feature, however, is not used by default because the defaults are set as γ_1L = γ_1H and γ_2L = γ_2H

It is important to note that, like in Detection *p*-value calculation, the level of photomultiplier saturation for each probe pair is evaluated. In the computation of Change *p*-value, any saturated probe cell, either in the Perfect Match or Mismatch, is rejected from analysis. The number of discarded cells can be determined from the Stat Common Pairs parameter.

In summary, the Change algorithm assesses probe pair saturation, calculates a Change *p*-value and assigns an Increase, Marginal Increase, No Change, Decrease, or Marginal Decrease call.

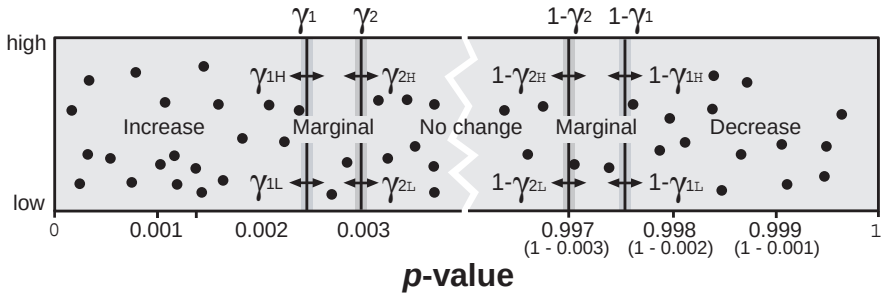


Figure 7. A representation of a range of *p*-values for a data set. The Y-axis is the probe set signal. The arrows on the vertical bars represent the adjustable γ values. The γ_1 value is a linear interpolation of γ_1L and γ_1H. Similarly γ_2 is derived from γ_2L and γ_2H.

Signal Log Ratio Algorithm

The Signal Log Ratio estimates the magnitude and direction of change of a transcript when two arrays are compared (experiment versus baseline). It is calculated by comparing each probe pair on the experiment array to the corresponding probe pair on the baseline array. This strategy cancels out differences due to different probe binding coefficients and is therefore more accurate than a single array analysis.

As with Signal, this number is computed using a one-step Tukey s Biweight method by taking a mean of the log ratios of probe pair intensities across the two arrays. This approach helps to cancel out differences in individual probe intensities, since ratios are derived at the probe level, before computing the Signal Log Ratio. The log scale used is base 2, making it intuitive to interpret the Signal Log Ratios in terms of multiples of 2. Thus, a Signal Log Ratio of 1.0 indicates an increase of the transcript level by 2 fold and -1.0 indicates a decrease by 2 fold. A Signal Log Ratio of zero would indicate no change.

The Tukey s Biweight method gives an estimate of the amount of variation in the data, exactly as standard deviation measures the amount of variation for an average. From the scale of variation of the data, confidence intervals are generated measuring the amount of variation in the biweight estimate. A 95% confidence interval indicates a range of values, which will contain the true value 95% of the time. Small confidence intervals indicate that the data is more precise, while large confidence intervals reflect uncertainty in estimating the true value. For example, the Signal Log Ratio for some transcripts may be measured as 1.0, with a range of 0.5 to 1.5 from low to high. For 95% of transcripts

with such results, the true Signal Log Ratio will lie somewhere in that range. A set of noisy experiments might also report a Signal Log Ratio of 1.0, but with a range of -0.5 to 2.5, indicating that the true effect could easily be zero, since the uncertainty in the data is very large. The confidence intervals associated with Signal Log Ratio are calculated from the variation between probes, which may not reflect the full extent of experimental variation.

Glossary

Biweight Estimate: An estimate of the central value of a sample using Tukey's Biweight method.

Change: A call indicating an Increase (I), Marginal Increase (MI), No Change (NC), Decrease (D), or Marginal Decrease (MD) in transcript level between a baseline array and an experiment array.

Change Threshold: A value used in place of the Mismatch intensity when Rules 2 and 3 are used in the Signal Algorithm.

Detection: A measurement indicating if the transcript was detected (Present), not detected (Absent), or marginally detected (Marginal).

Discrimination Score [R]: The relative difference between a Perfect Match and its Mismatch (R=PM-MM/PM+MM).

Empirical Algorithms: The algorithms contained in GeneChip® Analysis Suite and Microarray Suite versions prior to Microarray Suite version 5.0.

Latin Square: An experimental design used to monitor the detectability of a transcript accurately over a range of concentrations. It also allows the statistical analysis of patterns and variability in repeated measurements in a systematic fashion.

Mismatch (MM): A 25-mer oligonucleotide designed to be complementary to a reference sequence except for a single, homomeric base change at the 13th position. Mismatch probes serve as specificity controls when compared to their corresponding Perfect Match probe.

Nonparametric Test: A statistical test without the assumption of a particular distribution of the data, also known as a distribution-free test.

Normalization: Adjusting an average value of an experimental array equal to that of the baseline array so that they can be compared.

***p*-value:** The probability value of a certain statistic that is equal or more extreme to the observation when the null hypothesis is true. In the One-sided Wilcoxon Signed rank test, the statistic is the sum of positive signed ranks. A **Detection *p*-value** measures the probability that the discrimination scores of all probe pairs in the probe set are above a certain level (Tau), and that the target is likely to be Present. A **Change *p*-value** measures the probability that the expression levels of a probe set in two different arrays are the same or not. When the p-value is close to 0.5, they are likely to be the same. When the *p*-value is close to 0, the expression level in the experiment array is higher than that of the baseline array. When the *p*-value is close to 1, the expression level in the experiment arrays is lower than that of the baseline.

Parametric Test: A statistical test that assumes that the data sampled was from a population that follows a Gaussian or normal distribution.

Perfect Match (PM): A 25-mer oligonucleotide designed to be complementary to a reference sequence.

Perturbation: The range by which the normalization factor is adjusted up or down by the user.

Probe: A 25-mer oligonucleotide designed to be complementary to a reference sequence. It is synthesized on the surface of the array using photolithography and combinatorial chemistry. Hybridization to probes provides intensity data used in both empirical and statistical algorithms.

Probe Cell: A single square-shaped feature on an array containing probes with a unique sequence. A probe cell ranges in size from 18 to 50 microns depending on the array type.

Probe Pair: Two probe cells, designed as a Perfect Match (PM) and its corresponding Mismatch (MM).

Probe Set: A collection of probe pairs designed to detect a specific target sequence. A probe set may contain between 11-20 probe pairs.

Scaling: Adjusting the average intensity value of every array to a common value (target intensity), in order to make the arrays comparable.

Signal: A measure of the relative abundance of a transcript.

Signal Log Ratio: The change in expression level for a transcript between a baseline and an experiment array. This change is expressed as the $\log_2$ ratio. A $\log_2$ ratio of 1 is the same as a Fold Change of 2.

Signal Log Ratio Low: The lower limit of the Signal Log Ratio within a 95% confidence interval.

Signal Log Ratio High: The upper limit of the Signal Log Ratio within a 95% confidence interval.

Stat Common Pairs: The number of common probe pairs on two arrays (experiment versus baseline) after saturation across the probe set is determined.

Stat Pairs: The number of probe pairs in the probe set.

Stat Pairs Used: The number of probe pairs in the probe set used in the Detection call.

Statistical Algorithm: The algorithm contained in Microarray Suite version 5.0. This algorithm was developed using standard statistical methods.

Target: The sample applied as labeled (biotinylated) cRNA to a GeneChip® Probe Array for hybridization.

Wilcoxon's Signed Rank Test: A nonparametric test for paired data sets. It tests whether the locations (medians) of the two data sets are the same or not.

Reference

Affymetrix Technical Note: "Fine Tuning Your Data Analysis." (2001).

Affymetrix Technical Note: "New Statistical Algorithms for Monitoring Gene Expression on GeneChip® Probe Arrays." (2001).

Electronic Textbook, StatSoft, Inc., 1984-2000. www.statsoft.com/textbook/stathome.

Hoaglin, D.C., Mosteller, F., Tukey, J.W. Understanding Robust and Exploratory Data Analysis, John Wiley & Sons, New York (2000).

Hollander, M., Wolfe, D.A. Nonparametric Statistical Methods (second edition), John Wiley & Sons, New York (1999).

Liu, W.M., Mei, R., Di, X., Ryder T., Hubbell, E., Dee, S., Webster, T.A., Harrington, C.A., Ho, M., Baid, J., Smeekens, S. Analysis of high density expression microarrays with signed-rank call algorithms. In Preparation (2001).

Liu, W.M., Mei, R., Bartell, D.M., Di, X., Webster, T.A., Ryder, T. Rank-based algorithms for analysis of microarrays. Proceedings SPIE 4266, 56-67 (2001).

Motulsky, H. Intuitive Biostatisitcs, Oxford University Press, New York (1995).

Roderick, J.A., Little, D., Rubin, B., Statistical analysis with missing data, Wiley, New York (1987).

Wilcoxon, F. Individual comparisons by ranking methods. Biometrics 1, 80-83 (1945).

Appendix 1: Terminology Comparison Table (Statistical Algorithms versus Empirical Algorithms)

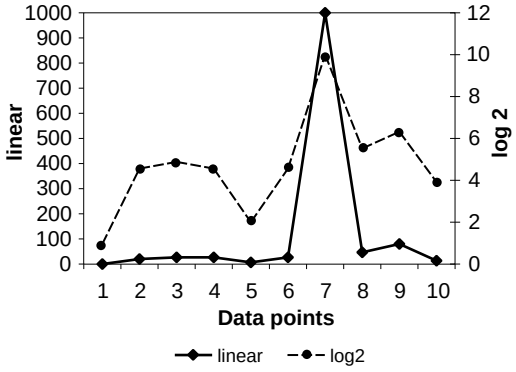
Statistical Algorithms	Empirical Algorithms
Signal	Average Difference
Detection	Absolute Call
Change	Difference Call
Signal Log Ratio	Fold Change

Appendix 2: The Logic of Logs

Quantitative changes in gene expression are reported as a Signal Log Ratio in the Statistical Algorithms as opposed to a Fold Change that was reported in the Empirical Algorithms.

— The Benefit of Logs:

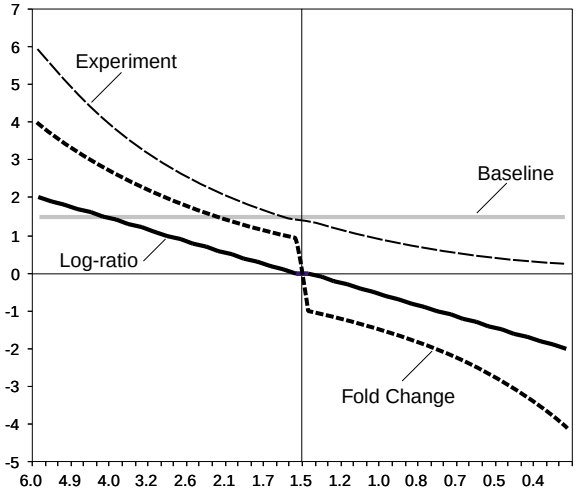
Hybridized probe intensities tend to be distributed over exponential space due to hybridization behavior that is governed by exponential functions of sequence-dependent base-pairing energetics. Thus, log transformation is an appropriate process for analyzing hybridization data. Some of the benefits are apparent in this graph where we show the same data set plotted on two scales. When the data is plotted on a linear scale (solid) the single, high data point (7) overwhelms the graph and obscures information contained in the low values. When the same data is plotted on a Log_2 scale (dashed line), we can see variations in the low values as well as the very high values.



— Signal Log Ratio vs. Fold Change

In this graph, Signal Log Ratio is compared to Fold Change in a hypothetical experiment. Baseline values were set to 1.5 and experiment values were reduced progressively from 6 to 0.375. The x-axis illustrates the values that were decreased in the hypothetical experiment. The Y-axis represents units (e.g., signal log ratio, fold change, or signal for baseline and experiment).

There is a discontinuous transition where the experiment and the baseline converge and the fold change approaches 1 or -1. At this point (smaller changes), the fold change is less sensitive. Since we use log_2, a Signal Log Ratio of 1 equals a Fold Change of 2 and a Signal Log Ratio of 2 equals a Fold Change of 4. Alternatively, use the following formula:



$$\text{Fold Change} = \begin{cases} 2^{\text{Signal Log Ratio}} & \text{Signal Log Ratio} \geq 0 \\ (-1) * 2^{-(\text{Signal Log Ratio})} & \text{Signal Log Ratio} < 0 \end{cases}$$

Appendix 3: Default Parameters

The default parameters were determined for probe sets with 15-20 probe pairs.

Detection Algorithm Defaults

Alphal:	0.04
Alpha2:	0.06
Tau:	0.015

Comparison Algorithm Defaults

Gamma1H:	0.0025
Gamma1L:	0.0025
Gamma2H:	0.003
Gamma2L:	0.003
Perturbation:	1.1

AFFYMETRIX, INC.
3380 Central Expressway
Santa Clara, CA 95051 USA
Tel: 1-888-362-2447 (1-888-DNA-CHIP)
Fax: 1-408-731-5441
sales@affymetrix.com
support@affymetrix.com

AFFYMETRIX UK Ltd.,
Voyager, Mercury Park,
Wycombe Lane, Wooburn Green,
High Wycombe HP10 0HH
United Kingdom
Tel: +44 (0)1628 552550
Fax: +44 (0)1628 552585
saleseurope@affymetrix.com
supporteurope@affymetrix.com

www.affymetrix.com

For research use only.
Not for use in diagnostic procedures.

Part No. 701110 Rev 1