AFFYMETRIX®

Technical Note

Guide to Probe Logarithmic Intensity Error (PLIER) Estimation

This guide is a reference tool for biologists utilizing the new probe logarithmic intensity error (PLIER) method for calculating signal. Additional detail regarding the algorithm appropriate for statisticians and bioinformaticists is provided in the appendix.

Introduction

The probe logarithmic intensity error (PLIER) method produces an improved signal (a summary value for a probe set) by accounting for experimentally observed patterns for feature behavior and handling error appropriately at low and high abundance. Resulting benefits include:

- Higher reproducibility of signal (lower coefficient of variation) without loss of accuracy
- Higher sensitivity to changes in abundance for targets near background
- Dynamic weighting of the most informative probes in an experiment to determine signal

This method was developed by building upon many of the concepts that have been published recently within the field of GeneChip® microarray data analysis, including model-based expression analysis and robust multichip analysis. It also builds upon the summarization algorithm provided in Affymetrix® Microarray Suite 5.0 (MAS 5) by taking into account the experimentally validated value of weighting feature intensities to determine an overall probe set summary.

Similar to other model-based approaches, PLIER accounts for the systematic differences in intensity between features by including parameters describing these differences. These parameters are termed "feature responses,[1]" and one such parameter is included in the model for each feature (or pair of features, when subtracting Mismatch (MM) intensities). Feature responses represent the relative differences in intensity between features hybridizing to a common target (Figure 1).

Figure 1: A probe set containing probes with varying feature response is helpful in detecting a range of abundance, as illustrated by spiked-in concentrations. For example, strongly responsive features are informative at the low end, but saturate at the high end. Weakly responsive features are uninformative at the low end, but are informative at the high end of the abundance range.

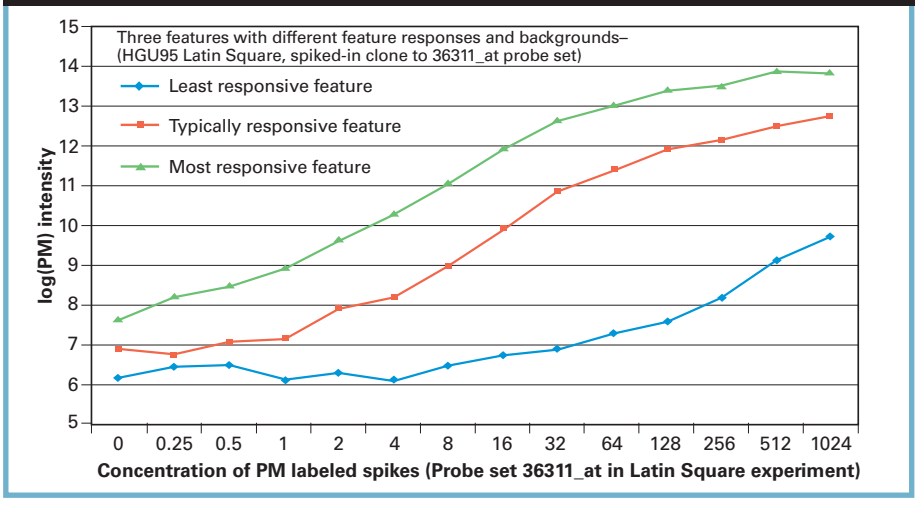
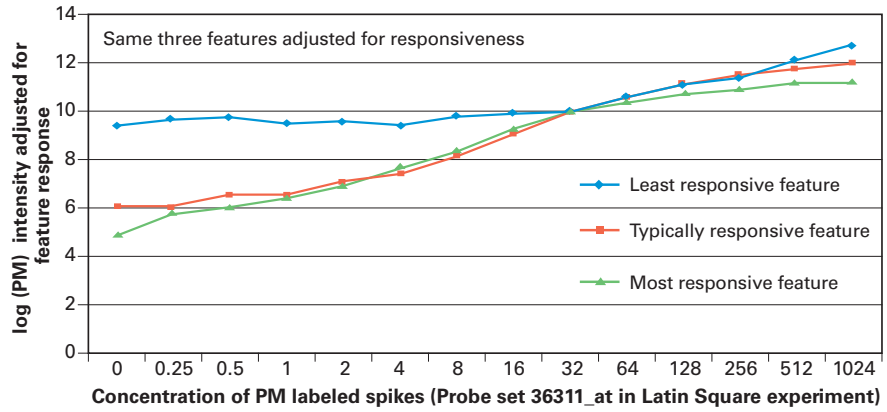


Figure 2: Once intensities are scaled using feature response, it becomes easier to detect non-systematic differences among features, and down-weight low performance features. Only the rescaled Perfect Match probes intensity is plotted here, and the actual algorithm incorporates background information. Note that background plays a large effect in the least responsive feature, and that the most responsive feature saturates first.



PLIER produces a more accurate probe set signal by utilizing these feature responses to interpret intensity data, dynamic weighting by empirical feature performance, and handling error appropriately across low and high target abundance.

Feature responses are calculated using experimental data across multiple arrays. PLIER also uses an error model that assumes error is proportional to the observed intensity, rather than to the background-subtracted intensity. This ensures that the error model can adjust appropriately for relatively low and high abundance of target nucleic acids.

[1] In Irizarry et al, this term is called the "probe affinity", by analogy to target binding. This is not physically accurate, due to the many factors which interact to produce measured intensity.

NOTE: Refer to Table 2 for a comparison of PLIER to other analysis methods.

Calculation of Signal

PROBE BEHAVIOR AND WEIGHTING

The PLIER algorithm utilizes experimental data generated across multiple arrays in order to identify and account for observed patterns in feature intensities. Feature responses can be fit from experimental data, which allows good calibration of performance in different ranges of abundance, as well as making it easy to identify poorly performing features with erratic hybridization behavior. Feature responses are a measure of how much the relative intensity of a feature is due to the feature itself, as opposed to the common target of a probe set. This relative response is influenced by how likely a probe is to bind to a complementary sequence across a range of abundance, as all probes have different thermodynamic properties and binding efficiencies, but also is influenced by many other factors, such as non-equilibrium washes, labeling, and density of synthesis. Feature response is therefore an empirical factor local to each probe set, not a global measure of probe binding. Note that it is not necessary to know the sequence of the probes within a feature to fit feature responses.

By using a scaling factor (feature response) to account for this difference in intensities, the intensity of all of the features within a probe set can be easily compared (Figure 2). For example, if one feature (or feature pair) is consistently twice as bright as others in the set, accounting for this enables the intensity data of that probe to be analyzed consistently with the others for their response to the common target. In the case of a probe set, this enables all set members to be compared and combined accurately.

Once the systematic differences between features have been accounted for in this way, it becomes easier to detect non-systematic differences. Features can then be classified as high or low performance fea-

Figure 3: The information from all three feature pairs is combined to generate the signal estimate. Each feature pair "votes" for a particular target response value (lines 1-3). Each possible value for the target response (plotted along the x-axis) has a penalty (lack of fit) when compared to the intensity of a given feature pair. The three individual penalties are added up (line 4), and the PLIER estimate of signal is the target response value with the lowest penalty (best fit to data). Note that the feature with the lowest feature response is only useful in this instance for excluding high signal values (because the feature reponse is so low that at a 4 pM spiked-in concentration the perfect match intensity is mostly background). The remaining two features estimate slightly different values individually due to noise. Note that the x-axis (possible target response values) is logarithmically scaled, so that 5 is 2^5.

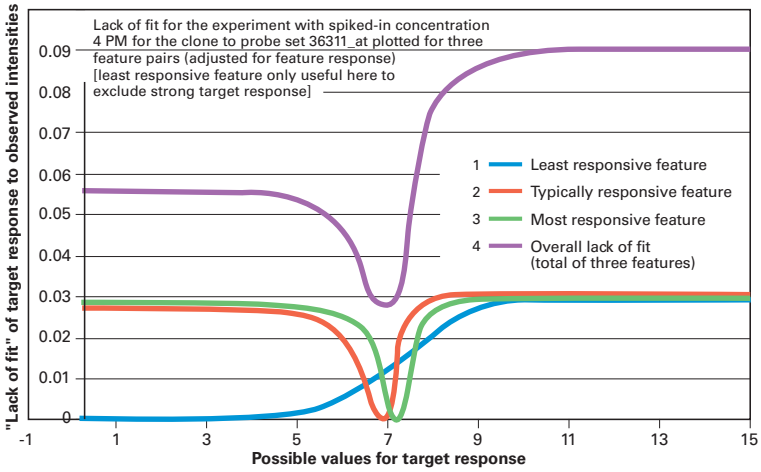
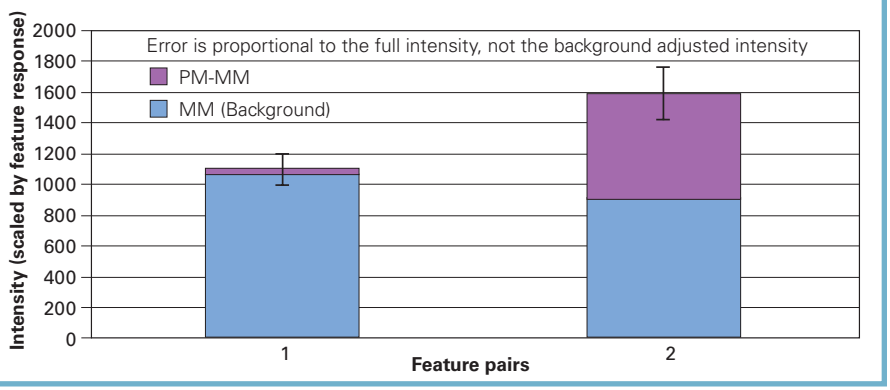


Figure 4: The amount of uncertainty (or potential error in estimation) present is dependent upon the total intensity, not just the background-adjusted intensity. The proportion of intensity represented by background is dependent upon the abundance of the complementary target as well as the feature response. Thus, the small PM-MM difference in the first feature pair has an error bar comparable to that of the large PM-MM difference in the second feature pair, and so is much less informative than might be expected. Small differences between relatively large numbers contain very little information, and therefore are effectively down-weighted by PLIER.



tures. For example, a feature that is consistently twice as bright as the median feature intensity is considered a high performance probe (and strongly responsive). However, a feature that is twice as bright only half the time but is the same as the median feature intensity the other half of the time is considered a low performance feature due to inconsistent behavior. This type of observation is often due to cross-hybridization effects creating erratic patterns.

Once the scaling factor is applied, the intensity data from each feature can be combined to generate a "goodness of fit" curve for possible signal values (target responses, see glossary). In this process, the most informative features provide the strongest contribution to the signal. At the low end of target abundance, where background dominates the intensity, these are most likely the probes with the highest feature response. Inconsistent features are down-weighted, using the Geman-McClure function. Through the scaling and down-weighting process, the features that are most informative in each experiment contribute the most to the final signal estimate (Figure 3).

ERROR MODEL

The challenge with error estimation is developing a model that fits both low and high abundance because the error is directly proportional to the largest component of intensity. At high abundance, error is approximately proportional to background-adjusted intensity, since most of the intensity is due to the response to the specific target. At the low end, however, error is approximately proportional to background intensity (which varies from feature to feature), as it is the largest component of the observed intensity (Figure 4). Due to this latter effect, it is inaccurate to treat error in target response as a proportion of background-adjusted intensity in all circumstances. Therefore, the error model must be able to estimate error in the target response depending on the abundance. If not, the amount of error at the low end will be underestimated.

The PLIER error model smoothly transitions between the low end, where error is dependent upon background, and the high end, where error is dependent on the response to the target.

Input Data

FEATURE INTENSITY

The algorithm requires feature-level intensity data as the input. From these feature-level data across many experiments, the algorithm calculates feature response terms. Once generated, these terms may be stored in a file, termed a model file, for repeated use on different data sets that are sufficiently similar to the reference set.

MODEL FILES

As explained above, a model file can be generated to store feature responses (scaling factors) once the feature responses have been generated using selected data. The most informative model files are those that have been generated using a large experimental data set across multiple relevant samples. Even though a model file can technically be generated from as few as two arrays, the minimum recommended number of arrays is five, although this is dependent upon the specific experiment. Additional recommendations include having:

- A broad range of abundance across experiments for each target of interest
- Comprehensive coverage across genes of interest

These criteria provide the algorithm with a reasonable amount of data to estimate both feature and target response. The higher the number of reliable input data points, the higher the accuracy of the resulting signal calculations. For example, informative model files can be generated using data from a control tissue panel that contains a wide variety of data points that represent genes of interest. If genes of interest are not well represented within the model file data set, the algorithm cannot model feature response accurately for these genes. If an experiment uses highly different tissues, these can be added to the tissue panel and a new model file can be generated.

If a model file is being generated from a small number of arrays or set of arrays with poor or uncertain abundance or gene coverage, the addition of a Bayesian probe penalty is recommended. This term informs the algorithm how heavily to weight the input experimental data. For example, a maximum value for this term will instruct the algorithm to ignore empirical data and treat all probes as though they were equally responsive. Use of this penalty helps mitigate the risk of treating anomalies, such as scratches or cross-hybridizations, as accurate data. The challenge with a small data set is not having enough data to recognize these types of problems as outliers.

Step-by-Step Selection of Options

NOTE: In the PLIER SDK, model file input/output, normalization, and background calculation are left for implementation wihtin the hosted application.

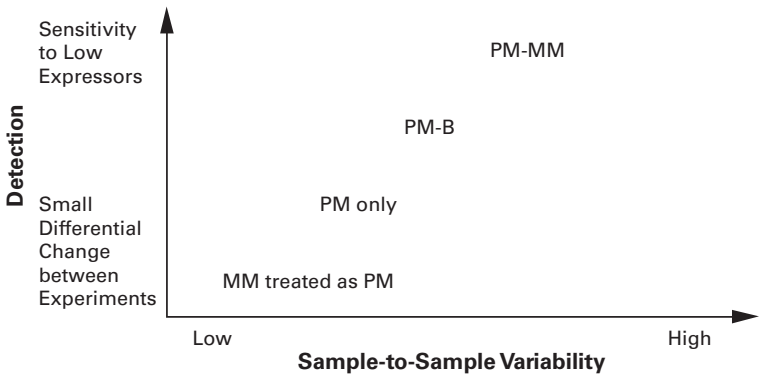
DATA PRE-PROCESSING

A normalization procedure should be applied either to intensity data before running the PLIER algorithm or a normalization procedure can be applied to the signal after the signal is generated. One common, powerful normalization procedure that operates on the raw intensity data is quantile normalization. Quantile normalization assumes a common distribution of intensities across arrays. Because of this assumption, it is designed for use across reasonably homogenous samples. As this is most often the case, quantile normalization is the preferred option. If samples are derived from significantly different tissues, quantile normalization should not be applied at this stage and normalization should be performed on the summary values after running PLIER.

PROBES AND BACKGROUND

Input probe types and background are the next primary consideration. By adjusting the input probe options and type of background calculated, the algorithm can be tuned to target sensitivity to low expressors or identification of small differential change.

Figure 5: Probe and background comparison



However, this selection should be guided by the expected sample-to-sample variability. MM probes as background, a global uniform background, or a GC-content-based background can be provided as the user requires. Recommended options are discussed in detail below.

PM-MM

As the most conservative approach, the default option is the perfect match probe feature intensity minus a corresponding mismatch probe (PM-MM) feature intensity. Since perfect match features and mismatch features are similar in both location and sequence, the difference between the intensities of a matched pair isolates target response from many background effects. This option is designed to handle high sample-to-sample variability and maximize sensitivity to low expressors by minimizing bias. For example, a neurogenesis study utilizing samples from different locations investigating low-level early responses would utilize the PM-MM option.

PM-B

The perfect match minus background option (PM-B) is useful for moderate sample-to-sample variability, and moderate sensitivity to low expressors. For example, an experiment utilizing different tissues from the same lab assessing low to medium expressors would leverage this option. For this option, the background subtraction can be calculated using any standard host pack-

Table 1: Summary Table

Method	Assumptions	Benefits	Drawbacks
PM-MM	Background effects are large and potentially variable between features across experiments relative to effects of interest	Background effects minimized due to low bias Sensitivity to low expressors	Slightly noisier when signal is higher than background
PM-B	Features have approximately the same background	Low noise	May not represent all probe sets accurately, typically leading to underestimated differential change
PM Only	Background variation is insignificant	Low noise Approximately constant CV	All probe sets biased Compression of differential change at the low end
MM treated as additional PM	Background variation is insignificant Abundances moderate to large	Added statistical power Low noise Constant CV	All probe sets biased Compression of differential change at the low end

Table 2: Other analysis methods

Method	Assumptions	Benefits	Drawbacks
PLIER	Multiple array analysis Mixed error model PM-MM, PM only etc. Multiple background options Smoothly handles intensities below background	Higher reproducibility of signal (lower coefficient of variation) without loss of accuracy relative to single array analysis Higher differential sensitivity for low expressors Lack of bias	Computationally intensive In cases where feature intensities disagree, may have more than one solution Performance relative to amount of model data provided Variance not stable on log scale
dCHIP	Multiple array analysis Arithmetic error model PM only (stanardly) Multiple background options (no background typical)	Higher reproducibility of signal over single array analysis Good differential change detection Variance stable on log scale with no background	In cases where feature intensities disagree, may have more than one solution Performance relative to amount of model data provided Positive bias at low end (compression of Fold Change)
RMA	Multiple array analysis Multiplicative error PM only Attenuated global background (single global background used to adjust for each intensity)	Higher reproducibility of signal over single array analysis Good differential change detection Variance stable on log scale	In cases where feature intensities disagree, may have more than one solution (mitigated by median polish) Performance relative to amount of model data provided Positive bias at low end (compression of Fold Change)
MAS 5	Single array analysis Multiplicative error PM-MM Background imputed to handle negative differences	Conservative Smooth down-weighting of outliers Positive output values Minimal bias	Limited by single array analysis Variance not stable on log scale Some positive bias

age method, such as a uniform percentage or spatially smoothed percentage method.

PM ONLY

The Perfect Match only option (PM Only) is most often utilized for experiments where the background is assumed to be minimally variable across experiments. Identical tissues, organisms, or cell lines are all examples of experimental designs that would benefit from this analysis option.

MM TREATED AS ADDITIONAL PM

The Perfect Match added to the Mismatch option (PM+MM) can be applied in cases where background can be assumed to be irrelevant to target response. In these cases, the mismatch probe can be utilized to increase statistical power by doubling the number of data points. The use of feature response terms in the algorithm enables the use of the mismatch probe features as either a measurement of background or a measurement of signal, as applied here. When used as a measurement of signal, the mismatch is simply treated as a less responsive feature. An example of this type of input sample set would be tissues from the same organisms from the same lab, such as inbred mice. This option is suitable for detecting small differential changes in expression value.

Figure 5 summarizes the way in which these options can be tuned. Table 1 summarizes the benefits and drawbacks of each option. The description of assumptions helps define the appropriate context for application.

The last option is quick or full optimization. Full optimization provides better sensitivity for low expressors but requires a slightly longer computational time. The quick version provides an output similar to RMA, and so is quite robust and requires less computational time.

Post-Processing

Once the PLIER algorithm has been run, the appropriate scaling and normalization options can be applied.

INTERPRETING OUTPUT

The signal value provided by PLIER is an estimate of the common response of the features in a probe set to a given target. Some transcripts are absent, and by design, PLIER provides near-zero values for targets corresponding to such transcripts. One of the most common ways to analyze these values is by using ratios. Ratios with denominators near zero are inherently unstable. By design, the raw PLIER signal values are not variance stabilized. To calculate ratios, a variance stabilizing transformation, such as log of signal plus a constant, should be applied to avoid unstable values. For example, a small constant value, such as 16, can be added to all values before taking the logarithm.

Comparison to Affymetrix® Microarray Suite 5.0

While Affymetrix® Microarray Suite 5.0 (MAS 5) treats all features equally, PLIER utilizes experimental evidence to 1) weight features based on consistency, and then 2) utilize the features that are most consistent for that experiment to estimate signal. Arrays analyzed using the MAS 5 algorithm can be re-analyzed using the PLIER algorithm.

Other Popular Methods

PLIER is designed to maximize sensitivity for low expressors. As explained above, it builds upon concepts used in dChip and RMA. Table 2 summarizes how these methods compare to one another. For more detailed information, please visit: affycomp.biostat.jhsph.edu.

Versions of PLIER are also provided that generate outputs similar to dChip and RMA for experiments that would benefit from either of those approaches. Table 3 summarizes the options for either output.

Conclusion

PLIER provides an advanced, flexible framework for primary analysis of GeneChip microarray data incorporating recent advances in expression analysis. These advances include the use of feature response terms to improve the interpretation of intensities, a robust M-estimator to reject outliers, and an improved model for intensity level errors. The PLIER error model is the recommended option, as it simultaneously incorporates a smooth transition between analysis of low and high expressors, rescaling by feature responses, and resistance to outliers. However, options within the PLIER framework allow simpler models to be used that can replicate the features of other popular low-level analysis techniques, including the analysis of perfect-match only arrays.

Table 3: Options

Options	RMA Like Methods	dChip Like Methods
Data Pre-Processing	Quantile normalization	Quantile normalization
Probe Type	Perfect match only	Perfect match only
Background	% Background (Uniform)	None
Optimization	Quick	Full

REFERENCES

Li, C. and W.H. Wong. Model based analysis of oligonucleotide arrays: Expression index computation and outlier detection. *PNAS* **98**:31-36 (2001).
www.biostat.harvard.edu/complab/dchip

Rafael A. Irizarry, Benjamin M. Bolstad, Francois Collin, Leslie M. Cope, Bridget Hobbs and Terence P. Speed. Summaries of Affymetrix GeneChip® probe level data. *Nucleic Acids Research* **31**(4):e15 (2003).

Glossary

M-Estimator – An estimator obtained by minimizing a function. Similar to maximum-likelihood estimates.

Mean – The familiar arithmetic mean of several numbers. Minimizes the sum of squares.

Median – The middle rank value (or average of the two such values, if there are an even number of data points). Minimizes the sum of absolute values.

Robustness – Ability to produce reasonably accurate results in the presence of outliers.

Sensitivity – The ability to detect an entity when it is present. Two types of sensitivity are often discussed for microarrays. The first is the ability to detect a (specified) transcript as being present in a sample. The second is the ability to detect a (specified) differential change between two experiments when there is a change.

False positive – When an entity is detected and it is not present. For microarrays, there are two types often discussed. First, when an absent transcript is reflected as a significantly present output. Second, when a differential change is falsely indicated when there is no change in an expression level.

False negative – When an entity is not detected, and it is present. For microarrays, there are two types of false negative rates. First, when a transcript is indicated as being absent when it is expressed in the sample. Second, when a (specified) differential change is falsely indicated as unchanged.

Specificity – The ability to exclude an entity when it is not there. For microarrays, there are two types of specificity. First, when an absent transcript is correctly indicated as absent. Second, when an unchanged expression level is correctly indicated as unchanged.

Background – Unwanted intensity observed on an array. Sources of background include fluorescence of the glass, stray DNA, and many other sources. Varies with sequence of DNA in the probe.

Residuals – Real experiments have noise, and therefore any estimates of true values will differ slightly from those observed in each data point. The difference between the observed values and the predicted values for a particular estimate.

Error model – A means of interpreting how well a model explains the observed data.

Signal – A summary value for the observed intensities in a probe set reflecting a common transcript. In PLIER, estimated as the target response that best fits the data given the feature responses.

Total response – The intensity due solely to the target of interest interacting with a given feature. Assumed in the PLIER model to consist of two components: feature response and target response, multiplied together.

Feature response – A relative measure within a probe set of the (multiplicative) difference in intensity due to a given feature being different (in location, probe sequence, etc.) than another. Assumed to be invariant across experiments for a given feature. Example: one PM-MM difference is reliably twice as bright as another PM-MM difference across experiments. The pairs differ in feature response. Feature responses cannot be compared across probe sets due to their relative nature. Feature response is a dimensionless scaling factor.

Target response – A measure within an experiment of the (multiplicative) difference in intensity due to a given experiment having a different target abundance. Assumed to be common to all features (background adjusted) within a probe set. Example: one PM-MM difference is twice as bright in one experiment as another experiment. The experiments differ in target response for this feature. Target responses cannot be compared across probe sets due to the relative nature of feature responses.

Appendix: Additional Statistical Details for PLIER

At a high level, the output of the PLIER algorithm is an improved estimate of target variation across experiments (signal) for each probe set. This signal improvement comes about in two ways:

1. By incorporating information about the individual feature(s) or feature-pairs, in particular, exploiting the fact that different features have different feature responses.
2. By incorporating an improved error model that smoothly transitions between the "arithmetic" regime in which feature intensities are near background, to the "multiplicative" regime in which feature intensities are far from background.

The algorithm can be run on a group of experiments. The feature response may be provided in a model file, in which case the program will simply generate signal estimates for each experiment, or the feature response may be calculated for the current group of experiments and saved for future use in a model file.

PLIER drafts an initial estimate of signal (target response) and feature response for each experiment and feature using the data provided, and then attempts to find estimates that more closely fit the data given an approximate likelihood. The algorithm stops when it finds an estimate that cannot be made to fit the data more closely.

Justification of the PLIER M-estimator

Estimators obtained by finding estimates that minimize (or maximize, depending on sign) some function are known as M-estimators. Familiar examples of estimators falling into this class are means (minimize the sum of squares) and medians (minimize the sum of absolute values). The function that an estimator minimizes can be chosen to have desirable properties, such as computational convenience and resistance to outliers, while still usefully approximating the fit of a chosen error model.

PLIER is based on the following simple assumptions about the behavior of probes and targets. First, the abundance of a target is never negative, but can be zero, therefore the target response is always non-negative ($t \geq 0$). Second, there is a linear link between intensity (total response) and target response ($T \sim f*t$), with the feature response term (f) as the slope of this relationship. That is, the true underlying intensity is the product of the feature response (common across experiments for a given feature) and the target reponse (t) (common across intensities in a probe set).

Third, the multiplicative intensity error is the most significant source of variation, that is, the error in repeated experiments for a feature intensity is approximately log-normal ($\log(I) \sim$ normal). It is presumed that background adds to the total response to make the total intensity ($I \sim T+B$), but the background value can vary from feature to feature, experiment to experiment.

Further, the optimistic assumption that background is sequence and location dependent is made, so that mismatched probe sequences have backgrounds closely related to their corresponding perfect match probe (when they are sufficiently close), and the pessimistic assumption is made that the background varies across different samples and locations.

Thus, the reduced model ($PM-MM = f*t$) is considered, due to the variability of background between experiments and between features containing probes of very different sequence. Improving the behavior of estimates of signal for low expressors is of greatest interest, and therefore concentrating on background driven effects, such as the amount of positive bias in an estimate, and the uncertainty of values near background, is important. The mismatch features do have some positive feature response for the target, and will increase the variability of estimates far from background, but are still the most useful way of removing bias from intensity data found so far.

Considering multiplicative error on intensities, the third assumption, leads to assigning error terms to the observed intensities of both the Perfect Match and the Mismatch probe features, that is, errors e_1 and e_2 should be found that satisfy $e_1*PM - e_2*MM = f*t$. In general, good estimates for feature response and signal (target response) are expected to lead to estimates for e_1 and e_2 near 1 (that is, the data fit the observations with minimal errors).

If it were possible to have exactly log-normal errors for all the intensities on a microarray, optimizing the function $\log(e_1)^2 + \log(e_2)^2$ (by analogy to typical least-squares fitting) could be attempted. However, it is known that there are outliers on the array, and that the distribution is not precisely log-normal. Also, it is computationally inconvenient to find e_1 and e_2 given this constraint. It is stressed that e_1 and e_2 are not the actual errors on the array (since these are unknown) but simply values giving an estimate of "goodness of fit." Therefore, a simplified model of the error terms for computational convenience is sought, and the tails of the distribution are discounted to insulate estimates from outliers.

One natural simplification is to assume $\log(e_1)^2 = \log(e_2)^2$. There are two ways this relationship can hold, either $\log(e_1) = \log(e_2)$, or $\log(e_1) = -\log(e_2)$. In the first case (which corresponds to log-transformation), no solutions exist when $MM > PM$. Even in the case where MM<PM, it can be shown that the total squared error $\log(e_1)^2+\log(e_2)^2$ is larger under the first constraint than when errors are fit under the second constraint. Therefore, the second constraint $\log(e_1) = -\log(e_2)$ is placed on the errors, which results in solutions for all positive PM and MM values and all non-negative target responses (t) and feature responses(f). Recall that the approximate likelihood is being simplified for computational convenience, and these terms should not be interpreted as the actual physical errors on each observed feature intensity.

The reduced model equation with errors under this constraint is $e*PM-MM/e = f*t$, which is easily solved to produce $e=\frac{[y+\sqrt{y^2+4*PM*MM}]}{2*PM}$, where $y = f*t$. This is a very simple formula for an approximating the error in fitting estimates to the observed values. Thus, if there were no outliers, the natural estimate for signal would minimize the sum of squared residuals r^2, where $r = \log(e)$ is obtained from the formula. [Aside: Note that $r=\text{glog}(f*t, 4*PM*MM) - \text{glog}(PM-MM, 4*PM*MM)$, where glog denotes the generalized logrithm. This shows that the effective variance of the target response (t) is arithmetic when measured near background.]

Appendix: Additional Statistical Details for PLIER (continued)

This approximation is likely to be best when errors are small, and there will be outliers that do not fit this model on arrays. Therefore, outliers and areas where the model is problematic are discounted by using a function which looks like r^2 near $r = 0$, but downweights the tails. The off-the-shelf function of Geman and McClure,

$h(r) = \frac{r^2}{\left(1+\frac{r^2}{z}\right)}$, satisfies the requirements.

Thus, a simple formula for goodness of fit has been derived which can be used to estimate signal and feature response by finding those non-negative values of signal and feature response that best fit the observed data. This is the M-estimator called PLIER. Some performance metrics for PLIER (PLIER+16 - with simple variance stabilization applied) can be found at the AffyComp web site run by Rafael Irizarry. Other presentations on PLIER can be found at the Affymetrix web site, including the presentation at the Low-Level Analysis workshop, discussing bias and variance issues (www.affymetrix.com/corporate/events/seminar/microarray_workshop.affx).

Calculation of Signal

PLIER drafts an initial estimate of signal (target response) and feature response for each experiment and background-adjusted feature using the data provided. Please note that a reference table to variables used in the subsequent equations (Table 4) is provided at the end of this section.

PLIER operates by finding target responses (*t(i)*) for each experiment i and feature responses *f(j)* for each feature (pair) j that minimize the function *LL(t,f) = sum H(PM,MM, BKG, f(j),t(i))* over all *i,j*.

Define the following component functions:

$y(i,j) = f(j)* t(i)$

$q = \sqrt{(y*y) + 4*PM*MM}$

$e = \frac{(y + q)}{(2*PM)}$

or (if not using mismatch)

$e = \frac{(y + BKG)}{PM}$

$r = log(e)$

$h = \frac{r*r}{\left(1+ r*\frac{r}{z}\right)}$ [z a tuning constant for robustness]

$H(PM,MM,BKG, f(j),t(i)) = h(r)$

The algorithm will find a minimum of such a function by a high-dimensional search – starting at some assignment of values for target response and feature response, the algorithm proceeds to find better and better assignment of values.

A reference table of symbols utilized in the equations described in this and subsequent sections is provided below:

Table 4. Symbols used in the equations

Symbol	Description
t(i)	Target response at each experiment i (current signal estimate)
f(j)	Feature response for each feature (or feature pair) j
q	Intermediate value in likelihood computation
r	Residual on log-scale
PM(i,j)	Intensity of the j-th perfect match feature in the i-th experiment
MM(i,j)	Intensity of the j-th mismatch feature, in the i-th experiment
y	Total response estimate with no background
BKG	Background value (if not using MM)
h	German-McClure function for discounting residuals as applied to one probe pair
GT	Extra penalty term added to log likelihood for differences in target response
GF	Extra penalty term added to log likelihood for differences in feature response
LL	Log-Likelihood – approximation used by PLIER

Appendix: Additional Statistical Details for PLIER (continued)

Data Augmentation

Because it is possible for a zero value to be reported as a probe intensity, and it is assumed that there are multiplicative errors on probes, the first step is to add a small positive value to all the input intensity values to avoid problems.

Drafting an Initial Estimate

It is best to start the iterative procedure from a reasonably good initial estimate. For producing this estimate, the software uses Simplified Expression Analysis (SEA). One way of coming up with reasonable estimates of target response and feature response is to look at transformed data $\mu(i,j)$ and fit a simple model $\mu(i,j)=\log(t(i))+\log(f(j))+e$ by use of median polish (this is similar to how RMA operates).

One particularly simple transformation is

$$\mu(i,j) = \log\left(\frac{(PM - B) + \sqrt{(PM - B)^2 + 4*L*PM*B}}{2}\right)$$

where B is either *BKG* or *MM*, depending on which option has been selected by the user. L is an attenuation parameter ranging from 0.0 to 1.0 that controls how the data approach zero as *PM* approaches B (or becomes less than B).

Median polish is a procedure for fitting a simple additive model robustly by alternately taking the median of each column ($\mu(i,j)$-$\log(f(j))$ as the estimate of $\log(t(i))$, and then taking the median of each row ($\mu(i,j)$-$\log(t(i))$ as the estimate of $\log(f(j))$, until stable estimates of each value are reached.

This initial estimate of target response ($t(i)$) is output as signal if the user selects the "quick" option for PLIER. Note that these estimates do not minimize the PLIER goodness of fit, and are simply provided for completeness.

Finding a Minimum

One way of finding (a) minimum of such a function is to use Newton's method. Starting with an assignment of values to target responses (t) and feature responses (f), given the rate at which the function is varying, find a good step size to take to find the next assignment of values to target responses and feature responses. Newton's method turns out to be computationally intensive and requires inversion of large matrices, and therefore an approximation to the method is used instead.

Identifiability Constraint

Since the method deals with two variables multiplied by each other, it is always the case that feature responses can be multiplied by some constant and target responses can be divided by some constant and wind up with the same fit to the data. To resolve this ambiguity, it is required that $\text{sum}(f(j)) = n$, the number of features (not counting any used for bias removal). Note that feature responses are relative to the other features in a probe set and cannot be directly compared between probe sets.

Avoiding Local Minima

If good values for target and feature response have been found, so that there are no local improvements possible, it may be that the method is trapped in a local minimum, and possible improvements to the estimates should be checked for. The natural method for finding an improved estimate is to examine the values a variable can plausibly take on to see if any of them improve the estimate.

In particular, in this case, each feature pair in each experiment provides a natural estimate for alternative values for target and feature response that might be near to better minima. The reasoning is that a minimum should be near to the value that is "perfect" for some feature in some experiment. These two estimates are:

$$t'(i) = \frac{PM(i,j) - MM(i,j)}{f(j)}$$

$$f'(j) = \frac{PM(i,j) - MM(i,j)}{t(j)}$$

These values (if non-negative) are checked systematically to find if there is a possible improvement. If there is possible improvement, the search is continued from the new value using the Newton-like method. While it is possible to check all possible values, or do a line search, it is unlikely that a "good" minimum will be far from one of these estimates.

Penalties

It is often useful to modify the likelihood function by putting constraints on the variables, based on the prior knowledge that they are unlikely to be extremely different from each other. Therefore, a "roughness" penalty can be incorporated in the likelihood function that penalizes extremely unusual variable values. This is an advanced option, but included for completeness. The actual functions are the obvious:

$GT = \text{differential\_target\_response\_penalty}*(\text{sum}(\log(t(i))-\text{average}(\log(t))^2)$

$GF = \text{differential\_feature\_response\_penalty}*(\text{sum}(\log(f(j)-\text{average}(\log(f)))^2)$

$LL = \text{sum}(H)+\text{sum}(GT)+\text{sum}(GF)$

This reduces to the standard LL when the penalties are zero [no penalty for roughness].

AFFYMETRIX, INC.

3380 Central Expressway
Santa Clara, CA 95051 USA
Tel: 1-888-DNA-CHIP (1-888-362-2447)
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
UK and Others Tel: +44 (0) 1628 552550
France Tel: 0800919505
Germany Tel: 01803001334
Fax: +44 (0) 1628 552585
saleseurope@affymetrix.com
supporteurope@affymetrix.com

AFFYMETRIX JAPAN K.K.

Mita NN Bldg., 16 F
4-1-23 Shiba, Minato-ku,
Tokyo 108-0014 Japan
Tel: +81-(0)3-5730-8200
Fax: +81-(0)3-5730-8201
salesjapan@affymetrix.com
supportjapan@affymetrix.com

www.affymetrix.com Please visit our web site for international distributor contact information.

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

Part No. 701903 Rev. 1