IDENTIFYING DIFFERENTIALLY EXPRESSED GENES FROM PROBE LEVEL INTENSITIES IN LONGITUDINAL AFFYMETRIX MICROARRAY EXPERIMENTS

Dongxiao Zhu [a,b] *and Alfred O Hero* [b]

[a]Bioinformatics Program,[b]Departments of EECS, Biomedical Engineering and Statistics
University of Michigan,Ann Arbor, MI 48105

ABSTRACT

Identifying differentially expressed genes over different physiological/genetic conditions is fundamental to microarray data analysis. Most of the traditional approaches do not consider the inherent correlation structure of the repeated measurements, and hence tend to give rise to inflated statistical significance of estimated treatment effects. We propose including dependency between time points and probes into a mixed linear model for gene microarray data. The approach can be viewed as an extension to existing linear model based approaches such as ANOVA, Li-Wong's Model and linear mixed effect model proposed by Chu et al. Model fitting diagnostics demonstrate significant performance improvement for longitudinal probe level data. We illustrate our approach for an aging experiment in a mouse model for quantifying retinal gene expression.

1. INTRODUCTION

Affymetrix GeneChip is a type of high-density oligonucleotide expression array that is widely used to measure tens of thousands gene expression levels simultaneously *in vitro*. Each gene, or more generally a DNA sequence of interest, is represented by a probe set that typically consists of 10 to 25 probe pairs. Each probe pair contains a perfect match (PM) probe and a mismatch (MM) probe. The former was designed to measure the true gene expression signal intensity, and the later was designed to measure the background noise [1]. The Affymetrix GeneChip design has been very successful in reliably extracting the inherently noisy gene expression signal from the gene microarray.

Gene expression signal intensities are represented in the GeneChip as numerical values that have to be pre-processed to be amenable to higher level statistical analysis. The typical pre-processing recipe consists of three steps [2]: background correction, normalization, and summarization of expression scores. There are a number of competing methods for each step, e.g. Robust Multi-Array Average (RMA) [2], Li-Wong Model [3]. Each method has advantages and disadvantages, see [2] and [4]. The first two steps are to reduce the unwanted system variation generated during the experimental process, and the third step is to extract an estimate of true signal intensities.

Detecting differentially expressed genes is a well-developed subfield in microarray data analysis. The common practice is to treat the summarized probe set expression scores as observational data to which to apply modeling and analysis [2], [3]. A more compelling integrated framework would be to estimate differential expression levels directly from the background corrected and normalized probe level intensities. Such approaches have recently been proposed [5], [6], [7]. Barrera et al. [5] blocked out probe effect and presented a simple parametric two-way ANOVA and nonparametric Mack-Skillings test in the framework of Randomized Complete Block Design (RCBD). Chu et al. [6] employed Linear Mixed Effect (LME) model to estimate the treatment effect(differential expression) and ranked genes based on the estimated effects. These studies can be viewed as extensions of the linear regression models employed in pre-processing (e.g. [3]) to estimate the treatment effect. Integrated analyses based on probe level intensities are substantially more powerful than those based on summarized probe set expression scores [5], [6], [7].

However, to the best of our knowledge, no previous approach has tried to model the hidden correlation structure of the probe level longitudinal and other effects, which may lead to bias and inflated levels of statistical significance. This kind of longitudinal microarray data is abundant, and screening differentially expressed genes over time is often of practical interest to biologists. Applying the existing linear model approaches with independence assumptions over time is not always justified. In addition, background correction is needed due to factors such as non-specific hybridization and instrument noise [2].

Here we propose an approach to estimate the longitudinal and other treatment effects from background corrected and normalized probe level intensities. Our algorithm falls into the theoretical framework of Linear Mixed Models that can be implemented using methods of Linear Mixed Effect

This work was supported by grants from the National Institute of Health (EY01115), The Foundation Fighting Blindness, Sramek Foundation and Research to Prevent Blindness.

(LME) modeling and Generalized Estimation Equations (GEE)[8]. The LME framework has been shown to have advantages. Our method can be viewed as an extension with random array effects to probe level intensity data collected over longitudinal conditions that introduce dependency [9].

We illustrate and compare our method to other approaches using retinal gene expression data obtained from our biology collaborators in the Kellogg Eye Center, University of Michigan, Ann Arbor. This data represents approximately 45101 probe set expressions on the Affymetrix mouse genome 430 2.0 array over five time points (E16, P2, P6, P10, Adult) with 4 replicates in each. The goal of this experiment is to identify genes that are differentially expressed over time. Since multiple comparisons can be reduced to a sequence of pairwise comparisons, we focus on identifying differentially expressed genes over only two postnatal time points (P2 and P10). These two time points are of significant interest to our collaborators as the developmental genes in the retina are differentially expressed during early postnatal stages of development.

2. PROPOSED APPROACH

2.1. Background correction and normalization

When analyzing raw probe level intensity data from image processing software, the total squared variation can be decomposed into biological variation and technical variation. While biological variations leading to differential expression are of interest, technical variations, such as array effects and background noise, are not of interest to the experimenter. Background correction and normalization are both necessary to reduce the technical variation without significantly affecting measured biological variation. We follow the algorithm described in Irizarry et al. 2003 [2]: the PM probe intensities are corrected by using a global model for the distribution of probe intensities, and then followed by quantile normalization. As in [2], the MM probes are not used in this analysis.

2.2. Experimental design and models

For each gene, there are two potential experimental effects, time (τ) at 2 levels and probe (ϕ) at 11 levels or 10 levels with 4 replicates at each level. One affymetrix probe set consists of 10-11 probe pairs for each gene in the mouse genome 430 2.0 array. Therefore, the total number of probe level intensities for each gene is 2*4*11 = 88 or 2*4*10 = 80. The experimental design is a balanced two-factor design (Table 1). In this design, two effects (time and probe) are blocked out, i.e. modeled and fitted as fixed effects, and the array effect is modeled as random effect to account for variation among replicates.

We first present a general mathematical model for probe level intensity data, which connects previous approaches to our approach. In a general model, both random and fixed effects can be estimated. In our case, microarray (α) variation is a random effect and includes the accumulation of small experimental sources of noise. Time (τ), on the other hand, is a fixed effect since it is due to a biological variation between the two time points in this study. The probe effect is also a fixed effect but it is not of direct interest to the experimenters.

Define the Affymetrix probe intensity y_{ijk} for a specific probe at time $i = 1, 2$, treatment $j = 1, \ldots, J, J = 10$ or 11, and replicate $k = 1, \ldots, K$. The general model for the probe set response at a particular oligonucleotide location on the GeneChip array is given by the Linear Mixed Effect (LME) model:

$$y_{ijk} = \mu + \tau_i + \phi_j + (\tau\phi)_{ij} + Z_{ij}\alpha_{ijk} + \epsilon_{ijk}, \qquad (1)$$

$$i = 1, 2, j = 1, \ldots, 11, k = 1, \ldots, 4,$$

where μ is a global offset affecting all replicates, probes, and time points for this oligonucleotide. The quantities τ_i, ϕ_j, $(\tau\phi)_{ij}$, are fixed (non-random) effects modeling temporal effect, probe effect, and mixed temporal-probe effect. The quantities α_{ijk} and ϵ_{ijk} are mutually independent zero mean random variables with variances σ^2_{ij} and η^2_{ij}, respectively. In this paper, these random effects will be assumed to be Gaussian-distributed. Z_{ij} is the fixed non-random covariate of α_{ijk}, which models possible longitudinal dependencies and also possible dependencies over 10 or 11 oligos in the probeset. Throughout this paper, it is assumed that different probesets have independent responses.

In the standard ANOVA model, the general model (1) reduces to a linear fixed effect model with no random effects $Z_{ijk} = 0$.

$$y_{ijk} = \mu + \tau_i + \phi_j + (\tau\phi)_{ij} + \epsilon_{ijk}, \qquad (2)$$

In this case, the two-sided paired-t test is the generalized likelihood ratio test (GLRT) for differential expression over time, which can be stated in terms of testing the hypotheses:

$$H_0 \ : \ \tau_1 \ = \ \tau_2 \quad \text{versus} \quad H_\alpha \ : \ \tau_1 \ \neq \ \tau_2. \qquad (3)$$

The paired-t test is one of the most widespread statistical tests used to detect differential expression. Note, that by ignoring random effects, the paired-t test does not account for possible dependency in probe response over time, probeset, or array.

To account for the array random effect, the full general model (1) will be employed. In this case the maximum likelihood estimates can on longer be found in closed form. These parameter estimates are approximated by a variety of approaches including: EM algorithm applied to the profile likelihood or to the restricted likelihood [10]. These approximate MLE's can be used in the likelihood ratio test of (2) leading to a test that accounts for fixed probe and temporal effects and random array effects. We will call this model LME1.

To account for all of the random effects, including both probe, time and array, the LME model (1) becomes:

$$y_{ijk} = \mu + \tau_i + \phi_j + (\tau\phi)_{ij} + \alpha + \epsilon_{ijk}, \qquad (4)$$

Note that α is a random variable with variance η^2 that correlates across all effects. Again, there is no closed form for the maximum likelihood estimates under this model. We will implement the method of Generalized Estimating Equations (GEE), described in more detail below, to obtain a test which we will call LME2.

The LME2 method approximates the maximum likelihood estimates by a iteratively reweighted least square approach [11] applied to the marginalized likelihood. Under the assumption of independence of different gene probes, the marginalized likelihood factors into a product of the densities:

$$\mathrm{f}(\mathbf{y}|\theta) = \int \mathrm{f}(\mathbf{y}|\theta, \alpha)\mathrm{f}(\alpha|\theta)\mathrm{d}\alpha, \qquad (5)$$

where $\mathbf{y}$ is a matrix consisting of y_{ijk} elements, $\theta = \{\mu, \tau_i, \phi_j, (\tau\phi)_{ij}, \eta\}$ are fixed effects, and $\mathrm{f}(\mathbf{y}|\theta)$ can be represented as a multivariate Gaussian density with mean θ and covariance matrix $\Lambda(\eta)$ of known form. The marginal log-likelihood is expressed as

$$\mathrm{logf}(\mathbf{y}|\theta) = -\frac{\mathbf{1}}{\mathbf{2}}\mathrm{tr}\{(\mathbf{y}-\mathbf{\Psi})^{\mathbf{T}}\mathbf{\Lambda}^{-\mathbf{1}}(\eta)(\mathbf{y}-\mathbf{\Psi})\} - \frac{\mathbf{1}}{\mathbf{2}}\mathrm{log}|\mathrm{det}(\mathbf{\Lambda}(\eta))|, \qquad (6)$$

where $\mathrm{E}[\mathbf{y}|\theta] = \mathbf{\Psi}$ is a matrix composed of the unknown fixed effects $\mu, \tau_i, \phi_j, (\tau\phi)_{ij}$. The LME2 model method alternates between estimating covariance $\mathbf{\Lambda}(\eta)$ and estimating $\mathbf{\Psi}$ [10]. The tests under LME1 and LME2 can be implemented using the R function `lme()` and `gee()` using the following inputs:

- The mean of y_{ijk}, $\mathrm{E}(y_{ijk}) = \mu_{ij}$,is related to the covariates for fixed effects by a Gaussian link

$$\mu_{ij} = \tau_i + \phi_j + (\tau\phi)_{ij} + \epsilon_{ij}, \qquad (7)$$

- The variance of each y_{ijk}, given the effects of covariates, is σ_j^2. Under a Gaussian assumption, the variance σ_j does not depend on the mean response. That is,

$$\mathrm{Var}(y_{ijk}|\alpha_{ijk}) = \sigma_j^2, \qquad (8)$$

- The temporal correlation over the five sample times is modeled as first-order autoregressive,

$$\frac{\mathrm{Cov}(y_{ijk}, y_{i'j'k'})}{\sqrt{\mathrm{Var}(y_{ijk})\mathrm{Var}(y_{i'j'k'})}} = \rho^{|i'-i|}. \qquad (9)$$

The correlation coefficient ρ is a nuisance parameter.

For the two sample comparisons considered here, equation (9) reduces to a single correlation coefficient having values ρ or 1.

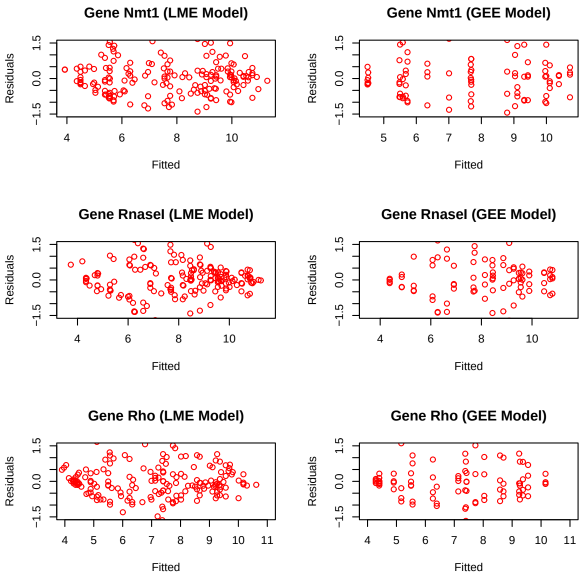


Fig. 1. *Model fitting diagnostics for three probesets. Each row shows residual plots in which longitudinal probe level intensities for two genes were fitted with the LME1 and LME2 model respectively.*

3. RESULTS

We distinguish two classes of "one-gene-at-a-time" methods for identifying differentially expressed genes: probe level intensity based methods and probe set expression score based methods. We aim to first show that our LME2 model for probe level intensities has equally strong discriminative power for estimating the biological main effect (τ) as the competing LME1 and Paired t methods [5], [6], [7]. We then show that for longitudinal data our LME2 model better estimates the main effect (τ) than does the LME1 model and simple ANOVA model.

Fig.1 shows residuals vs. fitted plot for the probe level data of two genes selected from a pool of 45101 genes. Following previous notations, the residual and fitted value for one probe level intensity observation are:

$$y_{ijk} - \widehat{y}_{ijk},$$

and

$$\widehat{y}_{ijk} = \hat{\mu} + \widehat{\tau}_i + \widehat{\phi}_j + \widehat{(\tau\phi)}_{ij},$$

Probeset ID	Gene Title	Gene Symbol	GEE Rank	LME Rank	Paired t Rank
1416309_at	nucleolar and spindle associated protein 1	Nusap1	1	4	413
1416474_at	neighbor of Punc E11	Nope	2	21	357
1417457_at	CDC28 protein kinase regulatory subunit 2	Cks2	3	6	671
1419998_at	Unknown	Unknown	4	9	403
1423774a_at	protein regulator of cytokinesis 1	Prc1	5	1	735
1426604_at	ribonuclease L (2', 5'-oligoisoadenylate synthetase-dependent)	Rnasel	6	11	199
1426936_at	cDNA sequence BC005512	BC005512	7	29	2658
1429051s_at	RIKEN cDNA 6230403H02 gene	6230403H02Rik	8	15	270
1434437x_at	ribonucleotide reductase M2	Rrm2	9	2	18
1434645_at	RIKEN cDNA C530008M17 gene	C530008M17Rik	10	10	244

Table 1. Top 10 down-regulated genes from Post-natal 2 days (P2) to Post-natal 10 days (P10) identified by LME2 model.

Probeset ID	Gene Title	Gene Symbol	GEE Rank	LME Rank	Paired t Rank
1419025_at	retinal S-antigen	Sag	1	3	506
1421084_at	retinoschisis 1 homolog (human)	Rs1h	2	2	40
1424256_at	retinol dehydrogenase 12	Rdh12	3	10	260
1424963_at	retinitis pigmentosa 1 homolog (human)	Rp1h	4	7	137
1425100a_at	phosphodiesterase 6G, cGMP-specific, rod, gamma	Pde6g	5	11	193
1425172_at	rhodopsin	Rho	6	8	215
1425696_at	thioredoxin-like 6	Txnl6	7	9	190
1427044a_at	amphiphysin	Amph	8	13	29
1428288_at	RIKEN cDNA 2310051E17 gene	2310051E17Rik	9	19	239
1430128a_at	deleted in polyposis 1-like 1	Dp1l1	10	4	394

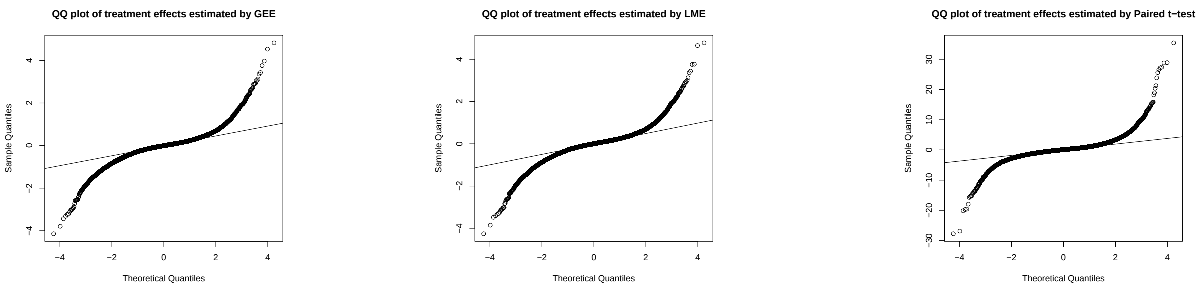
Table 2. Top 10 up-regulated genes from P2 to P10 identified by LME2 model. 7 (Sag, Rs1h, Rdh12, Rp1h, Pde6g, Rho, Dp1l1) out of 10 genes are well-known genes for the development of adult mouse retina. LME2 model misses the Pde6g, and has lower rank for the remaining five of six genes.

Probeset ID	Gene Title	Gene Symbol	LME Rank	GEE Rank	Paired t Rank
1423774a_at	protein regulator of cytokinesis 1	Prc1	1	5	735
1434437x_at	ribonucleotide reductase M2	Rrm2	2	9	18
1437750_at	RIKEN cDNA 2310037P21 gene	2310037P21Rik	3	12	277
1416309_at	nucleolar and spindle associated protein 1	Nusap1	4	1	413
1448698_at	cyclin D1	Ccnd1	5	15	71
1417457_at	CDC28 protein kinase regulatory subunit 2	Cks2	6	3	671
1438434_at	Rho GTPase activating protein 11A	Arhgap11a	7	14	543
1452242_at	RIKEN cDNA 1200008O12 gene	1200008O12Rik	8	17	1095
1419998_at	Unknown	Unknown	9	4	403
1434645_at	RIKEN cDNA C530008M17 gene	C530008M17Rik	10	10	244

Table 3. Top 10 down-regulated genes from Post-natal 2 days (P2) to Post-natal 10 days (P10) identified by LME1 model.

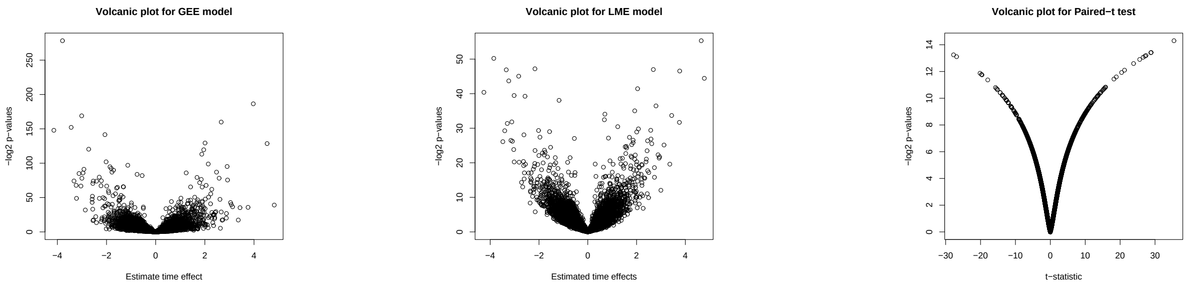
Probeset ID	Gene Title	Gene Symbol	LME Rank	GEE Rank	Paired t Rank
1434657_at	glutaminase	Gls	1	12	497
1421084_at	retinoschisis 1 homolog (human)	Rs1h	2	2	40
1419025_at	retinal S-antigen	Sag	3	1	506
1430128a_at	deleted in polyposis 1-like 1	Dp1l1	4	10	394
1438641x_at	RIKEN cDNA 1500016O10 gene	1500016O10Rik	5	14	92
1456341a_at	basic transcription element binding protein1	Gli3	6	19	178
1424963_at	retinitis pigmentosa 1 homolog (human)	Rp1h	7	4	137
1425172_at	rhodopsin	Rho	8	6	215
1425696_at	thioredoxin-like 6	Txnl6	9	7	190
1424256_at	retinol dehydrogenase 12	Rdh12	10	3	260

Table 4. Top 10 up-regulated genes from Post-natal 2 days (P2) to Post-natal 10 days (P10) identified by LME1 model. No additional relevant genes are identified.



(a) LME2 estimation (b) LME1 estimation (c) Paired-t estimation

Fig. 2. QQ plot of treatment (time) effects estimated by three different methods.



(a) Volcanic plot for LME2 estimation (b) Volcanic plot for LME1 estimation (c) Volcanic plot for Paired-t estimation

Fig. 3. Volcanic plot of treatment (time) effects and p-values estimated by three different methods.

respectively. The plot in Fig.1 is frequently used to check the goodness of fit of the model. For a good fit the distribution of the variance residuals should not be dependent on the fitted values. Overall, the residual fitting errors of LME1 models appear to have stronger heteroscedasticity (non-constant variance) than those of the LME2 models (Fig.1). The heteroscedasticity is seen from the fish shaped "residual" scatter on the left column of Fig.1. This suggests that LME2 models fit longitudinal data better than LME1 models do.

Looking into LME2 residual plots more carefully, we find that residuals of some fitted values show non random pattern. This is attributed to the fact that the total squared variation is mostly explained by only one effect, in our case, either time effect (τ) or probe effect (ϕ). The more obvious the non random pattern is, the more a single effect dominates over others. As mentioned before, in the GeneChip experimental design, probe effect is responsible for most of squared technical variation as observed by Li and Wong [3]. Biological effect such as time effect is responsible for most of biological variation, and it varies drastically from gene to gene. Therefore, we expect to see stronger non-random pattern in the LME2 residual plot for weakly differentially expressed genes while non-random patterns would occur for strongly differentially expressed genes. Fig.1 confirms this reasoning, e.g. LME2 model fitting of Gene Nmt1 has stronger non-random pattern since it is a house-keeping gene (constantly expressed gene) [12]; LME2 model fitting of Gene Rho has much weaker non-random pattern since its expression level greatly increases from P2 to P10 [13].

Fig. 2 presents the QQ plot of the estimated time effect (τ) distributions by three different methods vs. normal quantile distribution. The use of QQ plot is not to test whether the time effects follow a particular distribution, but as a visual aid for identifying genes with unusual test statistics. QQ plots informally correct for the large number of comparisons and the points which deviate markedly from an otherwise linear relationship are likely to correspond to those genes whose expression levels differ between the two time points, therefore, we can expect a large number of differentially expressed genes between two time points. Although the histograms of differential expressions estimated by the three different methods are quite similar across the three methods (Fig. 2), the relative ranks are very different as shown in Table 2 and 3 and discussed below.

We also used volcanic plots to visualize differentially expressed genes estimated from three different methods. In a volcanic plot, logs of raw p-values are plotted against the estimated fold change on a negative log scale. The p-values of the time effect in LME1 and LME2 models are calculated from Gaussian quantiles, and the p-values of the time effect in paired-t test are calculated from t quantiles. The volcanic plot better illustrates the time effect than the QQ plot since each gene is plotted in a two dimensional space. In a volcanic plot, (e.g. Fig.3a), each dot corresponds to a gene, the dots down at the bottom represent those house keeping genes, and the dots on the top (outliers) represent differentially expressed genes. Fig. 3a demonstrates that the LME2 is a more powerful test than Fig. 3b and Fig. 3c do since the differentially expressed genes in Fig.3a are better separated and have smaller p-values (see y-axis) than those in Fig.3b and Fig.3c.

In practice, biologists usually care about the genes in order of the strength of differential expression. Table 1 and Table 2 list top 10 up-regulated and down-regulated genes selected by the LME2 model, and the ranks of these genes in gene lists from other models are also shown for comparison. Similarly, Table 3 and Table 4 list top 10 up-regulated and down-regulated genes selected by the LME1 model. The top genes from probe level intensity based methods (LME2 and LME1) tend to be similar to each other but quite different from those from probe set expression score based method, i.e paired-t test.

In comparing different methods, the one that is able to pull out more relevant genes with relatively lower ranks is considered to be a better method. Much biological experimental data is available for genes that are up-regulated during retinal development. Therefore, we focus on the following discussion of the top 10 up-regulated genes (Table 2 and Table 4). Among the top 10 genes pulled out by the LME2 model, 7 genes (Sag, Rs1h, Rdh12, Rp1h, Pde6g, Rho, Dp111) (Table 2) are well known to our collaborators in Kellogg as genes governing the development of adult mouse retina as recognized by our collaborating biologists in Kellogg Eye Center. The LME1 approach missed an important one (Pde6g), which is rod specific [14]. Further, five of the remaining six genes had lower ranks (Table 2).

We also did the same comparison based on the top 10 up-regulated genes selected by the LME1 model. The results showed that the LME1 model did not identify additional relevant genes that were missed by the LME2 approach (Table 4). This confirms that the proposed LME2 approach has stronger discovery power. Compared with the paired-t test, the LME1 and LME2 approaches tend to generate closer and better results since none of these 7 genes are in the top 10 as pulled out by the paired-t test. Our analysis suggests that more extensive use of model-based approaches will be useful in identifying differentially expressed genes from Affymetrix GeneChip data.

4. DISCUSSION

Identifying differentially expressed genes from longitudinal microarray experiments remains a difficult problem. New approaches for analysis of this type of data are urgently needed. The structure of probe level intensity data fits very nicely in the framework of experimental design so that many well known methods such as mixed model analysis have potential for improving discovery from genomics data. In this paper, we demonstrated including of random effects in the probe response value can lead to significant improvement of differential expression analysis for gene microarray experiments. In particular, including the random probe and time effects can improve the array effect as proposed by [6]. The approach can be applied to both probe level (cDNA) and probe-set level (Oligo) data by properly adjusting model settings.

5. REFERENCES

[1] Lockhart D, Dong H (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays, *Nature Biotechnology*, 14, 1675–1680.

[2] Irizarry, RA, Hobbs, B, Collin, F, Beazer-Barclay, YD, Antonellis, KJ, Scherf, U, Speed, TP (2003) Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data, *Biostatistics*, 4, 249-264.

[3] Li C, Wong WH (2001) Model-based analysis of oligonucleotide arrays: expression score computation and outlier detection, *Proc Natl Acad Sci U S A*, 98, 31-36.

[4] Bolstad BM, Irizarry RA, Astrand M, and Speed TP (2003) A Comparison of Normalization Methods for High Density Oligonucleotide Array Data Based on Bias and Variance, *Bioinformatics*, 19, 185-193.

[5] Barrera, L, Benner, C et al. (2004) Leveraging two-way probe-level block design for identifying differential gene expression with high-density oligonucleotide arrays. *BMC Bioinformatics*, 5, 42.

[6] Chu, TM, Weir, B et al. (2002) A systematic statistical linear modeling approach to oligonucleotide array experiments. *Mathematical Biosciences*, 176, 35-51.

[7] Lemon WJ, Liyanarachchi S, et al. (2003) A high performance test of differential gene expression for oligonucleotide arrays. *Genome Biology*, 4:R67.

[8] Liang KY, Zeger SL (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13-22.

[9] Zeger SL, Liang KY et al. (1988) Models for longitudinal data. a generalized estimating equations approach. *Biometrics*, 44, 1049-1060.

[10] Demidenko, E (2004) Mixed Models: Theory and Applications. *Wiley & Sons*, Hoboken, New Jersey.

[11] Ralf Wolke (1992) Iteratively reweighted least squares: a comparison of several single step algorithms for linear models. *BIT*, 32:506-524.

[12] Farazi TA, Waksman G, Gordon JI (2001) The biology and enzymology of protein N-myristoylation. *J Biol Chem*, 276:39501-39504.

[13] Humphries MM, Rancourt D (1997) Retinopathy induced in mice by targeted disruption of the rhodopsin gene. *Nat Genet*, 15:216:219.

[14] Morin F, Vannier B et al. (2003) A proline-rich domain in the gamma subunit of phosphodiesterase 6 mediates interaction with SH3-containing proteins. *Mol Vis.*, 9:449-459.