On Finding and Interpreting Patterns in Gene Expression Data from Time Course Experiments

Yvonne E. Pittelkow and Susan R. Wilson

Mathematical Sciences Institute, Australian National University
Canberra ACT, Australia 0200

Abstract. Microarrays are being widely used for studying gene activity throughout a cell cycle. A common aim is to find those genes that are expressed during specific phases in the cycle. The challenges lie in the extremely large number of genes being measured simultaneously, the relatively short length of the time course studied and the high level of noise in the data. Using a well-known yeast cell cycle data set, we compare a method being used for finding genes following a periodic time series pattern with a method for finding genes having a different phase pattern during the cell cycle. Application of two visualisation tools gives insight into the interpretation of the patterns for the genes selected by the two approaches. It is recommended that (i) more than a single approach be used for finding patterns in gene expression data from time course experiments, and (ii) visualisation be used simultaneously with computational and statistical methods to interpret as well as display these patterns.

1 Introduction

DNA microarrays have enabled the simultaneous monitoring of the expression patterns of thousands of genes during cellular differentiation and response. A major challenge has been to find and interpret patterns in these massive data sets. Cluster analysis is one of the main methodologies used to study such data and find patterns. It is well known that there is no straightforward, rigorous way to quickly extract clusters from complex, high-dimensional data and hundreds of algorithms have been proposed [1]. As noted in [2] 'All the [cluster analysis] methods have their limitations and weak points. That is why it is so important to look at the clustering problem from multiple perspectives'.

One widely used application of DNA microarrays is for the study of the cell cycle transcriptome. For some time it has been clear that certain genes are expressed at specific stages of the cell cycle [3]. So these genes show a periodic pattern of expression when monitored during consecutive cell cycles. Clustering methods have been proposed for finding patterns in these time course data; including self-organizing maps [4], principal component analysis [5], amongst other methods (see [6]). Alternative approaches specifically for finding relevant, periodic patterns in time course data have been proposed; see, for example, [3], [7], [8], [9]. Generally, the foremost aim for these latter methods is to find cell cycle regulated genes.

M. Chetty, A. Ngom, and S. Ahmad (Eds.): PRIB 2008, LNBI 5265, pp. 276–287, 2008.

When searching for, and evaluating, patterns of gene expression, a statistical modelling approach has the following, major advantage compared with non-statistically based, computational modelling approaches that underpin most clustering methods. Namely, the type of patterns being found can be compared in a rigorous manner [10]. For time course data, recently we proposed a novel statistical model for selection of genes with different phases and/or amplitudes [9] and compared the results of applying this approach to the Cho *et al* data with the results previously published ([11], [3]).

As noted above, there are limitations for all approaches, and it is important to use and compare different perspectives. So we focus here on two approaches for finding relevant patterns in gene expression time course data, namely (i) Fisher's exact test for hidden periodicities of unspecified frequency [12], [7], and (ii) the absolute sine model [9]. A valuable step when analysing genome-wide expression is to visualise the results from the analysis in such a way as to facilitate interpretation of the data, including the patterns found, as well as those not found [6]. Here two useful visualisation tools for interpreting patterns simultanously in a large number of expression measures are used, the h-profile plot [9] and the *GE*-biplot [14].

The next section summarises the two approaches for finding patterns in time course data and the visualisation tools, and the Cho et al [11] data are described in section 3. The statistical approaches and visualisation tools are applied to these data in section 4, and the results presented there illustrate the usefulness of the two statistical approaches for finding different types of patterns in the time course data, and of the visualisation methods for interpreting these patterns.

2 Methods Summary

2.1 Fisher's Exact Test for Hidden Periodicities (FET)

Consider an observed time series $z_1, ..., z_N$ of gene expression values (possibly transformed). Fisher devised an exact procedure based on the periodogram to test the null hypothesis of Gaussian white noise against the alternative of an added deterministic periodic component of unspecified frequency. Basically, the null hypothesis will be rejected if the periodogram contains a value significantly greater than the average value; see 10.2 in [12]. Writing $[r]$ for the integer part of [r], the statistic is given by

$$\xi_r = \{\max_{1 \le k \le [r]} I(\omega_k)\} / \{\Sigma_{k=1}^{[r]} I(\omega_k)\}$$

where

$$I(\omega_k) = N^{-1} |\Sigma_{t=1}^{N} z_t e^{-it\omega_k}|^2, \ \omega_k = 2\pi k/N.$$

In [7], $[N/2]$ is used for $[r]$, but $[(N-1)/2]$ is the correct form; see [12], [13]. The significance level for the corresponding test is given by

$$P(\xi_r \ge x) = 1 - \Sigma_{j=0}^{[r]} (-1)^j \binom{[r]}{j} (1 - jx)_+^{[r]-1}$$

where $y_+ = \max(y, 0)$.

This procedure has been applied to multiple time series data derived from microarray experiments ([7], [15], [8]). For this application, it is referred to as the '(g)-statistic' and 'g test'. The challenge of multiple testing is addressed using the False Discovery Rate (FDR) that controls the expected proportion of false positives. If G genes are considered, first the corresponding p-values are ordered, $p_{(1)}, ..., p_{(G)}$ with corresponding genes $g_{(1)}, ..., g_{(G)}$, then j_q, the largest j such that $p_{(j)} \leq (j/G)q$, is determined. The null hypothesis is rejected for genes $g_{(1)}, ..., g_{(j_q)}$. This controls the FDR at level q. The GeneCycle package in R [16] implements the approach outlined in [7] that we refer to as FET-gs. Note that during revision of the paper, GeneCycle was updated to use the correct form of [r].

2.2 Absolute Sine Model (ASM)

Many genes, rather than having periodic behaviour instead may be following a different pattern in each cycle. Based on many of the profile plots that appeared to depict this behavior, we proposed the following model [9] for gene expression data

$$Z_t = |Asin(2\pi Kt + L)|,$$

where A is the amplitude, K the period and L the part of the cycle at time zero, i.e. related to phase. This model allows the selection of genes with different phases and/or amplitudes.

If $K = 1$ and time t is scaled to run from 0 to 1, there are exactly two cycles so that genes whose profiles complete two cycles and have approximately equal amplitude in both cycles will be selected. A scale free estimate of residual error $RSS = \sum_t ((z_t - \hat{z_t}^*)/\hat{A})^2$ is obtained, where $\hat{z_t}^* = |\hat{A}sin(2\pi t + \hat{L})|$. To assess fit based on RSS, simulation was used to determine the quantiles of the distribution of RSS, denoted by $\hat{c}$. Estimates of RSS_g for all genes, g, were calculated and compared to $\hat{c}$. All genes, where $\hat{RSS}_g \leq \hat{c}$ were selected as being compatible with the two cycle model, where different values of $\hat{c}$ select genes with more or less compatability with the model.

Values of $\hat{A}$ provide information on the extent of gene expression change (amplitude), and the corresponding value of t is an estimate of the time of maximal expression. The value of the intercept parameter $\hat{L}$ gives an estimate of the expression at the beginning of the cycle.

2.3 Visualisation Tools

It is advised that 'an essential *first* step [ ] when considering *any* time series is simply to plot the observation against time' [17]. This is straightforward if one is considering a handful of time series data, but when there are hundreds, or even thousands, of series it is more problematic. Two visualisation tools, the h-profile plot and the covariance-biplot (and a variant called the *GE*-biplot) are useful in this setting, and are now described.

Let $\mathbf{Z}$ be the matrix of expression values, or functions of gene expression values, with G 'genes' in the columns and N microarrays (one for each time point

in this application) in the rows, such that the column (gene) means are zero. To illustrate ideas, we describe the methods as if Z contained gene expression values and refer to the columns as 'genes', although this is not strictly correct.

Let the SVD of $\mathbf{Z}$ be $\mathbf{Z} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T$, where $\mathbf{U}$, size N by N, and $\mathbf{V}^T$, size G by G, are orthogonal matrices such that $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ and $\mathbf{V}^T\mathbf{V} = \mathbf{I}$ (where $\mathbf{I}$ is used to denote a conformable identity matrix).

In both plots, the coordinates for all genes, in d dimensions, are defined as

$$\tilde{\mathbf{G}}_d^T = \frac{1}{\sqrt{N-1}}\mathbf{\Lambda}_d\mathbf{V}_d^T,$$

where $\mathbf{U}_d$ and $\mathbf{V}_d$ are matrices comprising the first d columns of $\mathbf{U}$ and $\mathbf{V}$ respectively and $\mathbf{\Lambda}_d$ is a sub matrix of $\mathbf{\Lambda}$ formed from the first d columns and rows of $\mathbf{\Lambda}$. For two dimensional representation $d = 2$ and then $\tilde{\mathbf{G}}_2^T$ consists of pairs of co-ordinates, one for each gene, defining the location of gene points on the horizontal and vertical axes.

Genes, represented by gene points, have expression values which increase in variance as distance from the origin increases. The angular separations between the gene points are approximations to cosines of correlation between the gene expression profiles. So genes that are highly correlated will lie approximately on a line passing through the origin, with those that are positively correlated on the same side of the origin, while those that are negatively correlated lie on the opposite side.

In the h-profile plot reduced versions of a plot (thumbnail) for each gene is placed at the gene points. When the reduced plot is a time series graph, and Z contains the gene expression values, profiles of similar 'shape' are located together, while those of 'reversed' shape lie on the opposite side of the origin.

For the covariance-biplot, where microarrays and the genes are simultaneously displayed, the gene coordinates are as in the h-profile plot and the microarray coordinates are given by

$$\tilde{\mathbf{C}}_d = \sqrt{N-1}\mathbf{U}_d.$$

When Z consists of gene expression values that have been logged, standardized over each microarray, and finally column mean corrected, the covariance-biplot is called the *GE*-biplot [14].

In these biplots, the scalar product between the t^{th} row point (microarray point) and g^{th} column point (gene point) with respect to the origin is approximately equal to the $(t, g)^{th}$ element, $z_{t,g}$, of $\mathbf{Z}$. The juxtapositions of the gene points to the microarray points provides an approximation to the value of the (transformed) gene expression values on the microarrays. The inner product can be viewed geometrically as the product of the signed length of one of the vectors and the length of the projection of the other vector onto it. Thus if a gene point is close to a microarray point then the gene will be relatively up regulated in that microarray and if the gene point is on the opposite side of the plot to the microarray point then the gene will be relatively down regulated on that microarray. The accuracy of these predictions depends on how good the approximation is in the lower ranked space; two measures of fit, I_1 and I_2, ranging from 0 to 1, can be determined [14], [9]. R source code is available at [18].

For time course data, one would expect those microarrays falling in the same cell phase to be relatively close to one another, and the greater the separation between the different cell phases, the greater the distance between the corresponding microarrays.

A novel application of these plots for time series data demonstrated in this paper, replaces the gene expression values in the columns of Z by the residuals arising after first fitting 'some model or other' to the series. Such (initial) model fitting is often necessary in time series analyses [17].

3 Mitotic Cell Cycle Data

The aim of the time-course experiment described in Cho *et al* [11] was to characterize mRNA transcript levels during the cell cycle of the budding yeast S. *cerevisiae*. Synchronous yeast cultures were arrested in late G1 and the cell cycle re-initiated with cells collected at 10 minute intervals, covering two full cell cycles. The time course was divided into early G1, late G1, S, G2 and M phases based on the size of the buds, the cellular position of the nucleus, and standardization to known transcripts.

For our analyses of these data, the negative values were truncated to .01, the data logged using base 2 (as is commonly done, see for example [7]), and finally standardized so that, for each microarray, the mean over all values is zero and the corresponding variance is one. This latter transformation effectively 'normalizes' the distributions so that the first two moments of the distributions on each microarray agree. Control genes were removed leaving a total of 6565 genes. Although not technically correct the transformed gene expression is often referred to as simply 'gene expression' to avoid cumbersome phrases. Preprocessing can have a large impact on analyses, but such considerations are beyond the scope of this paper. The sample at time zero, that is immediately after arrest, was eliminated from the following analyses, leaving 16 microarrays, one at each 10 minute interval, from time 10 to 160 minutes.

4 Results

Using FET-gs in the GeneCycle package [16] with an FDR of 0.05, 532 genes were selected. On the left of Fig.1 is the *GE*-biplot where the genes are shown as symbols, marking their positions relative to the microarray points which are shown as numerals indicating the time in minutes. The microarrays are coloured according to their cell phase determined by Cho *et al*, and the same colouring is applied to the genes according to the phase in which the time of maximum (TOM) occurred. TOM was determined by averaging the four values in each phase and then finding the maximum. On the right of Fig.1 is the h-profile plot using a periodogram as the thumbnail [16], with different colours differentiating the estimates of k. The measures of fit are $I_1 = 0.65$ and $I_2 = 0.88$.

Previously, in the *GE*-biplot for these data using genes selected by ASM [9], microarrays allocated to the different coloured phases appeared in the same

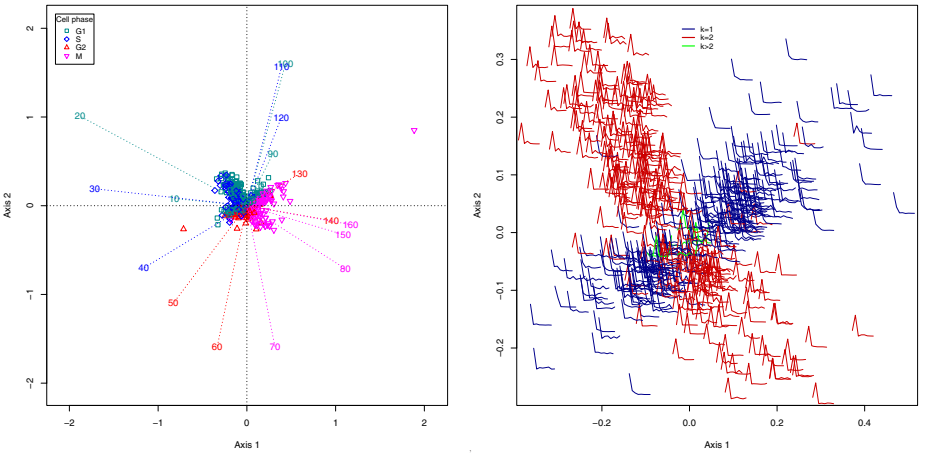

Fig. 1. On the left is a *GE*-biplot using the 532 genes selected by FET-gs (FDR=0.05). The microarray points are shown as coloured numerals (time in minutes) indicating the phase determined by Cho *et al* (see legend). The genes are shown as coloured symbols, marking their positions relative to the microarray points, where the colour (and symbol) analogously are according to the phase in which time of maximum occurred. On the right is a corresponding h-profile plot (outliers removed) showing the periodograms for 531 of these genes (after removing the outlier). The periodograms are coloured to differentiate the k values (see legend, and the equations in Section 2.1).

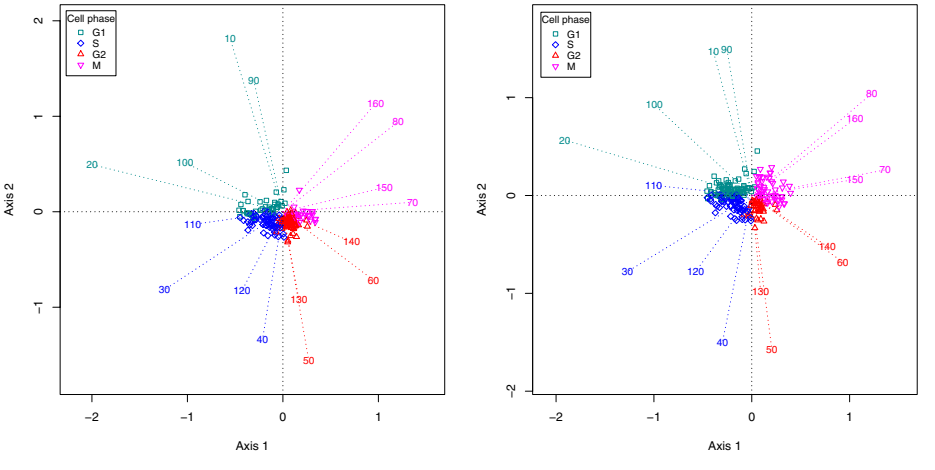

Fig. 2. On the left is the *GE*-biplot using the 221 genes determined by ASM ($\hat{c} = 0.6$) and on the right the *GE*-biplot using the 254 genes determined by FET-gs (FDR=0.05) restricted to $k = 2$. The colours, numbers and symbols are as described for the *GE*-biplot in the caption for Fig.1.

region and showed a strict ordering of the microarrays in time around the origin indicating strong cyclic behaviour in time. The plot on the left of Fig.2, that uses the 221 genes selected by ASM (with $\hat{c} = 0.6$), exemplifies this form of cyclic behaviour. Such clear, cyclic, behaviour is not apparent in Fig.1. For example, microarrays in the first G2 cell phase (50, 60 minutes) are quite separated from those in the second G2 phase (130, 140 minutes). In the plot on the left of Fig. 2, these microarrays are relatively close to each other.

In the h-profile plot in Fig.1, it is clear that only about half (254) of the 532 genes correspond to $k = 2$, the value one would expect for data collected for two cell cycles. The *GE*-biplot for these (254) genes (when $k = 2$) is shown on the right in Fig.2. Now it can be seen that the microarrays are positioned in such a way as to reflect the cell phases that are known for these yeast data. The two plots in Fig.2 are quite similar to each other, with clear separation of the microarrays and (most of) the genes into the four distinct phases. The measures of fit are essentially identical, and much better than those for Fig.1, namely $I_1 = 0.82$ and $I_2 = 0.99$. The corresponding h-profile plots are given in Fig.3.

From the FET-gs results, 7 genes had estimated k values greater than 2 (2 with a value of 3, 1 a value of 4, and 4 with value 8). The remaining 271 genes had an estimated k value of 1, and Fig.4 gives a *GE*-biplot and an h-profile plot for these genes. The measures of fit are essentially identical to those for Fig.2 The three groups of genes seen in the biplot of Fig.4 correspond to the first 6 time points (10 to 60 minutes), the next 5 (70 to 110 minutes) and the final 5 (120 to 160 minutes). From the h-profile plot in Fig.4, it appears that gene profiles towards the upper right corner might have an upward slope, while those

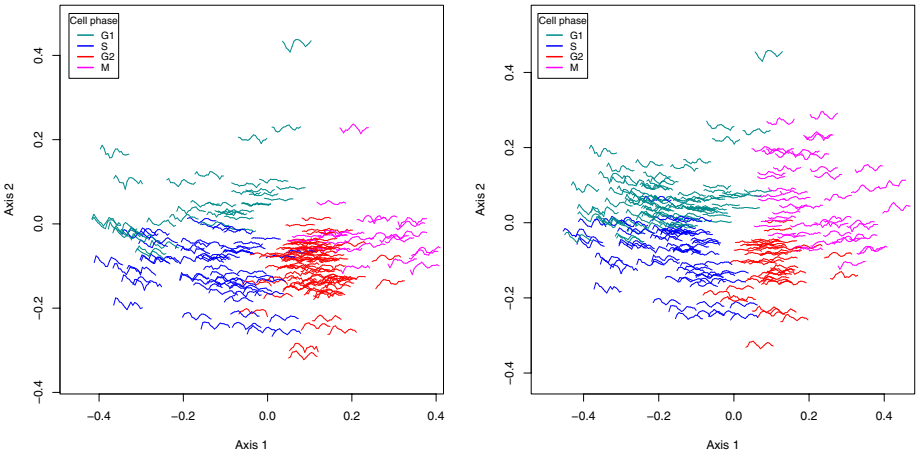

Fig. 3. On the left is the h-profile plot for the 221 genes determined by ASM ($\hat{c} = 0.6$), and on the right the corresponding plot for the 254 genes determined by FET-gs (FDR=0.05) restricted to $k = 2$. The h-profile plot uses the standard time series plots of the (transformed) gene expression values plotted over the 16 time points. The colours are described for the *GE*-biplot in the caption for Fig.1.

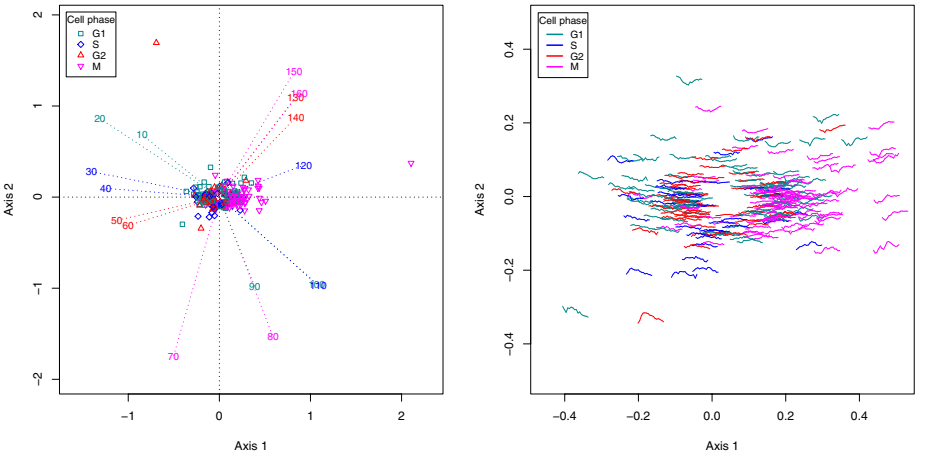

Fig. 4. On the left is the *GE*-biplot using the 271 genes determined by FET-gs (FDR=0.05) restricted to $k = 1$, and on the right is the corresponding h-profile plot with outlying genes removed. The colours, numbers and symbols are as described for the *GE*-biplot in the caption for Fig.1, with the profiles as described for Fig. 3.

towards the lower left might slope downward. These slopes are not so obvious in the h-profile plot on the right of Fig.3 for the 2-cycle genes found by FET-gs.

Now FET is a test for periodicity with the null hypothesis of randomness. Such a test is likely to be affected by other deviations from randomness such as a trend. So we detrended the (transformed) gene expression data, by applying ordinary least squares (linear) regression, and used the residuals in FET-gs (FDR=0.05). The outcome was quite stunning. The number of genes selected fell dramatically, from 532 to just 82. The I_1 and I_2 measures of fit improved slightly (now 0.74 and 0.9 respectively, compared with 0.65 and 0.88 previously). The covariance-biplot is given on the left in Fig.5. Amongst the 82 genes, 59 (72%) had a k-value of 2, 22 a value of 1 and one a value of 4. Amongst the 59 two-cycle genes, 21 were in the top 59 genes found using ASM. On the right of Fig.5, we give an h-profile plot showing these 21 genes as well as the 38 that were uniquely found for the detrended data using FET-gs, FDR=0.05, k=2, and the 38 that were unique to the top 59 found from fitting ASM.

We selected 8 genes for closer comparisons, namely 4 of the genes unique to FET-gs lying on the left hand side of the h-profile plot in Fig.5, and 4 that were unique to ASM and well separated from the first 4 genes. The genes are identified on the h-profile plot. In Fig.6 we give time series profiles for these 8 genes, distinguishing the 4 unique to FET-gs on the left plot, from the 4 unique to ASM on the right. Note the different shapes of the profiles for the genes selected by FET-gs, after accommodating the (slight) phase shift compared with the profiles for the genes selected by ASM. This demonstrates that different approaches are selecting genes with profiles that have distinct patterns.

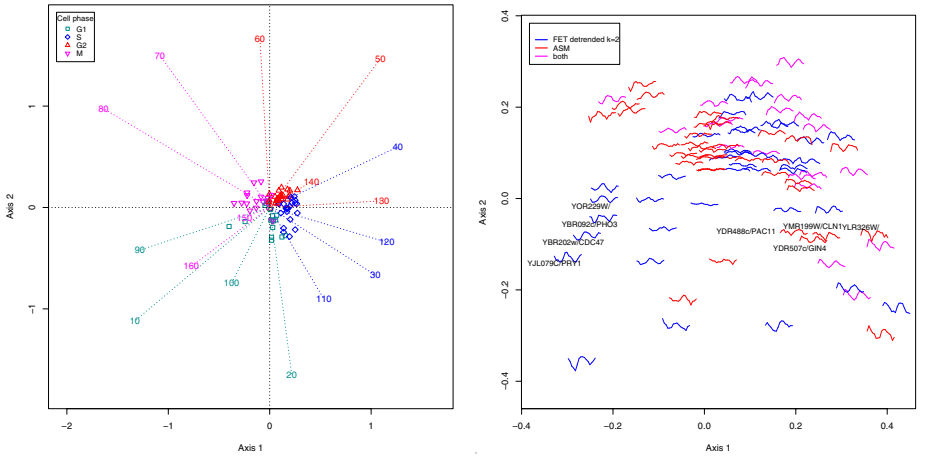

Fig. 5. On the left is the covariance-biplot using the 82 genes determined by FET-gs (FDR=0.05) after detrending (fitting a linear model to) the (transformed) gene expression values. The coordinates use the residuals from the linear model. On the right is an h-profile plot using the genes selected by different approaches; see legend. Eight of the genes selected for Fig.6 are identified.

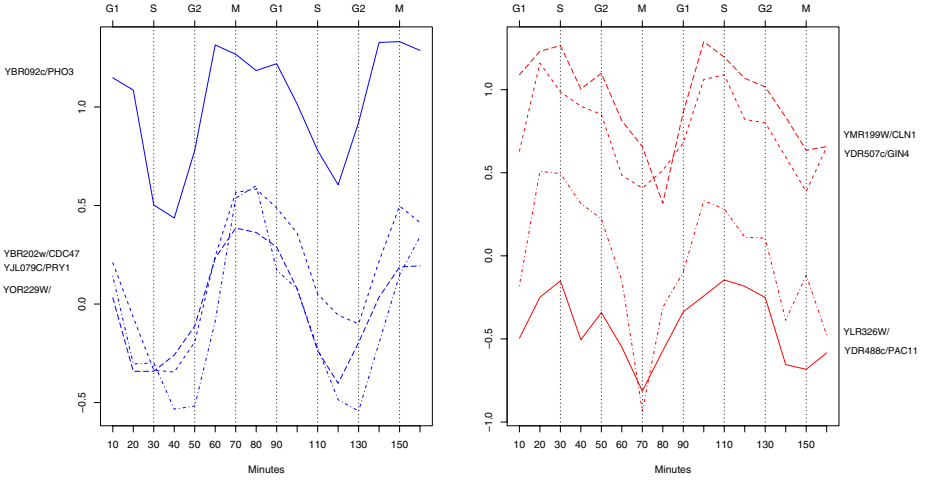

Fig. 6. Detailed time series profiles for 4 of the genes uniquely selected by FET-gs (using residuals, FDR=0.05) on the left, and 4 genes selected uniquely by ASM ($\hat{c} = 0.6$) on the right. The colours correspond to those used on the right in Fig.5.

Two of Cho *et al*'s [11] landmark genes are in our selection. One of these is CLN1 (found by ASM, and depicted in the plot in the right in Fig.6), and the other is CDC47 (found by FET-gs, and depicted in the plot on the left in Fig.6). Cho *et al* identified CLN1 as being characterized by the specific cell cycle

phase (late) G1. Previously we used CLN1 to show how one can determine those genes that are highly correlated with a (landmark) gene of interest [9], and found there, as here, that it peaks in the G1 phase. CLN1 was one of the two genes that Cho *et al* found to have the largest change of 25-fold. So it is interesting that this gene was not found using FET-gs (with the detrended data). In Fig.6, CDC47 has a profile that peaks in the M phase of the cell cycle. Cho *et al* used CDC47 as an early G1 phase landmark, although from their (Fig4C) plot it would appear to peak during phase M. In the space available, these results briefly illustrate that the ability to visualise the profiles allows the researcher to examine the approaches being used for finding genes and gives added insight into the findings from the analyses.

5 Discussion

There are no true gold standards for finding expressed genes that vary systematically during the cell cycle, and so absolute assessment of different methods is not possible. Until now, many approaches have been restricted to finding periodically expressed genes using time series methods. We advocate that alternative models should also be applied.

We note that complete evaluation also requires the integration of results from the genes found and the visual interpretation along with biological background. For example, it has been argued (e.g. [7], [3]) that due to the synchronisation technique used, the cells may be perturbed so some of the observed periodic genes are due to stress response rather than cell cycle activity. In other words some genes selected may be artifacts due to the treatment of the cells that would not occur in freely growing cells. Certainly this could explain the genes selected for $k = 1$.

On first consideration, one might think that the two approaches, ASM and FET, should be directly comparable. In general, they are not. ASM was developed to *model* the pattern of gene expression in a cycle in a specified manner (see section 2.2) while, on the other hand, FET is a *test* for periodicity with the null hypothesis of randomness and the (implicit) underlying model is quite different [12]. Further, the ASM could be generalised to allow a differently shaped function with a more pronounced peak where the gene is "switched on" compared with the expression value in the other phases, where say the gene is "switched off".

For FET, residuals from other detrending or modelling approaches, such as from fitting a quadratic function, could result in different genes being selected. We also note the possibility of different results from using FET-gs from [16], rather than FET in [12]. Further, the FDR assumes independence of the genes that obviously does not hold, as many genes are expected to be highly correlated with one another. Evaluation of this assumption is beyond the scope of the paper.

Preprocessing and transformation of the data has an effect whose size has not been consistently evaluated. Currently, there are no agreed guidelines, and consideration of the sensitivity of the results to this decision is beyond the scope of this paper. Further, we note that the common practice of 'norming' the gene

expression values for each gene will distort the plots as then the gene points will tend to lie on the circumference of a circle, as the variance has been set to 1 for all genes.

FET as proposed in [7] has been used for unevenly spaced time points [8], and detrending does not seem to have been considered. The ASM can be used whether the time points are evenly or unevenly spaced. A robust alternative to FET has been proposed [13] and this could be usefully evaluated, as could the application of methods for testing for fixed periodicities, as outlined in [12], when one was only interested in, say, finding all genes completing two cycles during the experiment.

In [7], averaging was recommended. We note that if there are approximately an equal number of genes in different phases then averaging over all the genes would result in no periodicity being able to be detected. Averaging would only determine periodically expressed genes if the number exhibiting the identical periodic behaviour were significantly greater than the remainder.

6 Conclusions

Fisher's exact test (FET) is being widely used for finding periodic patterns for gene expression data from time course experiments. The Absolute Sine Model (ASM) is an alternative approach to finding patterns during a cell cycle. Using a well-known yeast data set, we applied these two approaches, as well as two visualisation methods that enable (i) genes and arrays to be displayed simultaneously (the *GE*-biplot) and (ii) profiles for a large number of genes to be displayed (the h-profile plot). These visualisation tools enabled many insights to be obtained, including the differentiation of the genes into groups according to their periodicity, and the need to detrend the gene expression values first before applying FET.

This paper highlights the advantages of (i) using visualisation methods simultaneously with computational and statistical approaches, and (ii) using more than a single approach for finding patterns in gene expression data from time course experiments, as different approaches can highlight uniquely different aspects of the gene expression patterns.

References

1. Kettenring, J.R.: The practice of cluster analysis. J. Classif. 23, 3–30 (2006)
2. Kettenring, J.R.: A perspective on cluster analysis. Stat. Anal. Data Mining, 52–53 (2008)
3. de Lichtenberg, U., Jensen, L.J., Fausboll, A., Jensen, T.S., Bork, P., Brunak, S.: Comparison of computational methods for the identification of cell cycle-regulated genes. Bioinformatics 21, 1164–1171 (2005)
4. Tamayo, P., Slonim, D.S., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E.S., Golub, T.R.: Interpreting patterns of gene expression with self-organising maps: Methods and application to hematopoietic differentiation. Proc. Natl. Acad. Sci. 96, 2907–2912 (1999)

5. Yeung, K.Y., Ruzzo, W.L.: Principal component analysis for clustering gene expression data. Bioinformatics 17, 763–774 (2001)
6. Johansson, D., Lindgren, P., Berglund, A.: A multivariate approach applied to microarray data for identification of genes with cell cycle-coupled transcription. Bioinformatics 19, 467–473 (2003)
7. Wichert, S., Fokianos, K., Strimmer, K.: Identifying periodically expressed transcripts in microarray time series data. Bioinformatics 20, 5–20 (2004)
8. Liew, A.W.-C., Xian, J., Wu, S., Smith, D., Yan, H.: Spectral estimation in unevenly sampled space of periodically expressed microarray time series data. BMC Bioinformatics 8, 137 (2007)
9. Pittelkow, Y.E., Wilson, S.R.: h-Profile plots for the discovery and exploration of patterns in gene expression data with an application to time course data. BMC Bioinformatics 8, 486 (2007)
10. Pittelkow, Y.E., Rosche, E., Wilson, S.R.: Interpreting models in gene expression data. In: Francis, A.R., Matawie, K.M., Oshlack, A., Smyth, G.K. (eds.) Statistical Solutions to Modern Problems: Proceedings of the 20th International Workshop on Statistical Modelling, Sydney 2005, pp. 381–391 (2005)
11. Cho, R.J., Campbell, M.J., Winzeler, E.A., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T.G., Gabrielian, A.E., Landsman, D., Lockhart, D.J., Davis, R.W.: A Genome-Wide Transcriptional Analysis of the Mitotic Cell Cycle including control of mRNA transcription. Molecular Cell 2, 65–73 (1998)
12. Brockwell, P.J., Davis, R.A.: Time Series: Theory and Methods. Springer, New York (1991)
13. Ahdesmäki, M., Lähdesmäki., H., Pearson, R., Huttunen, H., Yli-Harja, O.: Robust detection of periodic time series measured from biological systems. BMC Bioinformatics 6, 117 (2005)
14. Pittelkow, Y.E., Wilson, S.R.: Visualisation of gene expression data - the GE-biplot, the Chip-plot and the Gene-plot. Stat. Appl. Genet. Mol. Biol. 2, 6 (2003)
15. Chen, J.: Identification of significant periodic genes in microarray gene expression data. BMC Bioinformatics 6, 286 (2005)
16. The GeneCycle Package, `http://www.strimmerlab.org/software/genecycle/`
17. Everitt, B.S.: Time Series. In: Armitage, P., Colton, T. (eds.) Encyclopedia of Biostatistics, 2nd edn., pp. 5451–5454. Wiley, Chichester (2005)
18. R source code for GE-biplot, `http://dayhoff.anu.edu.au/software.html`