

Comparison of computational methods for the identification of cell cycle-regulated genes

Ulrik de Lichtenberg^{1,†}, Lars Juhl Jensen^{2,†}, Anders Fausbøll¹, Thomas S. Jensen¹, Peer Bork² and Søren Brunak^{1,*}

¹Center for Biological Sequence Analysis, Technical University of Denmark, DK-2800 Lyngby, Denmark and ²European Molecular Biology Laboratory, D-69117 Heidelberg, Germany

Received on August 30, 2004; revised on October 1, 2004; accepted on October 5, 2004

Advance Access publication October 28, 2004

ABSTRACT

Motivation: DNA microarrays have been used extensively to study the cell cycle transcription programme in a number of model organisms. The *Saccharomyces cerevisiae* data in particular have been subjected to a wide range of bioinformatics analysis methods, aimed at identifying the correct and complete set of periodically expressed genes.

Results: Here, we provide the first thorough benchmark of such methods, surprisingly revealing that most new and more mathematically advanced methods actually perform worse than the analysis published with the original microarray data sets. We show that this loss of accuracy specifically affects methods that only model the shape of the expression profile without taking into account the magnitude of regulation. We present a simple permutation-based method that performs better than most existing methods.

Supplementary information: Results and benchmark sets are available at <http://www.cbs.dtu.dk/cellcycle>

Contact: brunak@cbs.dtu.dk

1 INTRODUCTION

It has been clear for many years that certain genes are expressed only at specific stages of the cell cycle (e.g. the cyclins and the histones). These genes consequently exhibit a periodic pattern of expression when monitored during consecutive cell cycles. In 1998, the first genome-wide DNA microarray studies were conducted in *Saccharomyces cerevisiae* (Cho *et al.*, 1998; Spellman *et al.*, 1998) to reveal a large number of periodically expressed genes which peak only once per cycle, also referred to as cell cycle-regulated genes. Similar investigations were later performed in human fibroblasts (Cho *et al.*, 2001) and HeLa cells (Whitfield *et al.*, 2002) respectively. Most recently, an extensive study has been carried out in *Schizosaccharomyces pombe* (Rustici *et al.*, 2004). Each of these studies have aimed at defining the cell cycle-regulated (or periodically expressed) subset of the genome in each organism. Although the periodic signal is strong in most data sets (Shedden and Cooper, 2002; Wichert *et al.*, 2004), the experimental noise is also considerable, as can be seen from the poor overlap between the gene sets identified as periodic in different experiments within the same organism (Zhao *et al.*, 2001;

Shedden and Cooper, 2002; Johansson *et al.*, 2003; de Lichtenberg *et al.*, 2003; Luan and Li, 2004) as well as between organisms (Rustici *et al.*, 2004).

The budding yeast data sets in particular have driven the development of various computational methods for identifying periodically expressed genes, most of which have concluded the periodically expressed subset of the yeast genome to comprise about 300–800 genes. However, the agreement is remarkably poor when different computational methods are applied to the same data. In total, nearly 1800 different genes have been proposed to be periodic—that is almost every third gene in the *S.cerevisiae* genome. Here, we provide the first benchmark of these computational methods by comparing the gene sets identified by the different methods when applied to three yeast experiments (Alpha, Cdc15 and Cdc28). A fourth, elutriation-based experiment in yeast (Spellman *et al.*, 1998) was not used, since most published methods were not applied to this data and because it only covers a single cell cycle. We benchmark the methods by measuring their ability to identify genes from three benchmark sets:

- (B1) A total of 113 genes previously identified as periodically expressed in small-scale experiments. The set encompasses the 104 genes used by Spellman *et al.* (1998) and nine genes added by Johansson *et al.* (2003).
- (B2) Genes whose promoters were bound (*P*-value below 0.01) by at least one of nine known cell cycle transcription factors in both of the Chromatin IP studies by Simon *et al.* (2001) and Lee *et al.* (2002). To obtain a benchmark set that is independent of B1, we removed all genes included in B1 (50). The resulting benchmark set, B2, consists of 352 genes of which many should be expected to be cell cycle regulated, since their promoters are associated with known stage-specific cell cycle transcription factors.
- (B3) Genes annotated in MIPS (Mewes *et al.*, 2002) as ‘cell cycle and DNA processing’. From these, we removed genes annotated specifically as ‘meiosis’ and genes included in B1 (67), leaving 518 genes. As a large number of genes involved in the cell cycle are not subject to transcriptional regulation (not periodic), and because B1 was explicitly removed, a relatively small fraction of these genes should be expected to be periodically expressed.

We thus define a good method as one that is able to reproduce previous findings (B1), extract genes whose promoters are

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

associated with stage-specific cell cycle transcription factors (B2), or enrich for genes that are known to play a role in the cell cycle (B3). Included in the benchmark are six published methods as well as a new permutation-based method that separately quantifies both the periodicity and amplitude of periodically expressed genes and estimates the time of peak expression for each gene. Our benchmark analysis enables us to understand the strengths and weaknesses of each computational approach, and to explain the obvious disagreement between methods. We show that amplitude-independent methods are outperformed by the amplitude dependent ones, and that those methods that perform well in the benchmark also provide a better overlap in genes identified in different experiments.

2 METHODS

2.1 Statistical tests for regulation

The standard deviation can be easily calculated for each log-ratio profile, giving a measure of the spread of the samples around the mean. Heavily regulated genes will thus have large standard deviations, whereas genes without significant regulation display little deviation from the mean. To test for the significance of regulation, we therefore compare the observed standard deviation for each expression profile to a randomly generated background distribution. A total of 1 000 000 random profiles were constructed by selecting at each time point the log-ratio from a randomly chosen gene. A P -value for regulation was calculated as the fraction of the simulated profiles with standard deviations equal to or larger than that observed for the real expression profile.

2.2 Statistical tests for periodicity

To estimate a P -value for periodicity, we compared the Fourier score of the observed gene expression profile for each gene to those of random permutations of the same gene. For each gene, i , a Fourier score, F_i , was computed as

$$F_i = \sqrt{\left(\sum_t \sin(\omega t) \cdot x_i(t)\right)^2 + \left(\sum_t \cos(\omega t) \cdot x_i(t)\right)^2}$$

where $\omega = 2\pi/T$ and T is the interdivision time. Similarly, scores were calculated for 1 000 000 artificial profiles constructed by random shuffling of the data points within the expression profile of the gene in question. The P -value for periodicity was calculated as the fraction of artificial profiles with Fourier scores equal to or larger than that observed for the real expression profile.

The P -value for regulation is thus a comparison between individual genes and the global distribution, whereas the P -value for periodicity is a comparison involving only data from the gene in question. To avoid any over-estimation of the significance of the P -values, we normalized all P -values within each data set by the median P -value (prior).

2.3 Combined tests for regulation and periodicity

For each gene, a combined P -value of regulation was calculated by multiplying the separate P -values of regulation from each of the three experiments. Analogously, a combined P -value of periodicity was calculated. Subsequently, the P -value of regulation and P -value of periodicity were multiplied to obtain the total P -value. An undesirable feature of the total P -value is that it may become very low (i.e. highly significant) due to only one of the tests. Genes that are strongly regulated but not periodic (or vice versa) will thus receive good scores. To address this, we multiply the total P -value with two penalty terms that weight down genes that are either not significantly regulated or not significantly periodic. The final score used for

ranking is:

$$P_{\text{total}} \cdot \left[1 + \left(\frac{P_{\text{regulation}}}{0.001}\right)^2\right] \cdot \left[1 + \left(\frac{P_{\text{periodicity}}}{0.001}\right)^2\right]$$

The calculation was done for each experiment separately as well as for the combined experiments.

2.4 Assigning the time of peak expression

Since we approximate each expression profile by a sine wave, the time of peak expression for a gene in a single experiment is trivially defined as the time where the sine wave is maximal. We refer to this as the peak time. Due to differences in experimental conditions, the time it takes the cell to complete a cycle (the interdivision time) varies greatly between the alpha, Cdc15 and Cdc28 experiments. In order to compare the timing of peak expression across experiments, we therefore transformed the timescales from minutes to percentage of the cell cycle by dividing with the interdivision times estimated by Zhao *et al.* (2001).

Subsequently, differences in release point of the synchronization techniques were corrected for by aligning the timescales of the three experiments. The optimal offsets for the experiments were determined by minimizing an error function, $E1 = \sum_i E1_i$, that measures the disagreement in the time of peak expression of the same gene in different experiments:

$$E1_i = w_i^{\text{alpha}} w_i^{\text{cdc15}} \text{dist}(t_i^{\text{alpha}}, t_i^{\text{cdc15}})^2 + w_i^{\text{alpha}} w_i^{\text{cdc28}} \text{dist}(t_i^{\text{alpha}}, t_i^{\text{cdc28}})^2 + w_i^{\text{cdc15}} w_i^{\text{cdc28}} \text{dist}(t_i^{\text{cdc15}}, t_i^{\text{cdc28}})^2$$

As weights (w_i^{alpha} , w_i^{cdc15} and w_i^{cdc28}), the negative logarithm of the respective total P -values were used. The function *dist* refers to the shortest possible distance between two points on a circle. This alignment results in an arbitrary zero time. We chose to define zero time as the M/G1 transition, i.e. the time of cell division. This time point was estimated from the combined peak times of 19 known M/G_1 genes (Fig. 5), all among the top 500 genes in our final gene list. The resulting offsets for the alpha factor, Cdc15 and Cdc28 experiments were 78, 84 and 93% respectively.

Combining peak times from different experiments into one is a non-trivial task, since the assignment should not be trusted in those experiments where the expression profile is not sufficiently periodic. To compensate for this, we weighted the individual peak times when computing the global, combined peak time. For each gene, a combined peak time (t_i) was calculated from the three individual peak times (t_i^{alpha} , t_i^{cdc15} , and t_i^{cdc28}) by minimizing the following error function:

$$E2_i(t_i) = \text{dist}(t_i^{\text{alpha}}, t_i)^2 w_i^{\text{alpha}} / W + \text{dist}(t_i^{\text{cdc15}}, t_i)^2 w_i^{\text{cdc15}} / W + \text{dist}(t_i^{\text{cdc28}}, t_i)^2 w_i^{\text{cdc28}} / W$$

where $W = w_i^{\text{alpha}} + w_i^{\text{cdc15}} + w_i^{\text{cdc28}}$ and the weights are defined as in $E1_i$.

2.5 Renormalization of Cdc28 data

Cho *et al.* (1998) used the temperature-sensitive mutant strain CDC28-13 to produce a synchronized cell culture from which 17 samples were taken at 10-min intervals and hybridized to Affymetrix chips. Getting the original images or scanner reads was not possible, but the final data were downloaded from http://yscdp.stanford.edu/yeast_cell_cycle/full_data.html. The paper contains no information on how these data were processed. We therefore renormalized the data with the signal-dependent non-linear Qspline method (Workman *et al.*, 2002). Multiple probes against the same gene were collapsed by taking the median of their intensities. To avoid negative expression values, the scale was shifted by +750. The data were \log_2 transformed and the mean expression value over all time points was subtracted to centre the profile at zero (as in Spellman *et al.*, 1998).

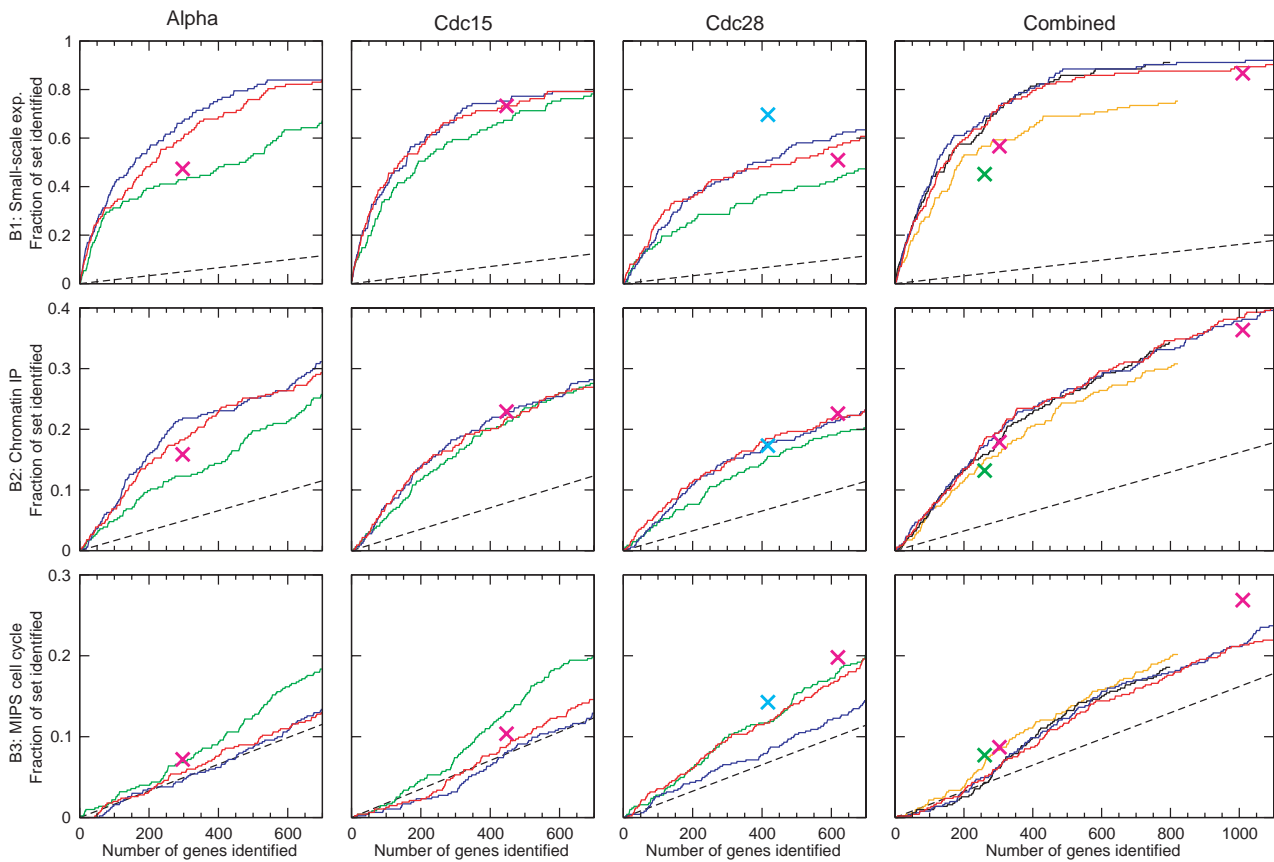


Fig. 1. Comparison of published methods. The fraction of the benchmark set that is identified is plotted as a function of gene rank for each method, experiment and benchmark set. The different methods are coloured as follows: Cho *et al.* (1998) (M1, cyan), Spellman *et al.* (1998) (M2, black), Johansson *et al.* (2003) (M3, blue), Zhao *et al.* (2001) (M4, green), Luan and Li (2004) (M5, magenta), Lu *et al.* (2004) (M6, orange) and our permutation-based approach (M7, red). Random performance is shown as a black dotted line.

3 RESULTS

In our benchmark, we included six methods that have all been applied to the data sets (Alpha, Cdc15 and Cdc28) published by Spellman *et al.* (1998):

- (M1) Cho *et al.* (1998) visually inspected the expression profiles of all genes regulated more than 2-fold during the Cdc28 experiment, classifying 421 of them as ‘periodic’.
- (M2) Spellman *et al.* (1998) computed a score for every gene that was based partly on the correlation to one of five idealized gene profiles, and partly on a Fourier-like score yielding the signal strength at a period similar to the interdivision time. The genes were ranked based on their combined score and truncated to 800 genes as this corresponds to a sensitivity of 90% on B1.
- (M3) Johansson *et al.* (2003) used partial least squares regression to analyse the three data sets individually and in combination. The approach was based on fitting of sine curves and thus also yields an estimate of the time of peak expression. Thresholds were estimated based on random permutations of the data.
- (M4) Zhao *et al.* (2001) reanalysed each of the three data sets individually using a statistical single-pulse model. The resulting

score, describing how well a profile fits the model, is independent of the magnitude of regulation. An appealing feature of the model is that it also estimates the time of activation and deactivation of the gene.

- (M5) Luan and Li (2004) used a modelling approach based on cubic splines, rather than sine waves. The sets were analysed individually and the statistical nature of the method enabled the authors to identify thresholds that satisfied a given false discovery rate.
- (M6) Lu *et al.* (2004) used Bayesian modelling techniques to estimate a periodic–normal mixture model based on five microarray time courses. The authors provide a ranked list of 822 genes based on all five data sets.

In addition to the methods M1–M6, we also include a new and simple statistical approach (M7) based on permutations. Its score is composed of two terms: one quantifies only the magnitude of regulation, whereas the other measures the periodicity of the expression profile. The combined score ensures that high-ranking genes are both significantly regulated and periodic (see Section 2 for details).

Figure 1 shows the performance of each method on each individual microarray expression data set as well as on all three sets in combination. Studies in which a ranked list was published are

plotted as curves, showing the percentage of genes in a benchmark set recovered (i.e. coverage) as a function of rank. Where only an unranked list was proposed, the performance is represented as a single point. The three benchmark sets and the lists of genes proposed by each computational method are all available from the web site <http://www.cbs.dtu.dk/cellcycle>

Although some methods are clearly better than others, there is no single best method. All methods perform significantly better than random on the two primary benchmark sets (B1 and B2), meaning that they all enrich for genes previously identified as periodic and genes associated with known cell cycle transcription factors. The best performing methods across all data sets are M2, M3 and our permutation-based approach M7. However, these methods all show close to random performance on benchmark set B3, where the methods M4–M6 seem to work slightly better. The sets of genes identified as periodic from each of the individual experiments are all highly enriched in genes from the benchmark sets B1 and B2, indicating that all three experiments are valuable. The enrichment of genes whose promoter regions are bound by known cell cycle transcription factors (B2) demonstrates that the genes identified as periodic in the microarray data are biologically meaningful. The chromatin IP technique measures physical association of transcription factors with the promoter regions of individual genes and the experiments were performed in freely growing cells, unlike the small scale and microarray experiments (Spellman *et al.*, 1998) which use various methods to synchronize the cells. The B2 set is thus independent of the microarray experiments on cell cycle regulation and the observed correlation between these two data sources is thus reassuring.

For benchmark sets B1 and B2, the curves rise steeply in the beginning with the fraction of benchmark genes gradually decreasing with rank. Judging from the shape of the curves, there does not seem to be one natural value for thresholding. Depending on the relative importance of accuracy versus coverage, anywhere from 300 to 800 genes should be included. Ranked lists should therefore generally be preferred over non-ranked ones. As a consequence, methods that provide a combined score based on all experiments are preferable over simple voting schemes.

The good performance of the method M2 on B1 was anticipated since the scoring scheme makes use of the correlation with the expression profiles of these genes. Surprisingly, Figure 1 demonstrates that this approach works equally well on the benchmark set B2, which includes no genes from B1. However, it remains unclear whether the performance should primarily be attributed to the Fourier term or to the correlation term.

The performance of the visual analysis of Cho *et al.* (1998) on the B1 and B3 sets is remarkable. None of the computational methods perform as well on these two benchmark sets, yet the results of computational and visual analyses are comparable on B2. The major difference between B2 and B1/B3 is that the latter sets consist of genes known to be involved in the cell cycle. In contrast, B2 is not biased towards genes of known function. The visual inspection was not done blindly (L. Steinmetz, personal communication), which has likely introduced an unintentional bias in the borderline cases towards genes with known roles in the cell cycle.

Most surprising is the poor performance of M4–M6 on B1 and B2. Particularly, the M6 analysis disappoints with respect to finding cell cycle-regulated genes, considering that additional experimental data were used. These methods differ from the better performing

methods in two respects: they model the peak shape more accurately than simple sine waves and their scores only depend on the shape of the expression profile, not on the magnitude of regulation. Whereas the first is likely to yield improvements, the second discards relevant information, which explains the poor performance (see below).

3.1 Shape versus magnitude

Our permutation-based scoring scheme combines two statistical tests. One term measures the significance of regulation, i.e. simply ranks genes according to the standard deviation of their expression profiles. The other term is a magnitude-independent test that compares the periodicity of the observed profile to that of permutations of the profile.

To explore the relative importance of profile shape versus magnitude of regulation, we benchmarked the regulation and periodicity terms separately (Fig. 2). The simple ranking by regulation yields strikingly good results and even outperforms the periodicity ranking on some sets. As expected, the periodicity term alone yields results that closely resemble those of the magnitude-independent methods (M4–M6). In general, combining both terms significantly improves the results, consistent with the better performance of all magnitude-dependent methods. It is thus clear that taking into account both regulation and periodicity is far more important than accurately modelling the shape of the expression profile.

Comparison of the regulation and periodicity terms also reveals compositional differences between the three benchmark sets. On B1, regulation works well on its own, indicating that the genes identified in small-scale experiments are, not surprisingly, biased towards genes that are both periodically expressed and strongly regulated. In contrast, the B3 set is composed of genes that are annotated as cell cycle related, but not necessarily periodically expressed. Additionally, genes already included in B1 were specifically removed from B3. On this set, periodicity alone works better than both regulation and the combined score, indicating a bias towards periodic genes that are only weakly regulated. In fact, regulation alone is poorer than random expectation, showing that the most strongly regulated genes that are not in B1 specifically belong to pathways not related to the cell cycle. The B2 set is probably the least biased of our three benchmark sets, since it is derived from genome-wide chromatin IP data. This set is not biased by current biological knowledge and contains a considerable proportion of genes of unknown function. From the performance of the various methods, benchmark set B2 indeed appears to be a compromise between the two other sets. To conclude, we regard B1 and B2 to be the most informative benchmark sets.

Clear differences between the three experiments can also be observed in Figures 1 and 2. The alpha factor and Cdc15 results are generally better than those of the Cdc28 experiment. On the alpha experiment, the regulation term alone significantly outperforms the periodicity term (and even the combined approach) on benchmark sets B1 and B2 (Fig. 2). Our interpretation is that only genes involved in the cell cycle are significantly regulated in this experiment. In contrast, many of the most highly regulated genes in the Cdc28 experiment are unrelated to the cell cycle. This is, to a lesser extent, also true for the Cdc15 experiment. In the two latter experiments, cells are released from arrest by an abrupt drop in temperature, which likely results in changes in expression of e.g. heat shock genes. Taken together, this suggests alpha factor synchronization to be the least perturbing of the three synchronization methods.

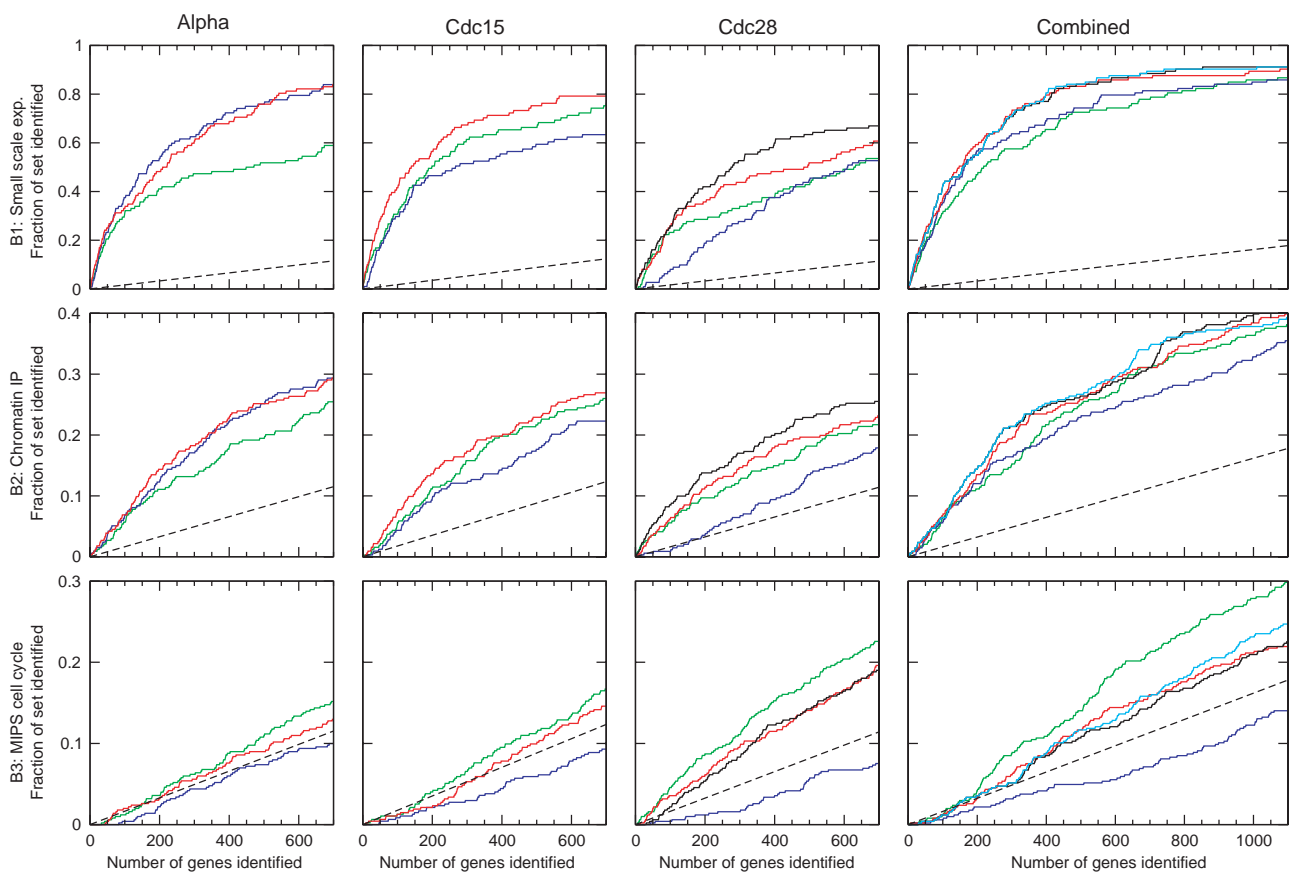


Fig. 2. Regulation versus periodicity. As in Figure 1, the fraction of the benchmark set that is identified is plotted as a function of gene rank for each method, experiment and benchmark set. For the original data sets, the regulation score alone (blue), periodicity score alone (green) and combined score (red, also shown in Fig. 1) are shown. Results using the renormalized Cdc28 data are shown in black. Also shown is the combined score with peak time consistency filter (cyan). Random performance is shown as a black dotted line.

It is clear that the treatment of the cells can introduce artefacts that would not occur in freely growing cells. These response genes are only up- or down-regulated in the beginning of the experiments and most likely specific to the arrest method used. Such profiles will not receive high scores from the computational methods, although some contamination with false positives should be expected for all methods. We have, however, found no indications or reasons to believe that synchronization artefacts manifest themselves as periodic patterns of expression reproduced over multiple cycles or in multiple experiments.

3.2 Renormalizing expression data

Since the publication of the cell cycle gene expression data, extensive research has gone into developing advanced bioinformatics methods for normalizing microarray data to correct for systematic biases. However, these techniques have not been utilized by any of the groups developing methods for identification of periodically expressed genes. To test the possible benefits of renormalizing the raw data prior to periodicity analysis, we applied the signal-dependent non-linear Qspline method (Workman *et al.*, 2002) to the Cdc28 data. As shown in Figure 2 (black curve), this considerably improved the performance of our permutation-based approach. It thus appears that many of the errors in the Cdc28 experiment are due

to systematic array biases rather than artefacts of the synchronization method.

We also attempted to renormalize the raw alpha factor and Cdc15 data as obtained from Stanford Microarray Database (Gollub *et al.*, 2003). Unfortunately, this analysis was inconclusive due to numerous discrepancies between the data deposited in SMD and those originally published (<http://genome-www.stanford.edu/cellcycle/data/rawdata/>).

3.3 Overlap between experiments

An alternative strategy for evaluating the performance of methods for extracting periodically expressed genes is to examine the overlap in genes identified in different experiments. For each of the methods that provide a ranked list of genes for each individual experiment, top 300 lists were extracted and their overlap visualized as Venn diagrams (Fig. 3). The results are consistent with those obtained from the benchmark sets. The results of M4, a representative of the magnitude-independent methods, show by far the least agreement between the three experiments. In comparison, M3 and M7 identify almost twice as many genes that agree in all three experiments. For all computational methods, the alpha factor experiment shows the best overlap with the two other experiments, while the Cdc28 experiment overlaps the least. This supports our previous observations with regard to the

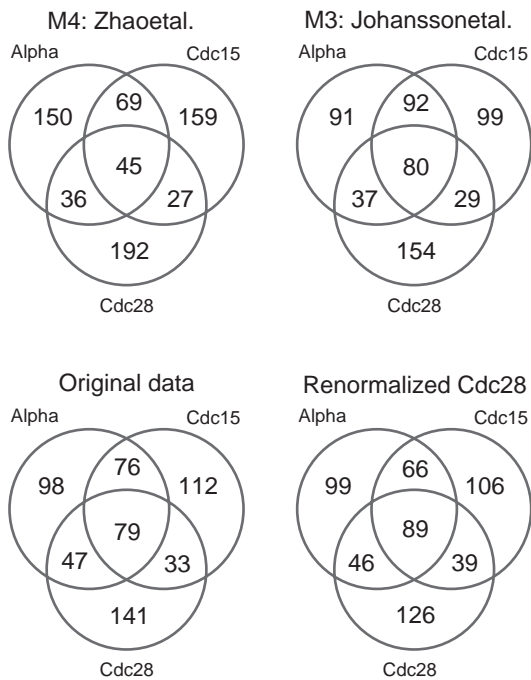


Fig. 3. Agreement across experiments. Venn diagrams based on the top 300 genes from each experiment are shown for the methods that provide ranked lists for the individual experiments. The lower two diagrams show the results of our analysis using the original and renormalized Cdc28 data respectively.

quality of the individual experiments. The lower right Venn diagram in Figure 3 illustrates that a renormalization of the Cdc28 data improves the agreement with the two other experimental data sets, in addition to improving performance on the benchmark sets (Fig. 2). When including the renormalized Cdc28 data in our analysis, two-thirds of the genes detected from the alpha factor synchronization are confirmed by at least one other experiment. The fact that the overlap can be improved through renormalization and use of better computational methods suggests that experimental noise, rather than synchronization artefacts, accounts for much of the variation.

3.4 Consistent timing of expression

So far, we have addressed the issues of finding periodically expressed genes and asked if the same genes are identified in different experiments. It is, however, equally important to check that those genes behave similarly across experiments, i.e. that their expression profiles peak at the same stage in the cell cycle. As a consequence of differences in experimental protocols, the interdivision time varies between experiments. Furthermore, the synchronization methods release cells at different points in the cycle. To enable cross-experiment comparisons, we represent time in percentage of the cell cycle (0% being the time of cell division) and align the time axes of the experiments relative to each other (see Section 2). For each gene, we can thus assign a time of peak expression in each experiment and compare these across experiments. To examine the degree of consistency between these peak times in different experiments, we computed the largest peak time difference for each gene. Figure 4 shows the distribution of these differences for the set of 89 genes that were identified as periodic in all three experiments (see Fig. 3).

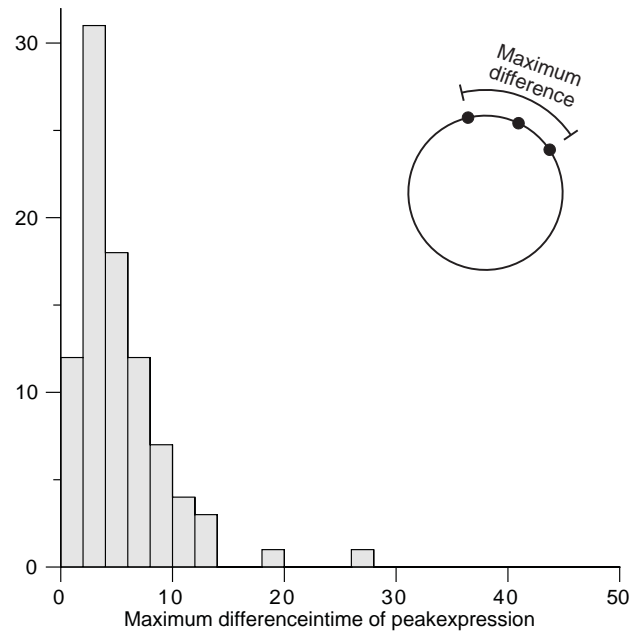


Fig. 4. Consistency of peak expression. For the 89 genes identified in all three experiments (see bottom right diagram in Fig. 3) the largest difference in peak time was computed and summarized as a histogram.

For 87 of the 89 genes, all three peak times fall within an interval of 15% of a cell cycle, clearly demonstrating the reproducibility of peak expression by different synchronization methods. For genes that appear periodic in all three experiments, an average peak time could easily be computed. At lower ranks, however, the genes may not appear equally periodic or regulated in all experiments. To account for this, we instead compute a weighted average of the three experiments. This procedure also allows us to calculate a variance measure to assess the reliability of each combined peak time.

As a final check of the reproducibility, we investigated the timing of peak expression for four sets of known phase-specific genes across the three experiments. B1 was subdivided by Spellman *et al.* (1998) into phase-specific groups according to their reported peak expression in the small-scale experiments. The distribution of peak times within each group is visualized in Figure 5, along with the distribution of our combined peak times. From this, it is clear that the phases occur in the same order, with the same length, and at the same time in all three experiments. It thus appears that the synchronization methods cause no abnormalities of the cell-cycle transcriptional programme. Together, Figures 4 and 5 show that the combined peak time is a meaningful measure that accurately describes when, in the cell cycle, a gene is expressed.

At lower ranks, we found a small number of genes with inconsistent peak times. Removing these from the ranked list based on the weighted variance score (see Section 2 for details) led to marginal improvements in performance (Fig. 2, cyan curve). Among the top 300 genes, only four were removed by the filter, while 27 of the next 200 genes were discarded. In addition to improving the performance, the filter ensures that all remaining genes on the list can be confidently assigned to a unique point in the cell cycle. This ranked list is available from the web site <http://www.cbs.dtu.dk/cellcycle>. It is based on renormalized Cdc28 data and only contains genes that

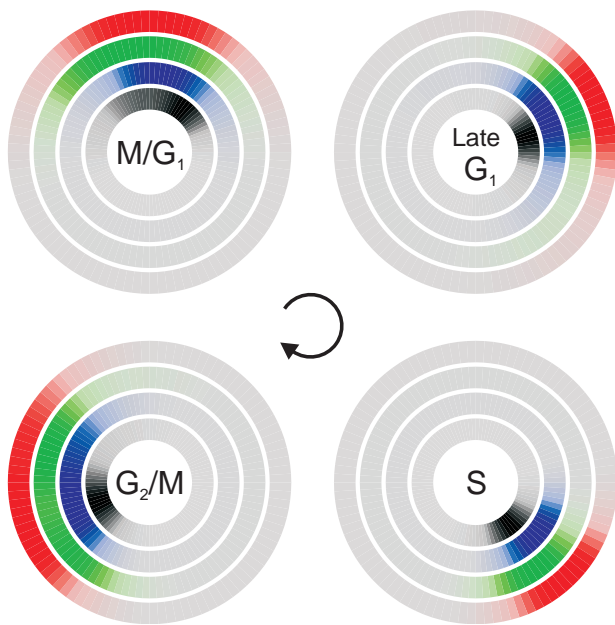


Fig. 5. Expression of phase-specific genes. For the phase-specific gene sets identified in small-scale experiments, the distribution of peak times is shown for each experiment: *cdc28* (black), *cdc15* (blue), alpha (green) and the three experiments in combination (red). The mode of the peak time distribution for the M/G1 group was chosen to define the zero time point. Within each experiment, only genes in the top 500 of our periodicity analysis were included.

show consistency in their time of peak expression. In the benchmarks analysis (Fig. 2), its performance is better than or comparable to all other published lists. This analysis shows a steep increase in the B1 and B2 curves for the first 300 genes identified. Beyond this point, a clear enrichment is also seen in B3, which continues until 500–600 genes are included, at which point the B1 also set saturates. Enrichment is still seen in B2, but dies out around 800. Based on these observations, we chose to define a high-confidence set as top-300, a medium-confidence at top-500, but also include the lower-confidence top-800 set. Using the total *P*-values, we employed the Benjamini-Hochberg multiple testing procedure (Reiner *et al.*, 2003) to estimate the false discovery rate (FDR) at different cut-offs. For the highest scoring 300, 500 and 800 genes we estimate the FDR to be $3 \cdot 10^{-6}$, $4 \cdot 10^{-4}$ and $8 \cdot 10^{-3}$ respectively.

4 DISCUSSION

Naturally, the results of our benchmark (Figs. 1 and 2) depend on the selected benchmark sets B1–B3. As none of the sets can be assumed to be perfect gold standards, it is not possible to assess how good each method is on an absolute scale. We are, however, able to evaluate the computational methods relative to each other by measuring their ability to recover genes from these sets. The methods that perform best on B1–B3 also provide the best overlap between genes identified across different experiments (Fig. 3). We thus believe that our analyses together provide a fair and unbiased assessment of the relative performance of the computational methods.

The budding yeast experiments (Alpha, *Cdc15* and *Cdc28*) used in our study were all synchronized by arrest-release techniques

(whole-culture synchronization). Such techniques block progression at a certain point in the cell cycle, from which all cells can later be released simultaneously, e.g. by lowering of the temperature. Recently, the interpretation and validity of such block–release experiments has been lively debated (Cooper, 2004a,b; Spellman and Sherlock, 2004a,b).

From the perspective of our analysis, we can conclude that there are many more genes that appear periodic than can be explained by mere chance. The genes are largely the same across different experiments, as shown in Figure 3 and by the overlap with B1 (Fig. 1). With few exceptions, the genes peak at the same point in the cell cycle across different experiments (Fig. 4). Also, the characteristic cell cycle phases occur in the same, correct order in all experiments and with the same relative length (Fig. 5). Finally, the genes that appear periodic in the synchronized cell cultures are heavily overlapping (Figs. 1 and 2) with those physically associated with cell cycle transcription factors in Chromatin IP experiments performed on freely growing cells (Simon *et al.*, 2001; Lee *et al.*, 2002).

In summary, the microarray data obtained from the synchronized cell cultures appear to draw a reproducible and biologically meaningful picture of the yeast cell cycle that is consistent with complementary experiments from freely growing cells. Both false positives and false negatives should be expected from the experiments and their subsequent analysis, given the considerable level of noise and the large number of expression profiles investigated. Artefacts that would not occur in freely growing cells may also be introduced by synchronization methods, e.g. induction of stress response genes. However, these would most likely be specific to each arrest method and manifest themselves in the beginning of an experiment. It is difficult to see how responses to different treatments of the cells could induce consistent, periodic patterns of expression that persists through several cycles, as proposed by Cooper (2004a,b).

5 CONCLUSIONS

Most surprisingly, our benchmark analysis reveals that most of the new and more mathematically advanced methods for identifying periodically expressed genes perform considerably worse than the early method by Spellman *et al.* (1998) (M2). These results should encourage developers of future computational methods to evaluate the performance of their methods carefully. We show that the performance gap is due to the magnitude independence of most newer methods. This independence may improve their ability to discover novel weakly regulated cell cycle genes. However, as the magnitude of regulation contains a large part of the signal, this should be exploited in order to derive the most accurate set of cell cycle-regulated genes.

Using the periodicity score alone yields results similar to those of other magnitude-independent methods, while our combined score performs as well as the best existing methods. In comparison to other methods, ours have two main advantages. First, genes ranked high on our list are guaranteed to be both significantly regulated and significantly periodic. Second, we require consistency in peak time across experiments, which allows us to assign a time of peak expression to each gene.

ACKNOWLEDGEMENTS

The authors would like to thank Henrik Bjørn Nielsen for valuable help on microarray data normalization. This work was supported by

grants from the Danish National Research Foundation, the Danish Technical Research Council and Novo Nordisk A/S.

REFERENCES

- Cho,R.J., Campbell,M.J., Winzler,E.A., Steinmetz,L., Conway,A., Wodicka,L., Wolfsberg,T.G., Gabrielian,A.E., Landsman,D., Lockhart,D.J., *et al.* (1998) A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell*, **2**, 65–73.
- Cho,R.J., Huang,M., Campbell,M.J., Dong,H., Steinmetz,L., Sapinoso,L., Hampton,G., Elledge,S.J., Davis,R.W. and Lockhart,D.J. (2001) Transcriptional regulation and function during the human cell cycle. *Nat. Genet.*, **27**, 48–54.
- Cooper,S. (2004a) Is whole-culture synchronization biology's 'perpetual-motion machine'? *Trends Biotechnol.*, **22**, 266–269.
- Cooper,S. (2004b) Rejoinder: whole-culture synchronization cannot, and does not, synchronize cells. *Trends Biotechnol.*, **22**, 274–276.
- Gollub,J., Ball,C.A., Binkley,G., Demeter,J., Finkelstein,D.B. and Hebert,J. (2003) The Stanford Microarray Database: data access and quality assessment tools. *Nucleic Acids Res.*, **31**, 94–96.
- Johansson,D., Lindgren,P. and Berglund,A. (2003) A multivariate approach applied to microarray data for identification of genes with cell cycle-coupled transcription. *Bioinformatics*, **19**, 467–473.
- Lee,T.I., Rinaldi,N.J., Robert,F., Odom,D.T., Bar-Joseph,Z., Gerber,G.K., Hannett,N., Harbison,C.T., Thompson,C.M., Simon,I. *et al.* (2002) Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science*, **298**, 799–804.
- de Lichtenberg,U., Jensen,T.S., Jensen,L.J. and Brunak,S. (2003) Protein feature based identification of cell cycle regulated proteins in yeast. *J. Mol. Biol.*, **329**, 663–674.
- Lu,X., Zhang,W., Qin,Z.S., Kwast,K.E. and Liu,J.S. (2004) Statistical resynchronization and Bayesian detection of periodically expressed genes. *Nucleic Acids Res.*, **32**, 447–455.
- Luan,Y. and Li,H. (2004) Model-based methods for identifying periodically expressed genes based on time course microarray gene expression data. *Bioinformatics*, **20**, 332–339.
- Mewes,H.W., Frishman,D., Güldener,U., Mannhaupt,G., Mayer,K., Mokrejs,M., Morgenstern,B., Münsterkoetter,M., Rudd,S. and Weil,B. (2002) MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.*, **30**, 31–34.
- Reiner,A., Yekutieli,D. and Benjamini,Y. (2003) Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics*, **19**, 368–375.
- Rustici,G., Mata,J., Kivinen,K., Lió,P., Penkett,C.J., Burns,G., Hayles,J., Brazma,A., Nurse,P. and Bähler,J. (2004) Periodic gene expression program of the fission yeast cell cycle. *Nat. Genet.*, **36**, 809–817.
- Shedden,K. and Cooper,S. (2002) Analysis of cell-cycle gene expression in *Saccharomyces cerevisiae* using microarrays and multiple synchronization methods. *Nucleic Acids Res.*, **30**, 2920–2929.
- Simon,I., Barnett,J., Hannett,N., Harbison,C.T., Rinaldi,N.J., Volkert,T.L., Wyrick,J.J., Zeitlinger,J., Gifford,D.K., Jaakkola,T.S., *et al.* (2001) Serial regulation of transcriptional regulators in the yeast cell cycle. *Cell*, **106**, 697–708.
- Spellman,P.T. and Sherlock,G. (2004a) Final words: cell age and cell cycle are unlinked. *Trends Biotechnol.*, **22**, 277–278.
- Spellman,P.T. and Sherlock,G. (2004b) Reply: whole-culture synchronization—effective tools for cell cycle studies. *Trends Biotechnol.*, **22**, 270–273.
- Spellman,P.T., Sherlock,G., Zhang,M.Q., Iyer,V.R., Anders,K., Eisen,M.B., Brown,P.O., Botstein,D. and Futcher,B. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *S. cerevisiae* by microarray hybridization. *Mol. Biol. Cell*, **9**, 3273–3297.
- Whitfield,M.L., Sherlock,G., Saldanha,A.J., Murray,J.I., Ball,C.A., Alexander,K.E., Matese,J.C., Perou,C.M., Hurt,M.M., Brown,P.O., *et al.* (2002) Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol. Biol. Cell*, **13**, 1977–2000.
- Wichert,S., Fokianos,K. and Strimmer,K. (2004) Identifying periodically expressed transcripts in microarray time series data. *Bioinformatics*, **20**, 5–20.
- Workman,C., Jensen,L.J., Jarmer,H., Berka,R., Gautier,L., Saxild,H.-H., Nielsen,C., Brunak,S. and Knudsen,S. (2002) A new non-linear normalization method to reduce variability in DNA microarray experiments. *Genome Biol.*, **3**, 48.1–48.16.
- Zhao,L.P., Prentice,R. and Breeden,L. (2001) Statistical modeling of large microarray data sets to identify stimulus-response profiles. *Proc. Natl Acad. Sci. USA*, **98**, 5631–5636.