Yeast
Yeast 2006; **23**: 261–277.
Published online in Wiley InterScience (www.interscience.wiley.com). **DOI:** 10.1002/yea.1351

Research Article

The more the merrier: comparative analysis of microarray studies on cell cycle-regulated genes in fission yeast

Samuel Marguerat[1#], Thomas S. Jensen[2#], Ulrik de Lichtenberg[2], Brian T. Wilhelm[1], Lars J. Jensen[3] and Jürg Bähler[1*]

[1] *Cancer Research UK Fission Yeast Functional Genomics Group, Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK*
[2] *Center for Biological Sequence Analysis, BioCentrum-DTU, Technical University of Denmark, DK-2800 Lyngby, Denmark*
[3] *EMBL Heidelberg, Meyerhofstrasse 1, D-69117 Heidelberg, Germany*

**Correspondence to:*
Jürg Bähler, Cancer Research UK Fission Yeast Functional Genomics Group, Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK.
E-mail: jurg@sanger.ac.uk

[#] *These authors equally contributed to this study.*

Received: 30 November 2005
Accepted: 15 January 2006

Abstract

The last two years have seen the publication of three genome-wide gene expression studies of the fission yeast cell cycle. While these microarray papers largely agree on the main patterns of cell cycle-regulated transcription and its control, there are discrepancies with regard to the identity and numbers of periodically expressed genes. We present benchmark and reproducibility analyses showing that the main discrepancies do not reflect differences in the data themselves (microarray or synchronization methods seem to lead only to minor biases) but rather in the interpretation of the data. Our reanalysis of the three datasets reveals that combining all independent information leads to an improved identification of periodically expressed genes. These evaluations suggest that the available microarray data do not allow reliable identification of more than about 500 cell cycle-regulated genes. The temporal expression pattern of the top 500 periodically expressed genes is generally consistent across experiments and the three studies, together with our integrated analysis, provide a coherent and rich source of information on cell cycle-regulated gene expression in *Schizosaccharomyces pombe*. The reanalysed datasets and other supplementary information are available from an accompanying website: http://www.cbs.dtu.dk/cellcycle/. We hope that this paper will resolve the apparent discrepancies between the previous studies and be useful both for wet-lab biologists and for theoretical scientists who wish to take advantage of the data for follow-up work. Copyright © 2006 John Wiley & Sons, Ltd.

Keywords: ***Schizosaccharomyces pombe*; cell cycle; transcription; microarray; cell division; periodic gene expression; *Saccharomyces cerevisiae*; computational biology**

Supplementary material for this article can be found at: **http://www.interscience.wiley.com/jpages/0749-503X/suppmat/**

Introduction

The terms 'cell cycle-regulated' and 'periodically expressed' are used interchangeably in the literature to describe genes that are expressed in a specific stage during the cell cycle. Since the pioneering work in budding yeast (Cho *et al*., 1998; Spellman *et al*., 1998), cell cycle-regulated gene expression has been studied at a genome-wide level in bacteria, plants and mammals (Laub *et al*., 2000; Ishida *et al*., 2001; Menges *et al*., 2002; Whitfield *et al*., 2002). Recently, three independent groups have used DNA microarrays to identify fission yeast genes that are periodically expressed as a function of the cell cycle (Rustici *et al*., 2004; Peng *et al*., 2005; Oliva *et al*., 2005). For *Schizosaccharomyces pombe* there are thus now more data available on cell cycle-regulated gene expression than for any other organism. This provides valuable biological information and a rich source for theoretical studies (Tyers, 2004; Bähler,

2005a; Gilks *et al.*, 2005; Wittenberg and Reed, 2005). As for other large-scale datasets (e.g. Cho *et al.*, 1998; Spellman *et al.*, 1998), there is only partial agreement between the three studies with regard to the number and identity of periodically expressed genes; together, the *Sz. pombe* studies proposed more than 1300 genes in total to be periodically expressed, but only 360 genes were reported in at least two of the three studies (Oliva *et al.*, 2005). Although such differences probably do not come as a surprise for experts on genomic approaches, they can be disconcerting for biologists who may be confused and lose trust in this type of data. These discrepancies, however, can be explained, and the data are quite consistent with each other when looking beyond a superficial comparison, as discussed below. We provide an overview of the data on periodic genes in fission yeast and focus on reconciling these data and reporting follow-up analyses that compare and integrate all three datasets. We identify the following main reasons for the discrepancies in the reported cell cycle-regulated genes: (a) differences in analysis methods; (b) choices of significance cut-offs; and (c) random experimental noise. Despite their differences, the three datasets are coherent and of comparable quality and, when combined, provide improved detection of periodically expressed genes.

Materials and methods

Microarray expression data

The normalized expression data from the three cell-cycle microarray studies (Rustici *et al.*, 2004; Peng *et al.*, 2005; Oliva *et al.*, 2005) were downloaded from the authors' web pages (Table 1). All values were converted to log-ratios and technical replicates (if present) were averaged. The expression profiles for each gene in each of the 10 experiments were normalized to a mean log-ratio of 0.

Analysis of cell-cycle periodicity

To rank genes, we used a scoring scheme that has been shown to be one of the best for finding cell cycle-regulated genes based on microarray data (de Lichtenberg *et al.*, 2005). Briefly, this scheme is based on two p values that measure the significance of regulation and of periodicity. The p value of regulation for a given expression profile was calculated as the fraction of 10^6 random profiles with a standard deviation above that of the observed profile. To evaluate the periodicity, the Fourier score was calculated for a given expression profile: $F_i = \sqrt{([(\Sigma \sin(\omega t) \cdot x_i(t)]^2 + [\Sigma \cos(\omega t) \cdot x_i(t)]^2)}$, where $\omega = 2\pi/T$, with T being the interdivision time. The optimal interdivision time for each experiment was estimated based on a reference set of 35 genes shown to be periodically expressed in small-scale experiments (Rustici *et al.*, 2004). The p value of periodicity was calculated for each gene by comparing its Fourier score to the Fourier scores of 10^6 random profiles constructed by shuffling the timepoints of the corresponding expression profile. To compensate for interdependencies among timepoints, all p values were normalized to a median of 1. A combined score was calculated by multiplying the p value for regulation ($p_{\text{regulation}}$) and the p value for periodicity ($p_{\text{periodicity}}$) for a given gene and applying penalty terms to ensure that a low score is only obtained if a gene is both significantly regulated and significantly periodic:

$$\text{score} = p_{\text{regulation}} \cdot p_{\text{periodicity}} \left[1 + \left(\frac{p_{\text{regulation}}}{0.001}\right)^2\right]\left[1 + \left(\frac{p_{\text{periodicity}}}{0.001}\right)^2\right].$$

To combine evidence from multiple experiments, the p values were multiplied to yield a total p value of regulation and a total p value of periodicity from which the combined score was calculated.

Calculation of peak times and alignment of time scales

Within a single experiment, the time of peak expression for a gene is determined by fitting its expression profile with a sine wave. We report this *peak time* as a percentage of the cell cycle to compensate for the difference in interdivision time between the experiments. Because different synchronization methods release cells from different points in the cell cycle, the timescales need to be aligned before peak times can be compared between experiments. To find the optimal alignment, we used a simulated annealing heuristic to minimize the total peak time difference between experiments for the top 500 genes. We arbitrarily defined the zero timepoint as the median peak time of the genes in Cluster 2 (M/G_1 phase) of Rustici *et al.* (2004). For each gene, a combined peak time was calculated as a weighted average (on a

Table 1. Overview of the three microarray studies on the fission yeast cell cycle

	Rustici *et al.* (2004)	**Peng *et al.* (2005)**	**Oliva *et al.* (2005)**
Array platform			
Microrray type	Spotted PCR array	Spotted oligo array	Spotted PCR array
Probe size	180–500 bp	50 bp	500–1000 bp
Spots/array	~13 000[a]	~10 000[b]	6720
Timecourse experiments			
Elutriation	3× (2 cycles)	1× (2 cycles)	2× (3 cycles)
cdc25 block-release	3× (2 cycles)[c]	1× (2 cycles)	1× (3 cycles)
Elu & *cdc10* block-release	1× (1 cycle)	—	1× (<1 cycle)
Elu & *cdc25* block-release	1× (1 cycle)	—	—
sep1 Δ *cdc25* block-release	1× (2 cycles)	—	—
Timepoint frequency	15 min	10 min	8–16 min
Timepoints per timecourse	18–22	33–38	12–52
Additional experiments			
sep1 Δ deletion mutant	4×	3×	—
sep1 overexpression strain	2×	—	—
*ace2*Δ deletion mutant	5×	3×	—
ace2 overexpression strain	2×	—	—
cdc10-C4 mutant	4×	3×	—
G_1 phase arrest: *cdc10-V50*/control	—	1× (24 timepoints)	—
G_1 phase arrest: -nitrogen	—	—	1× (7 timepoints)
G_1 phase arrest: *cdc10-M17*	—	—	2×
S phase arrest: hydroxyurea	1×	—	—
S phase arrest: *cdc22-M45*	—	—	2×
G_2 phase arrest: *cdc25-22*	—	—	2×
M phase arrest: *nuc2-663*	—	—	2×
Data and analysis			
Total number of arrays used	196	104	170
Arrays used to identify periodic genes	160	71	143
Identification of periodic genes	Fourier transform Determine *p* values Filter on amplitude Visual inspection	Gaussian smoothing Fourier transform CDC score	Fourier transform Determine *p* values
Proposed number of periodic genes	407	747	750
Overlap with other studies[d]	77.1%	48.3%	47.6%
Clustering of genes	Gaussian mixture model (ArrayMiner)	Hierarchical (Eisen *et al.*, 1998)	Hierarchical (Eisen *et al.*, 1998)
Access processed data	**http://www.sanger.ac.uk/PostGenomics/S_pombe**	**http://giscompute.gis.a-star.edu.sg/~gisljh/CDC**	**http://www.redgreengene.com**
Public repository	ArrayExpress (E-MEXP-54 to E-MEXP-64)	—	ArrayExpress (E-TABM-5 and E-TABM-8)

[a] All features are printed in duplicate to obtain two measurements (Lyne *et al.*, 2003).
[b] Two different oligos per gene.
[c] Two biological repeats, of which one also contains a technical repeat.
[d] Percentage of proposed periodic genes that were also proposed in at least one of the three studies (Figure 1B).

circle) of the peak time obtained in each of the 10 experiments (for details, see de Lichtenberg *et al.*, 2005; **http://www.cbs.dtu.dk/cellcycle/**).

Benchmark sets

To evaluate the quality of any list of periodically expressed genes proposed based on microarray time series, we constructed three independent benchmark sets, each consisting of genes for which there is independent experimental evidence for cell cycle-regulated expression.

The first set (B1) consists of 40 genes, for which periodicity has been demonstrated in small-scale experiments; slight variations of this list have been used by all three groups to verify their data analyses. From the list of 35 genes used by Rustici

et al. (2004), we excluded the gene *suc22*, as this produces two transcripts of which only one is periodic. We then added five genes that have recently been reported to be cell cycle-regulated (Alonso-Nunez *et al*., 2005) and the gene *uvi31* (Kim *et al*., 1997).

The second set (B2) consists of genes whose promoters are bound by at least one of the known cell-cycle transcription factors Cdc10p, Res1p, Res2p or Fkh2p, based on ChIP-chip experiments in unsynchronized cells (B.T.W., unpublished data). In cases of divergently transcribed genes, where binding is observed between the genes, both flanking genes are included in the set. Although false positives will be detected in these experiments, the set should be rich in genes that are truly regulated during the cell cycle. Genes also present in set B1 were excluded to ensure independence between the benchmark sets, leaving 188 genes in set B2.

The third set (B3) consists of genes that are differentially expressed in microarray experiments using unsynchronized strains with genetic perturbations of the genes *ace2, sep1* or *cdc10*, encoding transcription factors as well as S-phase arrested cells (Table 1; Rustici *et al*., 2004). All genes present in sets B1 and B2 were removed to ensure independence of the benchmark sets, leaving 321 genes in set B3.

Results and discussion

Overview of microarray papers analysing the fission yeast cell cycle

Table 1 provides a comparison of experimental platforms and designs of the microarray studies addressing cell cycle-regulated gene expression in fission yeast. All three studies used cells synchronized by centrifugal elutriation (selective synchronization) as well as cells synchronized using the temperature-sensitive cell-cycle mutant *cdc25-22* (whole-culture synchronization), with different array platforms and differing numbers of timepoints and biological repeats. The papers also include additional experiments to address the regulation of periodic transcription and/or to analyse specific cell-cycle phases in more detail (Table 1). The three studies propose different numbers of periodically expressed genes: Rustici *et al*. (2004) suggested 407 genes based on five experiments, whereas Peng *et al*. (2005) and Oliva *et al*. (2005) proposed 747 and 750 genes based on two and three experiments, respectively (Table 1, Figure 1A). When comparing the three proposed sets of genes, a striking and somewhat discouraging conclusion is the poor overlap between the genes reported as periodically expressed in the three studies (Figure 1A; Oliva *et al*., 2005). For the two papers that reported around 750 periodic genes, the overlap with the other gene lists is especially poor (Table 1; Figure 1A). When redoing this comparison, we noticed that some of the discrepancies arise as a consequence of using different (non-systematic) names for the same genes. Correcting for these gene-mapping problems improves the overlap between the studies (Figure 1B). As shown below, however, the main reasons for the poor overlap are differences in data interpretation, while the data *per se* show quite good agreement with each other. To assess these issues, we first evaluate the quality of the published datasets and analyses and then go back to discuss what the different experimental data show when analysed with the same computational method.

How best to detect periodic gene expression?

Genes that are periodically expressed as a function of the cell cycle are defined as those that change in expression levels with a period equal to the inter-division time. Various algorithms have been developed for identifying periodically expressed genes, and the choice of method can have a profound impact on the interpretation of cell-cycle microarray data. In budding yeast, for example, widely different sets of genes have been proposed, based on analysing the same microarray data with different computational methods (Zhao *et al*., 2001; de Lichtenberg *et al*., 2003, 2005; Johansson *et al*., 2003; Luan and Li, 2004; Ahdesmäki *et al*., 2005; Willbrand *et al*., 2005). While single studies identified between 150 and 1000 periodically expressed genes, in total over 1800 different genes have been proposed to be periodic. A recent comparison of the available computational methods showed that some methods simply work better than others in identifying truly cell-cycle-regulated genes and that the better methods yield more reproducible results when applied to different microarray datasets (de Lichtenberg *et al*., 2005). Thus, a large part of the differences between the lists of periodic genes in

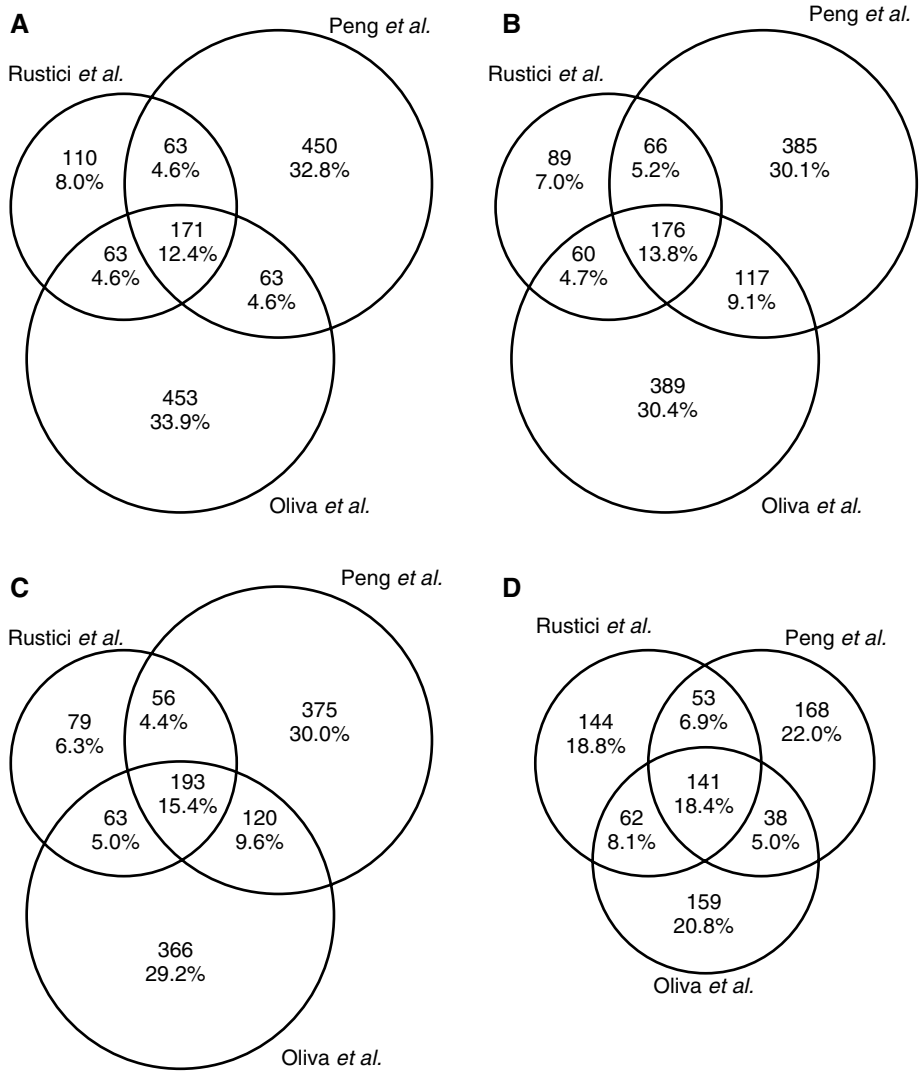

Figure 1. Overlap between genes identified as cell cycle-regulated in the microarray studies by Rustici *et al.* (2004), Peng *et al.* (2005) and Oliva *et al.* (2005). (A) Venn diagram showing the numbers originally reported by Oliva *et al.* (2005). (B) Correcting for the use of alternative gene names in the three studies improves the overlap. Not identifiable genes, non-coding RNAs and pseudogenes have been removed (see **http://www.cbs.dtu.dk/cellcycle/** for details). (C) Re-ranking the genes based on the scoring scheme by de Lichtenberg *et al.* (2005) and selecting the same number of genes as originally proposed by each group further improves the agreement between the three experiments. (D) The best relative overlap is attained by also limiting the comparison to more conservative gene lists consisting of only the top 400 periodic genes from each study

the *Sz. pombe* microarray studies could be due to differences in how the data were analysed.

In all three *Sz. pombe* studies, the identification of periodic genes was based, in part, on Fourier analysis. Rustici *et al*. (2004) and Oliva *et al*. (2005) then calculated probabilities for the oscillations to arise from random fluctuations by shuffling the data for each gene within each experiment, identifying more than 1000 genes each with apparently significant periodicity. Oliva *et al*. (2005) ranked the genes by their p values and proposed a list of 750 periodically expressed genes, whereas Rustici *et al*. (2004) filtered out genes with only subtle changes in expression levels and then visually inspected the remaining profiles to arrive at a smaller, more conservative list of 407 genes.

Peng *et al*. (2005) instead ranked the genes by a CDC score, which combines Fourier analysis with additional terms; their threshold (747 genes) and false-discovery estimates were based on randomly shuffling the data.

To evaluate the different proposed lists of periodically expressed genes, we compared them with independent experimental evidence for cell-cycle regulation using the three benchmark sets described in Materials and methods. In Figure 2 and Supplemental Figure S1, the number of genes retrieved from a given benchmark set is shown as a function of the number of genes included from each ranked list, whereas each non-ranked list is shown as a single point. Reassuringly, all proposed gene lists show much better than random overlap with the genes from all three benchmark sets. The enrichment over randomness (the slope of the curves) is also strongest for the highest ranked genes that scored best in the original analyses. As one goes down the ranked lists, however, the slopes of the curves eventually become comparable to that of the line representing random expectation. After the first 500 genes or so, there is no further enrichment of

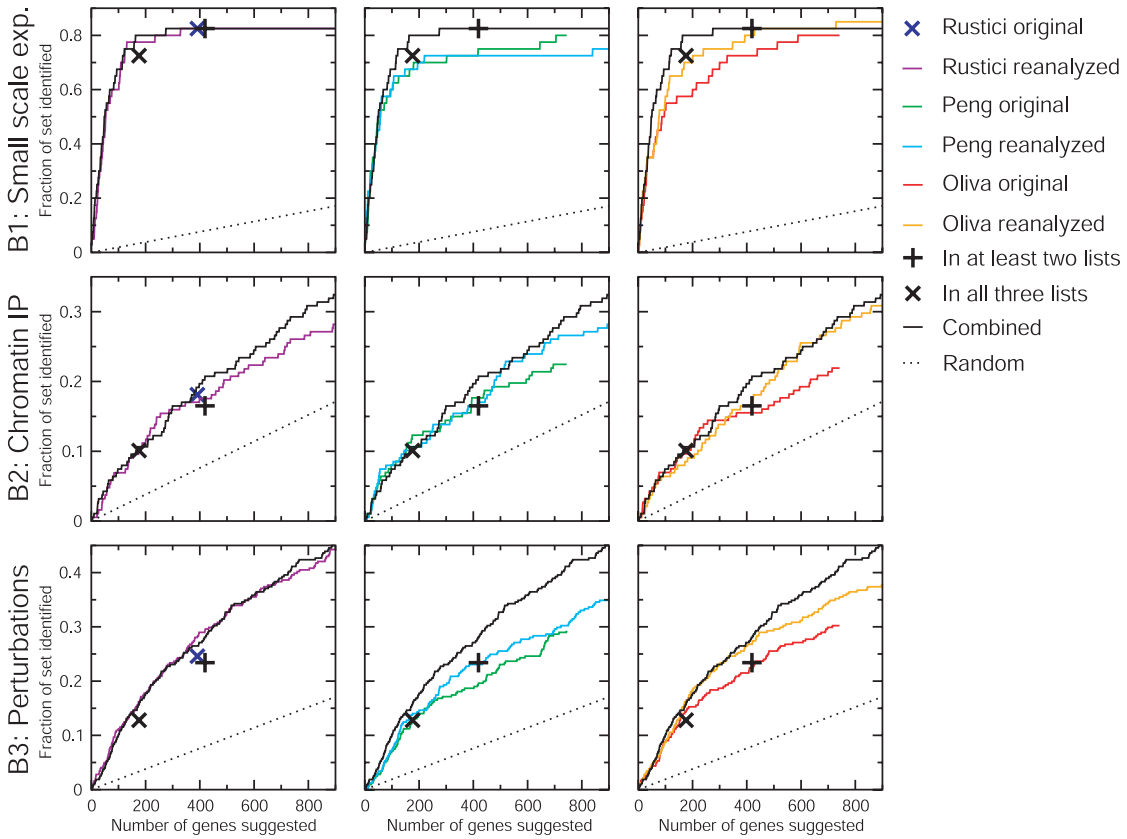

Figure 2. Benchmark analyses of the different proposed lists of periodic genes. The fraction of genes retrieved from each benchmark set is plotted against the gene rank (number of genes suggested). A steeper curve is equivalent to a better correspondence with the independent evidence for cell-cycle regulation and thus with a better gene list. Non-ranked lists are represented in the diagram as crosses. The list named 'in all three lists' is composed of the 176 genes proposed to be periodic in all three previous studies, whereas the list named 'in at least two lists' is made up by the 419 genes proposed by at least two of the three groups (Figure 1B). All curves eventually reach either saturation (B1) or a slope similar to the random expectation line, from which point on there is no enrichment of genes from that benchmark set. Note that none of the three microarray studies can detect periodicity for several of the previously reported cell cycle-regulated genes; in fact, five of these genes (*ppb1, uvi31, cmk1, rrg1* and *mcm2*) were not identified as cell cycle-regulated by any of the three studies or in our combined analysis. Indeed, periodicity of at least some of these genes looks questionable even from data in the original publications, or published data on the same gene are conflicting (Forsburg and Nurse, 1994; Plochocka-Zulinska *et al*., 1995; Kim *et al*., 1997; Anderson *et al*., 2002)

genes from the benchmark sets, and selecting more genes from the ranked lists is therefore no better than picking additional genes at random from the genome. Figure 2 can also be used to compare the performance of the three analyses relative to each other: the Rustici *et al*. list of 407 genes shows a better overlap with the benchmark sets than the highest scoring 407 genes from the lists of Oliva *et al*. and Peng *et al*. The benchmark set B3 might slightly favour the Rustici *et al*. list, as it is based on data from the same array platform. At this point, however, it is not clear to what degree these results are influenced by the number of experiments made by each group or by the methods used to measure periodicity in expression.

To better compare the different datasets, we reanalysed the data from all three groups using the method described by de Lichtenberg *et al*. (2005). In all cases, our reanalysis performs at least as well as the original analyses published (Figure 2). In brief, our analysis method combines a p value for regulation with a p value for periodicity, to ensure that top-ranking genes exhibit both a significant regulation and a periodic pattern of expression. On *S. cerevisiae* data, this approach has been shown to perform better than other methods for the identification of periodic genes, especially compared to those modelling only the shape of the expression profile without taking into account the magnitude of regulation (de Lichtenberg *et al*., 2005). The latter could explain the slightly poorer performance of the analysis by Oliva *et al*. (2005), who ranked the genes based on a score that is independent of the magnitude of regulation. Fewer or no improvements in performance are observed when reanalysing the data by Rustici *et al*. (2004) and Peng *et al*. (2005), who both used methods that take into account the magnitude of regulation. In accordance with the improved performance on the benchmark sets (Figure 2), our reanalysis also improves the agreement among the three datasets (Figure 1B, C). The apparent discrepancies between the datasets are thus in part explained by the use of different and less accurate analysis methods.

The relative performance of the reanalysis of data from the three groups (Figure 2) also shows that the best lists are derived from datasets that include more timecourse experiments (Table 1). This finding is confirmed when applying the de Lichtenberg *et al*. (2005) analysis method, either to all 10 experiments individually or to all 10 experiments in combination. Reanalysing each of the individual experiments (Supplemental Figure S1) demonstrates only minor differences in performance, which suggests that all timecourse data are of comparable quality. It is therefore not surprising that the best results were obtained when applying our analysis method to all 10 experiments in combination (black curves in Figures 2, S1). This is even better than taking the 176 genes included in all three published lists or the 419 genes included in at least two of the original lists (Figures 1B, 2). This shows that our integrated analysis of all data is superior to simple voting schemes at combining the signals from the 10 experiments, which, although being of comparable overall quality, each make independent and complementary contributions and together improve the identification of cell cycle-regulated genes.

How many genes are periodically expressed in fission yeast?

Peng *et al*. (2005) and Oliva *et al*. (2005) suggested almost twice as many periodically expressed genes as Rustici *et al*. (2004) (Table 1; Figure 1A). As already pointed out, the microarray expression data reveal no natural, distinct threshold between periodically expressed genes and genes expressed at constant levels throughout the cell cycle (de Lichtenberg *et al*., 2005; Oliva *et al*., 2005). Instead, there is a continuum from clearly periodic genes to genes that do not seem to fluctuate as a function of the cell cycle, with a large grey zone in between. This could suggest that many genes are only weakly cell cycle-regulated (<1.5-fold change in expression levels) as well as noise in the microarray data. The transition can be seen in the benchmark analyses as a gradual decrease in the slope of the curves as more genes are included (Figures 2, S1). The decision on the number of genes that are deemed periodic is thus ultimately based on a somewhat arbitrary threshold. However, the slope of every curve eventually becomes comparable to that of random expectation, from which point on the available benchmark sets cannot justify the inclusion of more genes, and the threshold should therefore be set before this point. Not surprisingly, gene lists based on smaller numbers of experiments reach this limit earlier. In the best-case scenario, where all 10 timecourse experiments are combined, the enrichment over random

is strong for the first 300 genes, then gradually decreases and is essentially lost altogether beyond the first 500 genes (Figure 2). These analyses thus lend little support to the proposition of ~750 cell cycle-regulated genes, particularly not when based on only two or three experiments. Indeed, both the original lists and the reanalyses of the datasets by Peng *et al*. (2005) and Oliva *et al*. (2005) display hardly any enrichment beyond the first 400 genes.

To test whether this lack of enrichment is due to limitations of the benchmark sets, we determined reproducibility by comparing the ranked lists obtained from our reanalysis of any two of the 10 individual experiments (Figure 3). When selecting the top 300 genes from each list, the average overlap is 121 genes. However, when comparing the next 300 genes (ranks 301–600), the reproducibility drops dramatically to only 31 genes on average. In comparison, the expected overlap between two randomly selected lists of the same size is 19 genes. At ranks 601–900, there is essentially no enrichment over random expectation. This demonstrates that only about the first 300 genes are reasonably reproducible between any two of the 10 experiments, consistent with the observations made from Figures 2 and S1. This drop in reliability for lower ranked genes is also confirmed by visually inspecting Figure 4, which shows the expression profiles of the same three sets of genes used in Figure 3. The top 300 genes show clear periodicity and large amplitudes, whereas these properties are less apparent to the eye in the other two groups. Similar conclusions are reached when comparing the set of 176 genes proposed in all three original studies to those included in at least two of the studies (243 genes) or those only proposed by one study (863 genes). Only the genes proposed by all three groups show a clear periodic pattern of expression (Supplemental Figure S2).

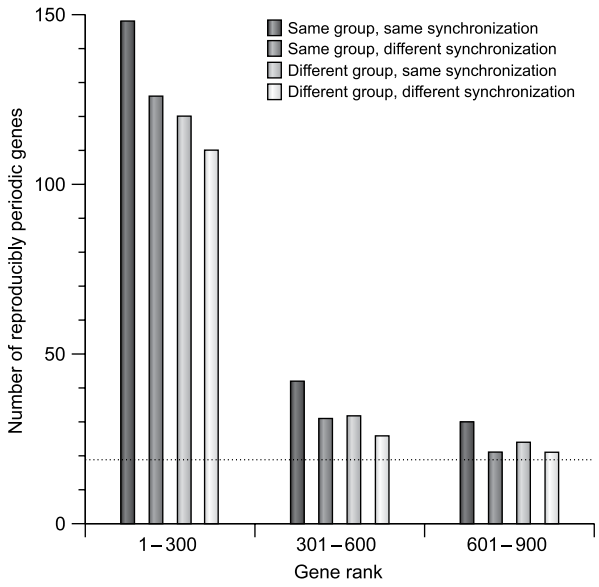

Figure 3. Reproducibility of genes identified in two experiments analysed by the method of de Lichtenberg *et al.* (2005). Each bar shows the average number of overlapping genes among two different experiments analyzed individually when using the 300 highest ranking genes from each experiment (left), or using the genes ranked from 301–600 (middle) and 601–900 (right). The comparisons are subdivided based on whether the experiments were performed in the same laboratory and by using the same protocol for cell-cycle synchronization. There is good reproducibility among the 300 highest ranking genes, but the reproducibility drops close to random expectation (19 genes) for genes in the second and third sets

The gene sets visualized in Figure 4 are sorted by their peak time, whereby the pattern of periodicity stands out very clearly across a group of genes. Although a periodic pattern is seen even for the two bottom panels in Figure 4, this periodicity is not reproducible at the single-gene level when comparing individual experiments (Figure 3). The patterns of periodicity among the lower ranked genes indicate that there are truly periodically expressed genes beyond the highest ranking 300–400 genes, but identification of these requires many independent datasets and even then comes at the price of including an increasing number of false positives as one goes down the ranks.

Together, the analyses shown in Figures 2–4, S1 and S2 demonstrate that only for the most significant 300–400 genes is the signal strong enough to deem periodicity based on a single timecourse experiment; by combining 10 timecourses, some 500 periodically expressed genes can be identified with reasonable confidence. Beyond that, regulation becomes weaker, noisier and/or less reproducible between experiments and therefore more questionable. Notably, many of the profiles of lower ranking genes look, at best, marginally periodic to the eye and would probably not be judged as cell cycle-regulated based on traditional methods (e.g. Figure 8D). A major reason for the poor overlap between the originally reported gene lists is thus that the studies by Peng *et al*. (2005) and

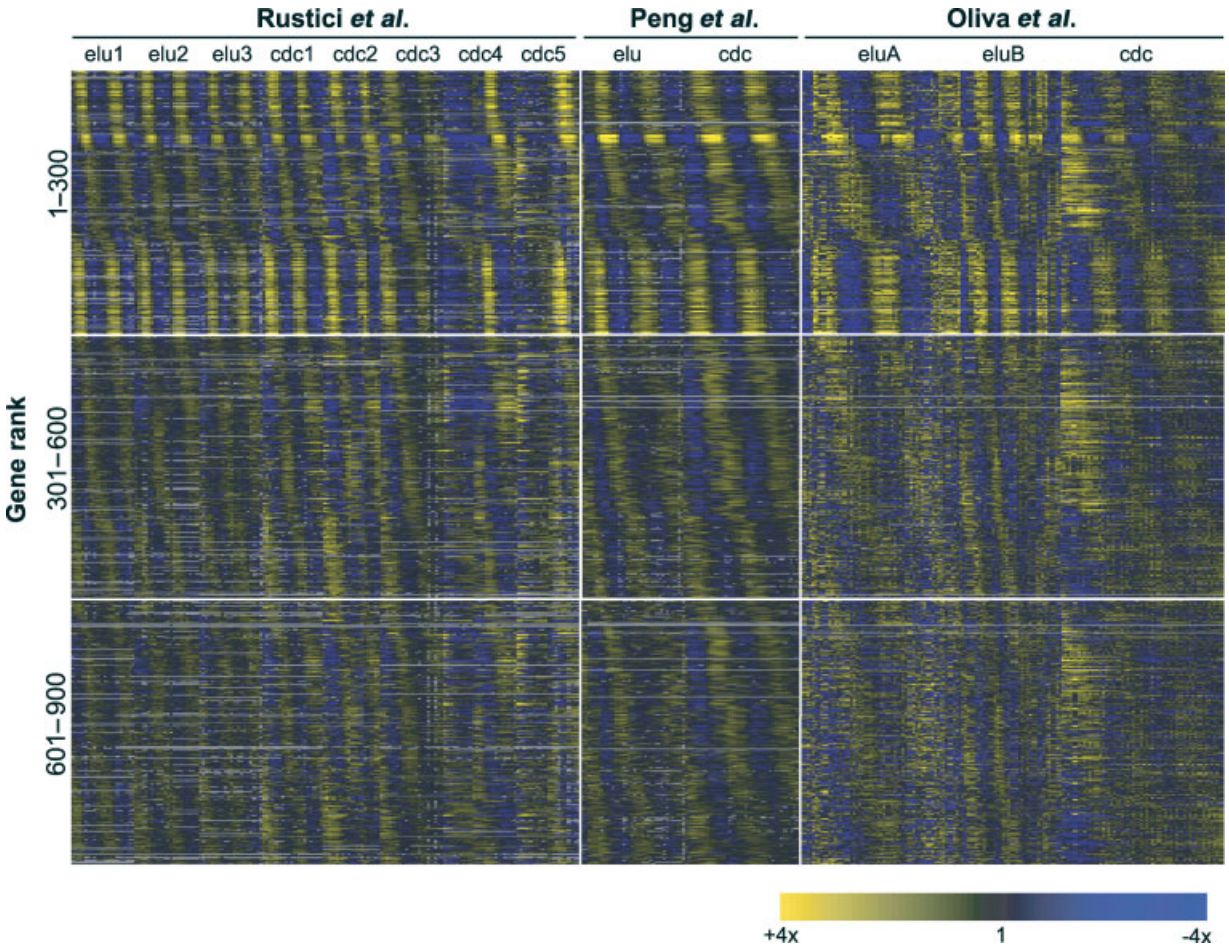

Figure 4. Diagram of gene expression profiles as a function of gene ranking. Each of the three panels shows the expression profiles for sets of 300 genes, ordered by their average peak times. The first panel contains the 300 highest ranking genes from our combined analysis of all 10 experiments, with the next two corresponding to genes ranked 301–600 and 601–900, respectively. Columns represent experimental timepoints of the timecourse experiments indicated. Experiments cdc1–cdc5 refer to the following experiments of Rustici *et al.* (2004): *cdc25* block-release 1 and 2, *sep1*Δ *cdc25* block-release, elutriation and *cdc10* block-release, and elutriation and *cdc25* block-release, respectively. All timecourse experiments that have been used to identify periodically expressed genes in the original studies are shown. The mRNA levels (fold change) at each timepoint relative to levels in unsynchronized cells are colour-coded, as indicated at the bottom, and missing data are shown in grey. The expression profiles for the top 300 genes appear to have a higher magnitude of regulation and better periodicity in all experiments than genes in the second and third panel. Although there is an overall periodic pattern in the ordered expression profiles for the lower panels, this pattern is largely unreproducible at the single-gene level (Figure 3)

Oliva *et al*. (2005) suggest many more cell cycle-regulated genes than can reliably be detected from their data. As shown in Figure 1D, the relative agreement between the three studies can be further improved if smaller, more conservative lists of periodic genes are compared. As will be shown below, most of the remaining discrepancies are explained by the general noise level in the microarray data, which leads to different genes that make it into the different top 400 lists, together with the fact that there is a continuum between cell cycle-regulated and non-regulated genes. In fact, differences in the array platforms and experimental protocols only account for a minor part of the apparent discrepancies in Figure 1D (see below). These discrepancies are expected when comparing conclusions from noisy datasets that are each based on only few replicates, and should not be interpreted as a

lack of congruence between the data from different groups.

Why do statistical tests suggest too many periodically expressed genes?

Since only a small fraction of the cell cycle-regulated genes have been identified through small-scale studies, it is difficult to assess the number of false positives in a proposed list of genes. In contrast, it is easy to count how many of the known periodic genes are confirmed by microarray analysis. This has led researchers analysing cell-cycle microarray expression data in different organisms to propose quite inclusive gene lists that have good sensitivity (including most of the known genes) but an unknown false positive rate. Peng *et al*. (2005) and Oliva *et al*. (2005) employed permutation-based statistical tests and estimated their false discovery rates to be 1.1% and 0.022%, respectively. These exceptionally low error rates are difficult to reconcile with an overlap of only 293 genes between lists of ~750 genes each (Figure 1B).

Peng *et al*. (2005) and Oliva *et al*. (2005) suggested higher sensitivity or better cell-cycle synchrony as reasons why they identified more periodic genes than did Rustici *et al*. (2004), although this is not supported by our reanalyses described above. In fact, when using an automated method, Rustici *et al*. (2004) identified >1000 'significant' periodic genes with p values < 0.01 in their data but decided to propose a smaller, more conservative list of cell cycle-regulated genes. It is important to realize that random permutation of timecourse data may overestimate the statistical significance of periodicity, and hence lead to an overly optimistic false discovery rate. This is because successive timepoints are not guaranteed to be independent of each other, thereby violating the underlying assumption of the statistical tests (Kruglyak and Tang, 2001). This problem is increased if samples are collected at a higher frequency and is particularly true for the data by Peng *et al*. (2005), who applied Gaussian smoothing to their expression profiles, thus artificially enhancing dependency between neighbouring timepoints. While p values are useful for judging the relative periodicity of a set of genes (ranking), it is problematic to rely on their absolute values. When reanalysing the data, we have found that the raw p values calculated based on random permutations are overestimated by about an order of magnitude, meaning that the false positive rates reported in the three original studies are probably underestimated accordingly. Using statistics alone to set the threshold, two of the groups suggested roughly twice as many genes as their data can support, as judged from the reproducibility between replicate experiments (Figure 3) and consistency with independent sources of evidence for cell-cycle regulation (Figures 2, S1). The only alternative explanation is that well over 1000 genes are periodically expressed and that each study simply detects a different subset of these, although this would contradict the claim of less than 20% false negatives by Peng *et al*. (2005). In any case, even if there were many more periodically expressed genes, our analyses show that their profiles are not reproducible between experiments (Figure 3).

Do microarray or synchronization methods give rise to biases?

In Figure 3, we have subdivided the pairwise comparisons of gene lists from different experiments into four classes, based on whether the two experiments were performed by the same group and based on the same synchronization method. This subdivision demonstrates that experiments performed by the same group tend to be more similar, as do experiments using the same synchronization method. For instance, experiments performed by the same group and using the same synchronization technique on average have 148 genes in common among the top scoring 300 genes, compared to 110 genes among experiments performed by different groups with different synchronization techniques. We speculate that the lab bias is largely due to differences in probe and chip design that may cause some genes to be detected less well on some arrays. One should note, however, that these biases are small and rather insignificant in comparison to the general level of reproducibility of only around 50% between the top 300 genes from any two experiments. Figure 3 thus contradicts the proposition that biases from different synchronization methods give rise to widely different, and spurious, results (Cooper and Shedden, 2003). Instead, the primary source of variation seems to be random, experimental noise rather than systematic experimental biases.

Minor variations in the data leading to different genes that make it into the different top 300 lists are the main reason for the small overlap between any two experiments, as the ranking in any single experiment is influenced by subtle differences in periodicity and regulation. We can therefore conclude that the data from the three groups are of similar overall quality (Figures 2, S1), and they are congruent (Figure 3). These findings also show that the poor overlap observed in Figure 1D is simply a consequence of comparing three lists, which have each been derived from too few experiments to eliminate random, experimental noise. Since many independent experiments are needed to extract the underlying signal from noisy data, it is no surprise that our combined analysis of all 10 experiments yields the best results. The differences in synchronization techniques, microarray design and laboratory protocols among the 10 experiments therefore make the entire dataset more information-rich than would have been the case had all the experiments been performed in the same laboratory using the same method.

Do periodically expressed genes peak at the same time in different experiments?

Agreeing on the cell cycle-regulated genes is one part of the problem; in principle, the time of expression of a gene could still vary between experiments. To examine this in more detail, we assigned a time of peak expression for each periodic gene in a given experiment by fitting its expression profile with a sine wave. These *peak times* were made comparable across experiments by converting the time scales from minutes to percentages of the cell cycle and subsequently aligning the scales with each other (for details, see de Lichtenberg *et al*., 2005). For the four phase-specific gene clusters defined by Rustici *et al*. (2004), we calculated the smoothed distribution of peak times for each of the 10 individual timecourse experiments (Figure 5). Reassuringly, we found that each gene cluster peaked at roughly the same time and occupied a similar fraction of the cell cycle in all experiments. As expected, the G_2 phase constituted about 60–70% of the cell cycle of fission yeast, in contrast to budding yeast, where the four cell-cycle phases are of similar length. Importantly, the different synchronization techniques led to similar results, although the distribution of peak times for the S phase genes was slightly delayed for *cdc25* block-release experiments compared to the elutriation experiments, indicating that the relative lengths of cell-cycle phases differed somewhat between these types of experiments.

Given the reproducibility of peak times between the different experiments (Figure 5), a single gene-specific peak time can be calculated that summarizes the expression across all 10 experiments by weighing the individual peak times relative to each other based on the periodicity of the gene in each given experiment (de Lichtenberg *et al*., 2005). A nice feature of this scheme is that the average peak time is associated with a standard deviation that quantifies the consistency (or spread) in the temporal expression for each gene. We can thus show that the great majority of the top 500 periodic genes exhibit highly consistent peak times across all experiments (Supplemental Figure S3).

How is periodic gene expression distributed across the cell cycle?

A simple way to globally view the temporal behaviour of gene expression during the cell cycle is to plot the distribution of peak times (Figure 6). This reveals two major waves where gene expression peaks are concentrated, one in M phase and one in early G_2 phase, as also observed by Oliva *et al*. (2005). Although there are genes peaking in expression at all stages of the cell cycle, there is a clear drop in the later half of G_2 phase before the largest wave is initiated at the G_2–M transition. The numerous genes peaking in early G_2 phase are generally much weaker regulated than those peaking during M to S phases (Rustici *et al*., 2004; Figures 4, 8) and show poor reproducibility between experiments (see below); their enrichment in functions such as ribosome biogenesis (Oliva *et al*., 2005) suggests that this surge in cell cycle-regulated gene expression may prepare the cell for the increased growth during G_2 phase (Mitchison and Nurse, 1985). Despite the two stages enriched in periodically expressed genes, the overall timing of peaks is quite continuous across the cell cycle, rather than in discrete steps (Figure 4), probably reflecting regulatory fine-tuning and/or differences in mRNA stability.

Based on their estimated p values, Oliva *et al*. (2005) proposed that as many as 2000 genes are weakly but significantly periodic. They supported

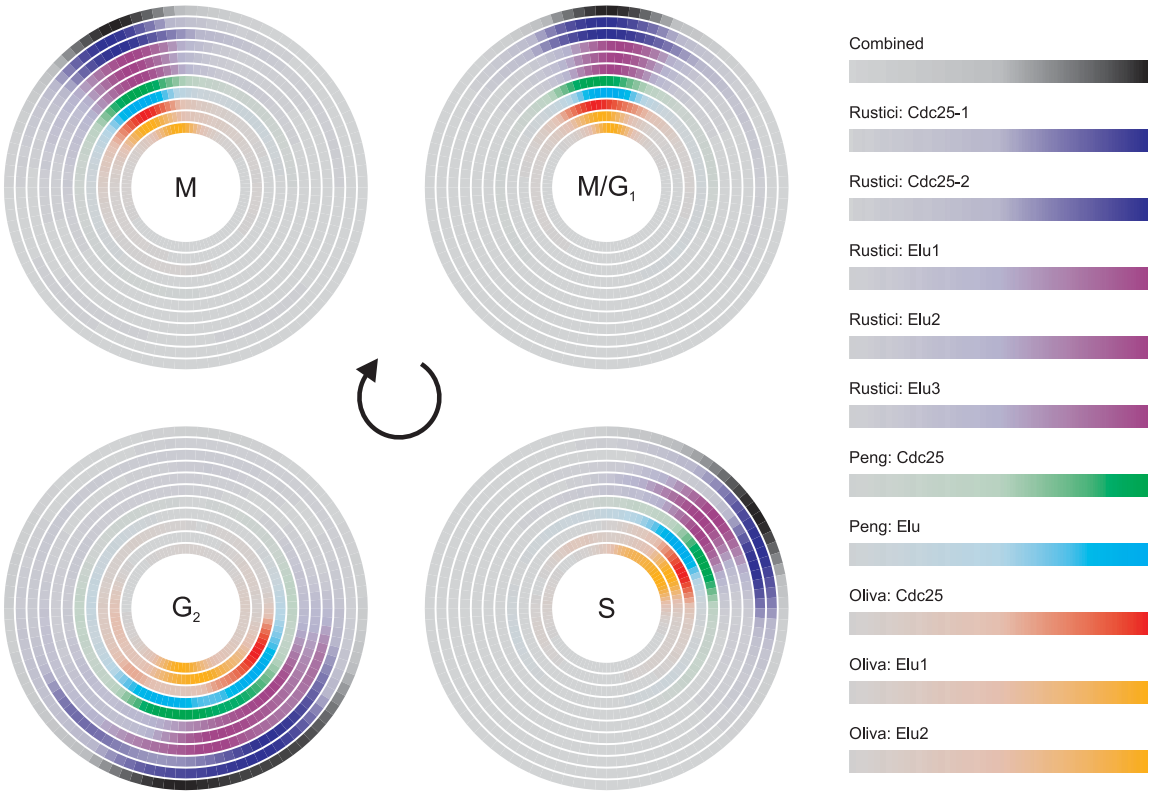

Figure 5. Distribution of peak times for four phase-specific clusters defined by Rustici *et al.* (2004). Each circle represents an experiment and visualizes the distribution of peak times for a cluster of genes peaking at the indicated cell-cycle phases. For each experiment the time-scale was normalized, experiments were aligned relative to each other and a smoothed distribution was made for each cluster of genes to assess the duration of phases (only genes among the top 500 in our reanalysis were included; for details, see de Lichtenberg *et al.*, 2005; and **http://www.cbs.dtu.dk/cellcycle/**). The duration of each phase is similar in all 10 experiments and the combined peak time is in good agreement with that from the individual experiments

this by showing that when analysing the 4000 lowest ranked genes in their study, the same two major waves of transcription were observed as for their 750 most regulated genes. When plotting the distribution of peak times for the 2000 least periodic genes according to our combined analysis of all 10 timecourses (Supplemental Figure S4), we generally cannot reproduce the distribution seen for the highest scoring 500 genes (Figure 6). We too observe a tendency for more genes to be assigned to early G_2 phase, but late G_2 is also rich in expression peaks, which is the opposite of what is observed for the highly scoring genes. Furthermore, we see no sign of a second wave in M phase among the 2000 lowest scoring genes (Supplemental Figure S4). This analysis therefore does not support the periodicity of genes far down the list, but reflects that if one fits sine curves to the profiles regardless of how random they look, the overall pattern shows a tendency for clustering in G_2 phase. Although there may be subtle fluctuations among the low ranking genes, the data presented here (Figures 2–4, S1, S2) indicate that these fluctuations do not arise from active regulation of these genes during the cell cycle. The fluctuations are not reproducible at the level of single genes, and the genes that are fluctuating show no significant overlap with any of the benchmark gene sets for which cell-cycle regulation is supported by other sources. Although the phenomenon as such might be interesting, more work would be required to clarify the biological relevance of these subtle oscillations. At this point, it is not even clear whether they should be viewed

as a real biological phenomenon or as a bias introduced by the treatment of the microarray data (e.g. normalization).

What do the three microarray papers tell us about the control of periodic gene expression?

Despite the poor overlap between the proposed periodically expressed genes, the three cell-cycle studies report a coherent picture of gene expression regulons. All three papers defined groups of genes that behave in a similar way across experimental conditions using different clustering algorithms (Table 1). Whereas the peak times define the timing of expression for each gene (Figure 6), the clustering analyses also take into account the shape of the expression profiles and incorporate additional experiments (e.g. transcription factor mutants). Rustici *et al*. (2004) describe four large clusters, which together contain almost all periodic genes, while Peng *et al*. (2005) and Oliva *et al*. (2005) examined eight smaller clusters each, which together cover only a fraction of the periodic genes. The genes within each cluster peak at a similar time during the cell cycle, reflecting the intuitive notion that peak time of expression is a critical feature of periodic transcription. The different clusters can be divided in three main groups: M/G_2 phase, S phase or G_2 phase. Reassuringly, different clusters within the same group share many genes, while clusters from different temporal groups show little overlap (Figure 7).

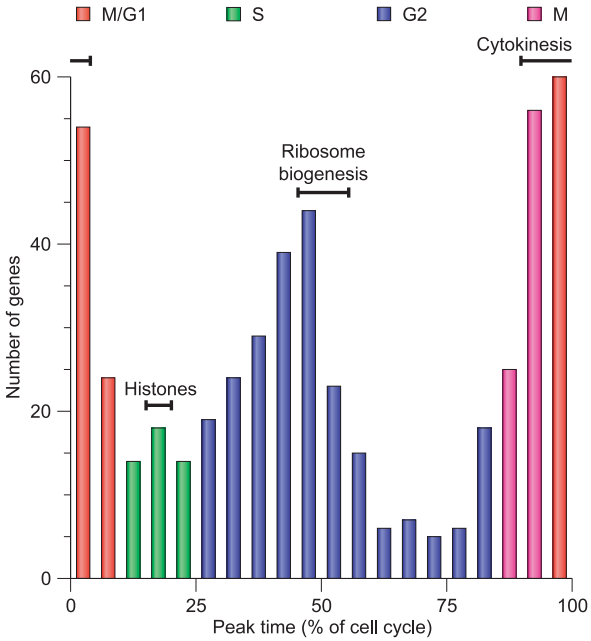

Figure 6. Histogram showing the distribution of average peak times for the highest ranking 500 genes from our analysis of all 10 experiments in combination. The duration of phases is based on Figure 5 and the distribution for histone genes, ribosome biogenesis genes and genes involved in cytokinesis is included as bars to aid visual interpretation (see **http://www.cbs.dtu.dk/cellcycle/** for more details). In M- and M/G_1 phases, the number of genes peaking in expression is far higher than average. In early G_2 phase, there is another burst in cell cycle-regulated genes, while few genes are periodically expressed during late G_2 phase

The M/G_1 phase includes the highest numbers of clusters: Clusters 1 and 2 (Figure 8A; Rustici *et al*., 2004), SFF(1), SFF(2), Ace2 and MCB (Peng *et al*., 2005) and Cdc15, Cdc18 and Eng1 (Oliva *et al*., 2005). There is good congruence between related clusters (Figure 7). Enrichment of regulatory motifs and genetic experiments agree that the M/G_1 clusters contain targets of Forkhead, Ace2p and MBF transcription factors, which regulate genes for mitosis, cell division and DNA replication, respectively. The data also support a model where a wave of transcription regulated by the Forkhead transcription factor Sep1p precedes and induces an Ace2p-dependent transcriptional wave, as is also emerging from other papers (Martín-Cuadrado *et al*., 2003; Dekker *et al*., 2004; Alonso-Nunez *et al*., 2005; Lee *et al*., 2005; Petit *et al*., 2005). Together, these findings define a transcriptional cascade for cell separation in fission yeast (Bähler, 2005b). Besides Sep1p and Ace2p, other regulators such as the Fkh2p forkhead transcription factor may be involved in this pathway (Buck *et al*., 2004; Bulmer *et al*., 2004; Rustici *et al*., 2004; Szilagyi *et al*., 2005). More work is required to understand how these regulators work together to control periodic transcription during mitosis. Detailed reviews and comparisons with the corresponding regulatory pathways in *S. cerevisiae* are available (Bähler, 2005a; Wittenberg and Reed, 2005).

The S phase is characterized by the strongly regulated and tightly co-expressed histone genes (Figures 7, 8B), the regulation of which is not understood. In addition, Rustici *et al*. (2004) reported a group of genes with lower amplitudes peaking during S phase, but these were not enriched for any functional category. Oliva *et al*. (2005) described a small cluster of genes close to telomeres, although most of these are almost

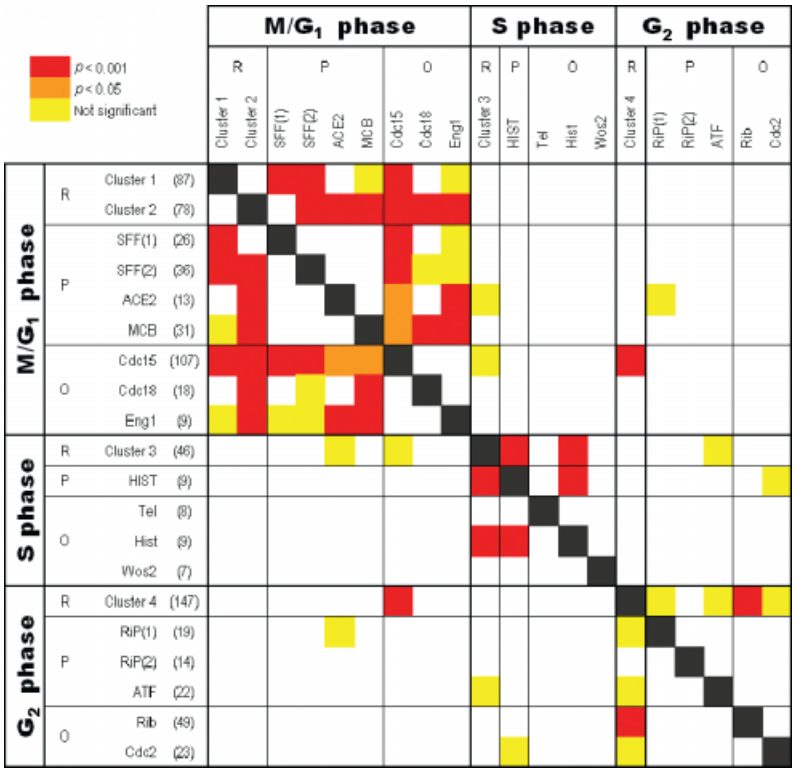

Figure 7. Comparison between the gene clusters described in the three cell-cycle microarray studies. The significance of the overlaps between clusters described by Rustici *et al.* (2004; R), Peng *et al.* (2005; P) and Oliva *et al.* (2005; O) is colour-coded, while a white space means no overlap. Numbers in parentheses indicate cluster size. The universe used as a reference for *p* value calculation is the 1282 genes found in at least one study (Figure 1B). Rustici *et al.* (2004) describe four large clusters, 1–4, containing genes peaking at successive times of the cell cycle. Peng *et al.* (2005) describe eight small clusters: the SFF(1), SFF(2), ACE2, MCB and ATF clusters are named after promoter motifs, the HIST cluster contains histones genes, and the RiP(1) and RiP(2) clusters contain ribosomal proteins. Oliva *et al.* (2005) describe eight clusters named Cdc15, Cdc18, Eng1, Tel (for telomeres), Hist (for histones), Wos2, Rib (for ribosome biogenesis) and Cdc2

identical in sequence, making it difficult to know whether all or just one of them is periodically expressed.

Genes peaking during G_2 phase are somewhat different, as they show less reproducible and generally much weaker regulation. Accordingly, the overlap between the different G_2 clusters is markedly lower than for the M/G_1 and S phase clusters; the only significant overlap is between Cluster 4 from Rustici *et al.* (2004) and the ribosome cluster (Rib) from Oliva *et al.* (2005), which is enriched for genes functioning in ribosome biogenesis (Figures 7, 8C). Peng *et al.* (2005) reported two small clusters containing ribosomal proteins. No promoter motifs were enriched in the ribosome cluster, and Oliva *et al.* (2005) proposed that global transcriptional repression during mitosis could account for the weak oscillation of these genes. This idea is supported by the observation that this cluster was repressed in *nuc2* mutants with condensed mitotic chromosomes (Oliva *et al.*, 2005), although the chromosome compaction in these mutants is stronger than during normal mitosis. Further experiments will be required to substantiate this interesting hypothesis.

Besides genes involved in cell growth, a number of stress genes peak during G_2 phase (genes in Cluster 4, the ATF cluster, and the Wos2 and Cdc2 clusters), which are induced in a range of environmental stresses (Chen *et al.*, 2003). Several of these genes seem at best marginally regulated as a function of the cell cycle (e.g. Figure 8D), but more than half of them are present in our top 500 list of periodic genes. Regulation of these genes could be caused by the synchronization methods, because they showed lower reproducibility across

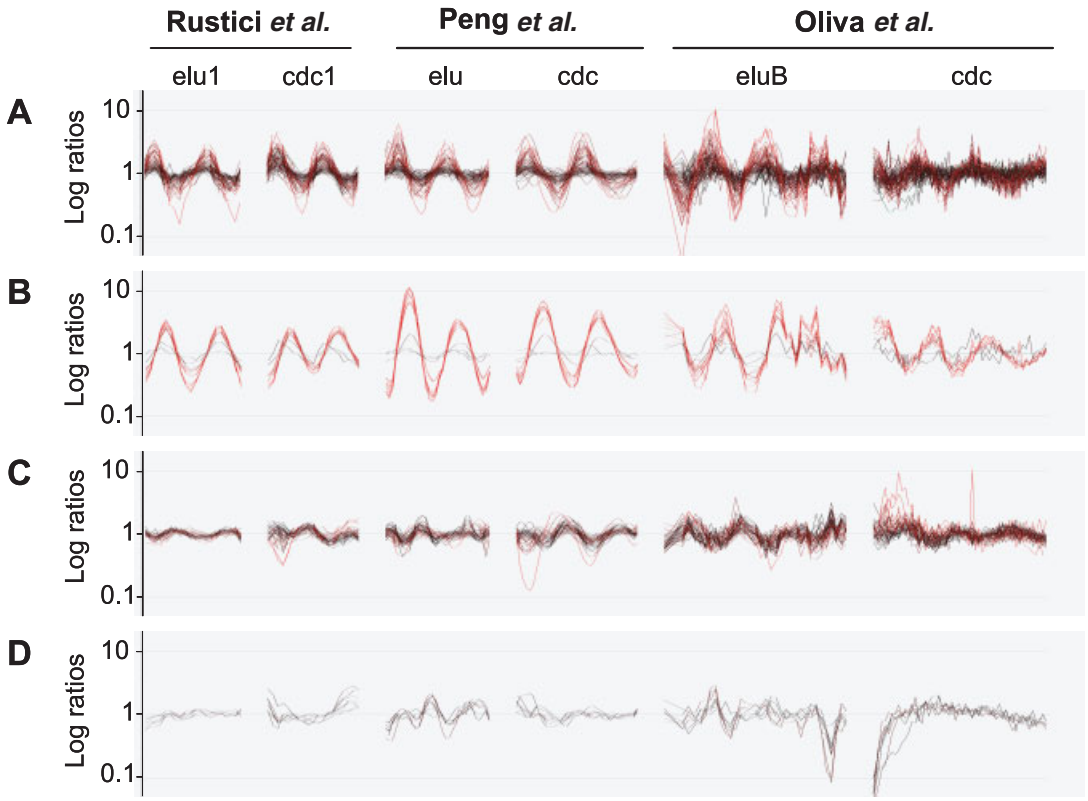

Figure 8. Comparison between the gene clusters described in the three cell-cycle microarray studies. Genes belonging to four clusters are shown in six different experiments. From left to right: elutriation 1 and *cdc25* block-release 1 from Rustici *et al.* (2004; two cell cycles), elutriation and *cdc25* block-release from Peng *et al.* (2005; two cell cycles), elutriation B and *cdc25* block-release from Oliva *et al.* (2005; three cell cycles). The distance on the *x* axis is proportional to time. The values on the *y* axis are normalized and zero centred log ratios between mRNA levels of synchronized cells at different timepoints and mRNA levels in unsynchronized cells. Peng *et al.* (2005) additionally applied Gaussian smoothing to their data. (A) Cluster 1 of Rustici *et al.* (2004), which contains the highly regulated genes forming the first forkhead-dependent transcriptional wave during M/G_1 phase. (B) Histone genes (S phase). (C) Ribosome biogenesis cluster of Oliva *et al.* (2005), which contains genes peaking with small amplitudes during G_2 phase, most of them functioning in ribosome biogenesis. (D) Wos2 cluster of Oliva *et al.* (2005), which contains genes with stress-response elements in their promoter regions peaking during early G_2 phase

experiments, and some of them were mostly regulated in the *cdc25* experiment, which requires a temperature shift. The periodicity of these genes suggests that the cell cycle and environmental stress response are linked, and two recent studies have started to shed light on how these processes are coordinated (Lopez-Aviles *et al.*, 2005; Petersen and Hagan, 2005).

Is cell cycle-regulated gene expression evolutionarily conserved?

The periodically expressed genes identified in fission yeast have been compared to those reported in budding yeast (Cho *et al.*, 1998; Spellman *et al.*, 1998). All three *Sz. pombe* cell-cycle studies agree that although there is a significant overlap in regulated genes, less than 50% of the orthologous gene pairs are periodic with high amplitude in both yeasts. Of our top 500 periodic genes identified by reanalysing all 10 experiments, 353 have an orthologue in budding yeast. 102 of these of are also among the top 500 periodically expressed genes in budding yeast microarray studies when applying the same computational method (de Lichtenberg *et al.*, 2005). Distinct regulatory patterns of cell-cycle genes between *S. cerevisiae, C. albicans* and *Sz. pombe* have recently also been reported

by Ihmels *et al*. (2005). Thus, cell-cycle regulation of gene expression is only partially conserved during evolution, although it does show a substantially higher conservation than the regulation of other processes, such as meiotic differentiation (Mata *et al*., 2002).

Conclusions

The three microarray expression studies of the fission yeast cell cycle together provide a wealth of data, including 10 time series experiments, which are of comparable quality according to our benchmark analyses. Yet rather poor agreement was observed when comparing the three published lists of periodically expressed genes (Oliva *et al*., 2005). We have revealed four primary causes for discrepancies between the proposed lists: (a) inconsistencies in gene naming; (b) use of different analysis methods for identifying periodic genes; (c) each individual experiment is subject to random noise; and, perhaps most importantly, (d) two of the three studies proposed more periodic genes than can reliably be detected from their data. We could detect only minor systematic differences between datasets produced by different laboratories or using different synchronization techniques. The data themselves are thus congruent, but subject to random experimental noise, which explains the remaining lack of overlap (Figure 1D). As demonstrated by our meta-analysis, the best results are obtained when using a powerful computational method to integrate all available data. The combination of all data from the three independent studies provides an information-rich dataset that is superior to the data from any single experiment or laboratory (hence 'the more the merrier' in the title). Based on benchmark and reproducibility analyses, we conclude that, even in this best situation, no more than about 500 periodically expressed genes can be reliably identified based on the available data. Although there may be more genes that are marginally cell cycle-regulated, increasing the list beyond the highest scoring 500 periodically expressed genes will come at a considerable cost of false positives. The temporal expression pattern of the top 500 genes is highly consistent across all 10 experiments, which shows that the three studies provide a coherent description of cell-cycle regulated gene expression in *Sz. pombe*. Accordingly, there has been good agreement between the three studies with regard to various gene expression modules and their regulation. We hope that our integrated analyses and datasets clarify the reasons for discrepancies between the original studies and that they will be useful for follow-up studies, both experimental and theoretical.

Acknowledgements

We thank Alvis Brazma and Juan Mata for comments on the manuscript. Research in the Bähler laboratory is funded by Cancer Research UK (CUK), Grant No. C9546/A5262 and by DIAMONDS, an EC FP6 Lifescihealth STREP (LSHB-CT-2004-512143). U.d.L and T.S.J. are supported by DIAMONDS and by grants from the Danish National Research Foundation and the Danish Technical Research Council (Systemic Transcriptomics in Biotechnology), L.J.J. by BioSapiens, an EC FP6 Lifescihealth NOE (LSH6-CT-2003-503265), S.M. by a fellowship from the Swiss National Science Foundation, and B.T.W. by Sanger Postdoctoral and Canadian NSERC fellowships.

References

Ahdesmäki M, Lahdesmäki H, Pearson R, Huttunen H, Yli-Harja O. 2005. Robust detection of periodic time series measured from biological systems. *BMC Bioinf* **6**: 117.

Alonso-Nunez ML, An H, Martin-Cuadrado AB, *et al*. 2005. Ace2p controls the expression of genes required for cell separation in *Schizosaccharomyces pombe*. *Mol Biol Cell* **16**: 2003–2017.

Anderson M, Ng SS, Marchesi V, *et al*. 2002. *plo1*$^+$ regulates gene transcription at the M–G_1 interval during the fission yeast mitotic cell cycle. *EMBO J* **21**: 5745–5755.

Bähler J. 2005a. Cell-cycle control of gene expression in budding and fission yeast. *Annu Rev Genet* **39**: 69–94.

Bähler J. 2005b. A transcriptional pathway for cell separation in fission yeast. *Cell Cycle* **4**: 39–41.

Buck V, Ng SS, Ruiz-Garcia AB, *et al*. 2004. Fkh2p and Sep1p regulate mitotic gene transcription in fission yeast. *J Cell Sci* **117**: 5623–5632.

Bulmer R, Pic-Taylor A, Whitehall SK, *et al*. 2004. The forkhead transcription factor Fkh2 regulates the cell division cycle of *Schizosaccharomyces pombe*. *Eukaryot Cell* **3**: 944–954.

Chen D, Toone WM, Mata J, *et al*. 2003. Global transcriptional responses of fission yeast to environmental stress. *Mol Biol Cell* **14**: 214–229.

Cho RJ, Campbell MJ, Winzeler EA, *et al*. 1998. A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol Cell* **2**: 65–73.

Cooper S, Shedden K. 2003. Microarray analysis of gene expression during the cell cycle. *Cell Chromosome* **2**: 1.

de Lichtenberg U, Jensen LJ, Fausboll A, *et al*. 2005. Comparison of computational methods for the identification of cell cycle-regulated genes. *Bioinformatics* **21**: 1164–1171.

de Lichtenberg U, Jensen TS, Jensen LJ, Brunak S. 2003. Protein feature based identification of cell cycle regulated proteins in yeast. *J Mol Biol* **329**: 663–674.

Dekker N, Speijer D, Grun CH, *et al*. 2004. Role of the alpha-glucanase Agn1p in fission–yeast cell separation. *Mol Biol Cell* **15**: 3903–3914.

Eisen MB, Spellman PT, Brown PO, Botstein D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* **95**: 14 863–14 868.

Forsburg SL, Nurse P. 1994. The fission yeast cdc19$^+$ gene encodes a member of the MCM family of replication proteins. *J Cell Sci* **107**: 2779–2788.

Gilks WR, Tom BD, Brazma A. 2005. Fusing microarray experiments with multivariate regression. *Bioinformatics* **21**(suppl 2): ii137–ii143.

Ihmels J, Bergmann S, Berman J, Barkai N. 2005. Comparative gene expression analysis by a differential clustering approach: application to the *Candida albicans* transcription program. *PLoS Genet* **1**: e39.

Ishida S, Huang E, Zuzan H, *et al*. 2001. Role for E2F in control of both DNA replication and mitotic functions as revealed from DNA microarray analysis. *Mol Cell Biol* **21**: 4684–4699.

Johansson D, Lindgren P, Berglund A. 2003. A multivariate approach applied to microarray data for identification of genes with cell cycle-coupled transcription. *Bioinformatics* **19**: 467–473.

Kim SH, Kim M, Lee JK, *et al*. 1997. Identification and expression of *uvi31*$^+$, a UV-inducible gene from *Schizosaccharomyces pombe*. *Environ Mol Mutagen* **30**: 72–81.

Kruglyak S, Tang H. 2001. A new estimator of significance of correlation in time series data. *J Comput Biol* **8**: 463–470.

Laub MT, McAdams HH, Feldblyum T, Fraser CM, Shapiro L. 2000. Global analysis of the genetic network controlling a bacterial cell cycle. *Science* **290**: 2144–2148.

Lee KM, Miklos I, Du H, *et al*. 2005. Impairment of the TFIIH-associated CDK-activating kinase selectively affects cell cycle-regulated gene expression in fission yeast. *Mol Biol Cell* **16**: 2734–2745.

Lopez-Aviles S, Grande M, Gonzalez M, *et al*. 2005. Inactivation of the Cdc25 phosphatase by the stress-activated Srk1 kinase in fission yeast. *Mol Cell* **17**: 49–59.

Luan Y, Li H. 2004. Model-based methods for identifying periodically expressed genes based on time course microarray gene expression data. *Bioinformatics* **20**: 332–339.

Lyne R, Burns G, Mata J, *et al*. 2003. Whole-genome microarrays of fission yeast: characteristics, accuracy, reproducibility, and processing of array data. *BMC Genom* **4**: 27.

Martín-Cuadrado AB, Dueñas E, Sipiczki M, Vázquez de Aldana CR, Del Rey F. 2003. The endo-β-1,3-glucanase eng1p is required for dissolution of the primary septum during cell separation in *Schizosaccharomyces pombe*. *J Cell Sci* **116**: 1689–1698.

Mata J, Lyne R, Burns G, Bähler J. 2002. The transcriptional program of meiosis and sporulation in fission yeast. *Nat Genet* **32**: 143–147.

Menges M, Hennig L, Gruissem W, Murray JA. 2002. Cell cycle-regulated gene expression in *Arabidopsis*. *J Biol Chem* **277**: 41 987–42 002.

Mitchison JM, Nurse P. 1985. Growth in cell length in the fission yeast *Schizosaccharomyces pombe*. *J Cell Sci* **75**: 357–376.

Oliva A, Rosebrock A, Ferrezuelo F, *et al*. 2005. The cell cycle-regulated genes of *Schizosaccharomyces pombe*. *PLoS Biol* **3**: e225.

Peng X, Karuturi RK, Miller LD, *et al*. 2005. Identification of cell cycle-regulated genes in fission yeast. *Mol Biol Cell* **16**: 1026–1042.

Petersen J, Hagan IM. 2005. Polo kinase links the stress pathway to cell cycle control and tip growth in fission yeast. *Nature* **435**: 507–512.

Petit CS, Mehta S, Roberts RH, Gould KL. 2005. Ace2p contributes to fission yeast septin ring assembly by regulating *mid2*$^+$ expression. *J Cell Sci* **118**: 5731–5742.

Plochocka-Zulinska D, Rasmussen G, Rasmussen C. 1995. Regulation of calcineurin gene expression in *Schizosaccharomyces pombe*. Dependence on the ste11 transcription factor. *J Biol Chem* **270**: 24 794–24 799.

Rustici G, Mata J, Kivinen K, *et al*. 2004. Periodic gene expression program of the fission yeast cell cycle. *Nat Genet* **36**: 809–817.

Spellman PT, Sherlock G, Zhang MQ, *et al*. 1998. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol Biol Cell* **9**: 3273–3297.

Szilagyi Z, Batta G, Enczi K, Sipiczki M. 2005. Characterisation of two novel fork-head gene homologues of *Schizosaccharomyces pombe*: their involvement in cell cycle and sexual differentiation. *Gene* **348**: 101–109.

Tyers M. 2004. Cell cycle goes global. *Curr Opin Cell Biol* **16**: 602–613.

Whitfield ML, Sherlock G, Saldanha AJ, *et al*. 2002. Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol Biol Cell* **13**: 1977–2000.

Willbrand K, Radvanyi F, Nadal JP, Thiery JP, Fink TM. 2005. Identifying genes from up–down properties of microarray expression series. *Bioinformatics* **21**: 3859–3864.

Wittenberg C, Reed SI. 2005. Cell cycle-dependent transcription in yeast: promoters, transcription factors, and transcriptomes. *Oncogene* **24**: 2746–2755.

Zhao LP, Prentice R, Breeden L. 2001. Statistical modeling of large microarray datasets to identify stimulus–response profiles. *Proc Natl Acad Sci USA* **98**: 5631–5636.

