

Research article

Open Access

# Phase Coupled Meta-analysis: sensitive detection of oscillations in cell cycle gene expression, as applied to fission yeast

Saumyadipta Pyne\*<sup>1,2</sup>, Roe Gutman<sup>3</sup>, Chang Sik Kim<sup>4</sup> and Bruce Futcher\*<sup>5</sup>

Address: <sup>1</sup>Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, MA 02142, USA, <sup>2</sup>Present address: Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02115, USA, <sup>3</sup>Department of Statistics, Harvard University, Cambridge, MA 02138, USA, <sup>4</sup>Department of Biological Sciences, RCWD, Sookmyung Women's University, Seoul, Republic of Korea and <sup>5</sup>Department of Molecular Genetics and Microbiology, Stony Brook University, Stony Brook, NY 11794, USA

Email: Saumyadipta Pyne\* - [saumyadipta\\_pyne@dfci.harvard.edu](mailto:saumyadipta_pyne@dfci.harvard.edu); Roe Gutman - [rgutman@fas.harvard.edu](mailto:rgutman@fas.harvard.edu); Chang Sik Kim - [cskim@kangwon.ac.kr](mailto:cskim@kangwon.ac.kr); Bruce Futcher\* - [bfutcher@ms.cc.sunysb.edu](mailto:bfutcher@ms.cc.sunysb.edu)

\* Corresponding authors

Published: 17 September 2009

Received: 29 May 2009

BMC Genomics 2009, **10**:440 doi:10.1186/1471-2164-10-440

Accepted: 17 September 2009

This article is available from: <http://www.biomedcentral.com/1471-2164/10/440>

© 2009 Pyne et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Abstract

**Background:** Many genes oscillate in their level of expression through the cell division cycle. Previous studies have identified such genes by applying Fourier analysis to cell cycle time course experiments. Typically, such analyses generate p-values; i.e., an oscillating gene has a small p-value, and the observed oscillation is unlikely due to chance. When multiple time course experiments are integrated, p-values from the individual experiments are combined using classical meta-analysis techniques. However, this approach sacrifices information inherent in the individual experiments, because the hypothesis that a gene is regulated according to the time in the cell cycle makes two independent predictions: first, that an oscillation in expression will be observed; and second, that gene expression will always peak in the same phase of the cell cycle, such as S-phase. Approaches that simply combine p-values ignore the second prediction.

**Results:** Here, we improve the detection of cell cycle oscillating genes by systematically taking into account the phase of peak gene expression. We design a novel meta-analysis measure based on vector addition: when a gene peaks or troughs in all experiments in the same phase of the cell cycle, the representative vectors add to produce a large final vector. Conversely, when the peaks in different experiments are in various phases of the cycle, vector addition produces a small final vector. We apply the measure to ten genome-wide cell cycle time course experiments from the fission yeast *Schizosaccharomyces pombe*, and detect many new, weakly oscillating genes.

**Conclusion:** A very large fraction of all genes in *S. pombe*, perhaps one-quarter to one-half, show some cell cycle oscillation, although in many cases these oscillations may be incidental rather than adaptive.

## Background

Cells reproduce and divide using an ordered set of processes. The cell division cycle is usually divided into four

phases, called G1 (Gap 1), S (DNA Synthesis), G2 (Gap 2) and M (Mitosis). In late G1 phase, cells commit to a round of cell division, and in other ways prepare for the upcom-

ing duplication; in S phase they replicate their DNA; in G2 they prepare for mitosis, and during mitosis they segregate their chromosomes, form two nuclei around these two sets of chromosomes, and finally the two new cells separate from one another. These ordered processes are extremely complex, involving hundreds if not thousands of proteins. These processes are regulated and assisted by changes in gene transcription: that is, many genes needed for DNA synthesis are transcribed just before S phase; many genes needed for mitosis are transcribed just before M phase, and so on. Genes regulated in this way - i.e., expressed at a particular time in the cell division cycle, with the effect of aiding progress through a particular part of the cell division cycle - are called cell cycle regulated genes.

In principle, there might be two kinds of genes whose expression oscillates as a function of progress through the cell cycle. What we will call adaptive cell cycle regulation refers to regulation that has the effect of aiding progress through a particular part of the cell cycle, for instance the up-regulation of DNA ligase or histone expression during the DNA synthesis phase. One might expect natural selection in favor of such regulation. In addition, however, there might be what we will call "incidental" cell cycle oscillation, where a gene oscillates in expression, not because the oscillation is directly helpful to the cell, but as a consequence of something else. For instance, chromatin typically condenses in preparation for mitosis, and this condensation might interfere with the transcription of some genes. Such genes would oscillate as a function of the cell cycle, but the oscillation would not be adaptive. In view of the fact that the cell division cycle entails massive changes in the conformation of the DNA, it might not be surprising if a large proportion of genes oscillated for incidental reasons. In this work, we will usually refer to changes in gene expression through the cell cycle as oscillations, and reserve the phrase "cell cycle regulation" for adaptive regulation leading to oscillation. That is, some cell cycle oscillations are due to cell cycle regulation, while other oscillations may be incidental.

The development of microarrays allowed changes in gene expression to be monitored for all genes in a genome, and microarray technology has been applied to the problem of identifying cell cycle oscillating genes in various organisms [1-5]. In a typical time course experiment, a population of cells is obtained such that all the cells are "synchronized" in one particular phase of the cell cycle, such as G1 phase. The synchronized population is allowed to grow and progress through the cycle, and samples of cells are taken at regular intervals as the cells pass through one, two, or even three consecutive, synchronous rounds of cell division. Messenger RNAs are extracted

from the cells in each sample, and these are analyzed using microarrays so that the pattern of expression over time can be determined for each gene. For a cell cycle oscillating gene, it is expected that gene expression will peak and trough once per cell cycle, and the oscillation can be modeled using Fourier analysis [4]. There are many variations of this approach, and these have been systematically compared by de Lichtenberg et al [6]. Using methods of this general kind, even the earliest studies found that a significant fraction of all genes oscillate through the cycle. For instance, Spellman et al. found 800 cell cycle oscillating genes in the yeast *S. cerevisiae*, which is more than 10% of the total genes (about 6,000). Because the number of cell cycle oscillating genes found is statistically limited by the amount of data and by the noise in the experiments, it is quite possible that the true number of cell cycle oscillating genes could be larger.

In recent years at least ten microarray time course experiments from three labs have studied cell cycle expression in the fission yeast *Schizosaccharomyces pombe* (Rustici et al. [7], Peng et al. [8], Oliva et al. [9]). This has made *S. pombe* currently the organism with the largest amount of cell cycle transcriptome data. These data sets can be combined by various techniques of statistical meta-analysis. The traditional method of combining cell cycle data sets has been to calculate a p-value (generally based on Fourier analysis and permutation testing) for oscillation of each gene in each individual cell cycle time course, and then combine these p-values with a traditional statistical method such as Fisher's inverse chi square method or Stouffer's sum of Z's method [10]. Marguerat et al. [11] analyzed all ten data sets, and by combining p-values (for periodicity of oscillation and also for expression regulation, see below), they found about 500 cell cycle oscillating genes (out of about 5000 total genes).

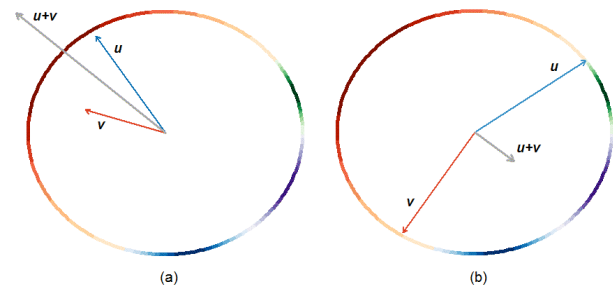
However, we argue here that simply combining p-values is an inefficient approach, which ignores a critical source of information in the available data. Consider a putative cell cycle regulated gene, perhaps involved in S-phase. There are two independent hypotheses about such a gene. First, it will peak and trough once per cell cycle, and thus it will generate a relatively large Fourier sum with a small p-value. This is the hypothesis tested by the standard methods of analysis, including the study by Marguerat et al. But in addition, it can also be predicted that in multiple independent time course experiments, the peak of expression will always be at a similar time in the cell cycle. For instance, a cell-cycle regulated gene involved in DNA synthesis will typically reach peak levels of expression just before S-phase (DNA synthesis), while a cell cycle regulated gene involved in mitosis will typically reach peak levels of expression just before M phase. The second

hypothesis - that there is a particular, consistent, time of peak expression for a cell cycle oscillating gene - is not tested by current methods; we show that, in conjunction with the earlier hypothesis, it can allow for a more powerful test of cell cycle oscillation.

There are major intuitive advantages of testing the peak phase consistency hypothesis. Microarray noise often affects gene expression readout; yet it is less likely to hide the peak signals in a time course. Moreover, if a gene peaks at the same phase in multiple independent experiments, it is highly unlikely for all the consistent peaks to be affected simultaneously by random noise.

To emphasize the phase consistency hypothesis in a different way, let us consider two different genes, and 10 independent cell cycle time course experiments. Suppose that the times of peak expression in the cell cycle are calculated for each gene in each time course. Let these times be converted to phase angle values in the circular range  $0^\circ$  to  $360^\circ$  (equivalently, the interval of  $[0, 2\pi)$  radians), which corresponds to one complete cycle, and assign  $0^\circ$  to the end of mitosis. For the first gene, let the times of peak expression in the 10 time course experiments occur at 10 different cell cycle times, scattered around the cell cycle. For the second gene, let the times of peak expression in the 10 time courses all occur in S-phase, at about  $90^\circ$  in this example. Then for these two genes, based on the variances of their phase angles, one could conclude that the second but not the first gene oscillates, even without knowing p-values.

Here, we propose a new method for the analysis of cell cycle oscillating genes. This new method takes into account both kinds of available information. That is, it uses the standard measures for determining the significance of a gene's oscillation, but in addition, it uses the reproducibility of the cell cycle phase of its peak expression. First, the time course of each gene in each experiment is represented as a vector whose magnitude captures the oscillation information and whose direction captures the peak phase information. Then we combine these two independent sources of information using vector addition; if a gene peaks or troughs in all experiments in roughly the same phase of the cell cycle (we call this property "phase consistency"), then the vectors add to produce a large final vector. Conversely, when the peaks of different experiments are in different phases of the cycle, vector addition produces a small final vector. Fig. 1 shows an example of this kind of analysis over two time courses. The gene shown in Fig. 1a has two time courses in which the phase of peak expression is similar between the time courses, and the resultant vector has large magnitude. The gene shown in Fig. 1b has two time courses in which the



**Figure 1**

**Vector sum based on oscillation and phase.** For a gene, each time course is represented by a vector such that the Periodicity and Expression scores determine the magnitude of the vector, and the cell cycle phase of peak expression gives the direction of the vector. Plot (a) shows two hypothetical time courses with similar phases of peak expression (both peaking in G2 phase). In (a), the sum of the two vectors  $u+v$  yields a vector (plotted in grey) with a greater magnitude than either of the component vectors. Plot (b) shows two time courses with dissimilar phases of peak expression, one peaking at the beginning of M phase, and the other peaking just after S-phase. The sum of these two vectors  $u+v$  yields a vector (in grey) with a lower magnitude than either of the component vectors. The colored circumference refers to the circular distribution of cell cycle phases in *S. pombe* (red-G2, green-M, purple-G1, blue-S) in both plots.

phase of peak expression is not similar, and the resultant vector has small magnitude.

To demonstrate the power of this approach, we re-consider the 10 cell cycle time course experiments of *S. pombe*. We use the same data as Marguerat et al., and we use roughly (though not exactly) the same approach to determine p-values for oscillations in the individual time courses. However, with this new meta-analysis, we find roughly 2,000 cell cycle oscillating genes in *S. pombe*, in contrast to the 500 found by more traditional methods (e.g. Marguerat et al. [11]), a difference we attribute largely to the increase in power due to consideration of peak phase consistency in addition to the p-value of the oscillation. We call this new method "Phase Coupled Meta-analysis", or PCM.

## Methods

Marguerat et al. [11] have already performed a meta-analysis of the *S. pombe* cell cycle data using methods that take into consideration the strength of the oscillations in gene expression. Because we particularly want to highlight the effect of adding phase information, in general we have used the same data and followed the same methods as Marguerat et al. for calculating the strength of oscillations.

### Experiments and data

The ten microarray cell cycle experiments used in the present study were from Rustici et al. 2004; Peng et al. 2005; and Oliva et al. 2005 [7-9], and were the same experiments used by Marguerat et al. [11]. Most of these time courses covered more than one round of cell division. In their meta-analysis, Marguerat et al. used a small set of experimentally validated cell cycle regulated genes to align the cell cycle stages of all ten experiments. They expressed the time of peak expression for each gene in each time course as a percentage of the elapsed cycle, after aligning the various time courses. Using their estimated cell cycle period for each experiment, we transformed these peak times of expression into angular values in the circular range  $[0, 2\pi)$  radians. This produced our input matrix of  $4990 \times 10$  angular data points where the  $(i, j)$ <sup>th</sup> entry represented the cell cycle phase of peak expression for gene  $i$  in experiment  $j$ . In addition, genome-wide data from ten transcription factor knockout and over-expression experiments were used for clustering (see below), which are described elsewhere [9].

### Periodicity and Expression scores

Following de Lichtenberg et al. and Marguerat et al. [6,11], we evaluated the Periodicity score ( $P_{gj}$ ) of each gene  $g$ 's time course in the  $j$ <sup>th</sup> experiment ( $j = 1 \dots 10$ ) using a Fourier sum, and its Expression score ( $E_{gj}$ ) with standard deviation. For assessment of global significance of the periodicity and expression scores, we used permutation testing. Since by shuffling only along the time axis, the standard deviation of a time course (and hence its expression score) does not change, we adopted a different but related strategy. We generated a joint null population of  $10^5$  permuted simulated time courses by shuffling the original time courses in each experiment both along the gene and time axes, thus destroying any signals from biological regulation. Then with respect to these  $10^5$  randomly-generated "genes" we computed the percentile scores  $qP_{gj}$  and  $qE_{gj}$  for every real gene  $g$ . That is, for every real gene, we compared its Periodicity score and Expression score to the scores of the joint null population. The scores for the real genes were then ranked. The percentiles  $qP_{gj}$  and  $qE_{gj}$  were computed as  $cdf_{P_j}(P_{gj})$  ( $P_{gi}$ ) and  $cdf_{E_j}(E_{gj})$  ( $E_{gi}$ ) respectively using the cumulative distribution functions ( $cdf_{P_j}(P_{gj})$  and  $cdf_{E_j}(E_{gj})$ ) of the null population's periodicity and expression scores for experiment  $j$ . The cumulative distribution functions were computed with the R function *ecdf*. As percentiles,  $qP_{gj}$  and  $qE_{gj}$  assumed values between 0 that represented minimal statistical significance and 1 that represented the gene that is

most strongly periodic or expressed. Note that the percentile scores allowed a direct genome-wide ranking mechanism with built-in nonparametric statistical significance owing to the background provided by the joint null distribution.

### Phase Coupled Meta-analysis score

Since a cell cycle oscillating gene with a true phase of peak expression is likely to peak consistently in multiple independent experiments, the phase consistency should cause the oscillatory signals to "add up" in a phase coupled meta-analysis measure. Thus we define our Phase-Coupled Meta-analysis (PCM) score for a gene  $g$  as the magnitude of a sum of  $L_g$  ( $\geq 5$ ) vectors,  $PCM(g) = \sqrt{\left(\sum_{j=1}^{L_g} M_{gj} \cos \theta_{gj}\right)^2 + \left(\sum_{j=1}^{L_g} M_{gj} \sin \theta_{gj}\right)^2}$ , where the  $j$ <sup>th</sup> vector  $(M_{gj}, \theta_{gj})$  of  $g$  has direction  $\theta_{gj}$  given by the phase angle of  $g$  in experiment  $j$  and magnitude  $M_{gj}$  defined as  $M_{gj} = \exp(qP_{gj}) \times \exp(qE_{gj})$  based on the percentile scores described above. The exponential spreading out of their percentile scores allowed the genes to be ranked more distinctively, in particular those with weak expression. For instance, the magnitude a gene that is both strongly periodic and strongly expressed is higher than that of a gene with strong periodicity but weak expression by an exponential factor. Moreover, the exponential function's lower bound of 1 (i.e. for the 0<sup>th</sup> percentile), instead of 0, allowed even the weak expression profiles to be included in the PCM score; only if a gene  $g$  is not present in some experiment  $j$ , the corresponding magnitude  $M_{gj}$  is zero. Finally all genes were ranked in the decreasing order of PCM scores, which we refer to as the PCM rank list (see Additional file 1).

### Circular statistics

Circular variance, test of circular uniformity and sampling from von Mises distribution were computed with the R package *CircStats*. P-values based on the test of circular uniformity were adjusted for multiple hypotheses testing by Benjamini-Hochberg procedure [12]. The median phase angle [13] of every gene was computed as a solution to the following optimization:

$\arg \min_{\theta} \pi - \frac{1}{L_g} \sum_{k=1}^{L_g} |\pi - |\theta_k - \theta||$ , where  $L_g$  is the number of experiments in which the gene  $g$  is present. Genes that were absent in more than five experiments (i.e.  $L_g < 5$ ) were excluded from the PCM analysis (marked as 'NA' in the PCM rank list).

To evaluate the extent of phase consistency among *S. pombe* oscillations, we compared every gene's PCM score with its simulated version (PCM\*) with known phase variance. The simulated score was computed for each gene  $g$  by using the original oscillation information  $\{M_{gj}; j = 1, \dots, L_g\}$  from every experiment but randomizing the original phase information  $\{\theta_{gj}; j = 1, \dots, L_g\}$ :  $PCM^*(g) = \sqrt{\left(\sum_{j=1}^{L_g} M_{gj} \cos \theta_{gj}^*\right)^2 + \left(\sum_{j=1}^{L_g} M_{gj} \sin \theta_{gj}^*\right)^2}$ , by taking the median over fifty random samplings of phases  $\{\theta_{gj}^*; j = 1, \dots, L_g\}$  from the von Mises distribution  $VM(\mu_g, k = 1)$  with mean  $\mu_g$  specified by circular mean of the observed phases  $\{\theta_{gj}; j = 1, \dots, L_g\}$  and unit circular variance ( $1/\kappa = 1$ ). Then we computed the difference  $PCM(g) - PCM^*(g)$  as a measure of deviation of the observed phase variance of a gene  $g$  from a corresponding distribution with fixed circular variance.

### Clustering

For clustering, we used the following two-step refinement strategy --

**Step 1:** we clustered the top 2,000 genes from the PCM rank list using data from the ten transcription factor knockout and over-expression experiments to identify genes with similar regulatory signatures. We used the Partitioning Around Medoids (PAM [14]) algorithm, a version of k-means clustering that is robust against outliers. The optimal number of clusters in this step (found to be 8) was determined by maximizing the Average Silhouette Width [15].

**Step 2:** we extended and used the recently developed Phase Synchronization Clustering (PSC [16]) algorithm on the ten time course experiments to classify the clusters from step 1 into groups of genes having specific peak phase expression. In particular, we focused on a cluster of 551 genes composed of two PAM clusters from step 1 having similar regulatory signatures and containing 43 out of 44 genes from the original ribosomal biogenesis cluster in Oliva et al. The extended PSC algorithm produced a phase-specific cluster of 103 genes, which peaked in mid G2 phase of the cell cycle, as an enhanced version of the original ribosomal biogenesis cluster of Oliva et al. (consequently we named it  $C_{ribo}$ ). The genes in  $C_{ribo}$  are indicated in the PCM rank list (Additional file 1).

In step 2, we extended the original multivariate PSC algorithm of Kim et al. [16] in two ways to construct the enhanced ribosomal cluster  $C_{ribo}$  with greater statistical power and precision. In the original PSC algorithm, the

strength of phase similarity (or "synchronization") between the time courses of two genes  $g$  and  $g'$  was measured by the mean phase coherence  $\rho_{gg'}$  (see [16] for details), which assumed values between 0 for no phase synchronization, and 1 for perfect synchronization. Then the expression  $\rho_{gg'}$  was extended to compute phase synchronization  $\rho_{gC}$  between a gene  $g$  and a cluster  $C$  by substituting the phase of  $g'$  with the weighted mean phase of  $C$ , denoted by  $\Phi_C$ , such that the weights measured how closely each gene  $g$  in cluster  $C$  followed the common phase  $\Phi_C$ . For details, see [16] and [17].

In the present study, we incorporated the idea of *consensus* about a gene's phases in multiple experiments by computing the mean phase coherence [16]  $\rho_{gC}$  for a given gene-cluster pair over five or more experiments. Another extension to the PSC algorithm was the idea of *seeding* the mean phase  $\Phi_{C_{ribo}}$  of the target cluster  $C_{ribo}$  with the phases of the original ribosomal biogenesis genes with the aim of identifying additional and possibly weakly oscillating genes which synchronized with the original ribosomal cluster in five or more experiments. The above-mentioned 551 genes (based on the clusters in Step 1), with regulatory signatures similar to the original 43 ribosomal genes, were further clustered with the extended PSC algorithm (with the cutoff parameter set to 0.6; see [16] for details on this parameter). Eight outlier genes with high phase variance were removed, resulting in a cluster of 103 genes (called  $C_{ribo}$ ).

### Motif analysis and orthology analysis

The enhanced cluster  $C_{ribo}$  allowed us to investigate the regulatory mechanism underlying it. No motif was reported for the ribosomal cluster in [9]. The powerful motif search tool BioProspector [18] was applied to the full upstream intergenic regions of the 103 genes in  $C_{ribo}$  as foreground DNA sequences, and all intergenic regions in *S. pombe* were used as the background DNA sequences. To identify phylogenetically conserved binding sites, the upstream intergenic sequences of the  $C_{ribo}$  orthologs in another fission yeast species *Schizosaccharomyces japonicus* were obtained from the *Schizosaccharomyces* group Database at the Broad Institute of MIT and Harvard University. The motif visualization program MotifViz [19] was used to detect the transcription factor binding sites with motif match p-value cutoff 0.01 and with all intergenic sequences of *S. pombe* and *S. japonicus* as background sequences for searching with the respective species. The TRANSFAC database [20] was searched for similar transcription factor binding sites in other species. The orthologs were identified by reciprocal best hits among species with the program Inparanoid [21]. For *S. pombe* orthologs of cell cycle oscillating genes in other species (human, *S. cerevisiae* and *Arabidopsis*) we integrated data from multiple sources (the databases YOGY [22] and

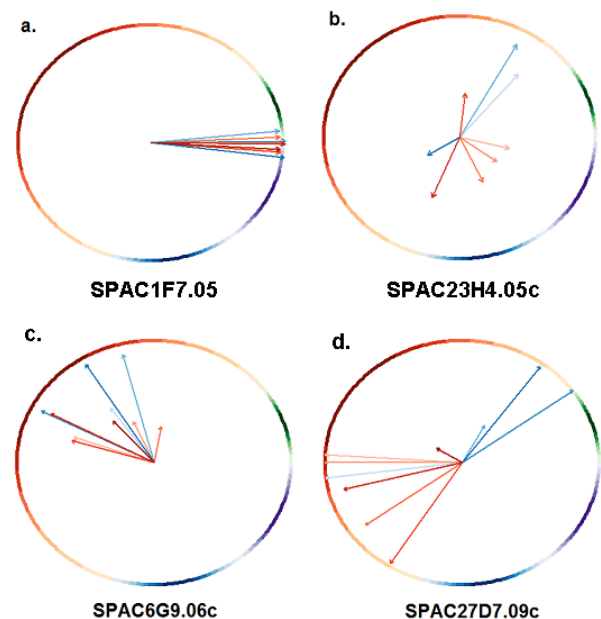
Cyclebase.org [23]; also unpublished orthology data from Chris Penkett).

## Results

For each gene and time course, we coupled cell cycle phase information with the standard measures of periodicity and expression to derive a new Phase-Coupled Meta-analysis measure (PCM score). The cell cycle oscillation of a gene  $g$  in a particular experiment  $j$  is represented by a vector  $\langle M_{gj}, \theta_{gj} \rangle$  with magnitude  $M_{gj}$  directly proportional to the gene's periodicity and expression scores, and direction  $\theta_{gj}$  given by the gene's peak phase angle within the cell cycle (Fig. 1). If  $g$  shows consistency in peak phase over many experiments, then its vector sum (which is also a vector) will be larger in magnitude than any component vector (Fig. 1). The PCM score for a gene is the magnitude of the vector sum over all the experiments (see Methods); it was computed for every gene that was present in at least five out of ten experiments. The genes were then ranked by their PCM scores, and these ranks were compared to the ranks of the same genes obtained by Margureat et al. [11]. The ranked list is available as Additional file 1. Fig. 2 shows results for four example genes. These demonstrate all four possible combinations of large or small vector magnitude, and high or low phase consistency. Fig. 3 shows the correlation between PCM ranks and Marguerat ranks. It is of course expected that they are well correlated, since the Marguerat Periodicity and Expression scores are used to help calculate the PCM score. Genes with relatively low magnitude vectors, but high phase consistency (Fig. 2c) have much better PCM ranks than Marguerat ranks (shown by dark points high above the diagonal line in Fig. 3). It is particularly genes of this kind that are found as cell cycle oscillating genes by our analysis, but missed by methods that focus only on the strength of the oscillation in expression. Genes with relatively large magnitude vectors, but poor phase consistency (Fig. 2d), have much worse PCM ranks than Marguerat ranks, and are found in the lower right quadrant of Fig. 3. These genes are not considered cell cycle oscillating by PCM analysis, but might be so considered by other methods. Fig. 4 shows the time course data for four example genes. The examples chosen had PCM ranks of 1 (i.e., the most strongly regulated gene, *cdc22*, also see Fig. 2a), 100, 500, and 1000. By rank 1000, oscillations (at least when presented in this format) are barely visible to the eye, but nevertheless are statistically detectable.

### Number of Genes showing Cell Cycle Oscillation

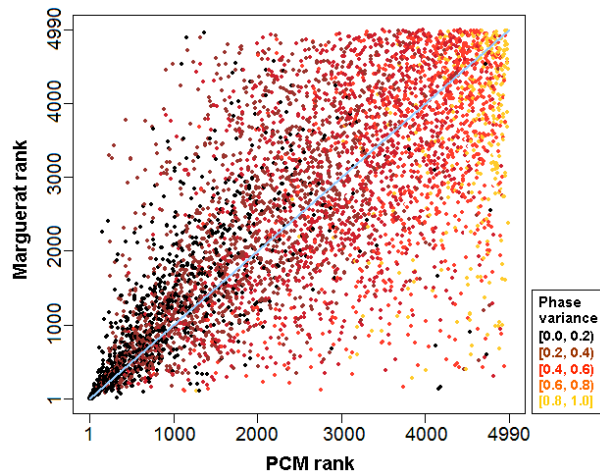
In principle, one could imagine that there might be genes regulated by the cell cycle, which might have strong or weak oscillations, and in addition there might be other genes that do not oscillate. Thus there might be a biphasic distribution of oscillation. However, previous genome-wide cell cycle transcriptional studies have not supported



**Figure 2**

**Four examples genes.** A variety of cases are observed: in (a), SPAC1F7.05 (*cdc22*, ribonucleotide reductase) has both high Periodicity and Expression scores, and also high phase consistency. It is the highest-ranked gene by PCM (rank 1). In (b), SPAC23H4.05c has low Periodicity and Expression scores, and also low phase consistency, leading to a very low PCM rank (4000). In (c), SPAC6G9.06c (*pcp1*) has mediocre Periodicity and Expression scores, but relatively high phase consistency (PCM rank of 259), and finally (d), SPAC27D7.09c (encoding a butI family protein) has mostly high Periodicity and Expression scores, but relatively low phase consistency (PCM rank of 1464). In some cases, vectors were intentionally offset by small amounts to avoid overlapping. The colored circumference refers to the circular distribution of cell cycle phases in *S. pombe* (red-G2, green-M, purple-G1, blue-S), while the vectors in colder (bluish) and warmer (reddish) hues represent *cdc25* and elutriation experiments respectively.

this [4,8,9]. Instead, these studies show a small number of genes with strong oscillation, a larger number of genes with slightly weaker oscillation, an even larger number of genes with still weaker oscillation, and so on, until the statistical evidence for cell cycle oscillation reaches the level of noise in the underlying experiments (e.g., Fig. 2 of ref [9]). That is, it appears there is just one population of genes, and it varies in a continuous manner from strongly oscillating to very weakly oscillating or not oscillating at all. Various authors have drawn a more-or-less arbitrary cut-off at some point on the distribution to define a set of "cell cycle regulated" genes. In *S. pombe*, this arbitrary number of genes thought to be cell cycle regulated was



**Figure 3**  
**Specificity and sensitivity of PCM ranks.** Each gene is plotted as a 2-dimensional point where the x-coordinate is its PCM rank and the y-coordinate is its Marguerat et al. rank. Genes with higher phase consistency are marked with darker points. Clearly there are many more darker points above the diagonal ( $y = x$ ) line, suggesting that genes with consistent phases across experiments get higher ranks (i.e. ranked closer to the top) by PCM than Marguerat et al. The sparse upper left quadrant of the plot shows that if genes did not receive a high rank by Marguerat et al., due to their poor Periodicity and Expression scores, then they did not get high a PCM rank either. However, many genes in the lower right quadrant received a high score from Marguerat et al. on the basis of good Periodicity and expression scores, but a low PCM score on the basis of poor phase consistency.

407 (Rustici), 747 (Peng), 750 (Oliva) or 500 (Marguerat).

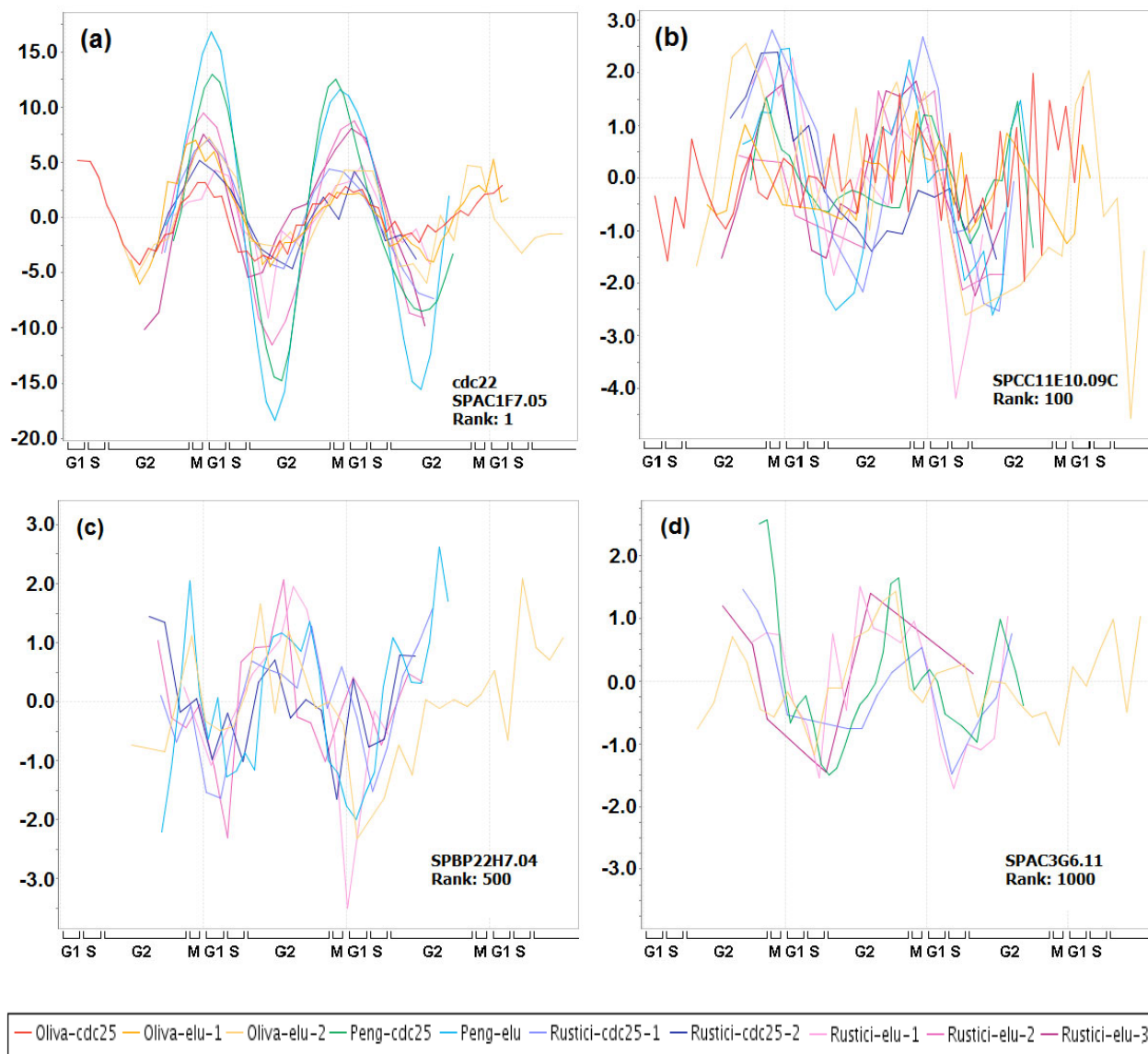
Here, we are using all available experimental data, and using more information from these data than previously considered, so it is expected that there will be statistical evidence for a larger number of cell cycle oscillating genes. Fig. 5, which uses only phase information, suggests there could be between 1400 and 3200 cell cycle oscillating genes. We applied a test of circular uniformity across experiments, using only peak phase information, and this test suggested around 1,900 cell cycle oscillating genes (Fig. 6). Finally, to most efficiently use all available information, we did permutation testing using PCM, as follows: for each gene, every time course experiment was shuffled with respect to the time axis and with the median expression level of a random gene (i.e.  $qE_{gi} = 0.5$ , see Methods). A "PCM" score was calculated based on the shuffled measurements in all experiments. This was repeated 1,000 times per gene to obtain the p-value based on its PCM scores from shuffled (i.e., randomized) data.

After FDR adjustment of the p-values, the randomization procedure suggests that there are about 2,554 cell cycle oscillating genes. In other words, we estimate about 40-50% of all *S. pombe* genes to be cell cycle oscillating to some degree, however slight. Many of these new, weakly-oscillating genes peak in G2 phase, but many peak at other times, especially at M/G1, the same time as most strongly-oscillating genes (Additional file 2; Fig. S1).

For visual confirmation of the idea that there could be as many as 2,000 weakly-oscillating genes, we took the top 2,000 genes in the PCM-ranked list, and removed all the genes that were designated cell cycle oscillating by either Oliva et al. or Marguerat et al., leaving about 1,275 genes (i.e., 1,275 genes hitherto considered "non-oscillating"). (Not all Oliva or Marguerat genes were in the top PCM 2000 genes, so the number of "non-oscillating" genes is more than 2000 minus the 750 genes of Oliva et al.) The expression patterns of these 1,275 genes are shown in Fig. 7. The genes are stacked on top of each other by time of peak expression. Fig. 7 clearly shows that to the eye, many or most of these genes appear to be cell cycle oscillating (high resolution version in Additional File 3). The statistical evidence for cell cycle oscillation of roughly 2,000 genes is strengthened by this visual evidence, even though the oscillation of the worst 1,000 of these genes is extremely weak. The significance, of any, of these very weak oscillations is addressed in the Discussion.

#### **An enhanced cluster of ribosomal biogenesis genes**

Given the increased statistical power to identify weakly-oscillating cell cycle genes, we performed clustering with the top 2,000 genes from the PCM-based rank list. Since this pool of genes includes many weakly-oscillating genes, we used a novel two-step refinement strategy to allow stepwise filtering to produce biologically meaningful clusters. In each step an algorithm that is robust and well suited for that step was applied (see Methods). We obtained an enhanced ribosomal biogenesis cluster  $C_{ribo}$  with 103 genes specific to early to mid-G2 phase (Fig. 8, high res version Additional file 4), which was a 137% increase with respect to the original ([9]) size of this cluster. Of the newly included genes, 52% had PCM ranks between 750 and 2000 (i.e. weakly oscillating) and many were annotated in GO as ribosomal proteins (e.g. *dbp10*, *SPBC3F6.04C*, *SPBC365.04C*, *SPAC823.04*, *SPBC1604.06C*, *fcf2*, *moe1*). Motif analysis of the intergenic regions upstream of these genes revealed a novel and statistically significant motif BYTCGTTA (where B is C or G or T, and Y is C or T) with p-value  $3.7 \times 10^{-18}$ . The motif was found in 56% (58/103) of all the genes in  $C_{ribo}$  as well as in 52% (47/91) of the known orthologs of  $C_{ribo}$  genes in *S. japonicus*. This motif could be the binding site of a DNA binding protein that helps regulate the genes of this cluster.



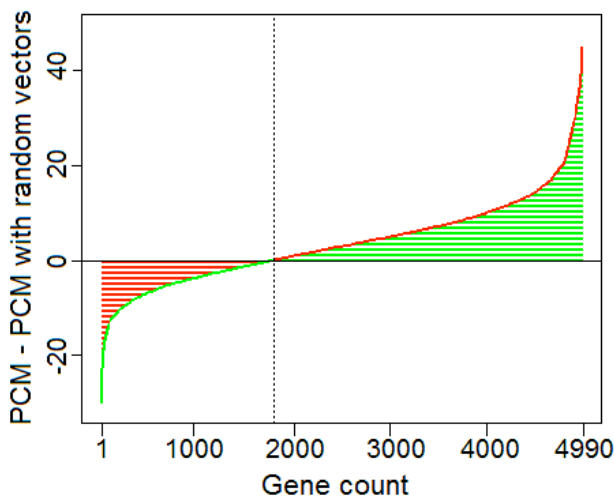
**Figure 4**  
**Variation in peak phase consistency captured by PCM ranking.** Time courses of genes with different PCM ranks are shown: (a) rank 1; (b) rank 100; (c) rank 500; and (d) rank 1000. High peak phase consistency among the ten independent time courses can be seen for the high PCM ranked genes in plots (a) and (b), while in the lower PCM ranked genes peak phase consistency is less, which can be observed among fewer time courses ((c) and (d)). (The plots in this figure were created with the help of Cyclebase.org due to Gauthier et al.)

**Discussion**

Most microarray cell cycle studies in the past decade have modeled the periodic time courses with Fourier summation of sinusoids [4-6,8,9,11]. However, in general such a Fourier score is independent of the time course's phase angle, and most microarray cell cycle analyses have not explored peak phase information in a systematic manner. While some approaches in the past have modeled a gene's

phase for detection of periodic oscillations [24,25] or used it to identify the bottleneck genes [26], others have explicitly ignored it [27]. Although Marguerat et al. noted the phase variation [11], no study that we are aware of has used it for combining multiple experiments [6,9,11,28]. The main contribution of our Phase-Coupled Meta-analysis is the use of peak phase information within a novel measure - the PCM score - to combine ten genome-wide



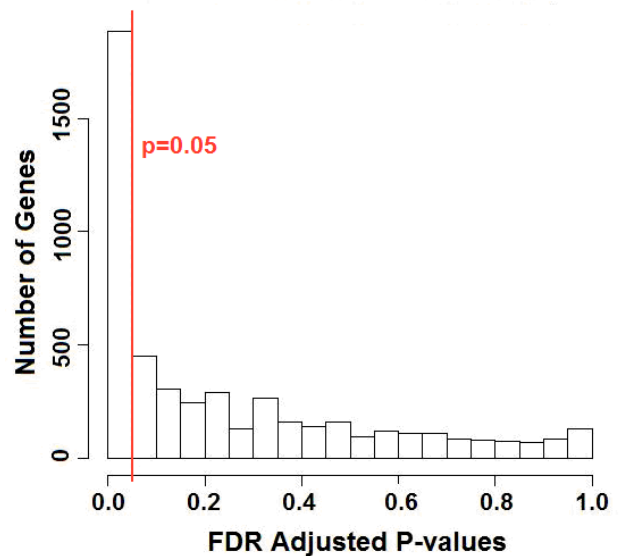


**Figure 5**  
**Difference of PCM scores with real and random phases.** For each gene, the PCM score was computed as sum of vectors (see Methods). Then for the same gene, a random-phase PCM score was computed as sum of vectors with the original magnitudes but randomly-chosen phase angles. For every gene, the difference between its original PCM score and its random-phase PCM score was then computed, and plotted in increasing order from left to right. To the right of the dotted line are 3,200 genes where the difference is positive; i.e., for these genes, the cell cycle phases of peak expression in the time courses are less variable than the randomly distributed phases.

studies for identification of cell cycle oscillating genes in *S. pombe* with substantial increase in power and prediction accuracy.

The premise of our study is simple: consistency in the cell cycle phase of a gene's peak expression across multiple independent experiments is an indicator of genuine cell cycle oscillation. Using information about the consistency of the phase of peak expression, and adding it to information about the strength of oscillation, will give a better result than the strength of oscillation alone. By design, the PCM score shows more specificity and sensitivity for identifying genes with high phase consistency than, for instance, the previous meta-analysis by Marguerat et al. which combined p-values for periodicity and expression (Fig. 3); the relatively lower phase variance of genes that received higher ranks by PCM than by Marguerat et al. is statistically significant (p-value  $< 2.2 \times 10^{-16}$  for two-sample *t* test). Thus our PCM meta-analysis helps to discover genes that are weakly but consistently oscillating.

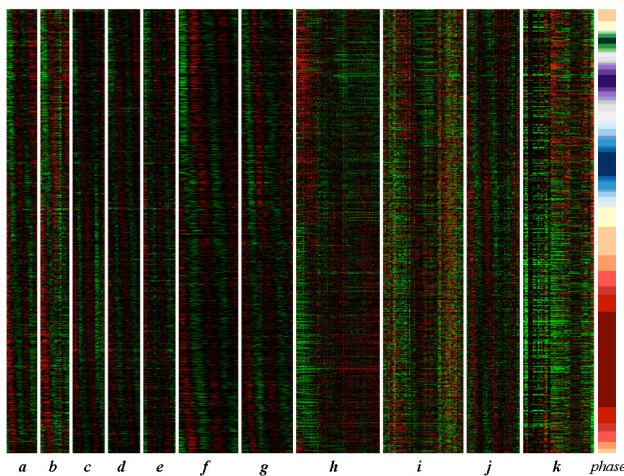
We observed low circular variance ( $< 0.4$ ) across 10 experiments for more than 50% of all genes in *S. pombe* (Addi-



**Figure 6**  
**Statistical significance of cross-experiment phase consistency.** Approximately 1,900 genes can be rejected at level 0.05 for the null hypothesis that their peak phases across experiments are distributed uniformly over a circular range.

tional file 2; Fig. S2a). In fact, the count increased to more than 2,900 genes when the experiment in which a gene deviated most from its median phase was excluded from variance computation (Additional file 2; Fig. S2b). Under the assumption that a gene which is not cell cycle oscillating should peak, with whatever amplitude, at random phases in different independent cell cycle experiments, we tested every gene for phase uniformity over a circular range [29], and for 1,884 genes, as shown in Fig. 6, the uniform phase distribution hypothesis was rejected at significance level 0.05 (of which 1,189 genes were rejected at level 0.01). Indeed a closer look at the peak phases of genes that were not identified as periodic by earlier studies show significant phase consistency over several experiments (Fig. 4). These observations motivated us to design the PCM score for systematic phase-coupled meta-analysis of multiple time course experiments, and to re-rank all genes in *S. pombe* based on it.

The effectiveness of PCM scoring can be illustrated using the contrasting examples of SPAC6G9.06C and SPAC27D7.09C. The former (a.k.a. *pcp1*) is a weakly oscillating but known cell cycle gene; Fig. 2c shows PCM vectors of moderate magnitude representing its weak oscillation in ten experiments. Marguerat et al. did not rank the gene among the top 1,000 nor included it as cell cycle oscillating. However, we note that the same vectors show high peak phase consistency with similarity of direc-



**Figure 7**  
**Periodic oscillation of 1,275 "non-periodic" genes.**

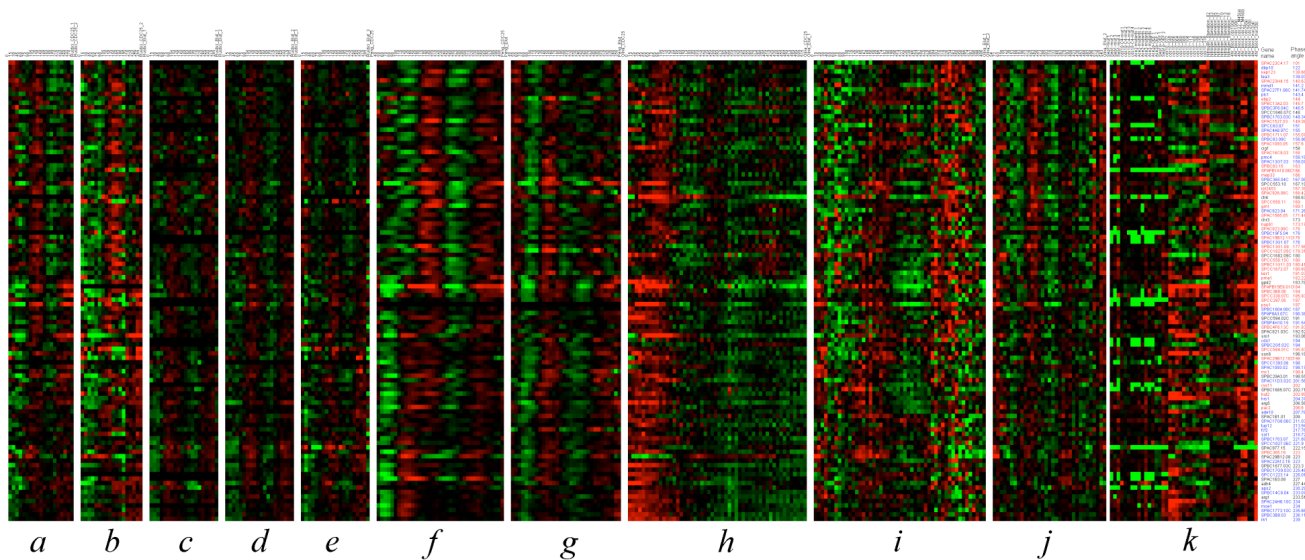
After removal of genes identified as periodic in the Oliva et al. and Marguerat et al. studies from the top 2,000 PCM ranked genes, there are 1,275 genes remaining. In this figure, these 1,275 genes are stacked top to bottom in phase order. Cyclic behavior is apparent. The time courses *a-e* are from Rustici et al., *f-g* from Peng et al., and *h-j* from Oliva et al. Block *k* consists of samples from transcription factor knock-out and overexpression experiments (i.e., this block does not contain a cell cycle experiment). The color band "Phase" marks the phase distribution (red-G2, green-M, purple-G1, blue-S). A high resolution version is available in additional file 3.

tion. As a result, the gene obtained PCM rank of 259. Indeed, the GO functional annotation for the gene is as follows: chromosome segregation; microtubule nucleation; mitotic cell cycle spindle assembly checkpoint; mitotic metaphase/anaphase transition. Moreover, its *S. cerevisiae* ortholog YDR356W is known to be phosphorylated by Mps1p in cell cycle-dependent manner. On the other hand, SPAC27D7.09C is in general strongly expressed, as seen in the form of large magnitude PCM vectors in Fig. 2d. However it is a known heat shock stress responder which has mediocre cell cycle peak phase consistency across ten experiments as can be observed by the inconsistent direction of the vectors. Marguerat et al. ranked it among their top 300 genes and included it as cell cycle oscillating while PCM penalized its phase inconsistency and assigned it a much lower rank of 1,464.

Many of the new cell cycle oscillating genes peak in early to mid G2 phase, and trough in mitosis. A cluster of 44 similar genes was described by Oliva et al. [9] as the "kap123" cluster, which contains many genes involved in ribosome biogenesis (note that this cluster does not contain many actual components of the ribosome; it contains

only a few actual ribosomal protein genes. The vast majority of the ribosomal protein genes form a separate cluster that we do not discuss here.) Using the new cell cycle genes, and a two-step clustering algorithm, we enhanced the "kap123" cluster to form the  $C_{ribo}$  cluster of 103 genes, many of which pertain to ribosome biogenesis. The presence of a larger number of genes allowed us to identify a statistically significant motif  $\{TCG\}\{TC\}TCGTTA$  in more than half of all the genes in  $C_{ribo}$ . The motif was found to be conserved in the regulatory sequences of a similar percentage of orthologs in fission yeast *S. japonicus*. Searching the TRANSFAC binding site database yields close matches with motifs for several Hox proteins: chicken HOXA4 (entry: TTCCTCGTTATCT) and human HOX11 (TGACCGGTCGTAA). Interestingly, HOX11 is a homeodomain transcription factor which is known to be cell-cycle regulated [30]. Further, it interacts directly with protein phosphatases that normally regulate cell cycle check point in G2-phase [31]. We note that in higher eukaryotes, ribosomal RNA transcription (which is dependent on RNA polymerase I) is inhibited during mitosis; this might also be true in *S. pombe*, and if so, the matching trough in the ribosomal biogenesis genes we see in  $C_{ribo}$  might be a response to a lack of ribosomal RNA.

Our work here carries on from the analysis of Marguerat et al. [11], in part because we wished to show the power of our new method by comparing our results to previous results for the same data. Thus, we have used the same data as Marguerat et al., and we have also used their "regulation" and "periodicity" scores to calculate the magnitude of our vectors. However, Marguerat et al. concluded that the data supported identification of about 500 cell cycle oscillating genes, while we believe it supports identification of about 2000 such genes. There are at least two reasons for this difference. First, because we are using consistency of peak phase in addition to regulation scores and periodicity scores, we are using more of the information inherent in the data, not just within-experiment but also across experiments, and this gives us greater power. Second, we make different assumptions than Marguerat et al. with regard to permutation testing. We use the null hypothesis that all variations in the experimental measurements in a time course are due to random noise; our p-values are relative to this null hypothesis. In contrast, Marguerat et al. believe that random noise null hypothesis may be naive; for instance, there could be effects such that experimental measurements in adjacent time points are correlated and thus not independent. If so, p-values calculated by permutation testing on the null hypothesis of completely random noise would be too small [32]. Marguerat et al. made an ad hoc adjustment for this possibility and then normalized (and raised) all their initially calculated p-value distributions around the median p-value of each distribution. We feel that this is a somewhat aggres-



**Figure 8**

**The enhanced ribosomal biogenesis cluster  $C_{ribo}$ .** The 103 genes constituting the " $C_{ribo}$ " cluster, expression of these genes peaks in G2 phase. The genes in this cluster are marked with ' $C_{ribo}$ ' in the PCM rank list (Additional file 1). The time courses a-e are from Rustici et al., f-g from Peng et al., and h-j from Oliva et al. Block k consists of samples from transcription factor knockout and overexpression experiments. A high resolution version is available in additional file 4.

sive adjustment that increases p-values significantly, thus significantly decreasing the apparent number of cell-cycle oscillating genes. While we completely agree with Marguerat et al., and Futschik and Herzel [32], that the null hypothesis of completely random noise may be too simple, which if true would lead to p-values that are too small, on the other hand, there is no objective, quantifiable alternative, and we do not wish to make an ad-hoc adjustment. Thus, at least for the time being, we feel that the null hypothesis of random noise is the best we can do. In any case, regardless of statistical arguments, it is visually and intuitively clear from Fig. 7 that there are many more than 500 cell cycle oscillating genes.

Marguerat et al. also argue for a relatively small number of cell cycle oscillating genes using benchmark sets of known cell cycle oscillating genes. They choose three sets of benchmark cell cycle genes, consisting of 40 very strongly oscillating genes whose periodicity has been demonstrated in small scale experiments (set B1); genes whose promoters are bound by the known cell cycle transcription factors Cdc10, Res1, Res2 (all three of these being components of MBF) or Fkh2 (set B2); and genes shown in expression experiments to be the targets of the transcription factors Ace2, Sep1, or Cdc10 (set B3). Marguerat et al. find that their top-ranked 500 cell cycle genes are enriched for the genes in these benchmark sets, but that there is little if any enrichment after the top 500. How-

ever, we feel this argument is unconvincing. The benchmark sets contain strongly oscillating cell cycle genes, or genes controlled by a small number of powerful cell cycle transcription factors. Cell cycle oscillating genes of distinctly different kinds (e.g., genes weakly regulated by other transcription factors, or by RNA half-life, or by chromatin condensation) are not in any of the benchmark sets at all. Thus, for example, many of the genes we find ranked between 500 and 2000 are in the  $C_{ribo}$  cluster described above. These are not strongly oscillating, and are not regulated by any of the transcription factors used for benchmark sets 2 or 3, and are not significantly (if at all) represented in any of the benchmark sets. In short, since the benchmark sets consist of strongly oscillating genes controlled by a few known transcription factors, it is not surprising that these genes are not enriched amongst weakly oscillating genes controlled in other ways.

#### **The biological relevance of 2000 cell cycle oscillating genes**

By "cell cycle oscillating" gene, we mean a gene whose expression oscillates up and down, however slightly, as a function of position in the cell cycle. Traditionally, one thinks in terms of genes whose cell cycle oscillation is adaptive; that is, genes where there is a selective advantage to the organism if gene expression oscillates. Examples of such genes are *cdc22*, encoding ribonucleotide reductase, and the histone genes, encoding histone proteins. Ribonucleotide reductase is needed only for making deoxyri-

bonucleotides in preparation for immediate DNA synthesis; otherwise it is a metabolically-expensive hindrance. Similarly, histone proteins are needed only for making nucleosomes during DNA replication; otherwise they interfere with proper DNA metabolism. We refer to such genes, where the oscillation is adaptive and purposefully controlled by the cell, as "cell cycle regulated" genes.

Recently, the issue of "how many genes are cell cycle regulated" has received some attention [33,34]. Although we believe that 2000 fission yeast genes, or even more, do oscillate in expression at least slightly as a function of cell cycle, the oscillations of many of these genes may not be adaptive. Instead, many oscillations may be indirect, unavoidable consequences of the cell cycle, and may have no adaptive significance. For instance, in preparation for mitosis, chromosomes condense. It seems likely that chromosome condensation might interfere with transcription. Indeed, in higher eukaryotes, bulk transcription is greatly inhibited during mitosis. Genes whose transcription is inhibited during mitosis, perhaps as either a direct or indirect effect of chromosome condensation, would be detected in microarray experiments as cell cycle oscillating genes.

Thus, we suggest that there may be two classes of cell cycle oscillating genes. First, there are what we call "adaptive" oscillations, i.e., oscillations that are a selective advantage for the organism; *cdc22* and histones are examples of genes with adaptive oscillations. Second, there might be what we call them "incidental" oscillations, i.e., oscillations that are not adaptive, but instead are a consequence of some other cell cycle event, such as chromosome condensation. At present there is no good way of distinguishing adaptive cell cycle regulation from incidental oscillation, but it seems likely that most strong oscillations, associated with a cell cycle transcription factor, are probably adaptive, while weak oscillations, not associated with a cell cycle transcription factor, may (or may not) be incidental.

Finally, we note that while evidence has accumulated for more and more cell cycle oscillating genes, parallel results have been obtained for genes oscillating with a circadian rhythm [35-37]. Exactly as in the case of cell cycle genes, microarrays have allowed the accumulation of large amounts of data for circadian oscillations, and exactly as in the case of cell cycle, increasingly powerful statistical methods have allowed the discovery of an increasingly large number of genes with a circadian oscillation. The most recent studies suggest that nearly all mammalian genes have at least a slight circadian oscillation [37]. As in the case of the cell cycle genes, these circadian rhythms may sometimes be adaptive, and sometimes indirect and incidental.

## Conclusion

In conclusion, we have developed a new meta-analysis method that uses more of the information inherent in microarray studies of cell cycle gene expression. This extra information allows us to detect a larger number of cell cycle oscillating genes. Because the proportion of oscillating genes is very large - one quarter to one half of all genes in the organism - we suggest that many of them may be oscillating for incidental reasons, rather than because the oscillation is necessarily adaptive.

## Authors' contributions

SP designed research, performed PCM analysis and wrote the manuscript; RG performed the motif analysis; CSK performed the clustering analysis; BF designed research and wrote the manuscript. All authors read and approved the final manuscript.

## Additional material

### Additional file 1

*PCM rank list of fission yeast genes.* The spreadsheet provides the genome-wide rank of fission yeast genes due to Phase-Coupled Meta-analysis (PCM). It also shows whether a gene was reported by Oliva et al. or Marguerat et al. as periodic, its median phase over all experiments and its circular variance, and to which cluster due to Oliva et al. it belongs. The last field also indicates membership in the new  $C_{nbo}$  cluster.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-440-S1.XLS>]

### Additional file 2

*Supplementary figures.* Fig. S1 plots the distribution of cell cycle phases for genes from the PCM rank-list. Fig. S2 plots the genome-wide distribution of the circular variance of every gene's phases across experiments.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-440-S2.PDF>]

### Additional file 3

*High resolution plot.* High resolution version of Fig. 7.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-440-S3.TIFF>]

### Additional file 4

*High resolution plot.* High resolution version of Fig. 8.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-440-S4.TIFF>]

## Acknowledgements

We would like to thank Steven Skiena for extensive discussions, ideas, and suggestions during the early part of this work. We would like to thank Thomas Skot Jensen and Lars Juhl Jensen for useful discussions of statistical approaches, and Chris Penkett for sharing orthology data. This work was funded by NIH RO1 GM 064813 and 039978 to BF.

## References

1. Cho RJ, Campbell MJ, Winzeler EA, Steinmetz L, Conway A, Wodicka L, Wolfsberg TG, Gabrieli AE, Landsman D, Lockhart DJ, et al.: **A genome-wide transcriptional analysis of the mitotic cell cycle.** *Molecular cell* 1998, **2(1)**:65-73.
2. Laub MT, McAdams HH, Feldblyum T, Fraser CM, Shapiro L: **Global analysis of the genetic network controlling a bacterial cell cycle.** *Science (New York, NY)* 2000, **290(5499)**:2144-2148.
3. Menges M, Hennig L, Gruissem W, Murray JA: **Cell cycle-regulated gene expression in Arabidopsis.** *The Journal of biological chemistry* 2002, **277(44)**:41987-42002.
4. Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B: **Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization.** *Molecular biology of the cell* 1998, **9(12)**:3273-3297.
5. Whitfield ML, Sherlock G, Saldanha AJ, Murray JI, Ball CA, Alexander KE, Matese JC, Perou CM, Hurt MM, Brown PO, et al.: **Identification of genes periodically expressed in the human cell cycle and their expression in tumors.** *Molecular biology of the cell* 2002, **13(6)**:1977-2000.
6. de Lichtenberg U, Jensen LJ, Fausboll A, Jensen TS, Bork P, Brunak S: **Comparison of computational methods for the identification of cell cycle-regulated genes.** *Bioinformatics (Oxford, England)* 2005, **21(7)**:1164-1171.
7. Rustici G, Mata J, Kivinen K, Lio P, Penkett CJ, Burns G, Hayles J, Brazma A, Nurse P, Bahler J: **Periodic gene expression program of the fission yeast cell cycle.** *Nature genetics* 2004, **36(8)**:809-817.
8. Peng X, Karuturi RK, Miller LD, Lin K, Jia Y, Kondu P, Wang L, Wong LS, Liu ET, Balasubramanian MK, et al.: **Identification of cell cycle-regulated genes in fission yeast.** *Molecular biology of the cell* 2005, **16(3)**:1026-1042.
9. Oliva A, Rosebrock A, Ferrezuelo F, Pyne S, Chen H, Skiena S, Futcher B, Leatherwood J: **The cell cycle-regulated genes of *Schizosaccharomyces pombe*.** *PLoS biology* 2005, **3(7)**:e225.
10. Pyne S, Futcher B, Skiena S: **Meta-Analysis of Microarray Data.** In *Bioinformatics Algorithms: Techniques and Applications* Mandoiu IaZ: Wiley-Interscience; Hoboken, NJ; 2008.
11. Marguerat S, Jensen TS, de Lichtenberg U, Wilhelm BT, Jensen LJ, Bahler J: **The more the merrier: comparative analysis of microarray studies on cell cycle-regulated genes in fission yeast.** *Yeast (Chichester, England)* 2006, **23(4)**:261-277.
12. Benjamini Y, Hochberg Y: **Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing.** *Journal of the Royal Statistical Society Series B-Methodological* 1995, **57(1)**:289-300.
13. Fisher NI: **Statistical Analysis of Circular Data.** Cambridge University Press; New York; 1996.
14. Kaufman L, Rousseeuw PJ: **Finding Groups in Data: An Introduction to Cluster Analysis.** Wiley, John & Sons, Incorporated; Hoboken, NJ; 2006.
15. Rousseeuw PJ: **Silhouettes - a Graphical Aid to the Interpretation and Validation of Cluster-Analysis.** *Journal of Computational and Applied Mathematics* 1987, **20**:53-65.
16. Kim CS, Bae CS, Tcha HJ: **A phase synchronization clustering algorithm for identifying interesting groups of genes from cell cycle expression data.** *BMC bioinformatics* 2008, **9**:56.
17. Allefeld C, Kurths J: **An approach to multivariate phase synchronization analysis and its application to event-related potentials.** *International Journal of Bifurcation and Chaos* 2004, **14(2)**:417-426.
18. Liu X, Brutlag DL, Liu JS: **BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes.** *Proceedings of the Pacific Symposium on Biocomputing* 2001:127-138.
19. Fu Y, Frith MC, Haverly PM, Weng Z: **MotifViz: an analysis and visualization tool for motif discovery.** *Nucleic acids research* 2004:W420-423.
20. Wingender E, Chen X, Fricke E, Geffers R, Hehl R, Liebich I, Krull M, Matys V, Michael H, Ohnhauser R, et al.: **The TRANSFAC system on gene expression regulation.** *Nucleic acids research* 2001, **29(1)**:281-283.
21. O'Brien KP, Remm M, Sonnhammer EL: **Inparanoid: a comprehensive database of eukaryotic orthologs.** *Nucleic acids research* 2005:D476-480.
22. Penkett CJ, Morris JA, Wood V, Bahler J: **YOGY: a web-based, integrated database to retrieve protein orthologs and associated Gene Ontology terms.** *Nucleic acids research* 2006:W330-334.
23. Gauthier NP, Larsen ME, Wernersson R, de Lichtenberg U, Jensen LJ, Brunak S, Jensen TS: **Cyclebase.org--a comprehensive multi-organism online database of cell-cycle experiments.** *Nucleic acids research* 2008:D854-859.
24. Lu X, Zhang W, Qin ZS, Kwast KE, Liu JS: **Statistical resynchronization and Bayesian detection of periodically expressed genes.** *Nucleic acids research* 2004, **32(2)**:447-455.
25. Liu D, Umbach DM, Peddada SD, Li L, Crockett PW, Weinberg CR: **A random-periods model for expression of cell-cycle genes.** *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101(19)**:7240-7245.
26. Caretta-Cartozo C, De Los Rios P, Piazza F, Lio P: **Bottleneck genes and community structure in the cell cycle network of *S. pombe*.** *PLoS computational biology* 2007, **3(6)**:e103.
27. Langmead CJ, Yan AK, McClung CR, Donald BR: **Phase-independent rhythmic analysis of genome-wide expression patterns.** *J Comput Biol* 2003, **10(3-4)**:521-536.
28. Tsjporkova E, Boeva V: **Fusing time series expression data through hybrid aggregation and hierarchical merge.** *Bioinformatics (Oxford, England)* 2008, **24(16)**:i63-69.
29. Rao JS, Sengupta A: **Topics in Circular Statistics.** World Scientific Publishing Company; Singapore; 2001.
30. Zhang N, Gong ZZ, Minden M, Lu M: **The HOX-11 (TCL-3) homeobox proto-oncogene encodes a nuclear protein that undergoes cell cycle-dependent regulation.** *Oncogene* 1993, **8(12)**:3265-3270.
31. Kawabe T, Muslin AJ, Korsmeyer SJ: **HOX11 interacts with protein phosphatases PP2A and PPI and disrupts a G2/M cell-cycle checkpoint.** *Nature* 1997, **385(6615)**:454-458.
32. Futschik ME, Herzel H: **Are we overestimating the number of cell-cycling genes? The impact of background models on time-series analysis.** *Bioinformatics (Oxford, England)* 2008, **24(8)**:1063-1069.
33. Jensen LJ, de Lichtenberg U, Jensen TS, Brunak S, Bork P: **Circular reasoning rather than cyclic expression.** *Genome biology* 2008, **9(6)**:403.
34. Lu Y, Mahony S, Benos PV, Rosenfeld R, Simon I, Breeden LL, Bar-Joseph Z: **Combined analysis reveals a core set of cycling genes.** *Genome biology* 2007, **8(7)**:R146.
35. Ptitsyn AA, Zvonik S, Conrad SA, Scott LK, Mynatt RL, Gimble JM: **Circadian clocks are resounding in peripheral tissues.** *PLoS computational biology* 2006, **2(3)**:e16.
36. Ptitsyn AA, Zvonik S, Gimble JM: **Permutation test for periodicity in short time series data.** *BMC bioinformatics* 2006, **7(Suppl 2)**:S10.
37. Ptitsyn AA, Zvonik S, Gimble JM: **Digital signal processing reveals circadian baseline oscillation in majority of mammalian genes.** *PLoS computational biology* 2007, **3(6)**:e120.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
http://www.biomedcentral.com/info/publishing\_adv.asp

