doi:10.1038/nature05186

LETTERS

Co-evolution of transcriptional and post-translational cell-cycle regulation

Lars Juhl Jensen[1]*, Thomas Skøt Jensen[2]*, Ulrik de Lichtenberg[2]*, Søren Brunak[2] & Peer Bork[1,3]

DNA microarray studies have shown that hundreds of genes are transcribed periodically during the mitotic cell cycle of humans[1], budding yeast[2,3], fission yeast[4–6] and the plant *Arabidopsis thaliana*[7]. Here we show that despite the fact the protein complexes involved in this process are largely the same among all eukaryotes, their regulation has evolved considerably. Our comparative analysis of several large-scale data sets reveals that although the regulated subunits of each protein complex are expressed just before its time of action, the identity of the periodically expressed proteins differs significantly between organisms. Moreover, we show that these changes in transcriptional regulation have co-evolved with post-translational control independently in several lineages; loss or gain of cell-cycle-regulated transcription of specific genes is often mirrored by changes in phosphorylation of the proteins that they encode. Our results indicate that many different solutions have evolved for assembling the same molecular machines at the right time during the cell cycle, involving both transcriptional and post-translational layers that jointly control the dynamics of biological systems.

To obtain comparable sets of transcriptionally regulated genes from distantly related eukaryotes, we reanalysed the existing cell-cycle gene expression data[1–7] and found 600 periodically expressed genes in *Homo sapiens*, 600 in *Saccharomyces cerevisiae*[8,9], 500 in *Schizosaccharomyces pombe*[10] and 400 in *Arabidopsis thaliana*. These gene lists are more conservative than those originally proposed, but achieve better sensitivity (estimated to be 80–90% for human and the two yeasts, but only 50% in the plant; see Supplementary Information). We further assigned genes with a common descent to orthologous groups, of which we estimate at least 80% to contain the correct set of functionally equivalent genes (see Supplementary Information). The complete list of orthologous groups with periodic members and their peak times is available at http://www.cbs.dtu.dk/cellcycle.

Of these groups, 381 contain orthologues from all four organisms and have at least one periodic member. This conserved set includes 550 of the total set of 2,100 periodically expressed genes, implying that most of the periodically expressed genes do not have orthologues in all four organisms. Even among this subset we found that periodicity is poorly conserved across the four organisms (Fig. 1), meaning that although the protein sequences are conserved through evolution, their transcriptional regulation during the cell cycle is not. This is compatible with the results of earlier comparisons of periodically expressed genes from two or three organisms[4,5,10–13]. Although the true overlap might be slightly better than is shown in Fig. 1, the large differences cannot be explained by the quality of the gene expression data or the orthology analysis (see Supplementary Information).

Only five orthologous groups are periodic in all four organism, namely a group of mitotic cyclins, three groups of histone proteins, and subunits of the ribonucleotide-diphosphate reductase (RNR) complex. The three additional orthologous groups that are periodic in all organisms except *A. thaliana* consist of the CDC20 recognition subunit of the anaphase-promoting complex/cyclosome (APC/C), the pre-replication complex component CDC6, and the sister chromatid cohesion protein MCD1. The genes for which periodicity is highly conserved thus encode either key regulators of the cell cycle or components that are needed for synthesizing the building blocks of new DNA and chromatin.

For the genes for which periodic expression *per se* is conserved, we compared whether their expression peaks during the same phase of the cell cycle in each organism (Fig. 1). Despite normalizations to correct for organism-specific differences in the relative lengths of the cell-cycle phases (see Supplementary Information), we found that the timing of expression is only conserved for 42% of the orthologous groups (allowing a relaxed 20% tolerance, Fig. 1). The widespread differences in both identity and timing of the transcriptionally

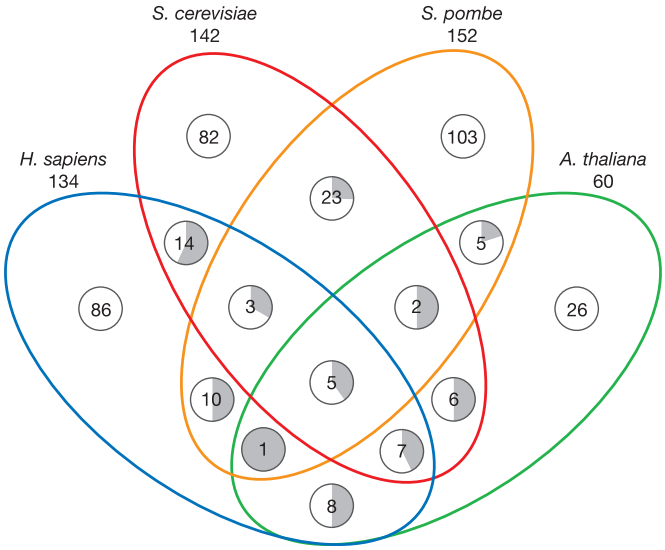

Figure 1 | Periodic expression is poorly conserved. The four-way Venn diagram summarizes the conservation of periodic expression for the 381 orthologous groups that contain at least one protein from each of the four organisms, and where one or more members are periodically expressed. The total number of such orthologous groups for each organism is given below its name. The pie charts show the fraction of orthologous groups for which the peak time is conserved (grey) or has changed by more than one-fifth of a cell cycle (white). Together, these analyses reveal that periodic expression is poorly conserved at the level of individual genes: conserved periodic expression across the four organisms is observed in only five cases and for only two of these is the timing conserved as well, namely histones H2A and H4 (see also Supplementary Figs 6 and 7).

[1]European Molecular Biology Laboratory, D-69117 Heidelberg, Germany. [2]Center for Biological Sequence Analysis, Technical University of Denmark, DK-2800 Lyngby, Denmark. [3]Max-Delbrück-Centre for Molecular Medicine, D-13092 Berlin, Germany.
*These authors contributed equally to this work.

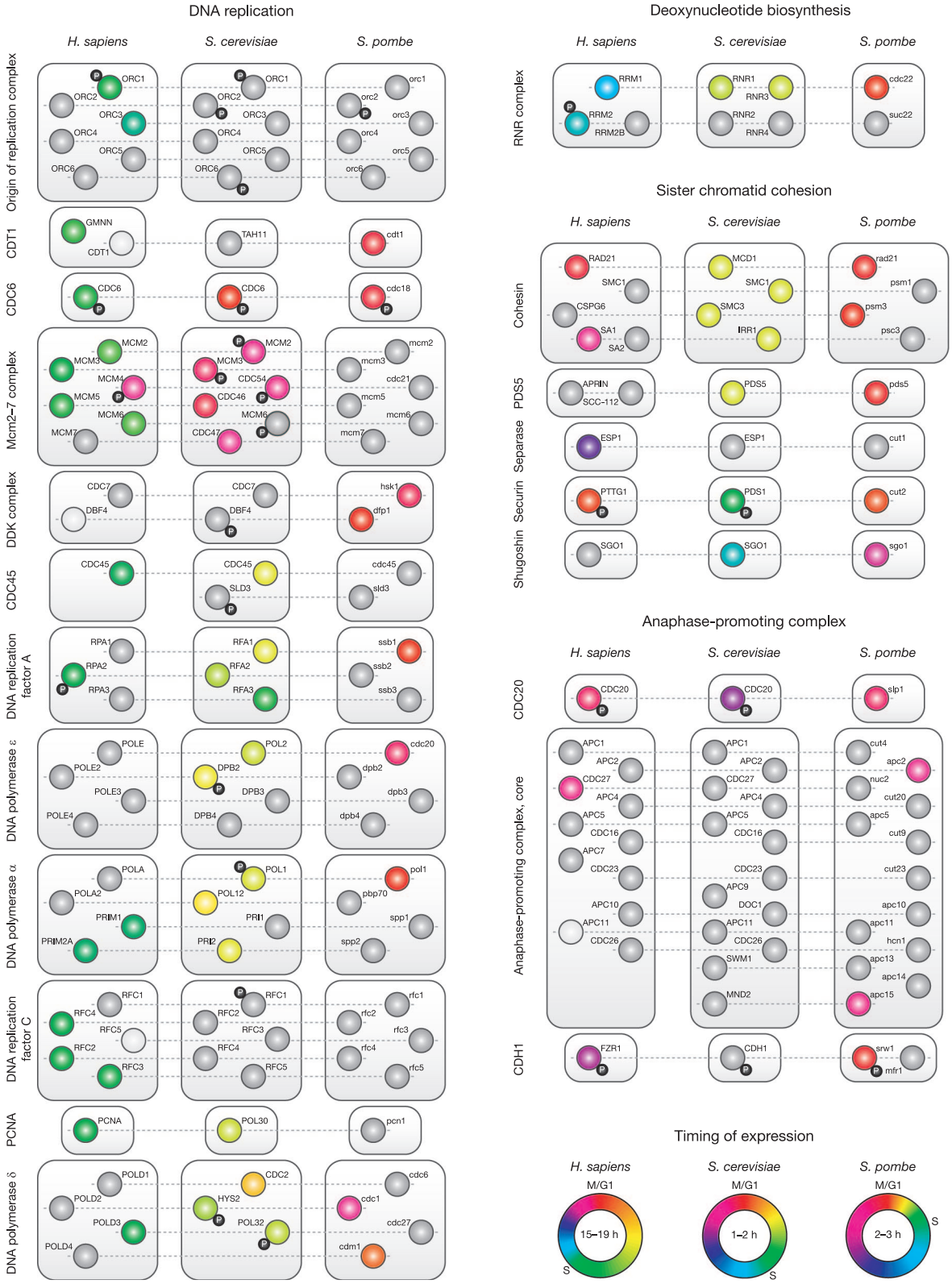

Figure 2 | Protein complexes are regulated through different subunits in each organism. The best-studied protein complexes involved in core cell-cycle processes were curated from the literature and public databases (see Supplementary Information). The subunits (circles) of each protein complex (boxes) are shown for three organisms (columns). Proteins on the same horizontal line are sequence orthologues and are believed to perform the same function. Dynamic subunits are coloured according to their time of peak expression, static subunits are shown in grey, and subunits for which no expression data were available are displayed in white. As the relative length of the phases varies between organisms, time warping was used to ensure that the same colour corresponds to the same stage of the cell cycle (see colour legends). The small black circles denote proteins that experiments indicate are phosphorylated by CDKs. Detailed discussions on the composition and regulation of these and other complexes can be found in Supplementary Information, along with details on the time warping.

regulated genes are at first glance surprising, considering the high degree of conservation of the core cell-cycle machinery.

However, protein complexes that are involved in the cell cycle often comprise a mixture of static (constitutively expressed) and dynamic (periodically expressed) subunits, the latter being expressed just before the stage of the cell cycle in which the complex is active ('just-in-time assembly' of complexes)[9]. To confirm the consistency of the periodic expression of complex subunits within each organism, and to test a previously stated hypothesis that the just-in-time assembly principle allows the identity of the regulated subunits to change during evolution[9], we carefully annotated the composition of some of the best-understood cell-cycle complexes in human and the two yeasts (Fig. 2; *A. thaliana* was omitted owing to the poor sensitivity of the expression data and the lack of knowledge on complex composition).

When mapping periodically expressed genes onto equivalent protein complexes, it becomes obvious that the dynamic subunits of individual complexes are consistently co-expressed, and that the timing of expression is similar for complexes that are involved in the same cellular process within each organism but differs between organisms (Fig. 2). However, our analysis also shows that the identity of the dynamic subunits within each complex has changed during evolution. The differences in temporal expression of individual, orthologous genes can thus be understood when viewed at the level of protein complexes and cellular processes (Fig. 2).

Furthermore, the differences in temporal expression are in remarkable agreement with current knowledge about the order and timing of assembly of the individual complexes that are involved in DNA replication[14,15] and mitosis[16], and even provides new insight into the regulation of these complexes. For example, both the sister chromatid cohesion complex and securin, which prevents separase from cleaving cohesin, are expressed during mitosis in human and fission yeast but at the G1/S transition in budding yeast. The same evolutionary changes in timing are observed for Pds5 and shugoshin, supporting the recent suggestions that Pds5 is a fifth subunit of cohesin and that shugoshin might help to protect mitotic cohesin from degradation (see Supplementary Information)[16]. The regulation of the DNA replication machinery has also evolved considerably. In fission yeast, it is expressed as early as M phase (Fig. 2) owing to the very short G1 phase[17,18]. Also, only two subunits of the pre-replication complex are subject to transcriptional regulation, namely cdt1 and cdc18, which when overexpressed can induce re-replication[18,19]. In humans, the pre-replication complex is regulated through many more subunits. Surprisingly, MCM4 is expressed in M phase, whereas the rest of the dynamic subunits appear in G1/S, which has to our knowledge never been described before. The distinct transcriptional regulation of MCM4 is intriguing, considering that phosphorylation of this subunit regulates the chromatin association of the entire Mcm2–7 complex[20,21]. Many more details on individual complexes are presented in Supplementary Information.

The agreement between the timing of transcription and the timing of action of the protein products is not necessarily to be expected, as phosphorylation is at least as important as transcription for regulating cell-cycle complexes. In this regard, we have noted that periodically expressed *S. cerevisiae* proteins tend also to be phosphorylated[9]. To generalize this observation, we annotated the complexes in Fig. 2 with information on phosphorylation by cyclin-dependent kinases, and found that the dynamic subunits are three times as likely to be targeted by phosphorylation as the static ones ($P < 10^{-3}$).

To assess the evolutionary significance of this correlation, we applied two proteome-wide statistical tests to several independent sets of known and predicted phosphoproteins from each organism (Fig. 3). We first compared the phosphorylation of dynamic and static proteins from each organism (Fig. 3a). This comparison revealed highly significant over-representation of dynamic proteins among human cyclin-dependent kinase (CDK) substrates identified by small-scale experiments[22], among human phosphoproteins identified by mass-spectrometry studies, and among *S. cerevisiae* Cdc28 substrates identified in at least one of two systematic screens[23,24]. Sequence-based predictions[25] provide yet another complementary, unbiased source of CDK phosphorylation sites, which supports the experimental results and shows that the correlation holds true also in fission yeast (Fig. 3a).

The fact that CDK substrates are enriched among the dynamic proteins in three organisms could be due to a core set of dynamic orthologues that are phosphorylated by CDKs in all organisms or, alternatively, it could reflect that loss or gain of transcriptional

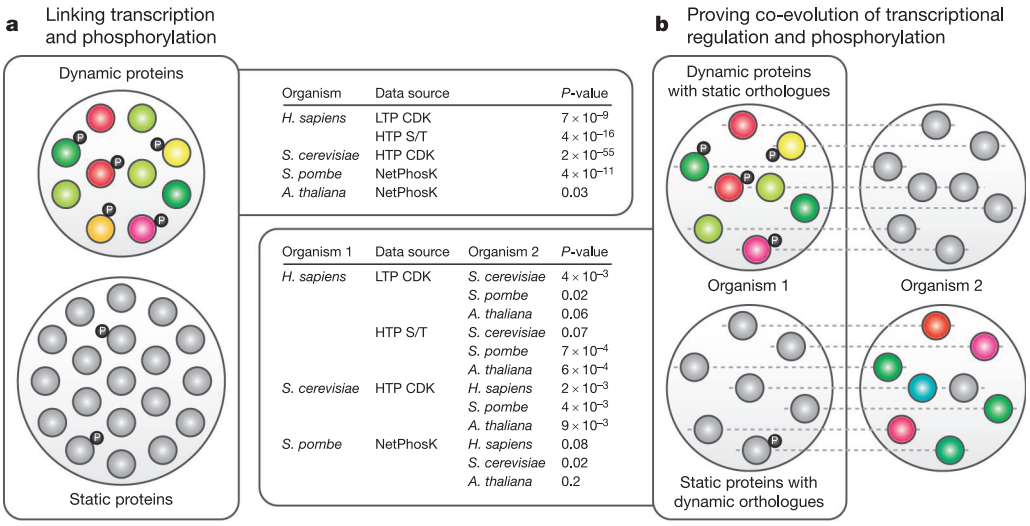

Figure 3 | Cyclin-dependent kinases preferentially phosphorylate dynamic proteins. a, **b**, Two statistical tests were used to show that transcriptional and post-translational regulation tend to control the same proteins. The test shown in **a** compares the dynamic proteins in an organism to all other (static) proteins encoded by the genome. The much stricter test shown in **b** instead compares only the dynamic proteins with static orthologues to the static proteins with dynamic orthologues, thereby directly showing that loss or gain of transcriptional regulation correlates with loss or gain of phosphorylation. Fisher's exact test was used to calculate the statistical significance of the over-representation among the dynamic proteins of phosphoproteins that are either known CDK substrates from low-throughput experiments (LTP CDK[22]), contain S/T phosphorylation sites according to high-throughput (HTP) mass-spectrometry studies (HTP S/T [22]), were identified as CDK substrates in systematic screens (HTP CDK[23,24]), or were predicted to contain CDK phosphorylation sites on the basis of their sequences (NetPhosK[25]). The lower significance in the second, stricter test mainly reflects lower counts, rather than changes in the fractions of phosphoproteins among the dynamic versus the static proteins (see Supplementary Information).

regulation of a gene is correlated with loss or gain of phosphorylation of the corresponding protein. Examples of the latter include the human RRM2 subunit of the RNR complex and the budding yeast DPB2 subunit of DNA polymerase-ε (Fig. 2), which are both dynamic and phosphorylated whereas their orthologues are neither (see Supplementary Information). Within each organism, we therefore compared the dynamic proteins with static orthologues to static proteins with dynamic orthologues (Fig. 3b) and found that transcriptional regulation and phosphorylation have co-evolved in the three organisms studied. Notably, this co-evolution is supported by all pairwise organism comparisons and by both experimental data and sequence-based predictions. Dynamic proteins are thus often controlled through both mechanisms (Fig. 2), although the identity of the dynamic proteins varies greatly between organisms (Fig. 1).

We have previously found in *S. cerevisiae*[9] that, like CDK phosphorylation, cell-cycle-regulated proteolysis preferentially affects the dynamic proteins. Our current results raise the intriguing possibility that all three levels of regulation have co-evolved. Unfortunately, experimental data on cell-cycle-regulated proteolysis is scarce, and we therefore applied the two statistical tests described above to sets of proteins whose sequences contain putative degradation signals. Although these data are too weak to prove co-evolution, the results show that targeted degradation preferentially affects the transcriptionally regulated proteins in all four organisms (see Supplementary Information).

It is tempting to speculate on the driving force that leads to the co-evolution of transcriptional and post-translational regulation, which we could prove in three of the four separate lineages studied. Requiring both transcription and phosphorylation of the same key components might increase robustness to prevent accidental activation. For other proteins, transcriptional regulation might be sufficient for activation, whereas phosphorylation might cause inactivation, for example, by targeting them for degradation.

Together, our results provide a first global view of the evolutionary dynamics of the transcriptional and post-translational regulation of a large and complex biological system. They clearly indicate that although the same general underlying principles, namely just-in-time assembly and multi-layer regulation of functional modules, are widely conserved in eukaryotes, the detailed regulation of individual genes and proteins varies greatly and thus generally cannot be inferred from distantly related organisms. This raises the question of how fast regulation evolves and how closely related two organisms have to be for regulatory details to be transferable. Although microarray expression studies from more closely related model organisms will be needed to investigate this globally, we show that changes in regulation can take place in the order of only a hundred million years (see examples in Supplementary Information), implying that even within vertebrates regulation might differ considerably.

METHODS

Analysis of microarray expression data. Periodically expressed genes were identified by applying the same computational method[8] to 19 cell-cycle microarray timecourses[1–7], and the results were benchmarked against genes for which other evidence indicates periodic expression. For each periodic gene, the time of peak expression was calculated, and the time scales were made comparable across organisms through time warping (see Supplementary Information).

Identification of orthologous proteins. All-against-all Smith–Waterman similarities were calculated, and the proteins were grouped into orthologous groups using a triangular linkage clustering similar to the original COG (Clusters of Orthologous Groups) procedure[26] (see Supplementary Information). The quality of these orthologous groups was assessed on the basis of the manually assigned orthology relationships for DNA replication complexes in Fig. 2.

Curation of complexes. The composition of the individual complexes in Fig. 2 was based on literature and annotations from the relevant model organisms databases (SGD for budding yeast; GeneDB for fission yeast; Ensembl, Uniprot and Reactome for human). Details on the annotations can be found in Supplementary Information and at http://www.cbs.dtu.dk/cellcycle.

Correlation analyses of transcriptional regulation and phosphorylation. Lists of human and budding yeast CDK phosphoproteins were compiled from the Phospho.ELM database[22] and large-scale screens[23,24]. In addition, CDK substrates were predicted from protein sequences using the NetPhosK cdk5 method[25], which we benchmarked on known CDK substrates (see Supplementary Information). A set of human phosphoproteins (not CDK-specific) was compiled from high-throughput mass-spectrometry studies[22]. For each of these sets of phosphoproteins, Fisher's exact test was used to assess the statistical significance of the correlation between transcriptional regulation and phosphorylation.

Received 23 May; accepted 18 August 2006.
Published online 27 September 2006.

1. Whitfield, M. L. *et al.* Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol. Biol. Cell* **13,** 1977–2000 (2002).
2. Cho, R. J. *et al.* A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell* **2,** 65–73 (1998).
3. Spellman, P. T. *et al.* Comprehensive identification of cell cycle-regulated genes of the yeast *S. cerevisiae* by microarray hybridization. *Mol. Biol. Cell* **9,** 3273–3297 (1998).
4. Rustici, G. *et al.* Periodic gene expression program of the fission yeast cell cycle. *Nature Genet.* **36,** 809–817 (2004).
5. Peng, X. *et al.* Identification of cell cycle-regulated genes in fission yeast. *Mol. Biol. Cell* **16,** 1026–1042 (2005).
6. Oliva, A. *et al.* The cell cycle-regulated genes of *Schizosaccharomyces pombe*. *PLoS Biol.* **3,** e225 (2005).
7. Menges, M., Hennig, L., Gruissem, W. & Murray, J. A. H. Genome-wide gene expression in an *Arabidopsis* cell suspension. *Plant Mol. Biol.* **53,** 423–442 (2003).
8. de Lichtenberg, U. *et al.* Comparison of computational methods for the identification of cell cycle regulated genes. *Bioinformatics* **21,** 1164–1171 (2005).
9. de Lichtenberg, U., Jensen, L. J., Brunak, S. & Bork, P. Dynamic complex formation during the yeast cell cycle. *Science* **307,** 724–727 (2005).
10. Marguerat, S. *et al.* The more the merrier: comparative analysis of microarray studies on cell cycle-regulated genes in fission yeast. *Yeast* **23,** 261–277 (2006).
11. Ota, K., Goto, S. & Kanehisa, M. Comparative analysis of transcriptional regulation in eukaryotic cell cycles. *Proc. Fourth Intl Workshop Bioinformatics Syst. Biol.* 26–27 (2004).
12. Sherlock, G. STARTing to recycle. *Nature Genet.* **36,** 795–796 (2004).
13. Dyczkowski, J. & Vingron, M. Comparative analysis of cell cycle regulated genes in eukaryotes. *Genome Inform. Ser. Workshop Genome Inform.* **16,** 125–131 (2005).
14. Bell, S. P. & Dutta, A. DNA replication in eukaryotic cells. *Annu. Rev. Biochem.* **71,** 333–374 (2002).
15. Kearsey, S. E. & Cotterill, S. Enigmatic variations: divergent modes of regulating eukaryotic DNA replication. *Mol. Cell* **12,** 1067–1075 (2003).
16. Nasmyth, K. & Haering, C. H. The structure and function of SMC and kleisin complexes. *Annu. Rev. Biochem.* **74,** 595–648 (2005).
17. Forsburg, S. L. & Nurse, P. Cell cycle regulation in the yeasts *Saccharomyces cerevisiae* and *Schizosaccaromyces pombe*. *Annu. Rev. Cell Biol.* **7,** 227–256 (1991).
18. Diffley, J. F. X. Regulation of early events in chromosome replication. *Curr. Biol.* **14,** R778–R786 (2004).
19. Yanow, S. K., Lygerou, Z. & Nurse, P. Expression of Cdc18/Cdc6 and Cdt1 during G2 phase induces initiation of DNA replication. *EMBO J.* **20,** 4648–4656 (2001).
20. Ekholm-Reed, S. *et al.* Deregulation of cyclin E in human cells interferes with prereplication complex assembly. *J. Cell Biol.* **165,** 789–800 (2004).
21. Zhu, Y., Ishimi, Y., Tanudji, M. & Lees, E. Human CDK2 inhibition modifies the dynamics of chromatin-bound minichromosome maintenance complex and replication protein A. *Cell Cycle* **4,** 1254–1263 (2005).
22. Diella, F. *et al.* Phospho.ELM: a database of experimentally verified phosphorylation sites in eukaryotic proteins. *BMC Bioinformatics* **5,** 79 (2004).
23. Ubersax, J. A. *et al.* Targets of the cyclin-dependent kinase cdk1. *Nature* **425,** 859–864 (2003).
24. Loog, M. & Morgan, D. O. Cyclin specificity in the phosphorylation of cyclin-dependent kinase substrates. *Nature* **434,** 104–108 (2005).
25. Blom, N., Sicheritz-Ponten, T., Gupta, R., Gammeltoft, S. & Brunak, S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* **4,** 1633–1649 (2004).
26. Tatusov, R. L., Koonin, E. V. & Lipman, D. J. A genomic perspective on protein families. *Science* **278,** 631–637 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. von Mering for assistance with detection of orthologues, and P. Nurse, E. Karsenti, J. Bähler, and members of the Bork and Brunak groups for comments on the manuscript. This work was supported by grants from the Danish National Research Foundation, the Danish Technical Research Council, and the European Commission FP6 Programme (grants DIAMONDS and BioSapiens).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.B. (bork@embl.de).