

**Supplementary Information for:
Co-evolution of transcriptional and posttranslational cell cycle
regulation**

**Lars Juhl Jensen^{1*}, Thomas Skøt Jensen^{2*}, Ulrik de Lichtenberg^{2*},
Søren Brunak², and Peer Bork^{1,3†}**

¹European Molecular Biology Laboratory, D-69117 Heidelberg, Germany

²Center for Biological Sequence Analysis, Technical University of Denmark, DK-2800 Lyngby, Denmark

³Max-Delbrück-Centre for Molecular Medicine, D-13092 Berlin, Germany

*These authors contributed equally

†To whom correspondence should be addressed; Email: bork@embl.de

Contents

Materials and methods	3
Microarray expression timecourses	3
Identification of cell cycle regulated genes	3
Benchmarking the gene lists	4
Assignment of peak times	6
Time warping	7
Detection of orthologous genes	7
Curation of protein complexes	7
Quality assessment of sequence-based orthology assignments	8
Known and predicted substrates of cyclin-dependent kinases	9
Prediction of ubiquitinated proteins	9
Global analyses and statistics	10
Conservation of cell cycle genes	10
Periodicity is rarely conserved	10
Timing of expression	11
Phosphorylation	13
Targeted protein degradation	13
Evolution of protein regulation	15
Details on individual protein complexes	17
DNA Replication	17
Pre-replication complex (pre-RC)	17
Dbf4-dependent kinase (DDK)	19
Pre-initiation complex (pre-IC)	19
DNA replication factor A (RFA)	20
DNA polymerase ϵ	20
DNA polymerase α	20
DNA replication factor C (RFC)	21
Proliferating cell nuclear antigen (PCNA)	21
DNA polymerase δ	21
Ribonucleotide-diphosphate reductase (RNR)	21
Cohesin	22
Separase and securin	23
Shugoshin	23
Anaphase promoting complex/cyclosome (APC/C)	23
Protein kinase A (PKA)	24
Cyclin-dependent kinases (CDKs)	24

Materials and methods

Microarray expression timecourses

In this study, we analyze microarray expression timecourses from four different organisms, namely *H. sapiens*, *S. cerevisiae*, *S. pombe*, and *A. thaliana*. The datasets are summarized in Table S1.

Two microarray studies have been published on the human cell cycle. The earliest dataset by Cho et al. studied the gene expression in human fibroblasts⁸. It has since been shown that the number of periodically expressed genes in this dataset is no higher than what is obtained for shuffled expression profiles, strongly suggesting that the cells were in fact not properly synchronized⁹. We thus excluded this dataset from our analysis. We instead use the data published by Whitfield et al. which contain four timecourses on HeLa cells synchronized by a double thymidine block or a thymidine-nocodazole block¹.

For *S. cerevisiae*, we based our analysis on the three microarray timecourses from the seminal papers by Cho et al.³ and Spellman et al.². In these experiments, the yeast cells were synchronized using three different synchronization methods, namely alpha-factor, Cdc15, and Cdc28 block-and-release. Spellman et al. also published a fourth timecourse in which a synchronized cell culture had been obtained using the elutriation protocol; however, the authors excluded this timecourse from their combined analysis². Indeed, it spans only a single cell cycle, and many of the genes which change in expression appear to be involved in stress response rather than the cell cycle¹⁰. For these reasons we too decided to exclude the elutriation timecourse from our analysis.

No less than 10 microarray timecourses on the *S. pombe* cell cycle have been published by three research groups⁴⁻⁶. The first and largest dataset was published by Rustici et al.⁴ and includes five timecourses. Peng et al.⁵ and Oliva et al.⁶ have since published another two and three timecourses, respectively. All three research groups synchronized their cell cultures by two different methods, namely block-and-release using a temperature sensitive *cdc25* mutant and elutriation. Here, we use the combined evidence from all ten experiments (see below).

Finally, we use two short timecourses (by Menges et al.) from *Arabidopsis thaliana* cell suspensions, which had been synchronized by aphidicolin block (APH) and sucrose starvation (SUC), respectively⁷. Both of these timecourses are very short compared to those of the other organisms, the longest covering only a single cell division cycle. The plant data is therefore by far the smallest data set, both in terms of the number of cycles

covered and in terms of the number of independent experiments.

Identification of cell cycle regulated genes

To enable a comparison of the subset of cell cycle regulated genes in the four organisms, we reanalyzed all available microarray expression timecourses to obtain the best possible list of periodically expressed genes for each organism. Importantly, using the same methodology to analyze the datasets for all four organisms also ensures that any observed differences between the organisms cannot be attributed to differences in choice of analysis method. For this purpose, we used a permutation-based method, which combines two statistical tests (one for regulation and one for periodicity) in a score for each gene¹¹.

The significance of regulation, p_{reg} , was estimated by comparing the observed variance for each expression profile to a background distribution, which was estimated based on 1,000,000 random profiles, each constructed by selecting for each timepoint the expression value of a random profile. The fraction of random profiles with a variance greater than or equal to that observed for the real expression profile was reported as the p-value for regulation.

To estimate the significance of periodicity, p_{per} , we first calculated the Fourier score for each expression profile:

$$F = \sqrt{[\sum \sin(\omega t) \cdot x(t)]^2 + [\sum \cos(\omega t) \cdot x(t)]^2}$$

where $\omega = 2\pi/T$, with T being the interdivision time. The p-value was estimated by comparing the observed score to that of 1,000,000 artificial profiles constructed by random shuffling of the data points from the expression profile in question. To compensate for interdependencies among the timepoints in an expression profile, p-values were scaled to a genome-wide median of 1 within each timecourse. For each expression profile, a combined p-value of regulation was calculated by multiplying the separate p-values of regulation from each of the timecourse. Analogously, a combined p-value of periodicity was calculated. Finally, a combined score was calculated for each expression profile:

$$S = p_{reg} \cdot p_{per} \cdot (1 + \lambda_{reg} \cdot p_{reg})^2 \cdot (1 + \lambda_{per} \cdot p_{per})^2$$

where λ_{reg} and λ_{per} are organism-specific parameters, used to favor genes which display both high amplitude and strong periodicity.

This method was applied to microarray data for the four organisms. The analysis of the *S. cerevisiae* data sets has previously been published

Table S1: **Summary of cell cycle microarray experiments used in this study.** The table summarizes who did the experiments, the type of microarrays used, and the length of each experiment. Two fundamentally different microarray platforms have been used, namely single-channel Affymetrix GeneChips (“Affymetrix”) and two-channel spotted cDNA microarrays (“Spotted”). The “Experiment” column specifies the method used to synchronize the cell culture and provides a unique identifier for each timecourse used throughout this supplement. “Time offset” specifies the value that was subtracted from the peak times obtained from each experiment in order to align the timecourses so that the zero timepoint always corresponds to the time of cell division (M/G_1).

Organism	Group	Microarray platform	Experiment	Time points	Interdivision time	Cell cycles	Time offset
<i>H. sapiens</i>	Whitfield et al. ¹	Spotted	Thy-Thy 1	11	15.5 hr	2	55%
			Thy-Thy 2	26	15.5 hr	3	46%
			Thy-Thy 3	47	15.4 hr	3	51%
			Thy-Noc	19	18.5 hr	2	3%
<i>S. cerevisiae</i>	Spellman et al. ²	Spotted	Alpha	18	58 min	2	78%
			Cdc15	24	115 min	2 ^{1/2}	84%
<i>S. pombe</i>	Cho et al. ³	Affymetrix	Cdc28	17	85 min	2	93%
	Rustici et al. ⁴	Spotted	Cdc25 1	20	142 min	2	71%
			Cdc25 2	18	137 min	2	63%
			Elutriation 1	20	158 min	2	78%
	Peng et al. ⁵	Spotted	Elutriation 2	20	154 min	2	69%
			Elutriation 3	20	144 min	2	57%
			Cdc25	37	182 min	2	55%
	Oliva et al. ⁶	Spotted	Elutriation	32	155 min	2	46%
			Cdc25	52	172 min	3	10%
			Elutriation 1	50	152 min	2 ^{1/2}	47%
Elutriation 2			33	156 min	3	4%	
<i>A. thaliana</i>	Menges et al. ⁷	Affymetrix	APH	10	22 hr	1	56%
			SUC	6	22 hr	1/2	—

and the present analysis is based on the 600 most periodically expressed genes¹¹. For *S. pombe*, we recently reanalyzed and combined the evidence from all ten experiments using the method described above¹². We here use the list of 500 periodically expressed genes, which was proposed based on this reanalysis. For *A. thaliana*, the raw Affymetrix probe set intensity values were log transformed and in the few cases where multiple probe sets existed for a single locus, these values were subsequently averaged for each timepoint. Furthermore, each expression profile was normalized to a mean expression of zero. Following these data transformations, the data were analyzed using the scoring scheme described above (with $\lambda_{reg} = 10$ and $\lambda_{per} = 100$) to obtain a score for each gene in the data set. Based on a benchmarking procedure (described below) we propose a list of 400 cell cycle regulated genes that we used in this work.

In each of the four *H. sapiens* experiments the expression values were averaged in the few cases where measurements were in duplicates. Expression profiles were subsequently normalized to a mean of zero and analyzed as described above. A combined score (with $\lambda_{reg} = 0$ and $\lambda_{per} = 0$) was calculated for each probe in the data set. Probes were, if possible, mapped to Ensembl genes and the highest score retained in the case of multiple entries for the same gene. We used the 600 most cell cycle regulated genes in the present analysis, based on the benchmark procedure described below.

Benchmarking the gene lists

For quality assessment and control, we used a benchmarking procedure where each list of periodic genes is compared to various sets of genes for which there is independent evidence or indication of cell cycle regulation. Most of these benchmark sets consist of genes identified either as periodically in classical small scale studies or as targets of known cell cycle transcription factors in ChIP-chip studies^{13;14}, although it should be noted that such sets can contain false positives.

The *S. cerevisiae* and *S. pombe* gene lists were benchmarked in earlier publications and were shown to be of better or equal quality as previously published lists of cell cycle regulated genes^{11;12}. Our reanalysis correctly identified 90% of the *S. cerevisiae* genes and 80% of the *S. pombe* genes that are known to be periodically expressed from small-scale studies, and we thus estimate the sensitivity of the *S. cerevisiae* and *S. pombe* gene lists to be at least 90% and 80%, respectively. In both yeasts, the evidence supporting some of the “known cell cycle regulated genes” is questionably and the sensitivity is thus likely better than what we estimate when assuming that all these are correctly assigned.

As with the two yeasts, the quality of the gene lists proposed for *H. sapiens* and *A. thaliana* was evaluated by their consistency with two benchmark sets for each organism. The primary *H. sapiens* benchmark set (B1) was based on a previously published list of genes, which are known from small-scale experiments to be peri-

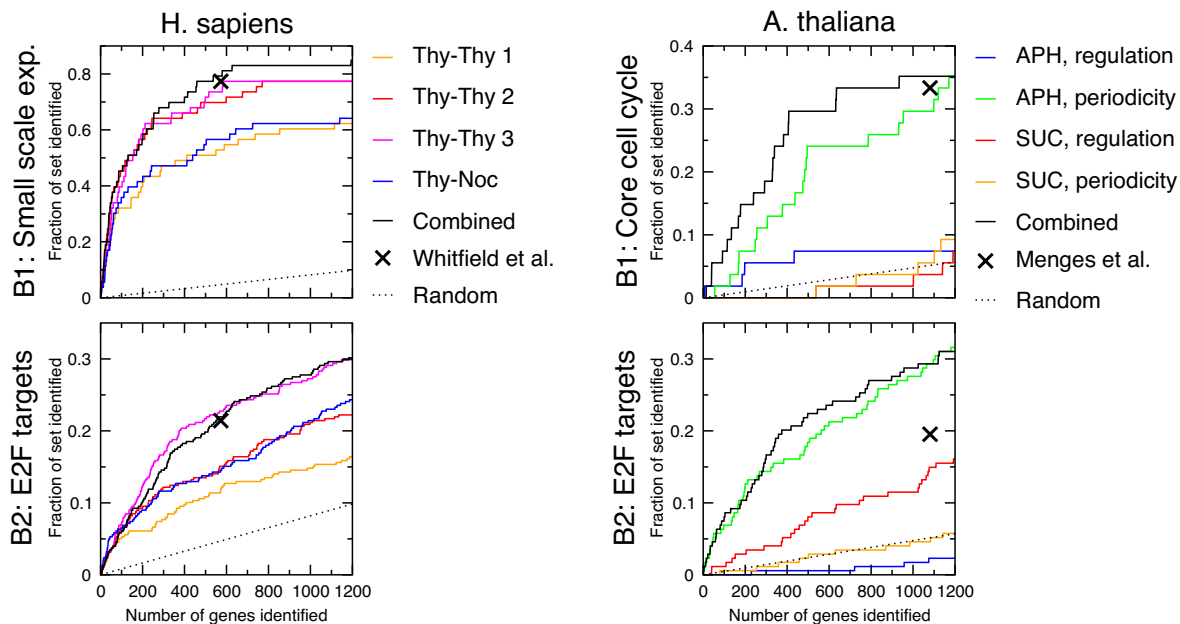


Figure S1: **Benchmarking of the *H. sapiens* and *A. thaliana* gene lists.** For each organism, the periodically expressed genes identified by our reanalysis were compared to each of two benchmark sets. The fraction of each benchmark set correctly identified is plotted as a function of the number of genes suggested to be periodically expressed. The gene lists proposed by Whitfield et al.¹ and Menges et al.⁷ are shown for comparison. Similar benchmark plots for *S. cerevisiae* and *S. pombe* have been published elsewhere^{11;12}.

odically expressed (including all histone genes)¹. A secondary benchmark set (B2) was compiled based on two lists of proposed E2F transcription factor targets^{15;16}. After mapping to Ensembl genes, the two *H. sapiens* benchmark sets consist of 63 and 439 genes, respectively. The two *A. thaliana* benchmark sets consist of 61 core cell cycle genes¹⁷ and 176 genes proposed to be regulated by E2F transcription factors¹⁸, respectively. These lists should contain a large fraction of truly cell cycle regulated genes, although one should expect some false positives among the transcription factor targets. So, while the benchmark sets are well-suited for comparing the quality of the gene lists that various methods propose as cell cycle regulated, the fraction of a benchmark set that is identified only gives a lower bound on the sensitivity. The benchmark sets for all four organisms can be downloaded from <http://www.cbs.dtu.dk/cellcycle>.

Figure S1 shows the fraction of each benchmark identified as a function of the number of genes suggested to be periodically expressed. For comparison, we also show the performance of the original analyses by Whitfield et al.¹ and Menges et al.⁷ as well as the performance that would result from selecting genes at random (dashed line). Based on where the curves break and/or show no further enrichment over random expectation, we selected the top-600 *H. sapiens* genes and the top-400 *A. thaliana* genes for use in our analysis.

Our reanalysis of the HeLa data results in a list of periodically expressed genes that is of very similar quality to the list proposed by Whitfield et al.¹. This was to be expected since Whitfield et al. used a Fourier-based analysis; such methods

generally give results that are only slightly worse than those obtained using our method^{11;12}. The top-600 list has an estimated sensitivity of 81% based on benchmark set B1, which is a conservative estimate since many of the histone genes included in benchmark set B1 have in fact never been shown to be periodically expressed¹.

Remarkably, our *A. thaliana* list contains about the same number of genes from the benchmark sets as the original list by Menges et al.⁷, although they proposed 2.7 times as many periodically expressed genes as us. This shows that our list has similar sensitivity but far better speci-

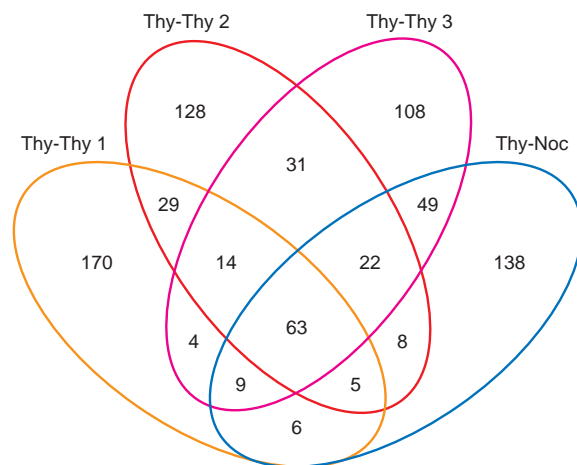


Figure S2: **Agreement across experiments.** The top-300 ranking genes identified based on each of the four human timecourses are compared in a Venn diagram. The highest overlap is obtained for the Thy-Thy 2 and Thy-Thy 3, whereas the Thy-Thy 1 gives the least overlap with other timecourses. These results are fully consistent with our benchmark (Figure S1).

ficity compared to the original list. The top-400 genes encompass 25% of the “core cell cycle genes” (B1), however, the sensitivity is expected to be about twice as high since many cell cycle genes are known to be static¹⁹. A sensitivity in the order of 50% is also consistent with what was observed for individual time courses from the other three organism. Moreover, it agrees well with the fact that five of the eight orthologous groups that contain a dynamic protein from each of the three other organisms also contain a dynamic protein from *A. thaliana* (Figure 1).

In addition to comparing the proposed gene lists to an external reference, the quality of the lists can also be assessed by comparing the highest ranking genes in each experiment internally. The Venn diagram in Figure S2 show such a comparison of the top-300 ranking genes based on each of the four individual *H. sapiens* timecourses. The best agreement is obtained between the Thy-Thy 2 and Thy-Thy 3 timecourses; indeed, these two timecourses also gave the best performance relative to external benchmark sets. Conversely, the Thy-Thy 1 timecourse shows the worst overlap in Figure S2 and also gave the worst performance on external benchmark sets. The results of internal and external benchmarking are thus fully consistent with each other.

The possible influence of synchronization methods on periodic gene expression has been intensely debated^{20–24}. In brief, a minority viewpoint is that different synchronization methods lead to widely different and spurious results^{20;21;23}. The fact that Thy-Thy 1 rather than Thy-Noc stands out among the human time courses hints that the discrepancies observed between different time courses are not primarily due to different synchronization methods being used (Figure S2). Recently, analyses of data in both budding and fission yeast have demonstrated that, although the level of noise is considerable, the experiments generally yield a consistent picture of the identity and transcriptional timing of cell cycle regulated genes, even though different synchronization methods release the cells from different points in the cell cycle^{11;12}. Most importantly, a pairwise comparisons of the ten microarray expression timecourses on the fission yeast cell cycle shows that the data produced with different synchronization techniques in the same laboratory are no more different than those generated with the same technique in different laboratories¹².

Assignment of peak times

The temporal profile of a periodically expressed gene can be summarized as a single number, the time of peak expression, which specifies at which point during the cell cycle the gene is maximally expressed. We report this *peak time* as a percentage of the cell cycle to compensate for the dif-

ference in interdivision time between the experiments.

Within a single experiment, the time of peak expression for a gene is determined by fitting its expression profile with a sine wave. Because different synchronization methods release cells at 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-ranking genes in each organism. A combined peak time was calculated for each gene as the weighted average (on a circle) of the peak time obtained in each of experiments (for details, see reference¹¹).

Aligning the time scales to each other still leaves one free parameter, namely the definition of “time zero”. In each organism, we arbitrarily defined the zero timepoint as the median peak time of genes believed to be expressed during cytokinesis (the M/G₁ transition). For this, we used the M/G₁ cluster from Whitfield et al.¹ for *H. sapiens*, the M/G₁ cluster from Spellman et al.² for *S. cerevisiae*, cluster 2 from Rustici et al.⁴ for *S. pombe*, and three periodically expressed cytokinesis genes (AT3G60840, AT4G32830, and AT5G67270) for *A. thaliana*. Although the latter number is based on just three genes, it should be noted that the result is consistent with Figure 3 in the paper by Menges et al.⁷. This definition of “time zero” in conjunction with the alignment procedure results in the time offsets listed in Table S1, which were subtracted from the peak times obtained from each individual timecourse.

To ensure that it is meaningful to calculate combined peak times across experiments, it is important to verify that each phase occupies approximately the same fraction of the cell cycle in each experiment. Figure S3 shows that this is indeed the case for the four *H. sapiens* timecourses. Similar consistency checks have been published for *S. cerevisiae*¹¹ and *S. pombe*¹², and no consistency check is needed for *A. thaliana* as the peak times are determined from a single time course only. The final, combined peak times for all periodically expressed genes from each of the four organisms are available at <http://www.cbs.dtu.dk/cellcycle/>.

To show that the peak time of a gene accurately summarizes its temporal expression in multiple time courses, we plotted the actual expression profiles of the periodically expressed genes, coloring each profile according to its combined peak time (Figure S4). These plots confirm that the genes with similar peak times indeed show similar expression patterns within each timecourse.

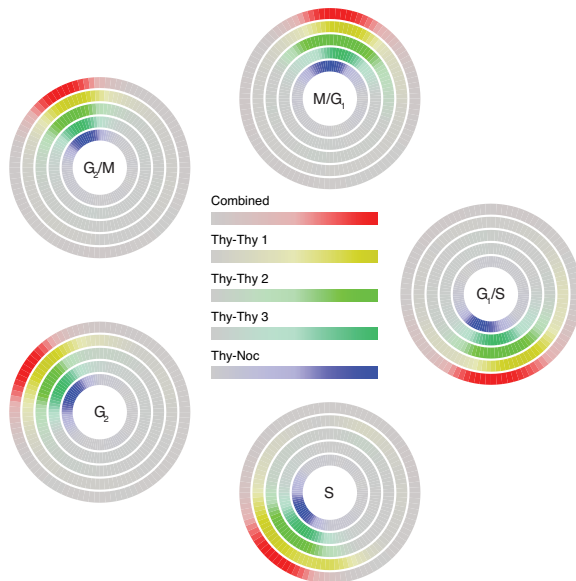


Figure S3: **Expression of human phase-specific genes.** The distribution of peak times in each experiment is shown for each of the phase-specific gene clusters identified by Whitfield et al.¹. The mode of the peak time distribution for the M/G₁ cluster was chosen to define the zero timepoint in each experiment. Equivalent figures for *S. cerevisiae* and *S. pombe* are available elsewhere^{11;12}.

Time warping

Although peak times are measured in percent of a cell cycle, they cannot be directly compared between organisms since the different phases do not have the same relative length in each organism. Especially G₁ and G₂ vary greatly in length; in *S. pombe*, G₂ accounts for around 60% of the cell cycle and G₁ is nearly non-existent whereas the two phases are about equally long in *S. cerevisiae*. To correct for this, we devised a time warping approach to align the time scales of the four organisms. Traditional time warping methods for microarray expression time courses rely on globally minimizing the peak time difference of individual genes between organisms^{25;26}. However, because of the small number of genes which are periodic in multiple species, and because many of these have changed peak times during evolution, such methods are not applicable.

Instead, we identified two well-defined timepoints during the cell cycle, which can be identified from expression profiles. One is the time of cell division (M/G₁), which was used to define the zero timepoint as described above. The other fixpoint is the middle of S phase, which was estimated based on the peak times of periodically expressed histone genes. Relative to the time of cell division, the histone genes are maximally expressed 62% (*H. sapiens*), 39% (*S. cerevisiae*), 18% (*S. pombe*), and 55% (*A. thaliana*) into the cell cycle. Between the two fixpoints, linear interpolation was used to transform the *S. pombe*, *H. sapiens*, and *A. thaliana* timescales to match that of *S. cerevisiae*. The latter was chosen as the reference timescale to allow direct comparison

with our earlier study¹⁹, and because each phase conveniently occupies roughly the same fraction of the *S. cerevisiae* cell cycle.

Detection of orthologous genes

Orthology assignment between genes in the four organisms is essential for any comparative study. We assembled proteins into orthologous groups using an automatic procedure similar to the original COG procedure^{27;28}. First, we compute all-against-all Smith-Waterman similarities for the proteins, using the longest isoform from each locus. We then group recently duplicated sequences within genomes into “in-paralogous groups”, to be treated as single sequences subsequently. For this, there is no fixed cutoff in similarity, but instead we start with a stringent similarity cutoff and relax it step-wise, until all in-paralogous proteins are joined, satisfying the following criteria: all members of a group have to be more similar to each other than to any other protein in any other species (very similar species are treated as one), and all members of the group have to have hits that overlap by at least 20 residues. After grouping paralogous proteins, we start assigning orthology between proteins, by joining triangles of reciprocal best hits involving three different species (here, paralogous groups were represented by their best-matching member). Again, a stringent similarity cutoff is used first and relaxed step-wise, and all proteins in a group are required to have hits overlapping by at least 20 residues. Finally, we join any remaining nodes by requiring only reciprocal pairs instead of triangles. This procedure resulted in 4631 orthologous groups containing a total of 21198 proteins.

Curation of protein complexes

For the detailed analysis of DNA replication and sister chromatid cohesion, we manually curated a set of protein complexes. The *H. sapiens* protein complexes were primarily obtained from Reactome²⁹ and were supplemented with information from Ensembl³⁰ and UniProt³¹. The Saccharomyces Genome Database (SGD)³² was used for curation of the *S. cerevisiae* version of the same complexes. The orthologous groups described above were used to identify orthologous genes in *S. pombe* and *A. thaliana*. We subsequently checked the functional annotation of these proteins in GeneDB³³ and The Arabidopsis Information Resource (TAIR)³⁴ to ensure that the proteins are not just sequence homologs but also functional homologs. Finally, the curated complexes were verified against the combined information obtained from several recent review papers^{35–44}.

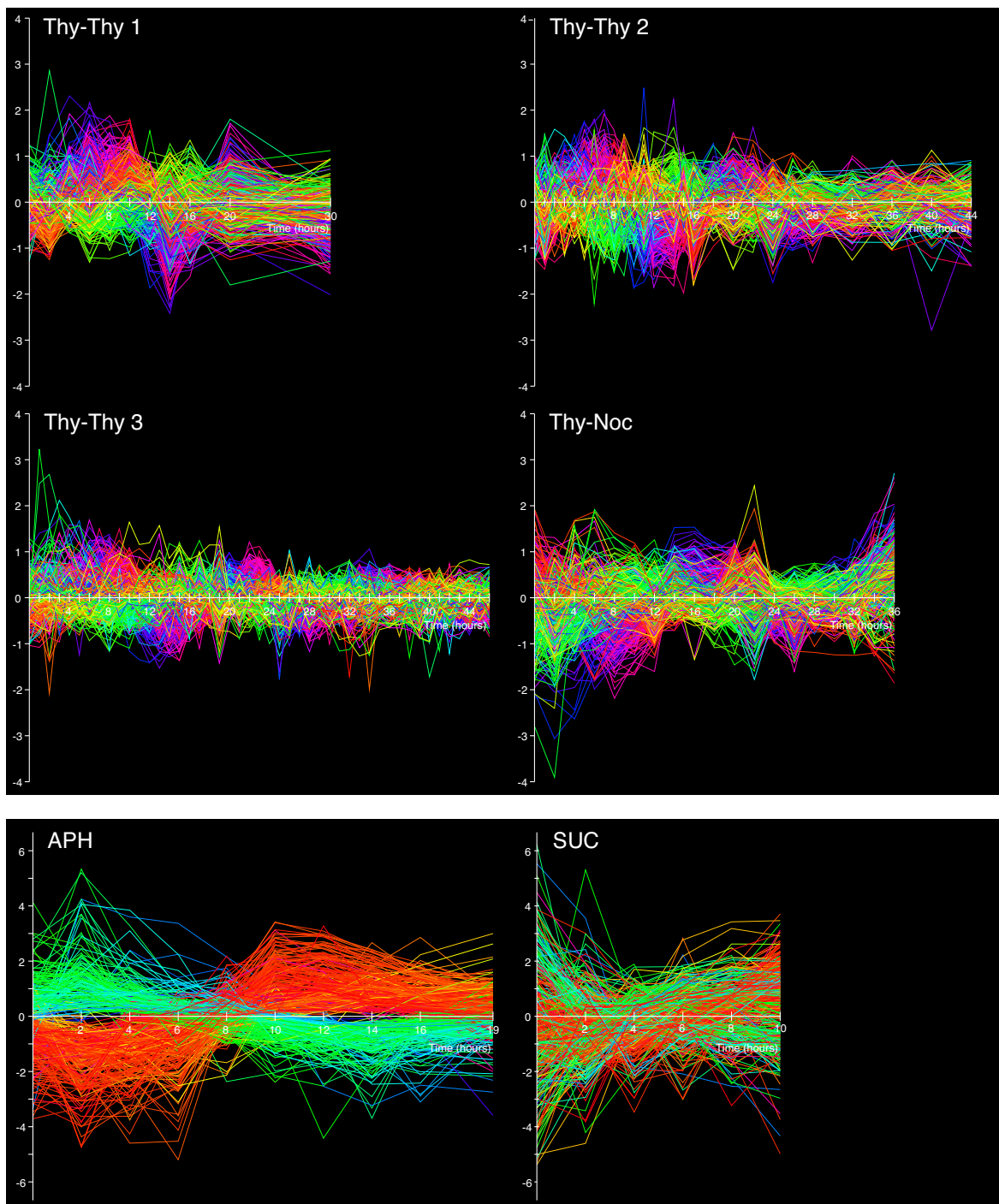


Figure S4: **Expression profiles of periodically expressed genes.** For the 600 and 400 periodically expressed genes from human and *A. thaliana*, respectively, the individual expression profiles are plotted. The profiles for each gene are colored according to its combined peak time using the same color scale as in Figure 2. As expected, the plots show that genes with similar peak time have similar expression profiles in each microarray timecourse.

Quality assessment of sequence-based orthology assignments

The curated DNA replication complexes described above consist of a total of 53 orthologous groups that contain proteins from all four species. We used these manually curated orthology relations to assess the quality of the orthologous groups that were automatically derived from sequence similarity. In only two cases did our automated procedure assign wrong proteins to an orthologous group, corresponding to a specificity of over 95%.

In an additional eight cases, the procedure failed to identify the orthologous protein in one or more organisms, which leads to an estimated sensitivity of 85%. We thus conclude that over 80% of the orthologous groups constructed by our automatic procedure are entirely correct, containing the functionally equivalent proteins from all four organisms.

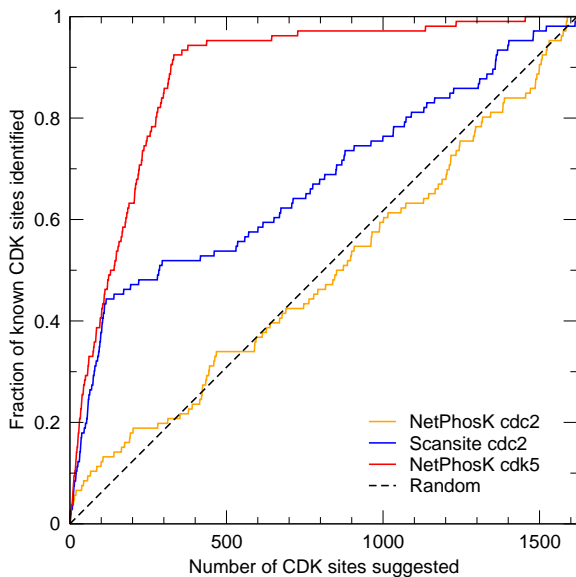


Figure S5: **Benchmark of methods for sequence-based prediction of CDK phosphorylation sites.** Three different methods for predicting CDK phosphorylation sites from sequence were benchmarked against the Phospho.ELM database⁴⁵. Each method was applied to all human phosphorylation in the database, and the predictions were sorted by their score. The fraction of known CDK phosphorylation sites identified by each method was plotted as function of the total number of such sites suggested by the method. According to this benchmark, the NetPhosK cdk5 prediction method yields the best results.

Known and predicted substrates of cyclin-dependent kinases

For our analyses, we made use of several types of phosphorylation data as well as sequence-based prediction. A list of 60 human CDK substrates identified in low-throughput experiments were obtained from the Phospho.ELM database⁴⁵. The human proteins shown in Figure 2 were furthermore checked in the PhosphoSite database⁴⁶; this did not reveal any additional CDK substrates. A possible caveat of this source is that known CDK substrates may be biased towards proteins, which are known to play a role in the cell cycle. A second set consisting of 454 human S/T-phosphoproteins was thus compiled from four mass spectrometry studies^{47–50} available from the Phospho.ELM database⁴⁵. Although tyrosine phosphorylation sites identified in these studies were specifically excluded, it should be noted that many of the 454 proteins are unlikely to be phosphorylated by a cyclin-dependent kinase but by some other S/T-kinase. A list of 340 *S. cerevisiae* substrates of the cyclin-dependent kinase Cdc28 was compiled based on two high-throughput screens^{51;52}. To the best of our knowledge, no databases or screens have been made for *S. pombe* and *A. thaliana* phosphoproteins. For these two organisms, we thus had to rely solely on sequence-based predictions.

We tested three different methods for predicting CDK phosphorylation sites from sequence, namely the cdc2 and cdk5 predictors from Net-

PhosK⁵³ and the cdc2 predictor from Scansite⁵⁴. To benchmark these methods, we applied each method to all phosphorylation sites in the Phospho.ELM database, sorted the predictions by score, and plotted the fraction of known CDK sites identified as a function of the total number of CDK sites suggested (Figure S5). The NetPhosK cdk5 predictor clearly outperforms both cdc2 predictors, which is surprising considering that CDK1 (also known as cdc2) substrates make up the majority of known CDK substrates. Based on the results of this benchmark, we chose NetPhosK cdk5 as our primary CDK prediction method and raised its threshold score from 0.5 to 0.6, since this considerably improves the specificity without notably affecting the sensitivity. Scansite cdc2 was used as a secondary prediction method to support the correlations observed using NetPhosK cdk5.

Prediction of ubiquitinated proteins

Ubiquitination is another posttranslational modification that regulates many proteins involved in the cell cycle; in particular in the form of polyubiquitination, which targets proteins for degradation by the proteasome. Two distinct ubiquitin ligase complexes are involved in regulating the cell cycle, namely the Skp1/Cullin/F-box protein complex (SCF) and the anaphase promoting complex/cyclosome (APC/C). SCF recognizes so-called PEST regions, which are only ubiquitinated once they have been phosphorylated^{55–58}. PEST regions were predicted using the PEST find program⁵⁶. The substrate recognition by APC/C is less well understood, with numerous different recognition motifs having been published^{37;59}. We based our prediction of possibly APC/C substrates on the two most prominent such motifs, namely the D box and the KEN box, which were predicted using the regular expressions $R..L\dots N$ and KEN , respectively. It should be noted, however, that several functional D boxes are known, which do not conform to the consensus sequence³⁷.

Global analyses and statistics

Conservation of cell cycle genes

It is generally accepted that the mitotic cell cycle is a highly conserved process throughout Eukarya. To quantify the conservation of the cell cycle, we first calculated the fraction of dynamic proteins for which orthologs could be identified in all four organisms. Of the $600 + 600 + 500 + 400 = 2100$ dynamic proteins, only 550 (26.2%) fulfill this requirement. This number is in line with the genome-wide fractions of orthology assignments, which vary from 16.2% of the *A. thaliana* proteins to 43.6% of the fission yeast proteins. A periodically expressed gene would thus appear to be just as likely to be lost or gained during evolution as any other gene.

The 550 dynamic proteins with orthologs in all four organisms fall into 381 orthologous groups of which only 128 (33.6%) are 1:1:1:1 orthologs, that is the orthologous group contain exactly one protein from each organism. For comparison, a genome-wide analysis reveals 1:1:1:1 relations for 774 (38.9%) of 1991 orthologous groups with members from all four organisms. Similar results were obtained when comparing the number of 1:1 orthologs to the total number of orthologous groups for each pair of organisms (Table S2). This suggests that cell cycle-related genes are no less likely to undergo gene duplication than other genes. In summary, we find no signs of cell cycle genes being more highly conserved than the rest of the genome.

Periodicity is rarely conserved

Most people would probably expect the cell cycle to involve at least 200 groups of orthologous proteins with conserved dynamic expression throughout Eukarya. In accordance with our benchmarks, we assume that 80% of the orthologous groups are entirely correct and that the lists of dynamic proteins have sensitivities of 80% (*H. sapiens*), 90% (*S. cerevisiae*), 80% (*S. pombe*), and 50% (*A. thaliana*). Given this, $200 \cdot 0.8 \cdot 0.8 \cdot 0.9 \cdot 0.8 \cdot 0.5 = 46$ of the 200 orthologous groups would be expected to occur in the overlap of our sets, that is almost an order of magnitude more than the five orthologous groups that we actually observe in Figure 1. Conversely, our observed overlap suggests that the actual number of conserved dynamic orthologous groups is only $5 / (0.8 \cdot 0.8 \cdot 0.9 \cdot 0.8 \cdot 0.5) = 22$. Given that each organism has 400–600 periodically expressed genes, we can thus conclude that periodicity is very poorly conserved at the level of individual orthologous genes. Although the true overlap may be larger than what is observed (5 orthologous

Table S2: **Orthology relations for pairs of organisms.** The number of 1:1 orthologs and the total number of orthologous groups is shown for each pair of organisms; these numbers are reported separately for orthologous groups that contain at least one dynamic protein (dynamic) and other orthologous groups (static). The numbers suggest that cell cycle genes are as likely to undergo gene duplication as other genes.

Organisms	Dynamic 1:1/all	Static 1:1/all
<i>H. sapiens</i> , <i>S. cerevisiae</i>	153/313	1489/2451
<i>H. sapiens</i> , <i>S. pombe</i>	190/367	1626/2602
<i>H. sapiens</i> , <i>A. thaliana</i>	99/278	1422/3343
<i>S. cerevisiae</i> , <i>S. pombe</i>	300/462	2424/2984
<i>S. cerevisiae</i> , <i>A. thaliana</i>	81/246	1138/2407
<i>S. pombe</i> , <i>A. thaliana</i>	106/269	1276/2550

groups), the general lack of conservation of cell cycle regulated transcription cannot be explained by the quality of the data, even when our most conservative sensitivity estimates are used.

Figure 1 only considers orthologous groups that contain proteins from all four species, and it thus does not encompass all the periodically expressed genes in each organism. To get a more complete picture of the conservation of periodicity, we constructed separate Venn diagrams for each of the six pairs of organisms (Figure S6). Notably, *A. thaliana* gives equally good overlap with each of the three other organisms, suggesting that the lists of periodically expressed genes in human and the two yeasts are of similar quality, which is consistent with the results of our benchmarks described above. These pairwise comparisons also confirm that periodic expression is indeed very poorly conserved at the level of single genes, also when taking into consideration the genes excluded in Figure 1. The only exception is, not surprisingly, a set of 42 orthologous groups that contain dynamic proteins from both yeasts but lack orthologs in *H. sapiens* and/or *A. thaliana*.

To summarize the functional similarities and differences between the organisms, we used the GOstat program⁶⁰ to search for overrepresented Gene Ontology terms among the genes that are periodically expressed in only one or two organisms. This revealed that many of the 75 orthologous groups that contain dynamic proteins from the two yeasts only are indeed involved in yeast-specific cell cycle processes, in particular related to the cell wall. Similarly, many of the conserved proteins that are dynamic only in *S. cerevisiae* are localized to the bud or involved in protein glycosylation or methionine metabolism. In contrast, the proteins that are dynamic only in human tend to localize to the nucleus where they are involved in DNA metabolism, in particular DNA repair.

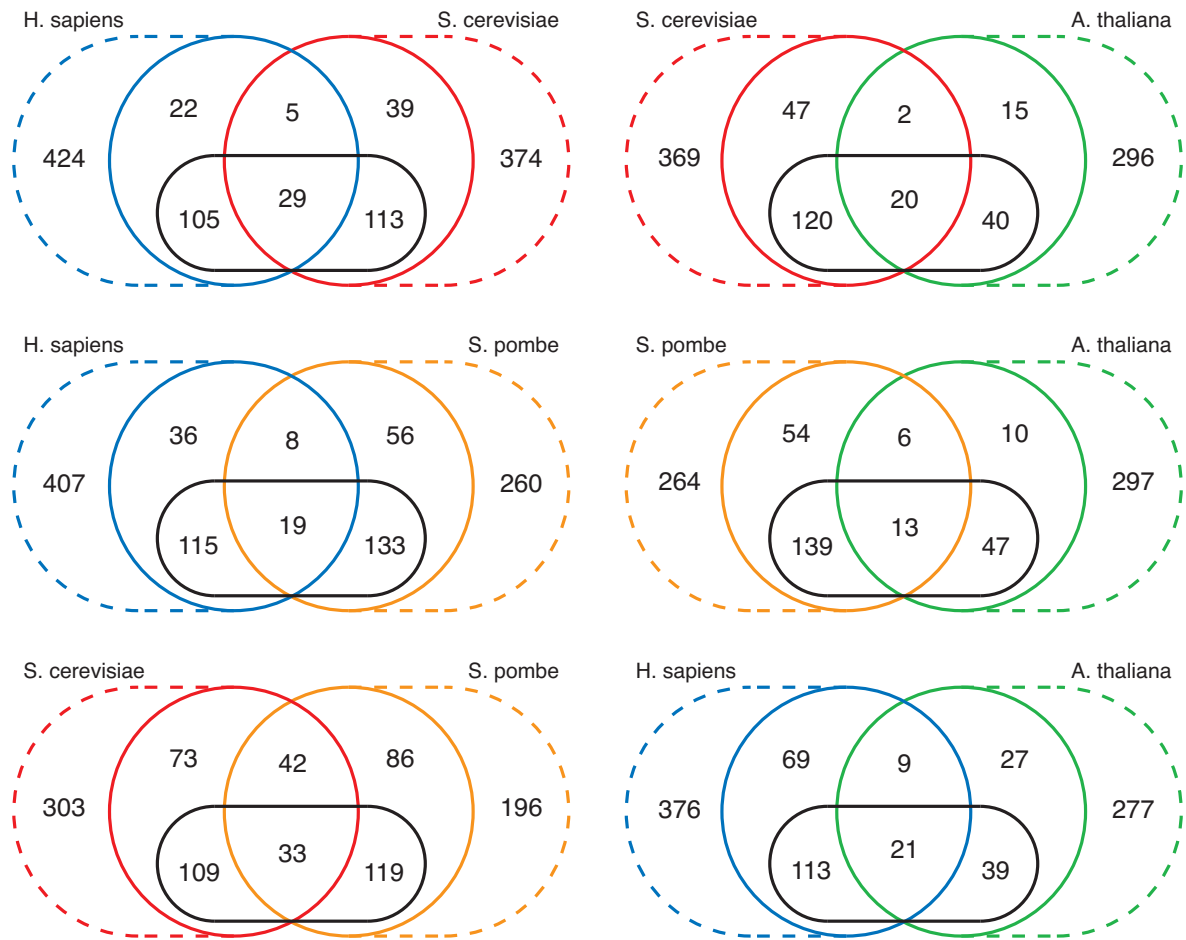


Figure S6: **Comparison of periodically expressed genes between pairs of organisms.** To assess the conservation of periodic gene expression in more detail than done in Figure 1, the Venn diagrams were constructed for each pair of organisms. The two colored circles in each diagram shows the number of orthologous groups that contain dynamic proteins from one or both organisms, and which have at least one member from each of the two organisms. The numbers inside the black line refer to the subset of orthologous groups shown in Figure 1, that is the ones which have members from all four organisms. The dashed extensions of each circle account for the additional orthologous groups that contain a dynamic protein from one organism but no proteins from the other organism. The analysis confirms that periodic gene expression is poorly conserved between orthologs, but also reveals a set of 42 yeast-specific orthologous groups that contain dynamic proteins from both yeasts.

Each organism also contains a considerable number of dynamic proteins for which no orthologs could be identified in any of the three organisms. The Gostat program⁶⁰ was again used to identify overrepresented Gene Ontology terms. Interestingly, many of the dynamic proteins only present in human are involved in regulation of cell proliferation, which makes biological sense since the proliferation of human cells, in contrast to yeast cells, is subject to regulation by external growth factors. Cell wall-related proteins are overrepresented among the dynamic proteins unique to the two yeasts and to *A. thaliana*, which is not surprising considering that the cell walls of the three organisms differ greatly. Moreover, many of the dynamic proteins unique to budding yeast localize to the site of polarized growth and are involved in budding.

Timing of expression

It is not only the identity of the dynamic proteins that varies between organisms; even when peri-

odic expression as such is conserved, the timing of expression has in many cases changed dramatically. This point is illustrated by the pie charts in Figure 1, which show the fraction of orthologous groups for which the timing of expression can be considered to be conserved. While this shows the general trend, it does not show the full picture because only the proteins that have orthologs in all for organisms are included in Figure 1, because conserved timing is required across all organisms in which a dynamic ortholog could be found, and because conservation is assessed at the level of orthologous groups rather than individual pairs of orthologous proteins.

These caveats are all addressed by the much more detailed Figure S7. This visualization also avoids the use of an arbitrary cutoff on when the timing is considered “conserved” and shows when in the cell cycle the genes are maximally expressed. The pairwise comparison of *S. pombe* and *A. thaliana* reveals that many genes are expressed during the same phase of the cell cycle in these two organisms. Conversely, large differ-

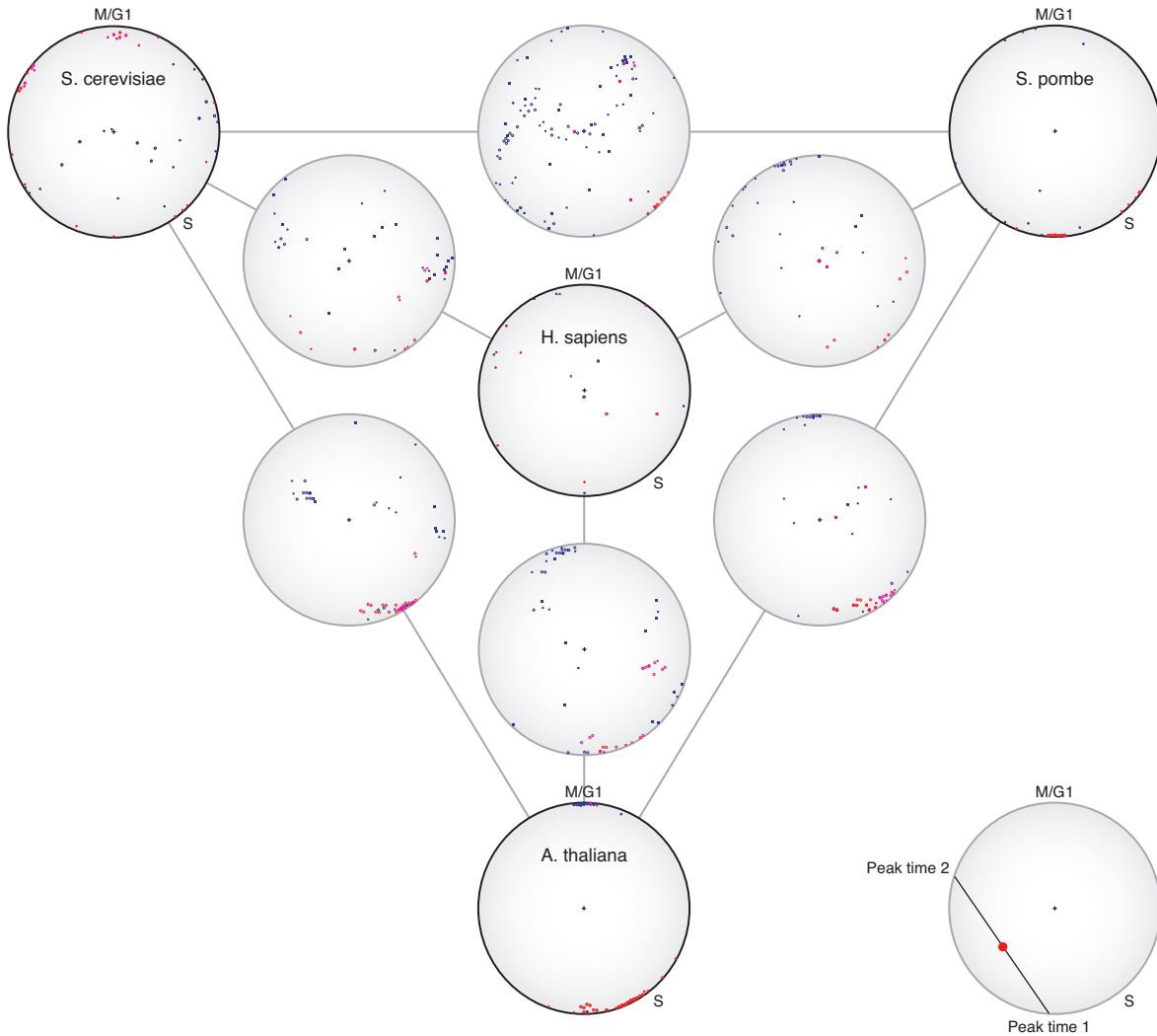


Figure S7: **Comparison of the timing of expression of orthologous and paralogous genes.** The large black circles compare the timing of expression for paralogous gene pairs within each of the four organisms, whereas the large gray circles show the results when comparing orthologous gene pairs from the two organisms specified by the gray lines. Each pair of periodically expressed orthologs or paralogs is represented by a symbol, which is placed at the middle of the straight line that connects the warped peak times of the two genes. For the orthologous gene pairs, the shape of the symbol specifies if the genes are 1:1 orthologs (filled squares), orthologs which have only one periodic gene from each organism (filled circles), or orthologs with multiple periodic paralogs in one or both organisms (open circles). Paralogous gene pairs are similarly classified based on whether there are two (filled circles) or more (open circles) periodically expressed paralogs in the organism in question. For each orthologous or paralogous pair, the color of the symbol signifies whether the protein sequence conservation is high (bits per residue > 1.4 , red), medium (bits per residue > 1.2 , magenta), or low (bits per residue ≤ 1.2 , blue). Many of these details are only visible if zooming.

ences in the timing of expression is observed for all other pairs of organisms. Consistent with these observations, paralogous genes typically have similar peak times in fission yeast or *A. thaliana* but not in budding yeast or human. This supports and extends the results by Maslov et al.⁶¹, who used protein–DNA interaction data (ChIP–chip) to show that the transcriptional regulation of paralogous genes diverges rapidly in budding yeast.

Phosphorylation

We show in the present study that the extensive changes in the transcriptional regulation of genes involved in the eukaryotic cell cycle are often mirrored by corresponding changes in protein phosphorylation by cyclin-dependent kinases. This correlation was shown to be statistically significant for several sets of known and predicted phosphoproteins in several species (Figure 3).

Table S3 elaborates on the results shown in Figure 3 by showing the counts underlying each statistical test, by including results for *A. thaliana*, and by adding additional sets of phosphoproteins. *A. thaliana* was excluded from Figure 3 because our benchmarks have revealed that the set of periodically expressed genes is of lower quality than for the other three organisms (Figure S1) and may be biased towards particular phases of the cell cycle (Figures S4 and S7). Despite these caveats, the tests show that CDK substrates are indeed overrepresented among dynamic proteins in *A. thaliana* as well as in the three other organisms, although we were not able to prove that transcriptional and posttranslational regulation has co-evolved in the *A. thaliana* lineage (Table S3). It should be noted that all but one of the correlations that are statistically insignificant involve either *A. thaliana*, for which we have very limited expression data, or sequence-based predictions of phosphorylation sites, which are inherently less reliable than experimental evidence. Conversely, the best P-values are obtained when analysing sets of experimentally determined CDK substrates.

Additional sets of known and predicted phosphoproteins are also included in Table S3. From the protein chip study by Ptacek et al.⁶², we extracted a set of 60 Cdc28 substrates. Reassuringly, this dataset confirms most of the correlations observed based on the datasets by Übersax et al.⁵¹ and Loog et al.⁵², although the statistical support is far worse due to the much smaller size of this dataset. To not base the *S. pombe* and *A. thaliana* correlation analyses on a single prediction method, tests were performed based on predictions by Scansite cdc2 as well as by NetPhosK cdk5. Scansite cdc2 confirms most of the correlations observed with NetPhosK cdk5, although it generally gives worse P-values as should be expected given the relative predictive performance of the two methods (Figure S5).

As noted in the main paper, the co-evolution test is much stricter and consequently leads to much lower counts—in some cases less than five proteins in a category (Table S3). Nevertheless, the fraction of phosphoproteins among the dynamic versus the static proteins is comparable between the first and the second test, meaning that the less significant p-values in the co-evolution mainly reflect the low counts. Each type of phosphorylation data used has its own strengths and weaknesses: low-throughput studies are highly accurate but are prone to study biases, high-throughput studies may suffer from biases towards highly expressed proteins, whereas sequence-based predictions are free of biases but have much lower accuracy. Importantly, all of these data types support the discovery that transcriptional regulation and CDK-dependent phosphorylation have co-evolved.

Targeted protein degradation

Transcriptional regulation can only control when a protein is synthesized; to fully control when in the cell cycle a protein is present, it must also be degraded at the appropriate timepoint. Protein degradation during the cell cycle is performed by the proteasome, which specifically targets proteins that have been polyubiquitinated by either the anaphase promoting complex/cyclosome (APC/C) or the Skp1/Cullin/F-box protein complex (SCF). The former complex targets proteins with D box and KEN box motifs^{37;59}, whereas the latter targets phosphorylated PEST regions^{55–58}.

Since many proteins are likely tightly regulated during the mitotic cell cycle, one would expect the ubiquitin ligases described to preferentially target the dynamic proteins in each organism. To test this hypothesis, we checked if PEST regions, D boxes, or KEN boxes are significantly overrepresented among dynamic proteins compared to static ones (Table S4). These correlation analyses reveal that PEST regions are indeed significantly overrepresented among the dynamic proteins in human, budding yeast, and fission yeast; this observation has earlier been made for budding yeast only¹⁹. Similarly, many more of the dynamic proteins from human and *A. thaliana* contain KEN boxes than would be expected by chance. In case of D-boxes, only weak correlations are observed, which is not surprising since the D-box motif is less well defined than the two other recognition sequences. Our correlation analysis thus support the hypothesis that ubiquitin ligases preferentially target dynamic proteins, which are thus subject to targeted protein degradation.

We went on to test if degradation motifs have co-evolved with periodic expression, however, we were not able to show that this is the case. We see two possible reasons for this: either co-evolution has simply not taken place or we are unable to detect it due to the limited predictive power of

Table S3: **Cyclin-dependent kinases preferentially phosphorylate dynamic proteins.** The tables provide additional details for the tests presented in Figure 3. Moreover, we here include results for *A. thaliana* as well as additional phosphorylation evidence. For each individual test, the number of phosphorylated proteins and the total number of proteins is listed for both the set dynamic and static proteins. Fisher’s exact test was used to calculate P-values, which are shown if only if they are below 20%. Several sets of phosphoproteins were used: known CDK substrates from low-throughput experiments (LTP CDK⁴⁵), S/T phosphoproteins from high-throughput mass-spectrometry studies (HTP S/T⁴⁵), CDK substrates from systematic screens (HTP CDK^{51;52}), *in vitro* CDK substrates identified by a protein chip study (Chip CDK⁶²), CDK phosphoproteins predicted by two computational methods (NetPhosK⁵³ and Scansite⁵⁴).

Organism	Data source		Dynamic proteins	Static proteins	P-value
<i>H. sapiens</i>	LTP CDK		16/600	40/11497	$7 \cdot 10^{-9}$
	HTP S/T		55/600	351/11497	$4 \cdot 10^{-16}$
	NetPhosK		278/600	4642/11497	$2 \cdot 10^{-3}$
	Scansite		276/600	4489/11497	$4 \cdot 10^{-4}$
<i>S. cerevisiae</i>	HTP CDK		116/600	224/5637	$2 \cdot 10^{-55}$
	Chip CDK		11/600	47/5637	0.02
	NetPhosK		189/600	1231/5637	$6 \cdot 10^{-8}$
	Scansite		186/600	1320/5637	$3 \cdot 10^{-5}$
<i>S. pombe</i>	NetPhosK		185/500	1060/4498	$4 \cdot 10^{-11}$
	Scansite		167/500	1119/4498	$3 \cdot 10^{-5}$
<i>A. thaliana</i>	NetPhosK		129/400	5724/20610	0.03
	Scansite		126/400	5605/20610	0.04

Organism 1	Data source	Organism 2	Dynamic proteins with static orthologs	Static proteins with dynamic orthologs	P-value
<i>H. sapiens</i>	LTP CDK	<i>S. cerevisiae</i>	5/136	0/292	$4 \cdot 10^{-3}$
		<i>S. pombe</i>	5/157	1/355	0.02
		<i>A. thaliana</i>	5/189	0/144	0.06
	HTP S/T	<i>S. cerevisiae</i>	12/136	13/292	0.07
		<i>S. pombe</i>	20/157	15/355	$7 \cdot 10^{-4}$
		<i>A. thaliana</i>	22/189	3/144	$6 \cdot 10^{-4}$
	NetPhosK	<i>S. cerevisiae</i>	45/136	81/292	0.2
		<i>S. pombe</i>	49/157	125/355	—
		<i>A. thaliana</i>	68/189	40/144	0.08
	Scansite	<i>S. cerevisiae</i>	50/136	87/292	0.1
		<i>S. pombe</i>	54/157	129/355	—
		<i>A. thaliana</i>	79/189	39/144	$4 \cdot 10^{-3}$
<i>S. cerevisiae</i>	HTP CDK	<i>H. sapiens</i>	32/171	10/147	$2 \cdot 10^{-3}$
		<i>S. pombe</i>	36/197	22/246	$4 \cdot 10^{-3}$
		<i>A. thaliana</i>	29/186	3/70	$9 \cdot 10^{-3}$
	Chip CDK	<i>H. sapiens</i>	2/171	1/147	—
		<i>S. pombe</i>	5/197	2/246	0.2
		<i>A. thaliana</i>	4/186	2/70	—
	NetPhosK	<i>H. sapiens</i>	47/171	35/147	—
		<i>S. pombe</i>	54/197	71/246	—
		<i>A. thaliana</i>	42/186	10/70	0.1
Scansite	<i>H. sapiens</i>	50/171	37/147	—	
	<i>S. pombe</i>	56/197	72/246	—	
	<i>A. thaliana</i>	45/186	15/70	—	
<i>S. pombe</i>	NetPhosK	<i>H. sapiens</i>	66/205	40/162	0.08
		<i>S. cerevisiae</i>	76/214	71/205	0.02
		<i>A. thaliana</i>	67/205	16/62	0.2
	Scansite	<i>H. sapiens</i>	56/205	48/162	—
		<i>S. cerevisiae</i>	68/214	58/205	—
		<i>A. thaliana</i>	64/205	15/62	0.2
<i>A. thaliana</i>	NetPhosK	<i>H. sapiens</i>	15/71	102/358	—
		<i>S. cerevisiae</i>	13/56	94/433	—
		<i>S. pombe</i>	15/59	138/454	—
	Scansite	<i>H. sapiens</i>	18/71	117/358	—
		<i>S. cerevisiae</i>	12/56	99/433	—
		<i>S. pombe</i>	15/59	129/454	—

Table S4: **Dynamic proteins are preferentially ubiquitinated and targeted for degradation.** Using the same statistical test shown in Figure 3A, dynamic proteins were shown to be enriched for sequence motifs recognized by the APC/C and SCF ubiquitin ligases. Fisher’s exact test was used to calculate P-values, which are shown if only if they are below 20%.

Organism	Sequence motif	Dynamic proteins	Static proteins	P-value
<i>H. sapiens</i>	PEST region	248/600	4355/11497	$2 \cdot 10^{-3}$
	D box	75/600	1244/11497	0.04
	KEN box	136/600	1423/11497	$7 \cdot 10^{-18}$
<i>S. cerevisiae</i>	PEST region	220/600	1579/5637	$2 \cdot 10^{-11}$
	D box	72/600	682/5637	0.2
	KEN box	89/600	859/5637	—
<i>S. pombe</i>	PEST region	154/500	1221/4498	$3 \cdot 10^{-4}$
	D box	60/500	532/4498	0.2
	KEN box	56/500	544/4498	—
<i>A. thaliana</i>	PEST region	91/400	5082/20610	—
	D box	46/400	1892/20610	0.05
	KEN box	58/400	1640/20610	$4 \cdot 10^{-7}$

the sequence motifs. The latter may very well be the explanation, considering that the co-evolution of phosphorylation and transcriptional regulation could in several cases also not be detected through correlation analysis of sequence-based predictions of phosphorylation sites. Until sufficient experimental evidence on ubiquitination has been accumulated, it is thus difficult to say if the sequence motifs governing transcriptional regulation and targeted protein degradation have in fact co-evolved.

Evolution of protein regulation

The four eukaryotes for which cell cycle microarray expression data are available are very distantly related: animals, fungi, and plants are all believed to have diverged more than 1 billion years ago⁶³. While the budding and fission yeast are more closely related, the two yeasts are as different from each other as either is from animals, presumably due to a higher evolutionary rate among fungi⁶³. As our analysis reveals, the detailed regulation of cell cycle proteins has changed dramatically over these long time spans.

To get a more fine-grained view of how rapidly protein regulation changes during the course of evolution, we studied two known CDK phosphorylation sites in more detail, namely serine-20 in the RRM2 subunit of human ribonucleotide reductase and serine-144 in the DPB2 of budding yeast DNA polymerase ϵ . Both human RRM2 and budding yeast DPB2 are periodically transcribed according to our analysis, and their protein products are known from small-scale experiments to be phosphorylated by CDK1 on a particular residue^{64,65}. Moreover, their orthologs in the three other organisms studied are static, not known to be phosphorylated, and the CDK phosphorylation sites are absent as shown by sequence alignments. Human RRM2 and budding yeast DPB2 are the only

two examples that fulfill all of these requirements. The most parsimonious explanation is thus that the transcriptional and posttranslational cell cycle regulation of RRM2 and DPB2 have evolved in the human and budding yeast lineages, respectively. To pinpoint when during evolution each of these phosphorylation sites arose, we aligned the full-length sequences of orthologous proteins from organisms at increasing evolutionary distances from human and budding yeast, respectively:

RRM2:

<i>Homo sapiens</i>	LQLSPLKG
<i>Pan troglodytes</i>	LQLSPLKG
<i>Macaca mulatta</i>	LQLSPLKG
<i>Bos taurus</i>	QQLSPLKG
<i>Canis familiaris</i>	LQLSPLKG
<i>Mus musculus</i>	LQLSPLKR
<i>Rattus norvegicus</i>	LHLSPLKR
<i>Monodelphis domestica</i>	LRLSPLKS
<i>Gallus gallus</i>	PRLSPLKN
<i>Danio rerio</i>	TISTKMNN
<i>Takifugu rubripes</i>	TLISDVTK
<i>Tetraodon nigroviridis</i>	TLVSDVRQ

DPB2:

<i>Saccharomyces cerevisiae</i>	SLSSPMRQ
<i>Candida glabrata</i>	QISSPARD
<i>Kluyveromyces lactis</i>	DSHSPMML
<i>Lachancea kluyveri</i>	VSISPVQD
<i>Eremothecium gossypii</i>	TATEPMDE
<i>Candida albicans</i>	SFEDKNME
<i>Schizosaccharomyces pombe</i>	SFSGSFSQ

For a CDK phosphorylation site to be conserved, the two highlighted residues in the alignment must be either SP or TP. From the RRM2 alignment, we thus conclude that CDK1 phosphorylation of RRM2 is conserved at least within *amniotes*, which encompasses mammals, reptiles, and birds. We further speculate that the transcriptional cell cycle regulation of RRM2 is also

conserved within this group of animals, which diverged 288–338 million years ago⁶⁶. Similarly, we conclude that the CDK1 phosphorylation site in DPB2 was present in the last common ancestor of *Saccharomycetaceae*, from which *Saccharomyces cerevisiae*, *Candida glabrata*, *Kluyveromyces lactis*, and *Lachancea kluyveri* diverged about 100 million years⁶⁷.

Based on these two examples, we tentatively suggest that evolutionary changes in the regulation of proteins can take place on a time scale in the order of only a hundred million years. This has major implications for the use of model organisms to unravel complex regulatory networks in humans, since even vertebrates such as zebrafish are too distantly related to humans for regulatory mechanisms to be inferred.

Details on individual protein complexes

In this section, we review the literature on individual complexes involved in DNA replication, sister chromatid cohesion, and mitotic protein degradation (by the APC/C). As most of the complexes are visualized in Figure 2 of the main paper, we encourage you to look at this figure while reading. The section focuses on the evolutionary differences in regulation, known or unknown, which are revealed by our analysis. It documents that the regulation does indeed happen through different components in each organism, and in the few cases where no subunits of a complex are dynamic, the transcriptional regulation acts through other complexes involved in the same process.

DNA Replication

In eukaryotes, DNA replication is a tightly controlled process that involves a large number of protein complexes and regulators that assemble on the DNA in a sequential manner. Many of these components are conserved through evolution, however, their regulation differs between organisms.

What is first noticed when looking at the left half of Figure 2, is that although only a subset of the subunits are dynamic in human, these all peak in expression just prior to S phase. The only major exception is MCM4, which surprisingly peaks during M phase. Budding yeast displays a pattern of expression similar to that of human cells, except that the pre-replication complex (Cdc6 and Mcm2–7) is expressed already in late M phase (Figure 2). In fission yeast, the entire machinery is expressed in late M phase, which is consistent with the finding that rapidly growing *S. pombe* cells are capable of initiating DNA replication before having completed cytokinesis^{68;69}. This phenomenon is likely caused by the very short short G₁ phase in this organism, which creates a need to start the synthesis earlier in order to have the components ready for S phase.

Pre-replication complex (pre-RC)

A large protein assembly known as the pre-replication complex (pre-RC) is required to recruit the replication machinery to specific locations on the DNA where they assemble in a step-wise fashion and initiate replication. Although the order in which the subcomplexes are assembled appears to be similar in all model organisms, the regulation differs between organisms⁶⁹. The pre-RC is thus an example of a very large protein complex that

is conserved throughout Eukarya but has changed a lot at the regulatory level.

Origin of replication complex (ORC). As the name suggests, the ORC is six-subunit protein complex which assembles at specific locations on the DNA, namely the *origins of replication*^{36;42}. The ORC complexes thus select the sites from which replication is later initiated. ORC can bind and hydrolyze ATP; a function important for binding to both DNA and other factors³⁶.

In *S. cerevisiae* and *S. pombe*, the ORC complex appears to be constitutively associated with the origins of replication throughout the entire cell cycle³⁶. In accordance with this, none of the ORC subunits are transcribed in a periodic fashion in the two yeasts (see Figure 2). In budding yeast, it is known that Orc2 and Orc6 are phosphorylated by CDK and that this is one mechanism which prevents reinitiation of DNA replication^{69;70}. Fission yeast Orc2 is phosphorylated by Cdc2 (Cdk1)³⁶, as is Orc1 in budding yeast. The ORC complex in the two yeasts is thus one of the exceptions from the general trend that phosphorylation preferentially affects the dynamic proteins.

In mammals, the composition of ORC seems to change during the cell cycle, mainly through Orc1, which is phosphorylated by CDK and believed to be targeted for ubiquitin-mediated degradation^{36;38}. The degradation is controversial⁶⁹, but in our analysis we find Orc1 (from human HeLa cells) to be dynamic with peak expression in G₁/S phase (Figure 2). The re-synthesis thus supports a model in which Orc1 is “the limiting factor” that controls pre-RC assembly⁷¹, whereas the other subunits seem to be stably associated with the origins⁷². The only other dynamic subunit in human is Orc3, which also peaks in G₁/S phase (Figure 2). This does not seem to have been noticed before and it is tempting to speculate that the regulation of ORC3 could reflect a special role for this subunit in humans. In many cases, we have observed that such periodic transcription is linked to posttranslational regulation the protein product, although we have not been able to find any information on CDK phosphorylation or degradation of human Orc3.

In summary, the mapping in Figure 2 reflects what is known from detailed biochemical studies: the ORC complex in the two yeasts is constitutively associated with the origins (they are static in Figure 2), whereas the human ORC composition changes (it has 2 dynamic subunits, which peak just prior to S phase).

Cdc6. Although the precise molecular mechanism is presently not clear, the protein Cdc6 associates with the ORC complex once it is assem-

bled on DNA and is required for subsequent loading of the Mcm2–7 complex^{36;42;73;74}. In the two yeasts, Cdc6/Cdc18 is phosphorylated by CDK, which marks it for ubiquitin-mediated degradation^{36;38;75}. In human cells, the phosphorylation instead leads to nuclear export; only later is Cdc6 degraded by APC/C³⁶. In fact, CDK phosphorylation protects human Cdc6 from APC/C-dependent degradation⁷⁶.

The mapping of dynamics in Figure 2 fits perfectly with the patterns of degradation and the timing of Cdc6 association with the ORC and MCM complexes. In human, CDC6 is co-expressed with the dynamic subunits MCM2, MCM3, MCM5, and MCM6 just prior to S phase, in agreement with other evidence on its re-synthesis⁷². Similarly, *S. cerevisiae* CDC6 is co-expressed with all six subunits of MCM (see below), albeit the expression occurs at the M/G₁ transition. Similar to in human, the CDC6 ortholog in *S. pombe*, cdc18, peaks at the same time as the rest of the replication machinery; however, the expression of the entire machinery is shifted to late M phase. Also, the Mcm2–7 complex is not subject to transcriptional regulation in fission yeast (Figure 2), which suggests that cdc18 and cdt1 (see below) may control the assembly of the pre-replication complex. Consistent with this view, overexpression of Cdc6/Cdc18 has been shown to induce re-replication in fission yeast, but not in budding yeast⁶⁹. This illustrates that our analysis of transcriptional regulation and phosphorylation can help to pinpoint the components in each organism responsible for regulating complex assembly.

Cdt1. The implication of Cdt1 in pre-replication complex assembly was first discovered in *S. pombe*⁷⁷, where the protein level peaks in G₁ and decays in S phase³⁶. The gene expression of Cdt1 follows that of the rest of the replication machinery in our analysis (Figure 2) and the seemingly conflicting timing therefore probably just reflect that the rapidly growing fission yeast cells shift expression from G₁/S to M phase. Also, the co-expression of cdc18 and cdt1 (Figure 2) is consistent with the finding that they are both targets of the cdc10 transcription factor responsible for M/G₁ transcription⁶⁹.

The regulation of human Cdt1 is presently not well understood and only recently was a functional homolog, Tah11, identified in budding yeast³⁶. Whereas *S. pombe* Cdt1 is subject to transcriptional control, the regulation of human CDT1 appears to be through proteolysis, CDK phosphorylation and ubiquitination by SCF^{Skp2} in S phase^{78–81}. The orthologous Tah11 protein in *S. cerevisiae* does not seem to be controlled neither transcriptionally nor through proteolysis³⁸. Its protein level is constant throughout the cell cycle. Similar to the Mcm2–7 complex, however, it is localized to the nucleus in G₁ and to cyto-

plasm later on^{36;69}.

What is described above agrees well with what we observe in our analysis: human CDT1 contains a PEST degradation signal, *S. pombe* cdt1 is regulated at the mRNA level, and *S. cerevisiae* Tah11 is neither periodically expressed nor does it contain a PEST degradation signal.

An additional layer of Cdt1 control is present in metazoans through the inhibitor geminin, which has no ortholog in yeasts³⁸. The binding of geminin is believed to stabilize Cdt1 by protecting it from ubiquitin-mediated proteolysis⁷². The protein level of geminin increases during S phase and accumulates until late M phase when geminin is degraded via APC/C³⁶. The net effect is that most Cdt1 is kept in a protected, but inactive, state from S phase until M phase⁷² (Figure 2). Consistent with this, both overexpression of the Cdt1 protein itself and loss of geminin lead to re-replication⁷². Metazoans thus have a special mechanism that prevents unwanted assembly of pre-replication complexes through tight regulation of Cdt1.

Mcm2–7 complex. The final step in the assembly of the pre-replication complex is the loading of the hexameric, ring-shaped Mcm2–7 complex. The loading is mediated by Cdc6 and Cdt1, although the detailed mechanism is unknown³⁶. The composition the complex appears to be conserved across all eukaryotes, however, the precise function of MCM is unclear, although it has been to act as a helicase^{36;42}.

In budding yeast, Mcm2–7 enters the nucleus at the end of mitosis and remains nuclear in G₁ and S phase until it is exported to the cytoplasm in G₂ and M^{36;69;70}. This mechanism fits perfectly with the observed expression of the MCM2–7 genes during M phase (Figure 2). It should be noted that although MCM6 falls just below the threshold for inclusion as a periodic gene, its peak time is similar to that of the other MCM genes, indicating that it too should be considered dynamic; to be stringent, however, we display it as static in Figure 2. The nuclear import appears to require the entire six-subunit complex³⁶ and to be independent of Cdc6⁶⁹, which indicates that the localization of the Mcm2–7 complex is independent of its loading onto DNA. The export is promoted by CDK activity, although it is presently unclear whether this is directly caused by phosphorylation³⁶. This is in our view likely to be the case, considering that four out of six subunits of the complex are phosphorylated by Cdc28.

In fission yeast, the evidence on Mcm2–7 complex regulation is less clear. Although the complex is presumably able to shuttle between the cytoplasm and the nucleus (mediated by the Crm1 nuclear export factor), the bulk of the MCM proteins appear to be constitutively nuclear³⁶. These findings are consistent with our analysis in which none of the six subunits in *S. pombe* are transcrip-

tionally regulated.

In mammals, the Mcm proteins are located in the nucleus throughout the cell cycle, however, their association with chromatin becomes progressively weaker as cells proceed through S phase³⁶. The literature is vague on the issue of transcriptional regulation of the complex. We find that although MCM7 is static, the majority of the subunits of the human Mcm2–7 complex (MCM2, MCM3, MCM5, and MCM6) are dynamic with peak times similar to those of the other pre-RC components. The regulation in human thus differs from both budding yeast (where all six components are dynamic) and fission yeast (where all subunits are static). In humans, MCM4 is also dynamic, however, its expression in late M phase is distinct from the rest of the complex. This timing of expression is highly surprising and suggests that Mcm4 has a central role in controlling the assembly of the human pre-RC, which is indeed known to assemble during telophase (late M phase)^{82;83}. In accordance with this, recent data have shown that Mcm4 is phosphorylated at multiple sites, most likely by Cdk2, and that the phosphorylation inhibits the association of Mcm4 with chromatin^{84–86}. DNA replication in humans seems to be dictated by the phosphorylation level of Mcm4, which appears to be the rate-limiting step in the loading of the other Mcm subunits onto chromatin^{84–86}. This model is in perfect agreement with our results, and again demonstrate that the timing of expression is closely linked to the time of action. To the best of our knowledge the unique pattern of expression of the human MCM genes, and in particular of MCM4, has not been reported by others.

Dbf4-dependent kinase (DDK)

Once the pre-replication complex (ORC, Cdt1, Cdc6 and MCM) has assembled, recruitment of the rest of the replisome and initiation of DNA replication is dependent on the action on a number of factors known as the pre-initiation complex (see below), as well as on the action of CDKs and another kinase called the Dbf4-dependent kinase (DDK). Like CDKs, DDK consists of a regulatory subunit (Dbf4/Dfp1), which regulates the activity of the kinase (Cdc7/Hsk1). The regulatory subunit in *S. pombe*, dfp1, peaks in late M phase like the rest of the replication machinery. The human ortholog, DBF4, was not measured in the experiments used in our analysis, and *S. cerevisiae* DBF4 appears to be static in our analysis, although it has been found in other analysis to be periodic with peak expression in late G₁, both at the transcriptional level^{87;88} and at the protein level⁸⁹. The evidence is thus conflicting, although the fluctuations at the protein level could be caused by SCF-mediated proteolysis after S phase, since Dbf4 is a Cdc28 substrate and also contains a PEST region.

The kinase subunit CDC7 is static in human and budding yeast, but dynamic in fission yeast with peak expression similar to dfp1, cdt1, cdc18, and the rest of the replication machinery (Figure 2). The DDK complex is thus clearly regulated in fission yeast through both subunits and possibly regulated in budding yeast through DBF4, whereas the regulation of the human complex remains unknown due to the lack of experimental data.

Pre-initiation complex (pre-IC)

Apart from the activities of CDK and DDK, Initiation of DNA replication is dependent on a number of factors collectively known as the pre-initiation complex (pre-IC). Relatively little is known about the order of assembly of these factors and their individual functions. The majority of these components were omitted in Figure 2 in order to restrict ourselves to well characterized complexes with independent evidence from all species. Most of the evidence on pre-IC stems from yeast and the components currently believed to be part of the initiation are shown in Figure S8.

Dpb11–Sld2. This complex binds DNA polymerases ϵ and α ⁴². Formation of the complex and initiation of DNA replication requires CDK phosphorylation of Sld2 in budding yeast^{52;90}. Both proteins exist in fission yeast, but SLD2 appears to have no ortholog in human. In human both subunits are dynamic with peak expression in late G₁/S, whereas only SLD2 is dynamic in budding yeast with maximal expression during G₁, and both subunits are static in fission yeast (Figure S8).

Mcm10. Mcm10 seems to play a role in stabilizing DNA polymerase α and targeting it to chromatin, at least in *S. cerevisiae*^{44;91}. Mcm10 has also been shown to interact with Cdc45 and has been considered a component of the pre-IC^{44;91}. Although phosphorylated by CDK in *S. cerevisiae*, Mcm10 is static in both yeasts and has not been measured in the human experiments (Figure S8).

Cdc45–Sld3. The best characterized component of the pre-IC, Cdc45, is required for loading of one or more of the polymerases prior to DNA unwinding. Evidence from budding yeast shows that Cdc45 throughout the cell cycle forms a complex with Sld3⁹², which according to our analysis is static in both yeasts (Figure S8); no human ortholog of Sld3 has been identified. Cdc45 and Sld3 are released from chromatin in late S phase like the MCM proteins³⁶, but only Cdc45 is incorporated in the replication fork^{93;94}. Consistent with this, we detect CDC45 to be maximally expressed in G₁/S phase in human as well as budding yeast. Interestingly, in fission yeast, where both subunits of DDK are dynamic, the cdc45 gene is static. Data from *S. cerevisiae* suggest that the protein level of Cdc45 is constant through the cell cycle⁹⁵.

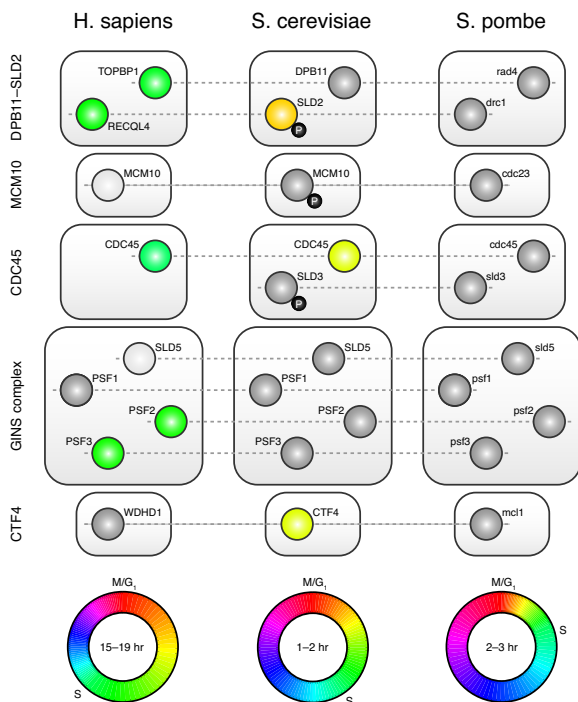


Figure S8: **The pre-initiation complex (pre-IC).** A number of largely uncharacterized proteins and complexes known to be important for recruitment of the replication machinery and initiation of DNA replication. Only the relatively well characterized Cdc45/Sld3 complex was included in Figure 2 of the main paper.

This seemingly conflicting observation can be explained by simultaneous degradation of the modified form and synthesis of new, unmodified Cdc45, or alternatively by excess of Cdc45, which could cause only minor fluctuations in the protein concentration that may have gone undetected.

GINS complex. The GINS (go ichi ni san) complex consists of four subunits (Sld5, Psf1, Psf2, and Psf3). GINS is required to maintain association of Cdc45 with moving replication forks^{93;94} and also appears to interact with CMCs, and DNA polymerase ϵ ⁴², but its role is presently not clear. Two of the subunits of the GINS complex (Psf2 and Psf3) are dynamic with peak expression right before S phase, whereas the entire complex is static in both yeasts (Figure S8). A recent protein interaction screen in budding yeast identified Ctf4 as a possibly fifth subunit of the GINS complex⁹⁶; this interaction has also been identified in a targeted screen for GINS interaction partners⁹³). CTF4 is periodically expressed with a peak time in late G₁ and has static orthologs in human as well as fission yeast (Figure S8). Its protein product is known to interact with DNA polymerase α and to be involved in both DNA replication and sister chromatid cohesion^{93;97;98}.

In summary, none of the pre-IC components are regulated in fission yeast and only Cdc45 is periodic in budding yeast. In contrast, all complexes and factors except Mcm10 are influenced by transcriptional control in human, where they all peak at the G₁/S transition.

DNA replication factor A (RFA)

For DNA replication to take place, the two strands of DNA must be separated at the replication origins and at the replication forks. RFA (also known as replication protein A, RPA) binds to the resulting single-stranded DNA (ssDNA) in order to protect it and possibly also to prevent it from forming secondary structure elements^{36;99}. In addition to these generic ssDNA-binding functions, RFA is required for loading of DNA polymerase α ³⁶ and hence for assembly of the replication fork. The RFA complex is a prime example of the evolution of transcriptional regulation: All three subunits are regulated in *S. cerevisiae*, whereas only Ssb1 is regulated in *S. pombe* and only Rpa2 is regulated in human (Figure 2). In all three organisms, the dynamic subunits of RPA are synthesized right before S phase.

DNA polymerase ϵ

Although it is clear that DNA polymerase ϵ is part of the replisomal particle, there is conflicting genetic evidence as to whether it is directly involved in DNA replication, serves as a checkpoint protein, or simply holds together other subcomplexes⁴². The complex is not transcriptionally regulated in humans, but controlled by this mechanism in both yeasts (Figure 2). Fission yeast only regulates *cdc20*, whereas budding yeast regulates both POL2 (the ortholog of *cdc20*) and DPB2. In addition to the two dynamic subunits identified in our analysis (Figure 2), budding yeast DPB3 has been found to be periodically expressed in a small-scale study¹⁰⁰. The *S. cerevisiae* DPB2 subunit is periodically expressed and has also been shown to be phosphorylated by Cdc28 in both large-scale^{51;52} and small-scale studies⁶⁵ (Figure 2). In contrast, the orthologs of DPB2 are static, and alignments reveal that the Cdc28 phosphorylation site, which has been mapped to serine-144⁶⁵, is absent in the orthologous proteins. DPB2 thus exemplifies that periodic expression and CDK phosphorylation have co-evolved.

DNA polymerase α

Although originally believed to be the primary polymerase responsible for DNA replication in eukaryotes, it is today clear that DNA polymerase α only serves to synthesize primers that are subsequently extended by DNA polymerase δ ^{36;42}. The complex consists of four subunits, of which at least one is dynamic in each of the three organisms shown in Figure 2; however, there is not a single subunit that is periodically expressed in all three organisms. The dynamic subunits are maximally expressed at the time when DNA replication is initiated in the each organism. DNA polymerase α is a good example of how just-in-time assembly can be achieved in many different ways and thus

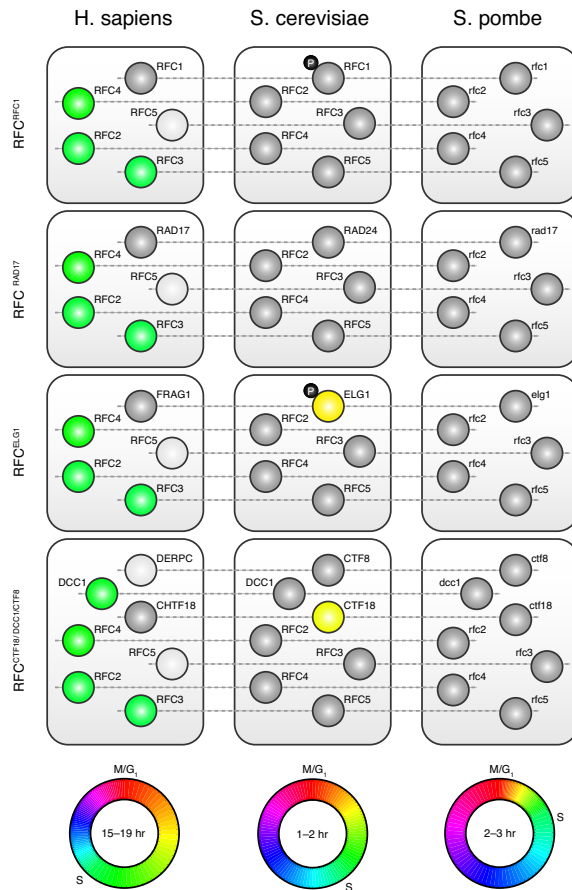


Figure S9: **DNA replication factor C (RFC) complexes.** Four variations of the RFC complex are visualized as described in Figure 2. The RFC1 subunit of the standard RFC have been replaced in the alternative complexes by RAD24/rad17, ELG1, and the CTF18–DCC1–CTF8 trimer, respectively. Notably, human DCC1 as well as *S. cerevisiae* ELG1 and CTF18 are periodically expressed with peak times similar to other DNA replication genes.

allows for great evolutionary flexibility in the detailed regulation.

DNA replication factor C (RFC)

The RFC protein complex is a so-called clamp loader, which is responsible for loading the PCNA clamp onto the DNA⁴². This protein complex consists of five subunits, RFC1–5, all of which are static in the two yeasts, whereas RFC2–4 (and possibly also RFC5) are dynamic in human. In addition to this clamp loader complex, at least three other forms of RFC exist (Figure S9), each of which consists of the RFC2–5 subunit and an alternative RFC1 subunit, which is either RAD24/rad17, ELG1, or CTF18. The latter form of RFC also contains two extra subunits, namely DCC1 and CTF8, and is known to be important for sister chromatid cohesion^{98;101}. However, the function of these alternative RFC complexes is still largely unknown, although they all appear to be involved in DNA replication and/or repair¹⁰². Human DCC1 is dynamic with a peak time similar to Rfc2-4. In budding yeast, ELG1 and CTF18 are periodically expressed, peaking in late G₁ like

many other genes involved in DNA replication. This is a strong indication that these alternative forms of RFC may play a role in the mitotic cell cycle of *S. cerevisiae*. Elg1 is particularly interesting, because this protein is both dynamic with expression just prior to DNA replication and phosphorylated by CDK in budding yeast (Figure S9).

Proliferating cell nuclear antigen (PCNA)

The PCNA protein forms a homo-trimer, which encircles a single strand of DNA and serves as a *sliding clamp*, which holds DNA polymerase δ in place and enables it to replicate long stretches of DNA⁴². PCNA is periodically expressed in budding yeast and human, being maximally expressed at the same time as the dynamic subunits of the DNA polymerases (Figure 2). In fission yeast, neither the clamp (PCNA) nor the clamp loader (RFC) are controlled transcriptionally.

DNA polymerase δ

This protein complex is the primary polymerase responsible for the DNA replication in eukaryotes^{36;42}. Although DNA polymerase δ consists of four subunits in most organisms, the *S. cerevisiae* version of the complex has only three subunits⁴². DNA polymerase δ can only assemble once the PCNA clamp has been loaded onto the DNA^{36;42}, and this crucial subcomplex is thus one of the last to join the replisomal particle³⁶. The clamp is believed to be positioned behind the polymerase⁴². As is the case for DNA polymerase α , the complex contains at least one dynamic subunit in each organism, yet no subunit is dynamic in all three organisms (Figure 2).

Ribonucleotide-diphosphate reductase (RNR)

The ribonucleotide-diphosphate reductase (RNR) complex is a tetramer consisting of two large and two small subunits. It is important for maintaining the proper dNTP pool for DNA replication and repair¹⁰³, and the RNR activity is maximal in S phase in most, if not all, organisms¹⁰⁴. Our results suggest that this regulation is primarily mediated through the large subunits, which are periodically expressed in all the four organisms that we have analyzed. The large subunits of RNR thus constitute one of the just five orthologous groups in Figure 1 for which periodic expression is fully conserved. Although periodicity as such is conserved, the time of peak expression differs greatly between organisms, as is evident from Figures 2 and S10. Notably, the *S. pombe* *cdc22* gene is strictly speaking not expressed in S phase but at the M/G₁ transition; this makes sense since it implies that the RNR activity is maximal during DNA replication.

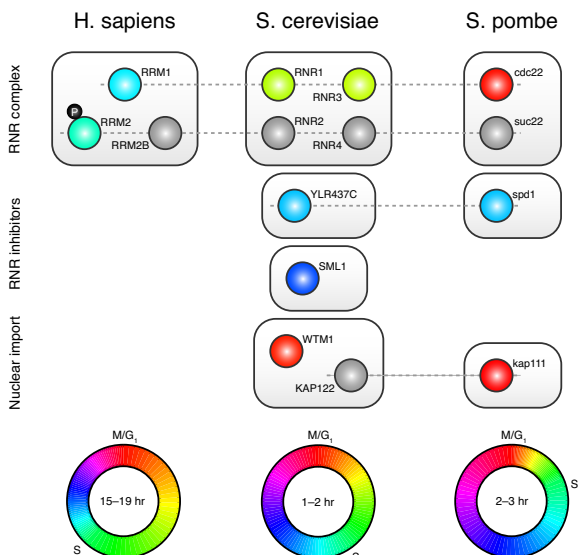


Figure S10: **Ribonucleotide-diphosphate reductase (RNR) and associated proteins.** The RNR complex is one of the most tightly regulated cell cycle complexes. Its large subunit is transcriptionally regulated in all organisms, and in human its small subunit is regulated both at the transcriptional level and through CDK phosphorylation. In the yeasts, the RNR complex is further controlled by inhibitors and through subcellular relocalization of its small subunit.

The human RRM2 subunit is the only small subunit of RNR to be periodically expressed according to our analysis. It is maximally expressed during S phase, which is consistent with what has been observed in mouse¹⁰⁵, and its peak time differs from that of the large subunit by only 3% of a cell cycle. Notably, RRM2 is not just the only RNR small subunit to transcriptionally regulated—it is also the only one to phosphorylated by cyclin-dependent kinase 1 (CDK1). RRM2 has been shown to be phosphorylated by CDK1 on serine-20⁶⁴, which may regulate APC/C-mediated proteolysis of RRM2 since a KEN box is located close to serine-20¹⁰⁵. A multiple alignment reveals that this phosphorylation site does not exist in the RNR small subunits from any of the three other organisms. Consistent with this, none of the large-scale screens in *S. cerevisiae* have suggested Rnr2 or Rnr4 to be Cdc28 substrates^{51;52}. The small subunit of RNR is thus a clear example of co-evolution of transcriptional and posttranslational regulation.

To prevent large amounts of dNTP from accumulating once DNA replication has been completed, the RNR complex must be deactivated shortly thereafter. In mammals, this is accomplished through allosteric inhibition by dATP¹⁰⁶ and presumably also through APC/C-mediated proteolysis of the RRM2 subunit as described above. In contrast, budding and fission yeast primarily rely on the non-homologous RNR inhibitors SML1^{107–109} and spd1^{110;111}, which are both expressed in early G₂ phase as should be expected given their function (Figure S10). Interestingly, spd1 has a putative ortholog in budding

yeast, YLR437C, which according to our analysis is expressed at exactly the same timepoint in the cell cycle as spd1 itself (Figure S10). We thus propose that YLR437C is indeed a functional homolog of spd1, which together with SML1 ensures that RNR activity is suppressed after DNA replication has been completed.

In the two yeasts, RNR activity is further prevented by relocalizing the small subunit of RNR to the nucleus. WTM1 and KAP122 have recently been shown to be required for nuclear import and anchoring of budding yeast RNR2 and RNR4^{112;113}. It is currently not known if kap111, the fission yeast ortholog of KAP122, is also involved in nuclear localization of suc22; however, it is tempting to speculate that this is the case considering that kap111, like WTM1, is expressed at the M/G₁ transition.

Cohesin

To facilitate proper alignment and separation of sister chromatids into two cells, the sister chromatids are glued together by cohesin complexes, which consist of the four subunits Scc1, Scc3, Smc1, and Smc3. The nomenclature of the Scc subunits is unfortunately rather confusing: Scc1 is called Rad21 in human and fission yeast and Mcd1 in budding yeast, whereas Scc3 is called SA1/2 in human, Irr1 in budding yeast, and Psc3 in fission yeast.

As was observed for many of the protein complexes involved in DNA replication, our analysis shows that the expression of cohesin subunits differs between organisms, both in terms of the identity of the dynamic subunits and their time of peak expression (Figure 2). Human and fission yeast each have two dynamic cohesin subunits, which are expressed during M phase; in contrast all four subunits are expressed at the G₁ transition in budding yeast. The latter is known to be the case for the Scc1 subunit, and it is consistent with the fact that cohesin assembles onto the chromosomes shortly before the initiation of DNA replication in budding yeast⁴³. Notably, the Scc1 subunit is resynthesized in all three organisms after having been cleaved to allow the sister chromatids to separate (see below).

In addition to the mitotic cohesin complex just described, meiotic cohesin complexes also exist in which an alternative Scc1 subunit, Rec8, replaces the Scc1 subunit of mitotic cohesin. Moreover, the SMC1L1 subunit is replaced by SMC1L2 in human and the Scc3 subunit is replaced by STAG3/rec11 in human and fission yeast⁴³. Reassuringly, none of these meiosis-specific cohesin subunits are cell cycle regulated according to our analysis.

Similar to Scc3, the protein Pds5 binds to the Scc1 cohesin subunit. Although the binding of Pds5 is too weak for it to copurify with cohesin, it

is necessary for maintaining sister chromatid cohesion and thus should possibly be considered a cohesin subunit^{43;114;115}. Pds5 is periodically expressed in budding and fission yeast with peak times matching perfectly to those of the dynamic cohesin subunits. Our analysis thus supports the view that Pds5 should be considered a fifth subunit of the mitotic cohesin complex and possibly of the meiotic cohesion complexes as well.

Separase and securin

Once the sister chromatids have been properly aligned, they must be released from the cohesin complex to allow them to separate. This is accomplished by the protease *separase* (Esp1), which cleaves the Scc1 subunit of cohesin. Separase is constitutively expressed in the two yeasts, whereas our analysis suggests that it is expressed at the G₂/M transition in human. The latter is to the best of our knowledge a new discovery, and may well be related to the observation that human separase is activated through autocatalytic self-cleavage¹¹⁶.

To prevent premature Scc1 cleavage by separase, it is bound by an inhibitor called *securin* (Pds1) through most of the cell cycle. Securing is also required in order to localize separase to the nucleus¹¹⁶. Sister chromatid separation is controlled by APC/C degradation of securin, hence allowing separase to cleave Scc1. Our analysis reveals that securin is resynthesized very shortly after it has been degraded in human and fission yeast; in contrast, it is known that budding yeast waits until the following S phase before resynthesizing securin⁴³. Either solution implies that securin is present to inhibit separase during the time when the sister chromatids should be held together by cohesin. While the timing of separase expression in human agrees well with its time of action, the self-cleavage and resynthesis of human separase most likely only serves as an additional layer of control, which may improve the robustness of the system.

Shugoshin

The *shugoshin* (Sgo1) protein is best known for its role in protecting centromeric Rec8 from the separase during meiosis⁴³. However, recent experiments suggest that shugoshin is also required during mitosis to sense tension between sister chromatids¹¹⁷ and to protect the Scc3 subunit (STAG2, IRR1, psc3) of centromeric cohesin complexes from phosphorylation and subsequent cleavage¹¹⁸; the latter has very recently been shown to be achieved through recruitment of protein phosphatase 2A^{119;120}. Indeed, the microarray expression data reveals it periodically expressed in both budding and fission yeast; in both organisms, shugoshin is maximally expressed shortly after securin. This observation supports

the idea that shugoshin, like securin, is involved in protecting the mitotic cohesin complexes from premature degradation.

Anaphase promoting complex/cyclosome (APC/C)

As its two names suggests, the anaphase promoting complex/cyclosome is both important for controlling the onset of anaphase, which it initiates by targeting securin for degradation by the proteasome, and for driving the cell cycle clock by mediating the degradation of cyclins^{41;43;121}. The APC/C is a ubiquitin E3 ligase consisting of at least 11 core subunits that are conserved through eukaryote evolution as well as several additional subunits that appear to be present in only some organisms (Figure 2)^{37;41;121;122}. Like other multi-subunit E3 complexes, APC/C contains a RING subunit (Apc11) and a cullin subunit (Apc2), which are responsible for the actual transfer of ubiquitin to the substrates^{35;37}. In fact, it has been shown that Apc2–Apc11 can assemble poly-ubiquitin chains *in vitro*, however, with no substrate specificity^{37;40}. Little is known about the molecular function of the other APC/C subunits.

The activity and substrate recognition of APC/C is regulated through binding of the regulatory proteins, Cdc20 and Cdh1^{37;121;123}. In budding yeast, Cdc20 is known to be transcriptionally upregulated in S or G₂ phase¹²¹; indeed, our analysis shows that Cdc20 is maximally expressed during G₂ phase in budding yeast whereas its expression peaks during early M phase in human and fission yeast (Figure 2). The latter is consistent with the timepoint from which Cdc20 is known to bind to the APC/C core complex until it is eventually targeted for degradation by APC^{Cdh1}¹²¹. Moreover, Cdk1 has been shown to phosphorylate Cdh1 in all three organisms^{51;124;125} and Cdc20 in human¹²⁶ and budding yeast^{51;52}.

Surprisingly, our analysis reveals that it is not only the substrate recognition subunits of APC/C that are subject to transcriptional regulation during the cell cycle; some of the core subunits are regulated as well. In human, the Apc3 (CDC27) subunit is periodically expressed, whereas the crucial cullin subunit, Apc2, and another core subunit, Apc15, are both periodically expressed in fission yeast. Notably, the three genes have very similar peak times that coincide with the onset of APC/C activity at the G₂/M transition. As is the case for most of the complexes described above, the APC/C thus appears to be subject to transcriptional regulation mediated by different subunits in each organism. To the best of our knowledge, this is the first time that transcriptional regulation of the core APC/C complex has been proposed.

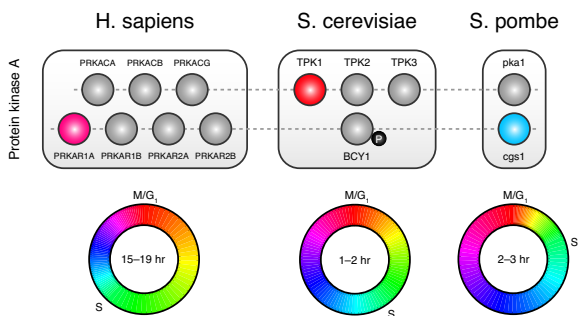


Figure S11: **Protein kinase A (PKA)**. Several alternative PKA complexes exist, each consisting of a catalytic (top row) and a regulatory (bottom row) subunit. PKA has been shown to regulate APC/C activity during the cell cycle through phosphorylation of APC/C core subunits³⁷. Mutations in the human regulatory subunit PRKAR1A have been associated with inherited disposition to certain cancers (Carney complex)^{127;128}.

Protein kinase A (PKA)

In addition to the regulatory mechanisms described above, APC/C is also regulated by protein kinase A (PKA, also known as cAMP-dependent protein kinase), which consists of a catalytic and a regulatory subunit^{37;129}. In budding yeast and human, however, several alternative PKA complexes can be formed due to the existence of paralogous catalytic and regulatory subunits (Figure S11), the latter being expressed in different human tissues^{128;130}.

PKA has been shown to inhibit APC/C activity in both budding and fission yeast and to phosphorylate core subunits of APC/C in human and mouse cell lines³⁷. Our analysis shows that PKA indeed has one periodically expressed subunit in each organism, whereas the identity of the dynamic subunit varies between organisms as is the case for most complexes. Consistent with the function as an APC/C inhibitor, the PKA activity in HeLa cells increases rapidly when the cells enter mitosis and falls sharply when they reach metaphase^{37;131}. We suggest that this is controlled via the regulatory subunit PRKAR1A, which is periodically expressed with a peak time in early M phase (Figure S11). Intriguingly, mutations in the PRKAR1A gene have been associated with Carney complex, a genetic condition leading to benign tumours in various tissues, whereas none of the other subunits of PKA have been associated with this or other forms of cancer^{127;128}.

Cyclin-dependent kinases (CDKs)

The cyclins, CDKs, and their related regulatory proteins are at the core of the eukaryotic cell cycle. The two yeasts have only a single cell cycle-related CDK, which serves multiple functions by interacting with different cyclins during different phases of the cell cycle. As with many proteins discovered in the early days of molecular biology, the nomenclature of this kinase is slightly confusing: while it is typically referred to as Cdk1, it

is officially named Cdc28 in budding yeast and Cdc2 in fission yeast and human. The two yeasts have two classes of cyclins, namely the G₁-type cyclins, which as the name suggests are active in G₁ phase, and the B-type cyclins, some of which are active during mitosis and others during G₁ and S phase. In mammals, including human, the situation is more complex, involving several CDKs that associate with different sets of cyclins: cyclin D–Cdk4/6 is involved in G₁ progression, cyclin E–Cdk2 in the G₁/S transition, cyclin A–Cdk2 in S phase progression, and cyclin A/B–Cdc2 in the G₂/M transition and intra-M control^{132–135}.

Budding yeast Cln3 (together with Cdc28) is responsible for progression through G₁ phase. In late G₁ phase, it activates SBF and MBF transcription factors, and thereby induces transcription of the cyclins Cln1, Cln2, Clb5, and Clb6^{136–139}. Although Clb5/6 are expressed at this early timepoint, they are inhibited by Sic1 until S phase when they trigger DNA replication, whereas the Cln cyclins are unaffected by Sic1^{138;140}. The timing of expression observed in our analysis (Figure S12) is thus in perfect agreement with what is known about these five cyclins.

Fission yeast has only a single G₁-type cyclin, puc1, which is most similar to budding yeast Cln3 (formerly known as Whi1)¹⁴¹ and is likewise involved in progression through G₁ phase^{142–144}. The primary cyclin responsible for promoting S phase is cig2, which like budding yeast Clb5/6 is a B-type cyclin regulated by MBF^{142;145}. Our analysis shows that it is expressed already during cytokinesis (Figure S12), which may at first seem surprising. However, it makes perfect sense since it implies that cig2 is expressed at the same time as the replication machinery, as is the case for Clb5/6 in budding yeast. This supports the view that cig2 is functionally equivalent to Clb5/6¹⁴².

The mammalian cyclin D1 (CCND1) was discovered in a screen using a budding yeast strain conditionally deficient for Cln3 function¹⁴⁶; this together with sequence similarity suggests that cyclin D–Cdk4/6 complexes are the human equivalents of the budding yeast Cln3–Cdc28 complex. Similarly, the cyclin E–Cdk2 complexes can be considered equivalent to Cln1/2–Cdc28 complexes in budding yeasts. Indeed, our analysis shows that cyclin E1 (CCNE1) is maximally expressed at the G₁/S transition, whereas cyclin E2 (CCNE2) is static (Figure S12). Knockout of the latter results in germ defects¹³⁵, and we thus speculate that cyclin E1 (CCNE1) may be the primary cyclin E in the mitotic cell cycle. The entry into S phase is further regulated by the phosphatase Cdc25A, which is an activator of Cdk2 and is expressed at the G₁/S transition. Curiously, CDC25A is in several databases wrongly annotated as an “M phase inducer phosphatase” due to its homology with other Cdc25 phosphatases, although the expression data clearly shows that it acts during S

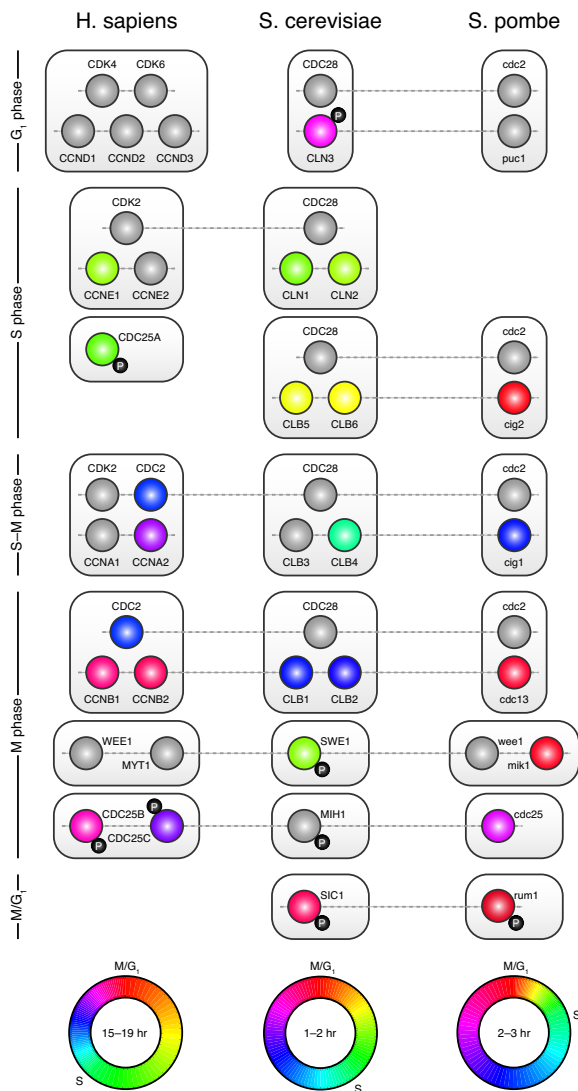


Figure S12: **Cyclin-CDK complexes and their activators/inhibitors.** Only the subset of cyclin-CDK complexes known to be involved in the mitotic cell cycle are shown. Functionally equivalent cyclins in the different organisms are placed next to each other, whereas dotted lines signify sequence homology.

phase, as is also known.

All three organisms have cyclins that accumulate during S phase, reaching their maximal expression level during G₂; these are Clb3/4, cig1, and cyclin A. In the two yeasts, Clb4 and cig1 are both dynamic with the expected peaktimes, whereas Clb3 appears static in our analysis (Figure S12). The latter may be due to problems with the microarray data, as earlier studies report that both Clb3 and Clb4 are periodic with a similar pattern of expression^{138;147}; however, ChIP-chip data suggest that the CLB3 promoter is indeed not bound by known cell cycle transcription factors^{13;14}. The situation is slightly more complicated in mammals, where the dynamic cyclin A2 (CCNA2) forms complexes with Cdk2 and Cdc2 during S phase and G₂/M, respectively. Similar to cyclin E2 (CCNE2), knockout of cyclin A1 (CCNA1) leads to germ defects, and its expression is normally restricted to germ cells¹³⁵; it is

thus no surprise that we observe it as static, since it is presumably not expressed at all in the cells studied.

The regulation of entry into M phase is the most well-conserved part of cyclin-CDK regulation across the three organisms studied. It is in all three organisms mediated by orthologous B-type cyclins that form complexes with Cdk1: Clb1/2-Cdc28 in budding yeast^{132-135;138}, cdc13-cdc2 in fission yeast¹⁴⁸, and cyclin B1/2-Cdc2 in human. The fission yeast and human cyclins all peak in expression during mitosis, while Clb1 and Clb2 are expressed already during G₂ phase (Figure S12). Since human Cdc2 is only active during late G₂ and M phase, it is, in contrast to its yeast orthologs, free from the evolutionary constraint that it must be static. Indeed, our analysis suggests that Cdc2 is expressed in G₂ phase, which offers an explanation why cyclin A2-Cdk2 changes to cyclin A2-Cdc2 at this point in the cell cycle.

The Wee1 kinases (WEE1/MYT1 in human, SWE1 in budding yeast, and wee1/mik1 in fission yeast) cause inhibitory phosphorylation of the Cdk1 kinases, which are countered by the Cdc25 phosphatases (CDC25B/C in human, MIH1 in budding yeast, cdc25 in fission yeast)^{133;138}. In human, the Wee1 kinases are static and thus by default inhibit Cdk1 activity, whereas the Cdc25 phosphatases are specifically expressed during M phase, thereby activating Cdk1. The same is true in fission yeast, except that the mik1 paralog of wee1 is expressed shortly before S-phase^{149;150}. In contrast, our analysis shows that budding yeast MIH1 is static, which is compensated by dynamic expression of SWE1 during G₁ and S phase, when high Cdk1 activity is unwanted (Figure S12). This budding yeast-specific difference in the core regulation of the cell cycle is yet another example of compensatory evolutionary changes in gene expression and is to the best of our knowledge not known. Moreover, Swe1 was recently discovered to be directly phosphorylated by Cdc28¹⁵¹, and thus also exemplifies co-evolution of transcriptional regulation and phosphorylation.

References

- [1] Whitfield, M. L. *et al.* Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Molecular Biology of the Cell* **13**, 1977–2000 (2002).
- [2] 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).
- [3] Cho, R. J. *et al.* A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell* **2**, 65–73 (1998).
- [4] Rustici, G. *et al.* Periodic gene expression program of the fission yeast cell cycle. *Nature Genetics* **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 Biology* **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] Cho, R. J. *et al.* Transcriptional regulation and function during the human cell cycle. *Nature Genetics* **27**, 48–54 (2001).
- [9] Shedden, K. & Cooper, S. Analysis of cell-cycle-specific gene expression in human cells as determined by microarrays and double-thymidine block synchronization. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4379–4384 (2002).
- [10] Jensen, L. J. & Knudsen, S. Automatic discovery of regulatory patterns in promoter regions based on whole cell expression data and functional annotation. *Bioinformatics* **16**, 326–333 (2000).
- [11] de Lichtenberg, U. *et al.* Comparison of computational methods for the identification of cell cycle regulated genes. *Bioinformatics* **21**, 1164–1171 (2005).
- [12] 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).
- [13] Simon, I. *et al.* Serial regulation of transcriptional regulators in the yeast cell cycle. *Cell* **106**, 697–708 (2001).
- [14] Lee, T. I. *et al.* Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298**, 799–804 (2002).
- [15] Cam, H. *et al.* A common set of gene regulatory networks links metabolism and growth inhibition. *Mol. Cell* **16**, 399–411 (2004).
- [16] Balciunaite, E. *et al.* Pocket protein complexes are recruited to distinct targets in quiescent and proliferating cells. *Mol. Cell Biol.* **25**, 8166–8178 (2005).
- [17] Vandepoele, K. *et al.* Genome-wide analysis of core cell cycle genes in arabidopsis. *Plant Cell* **14**, 903–916 (2002).
- [18] Vandepoele, K. *et al.* Genome-wide identification of potential plant E2F target genes. *Plant Physiol.* **139**, 316–328 (2005).
- [19] de Lichtenberg, U., Jensen, L. J., Brunak, S. & Bork, P. Dynamic complex formation during the yeast cell cycle. *Science* **307**, 724–727 (2005).
- [20] Cooper, S. & Shedden, K. Microarray analysis of gene expression during the cell cycle. *Cell Chromosome* **2**, 1 (2003).
- [21] Cooper, S. Is whole-culture synchronization biology’s ‘perpetual-motion machine’? *Trends in Biotechnology* **22**, 266–269 (2004).
- [22] Spellman, P. T. & Sherlock, G. Reply: whole-culture synchronization—effective tools for cell cycle studies. *Trends in Biotechnology* **22**, 270–273 (2004).
- [23] Cooper, S. Rejoinder: whole-culture synchronization cannot, and does not, synchronize cells. *Trends in Biotechnology* **22**, 274–276 (2004).
- [24] Spellman, P. T. & Sherlock, G. Final words: cell age and cell cycle are unlinked. *Trends in Biotechnology* **22**, 277–278 (2004).
- [25] Aach, J. & Church, G. M. Aligning gene expression time series with time warping algorithms. *Bioinformatics* **17**, 495–508 (2001).
- [26] Criel, J. & Tsiporkova, E. Gene expression warper: a tool for alignment, template matching and visualization of gene expression time series. *Bioinformatics* **22**, 251–252 (2006).
- [27] Tatusov, R. L., Koonin, E. V. & Lipman, D. J. A genomic perspective on protein families. *Science* **278**, 631–637 (1997).
- [28] Tatusov, R. L. *et al.* The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4**, 41 (2003).
- [29] Joshi-Tope, G. *et al.* Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.* **33**, D428–D432 (2005).

- [30] Hubbard, T. *et al.* Ensembl 2005. *Nucleic Acids Res.* **33**, D447–D453 (2005).
- [31] Bairoch, A. *et al.* The universal protein resource (UniProt). *Nucleic Acids Res.* **33**, D154–D159 (2005).
- [32] Christie, K. R. *et al.* Saccharomyces Genome Database (SGD) provides tools to identify and analyze sequences from *Saccharomyces cerevisiae* and related sequences from other organisms. *Nucleic Acids Res.* **32**, D311–D314 (2004).
- [33] Hertz-Fowler, C. *et al.* GeneDB: a resource for prokaryotic and eukaryotic organisms. *Nucleic Acids Res.* **32**, D339–D343 (2004).
- [34] Rhee, S. Y. *et al.* The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. *Nucleic Acids Res.* **31**, 224–228 (2003).
- [35] Tyers, M. & Jorgensen, P. Proteolysis and the cell cycle: With this RING I do thee destroy. *Curr. Opin. Gen. Dev.* **10**, 54–64 (2000).
- [36] Bell, S. P. & Dutta, A. DNA replication in eukaryotic cells. *Annual Review of Biochemistry* **71**, 333–374 (2002).
- [37] Harper, J. W., Burton, J. L. & Solomon, M. J. The anaphase-promoting complex: it's not just for mitosis any more. *Genes Dev.* **16**, 2179–2206 (2002).
- [38] Kearsey, S. E. & Cotterill, S. Enigmatic variations: Divergent modes of regulating eukaryotic dna replication. *Mol. Cell* **12**, 1067–1075 (2003).
- [39] Forsburg, S. L. Eukaryotic MCM proteins: Beyond replication initiation. *Microbiol. Mol. Biol. Rev.* **68**, 109–131 (2004).
- [40] Passmore, L. A. The anaphase-promoting complex (APC): the sum of its parts? *Biochem. Soc. Trans.* **32**, 724–727 (2004).
- [41] Castro, A., Bernis, C., Vigneron, S., Labbe, J. C. & Lorca, T. The anaphase-promoting complex: a key factor in the regulation of cell cycle. *Oncogene* **24**, 314–325 (2005).
- [42] Johnson, A. & O'Donnell, M. Cellular DNA replicases: Components and dynamics of the replication fork. *Annu. Rev. Biochem.* **74**, 283–315 (2005).
- [43] Nasmyth, K. & Haering, C. H. The structure and function of SMC and kleisin complexes. *Annu. Rev. Biochem.* **74**, 595–648 (2005).
- [44] Maiorano, D., Lutzmann, M. & Méchali, M. MCM proteins and DNA replication. *Curr. Opin. Cell Biol.* **18**, 130–136 (2006).
- [45] Diella, F. *et al.* Phospho.ELM: a database of experimentally verified phosphorylation sites in eukaryotic proteins. *BMC Bioinformatics* **5**, 79 (2004).
- [46] Hornbeck, P. V., Chabra, I., Kornhauser, J. M., Skrzypek, E. & Zhang, B. PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics* **4**, 1551–1561 (2004).
- [47] Salomon, A. R. *et al.* Profiling of tyrosine phosphorylation pathways in human cells using mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 443–448 (2003).
- [48] Beausoleil, S. A. *et al.* Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12130–12135 (2004).
- [49] Brill, L. M. *et al.* Robust phosphoproteomic profiling of tyrosine phosphorylation sites from human T cells using immobilized metal affinity chromatography and tandem mass spectrometry. *Anal. Chem.* **76**, 2763–2772 (2004).
- [50] Shu, H., Chen, S., Bi, Q., Mumby, M. & Brekken, D. L. Identification of phosphoproteins and their phosphorylation sites in the WEHI-231 B lymphoma cell line. *Mol. Cell. Proteomics* **3**, 279–286 (2004).
- [51] Ubersax, J. A. *et al.* Targets of the cyclin-dependent kinase cdk1. *Nature* **425**, 859–864 (2003).
- [52] Loog, M. & Morgan, D. O. Cyclin specificity in the phosphorylation of cyclin-dependent kinase substrates. *Nature* **434**, 104–108 (2005).
- [53] 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).
- [54] Obenauer, J. C., Cantley, L. C. & Yaffe, M. B. Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res.* **31**, 3635–3641 (2003).
- [55] Yaglom, J. *et al.* p34(Cdc28)-mediated control of Cln3 cyclin degradation. *Mol. Cell. Biol.* **15**, 731–741 (1995).
- [56] Rechsteiner, M. & Rogers, S. W. PEST sequences and regulation by proteolysis. *Trends Biochem. Sci.* **21**, 267–271 (1996).

- [57] Nash, P. *et al.* Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication. *Nature* **414**, 514–521 (2001).
- [58] Berset, C. *et al.* Transferable domain in the G1 cyclin Cln2 sufficient to switch degradation of Sic1 from the E3 ubiquitin ligase SCF(Cdc4) to SCF(Grr1). *Mol. Cell. Biol.* **22**, 4463–4476 (2002).
- [59] Pfleger, C. M. & Kirschner, M. W. The KEN box: an APC recognition signal distinct from the D box targeted by Cdh1. *Genes Dev.* **14**, 655–665 (2000).
- [60] Beissbarth, T. & Speed, T. P. GOstat: Find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* **20**, 1464–1465 (2005).
- [61] Maslov, S., Sneppen, K., Eriksen, K. A. & Yan, K.-K. Upstream plasticity and downstream robustness in evolution of molecular networks. *BMC Evol. Biol.* **4**, 9 (2004).
- [62] Ptacek, J. *et al.* Global analysis of protein phosphorylation in yeast. *Nature* **438**, 679–684 (2005).
- [63] Sipiczki, M. Where does fission yeast sit on the tree of life? *Genome Biology* **1**, reviews1011 (2000).
- [64] Chan, A. K., Persad, S., Litchfield, D. W. & Wright, J. A. Ribonucleotide reductase R2 protein is phosphorylated at serine-20 by P34cdc2 kinase. *Biochim. Biophys. Acta* **1448**, 363–371 (1999).
- [65] Kesti, T., McDonald, W. H., Yates, J. R., r. & Wittenberg, C. Cell cycle-dependent phosphorylation of the DNA polymerase epsilon subunit, Dpb2, by the Cdc28 cyclin-dependent protein kinase. *J. Biol. Chem.* **279**, 14245–14255 (2004).
- [66] Graur, D. & Martin, W. Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. *Trends Genet.* **20**, 80–86 (2004).
- [67] Berbee, M. & Taylor, J. Fungal molecular evolution: gene trees and geologic time. in *The Mycota*, (McLaughlin, D., McLaughlin, E. & Lemke, E., eds), volume VIIB, 229–246. Springer-Verlag, Berlin (2001).
- [68] Forsburg, S. L. & Nurse, P. Cell cycle regulation in the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *Annu. Rev. Cell Biol.* **7**, 227–256 (1991).
- [69] Diffley, J. F. X. Regulation of early events in chromosome replication. *Curr. Biol.* **14**, R778–R786 (2004).
- [70] Nguyen, V. Q., Co, C. & Li, J. J. Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. *Nature* **411**, 1068–1073 (2001).
- [71] Ohta, S., Tatsumi, Y., Fujita, M. & Obuse, C. The ORC1 cycle in human cells. *J. Biol. Chem.* **278**, 41535–41540 (2003).
- [72] DePamphilis, M. L. *et al.* Regulating the licensing of DNA replication origins in metazoa. *Curr. Opin. Cell Biol.* **18**, 231–238 (2006).
- [73] Speck, C., Chen Z, Li, H. & Stillman, B. ATPase-dependent cooperative binding of ORC and Cdc6 to origin DNA. *Nat. Struct. Mol. Biol.* **12**, 965–971 (2005).
- [74] Randell, J. C. W., Bowers, J. L., Rodríguez, H. K. & Bell, S. P. Sequential ATP hydrolysis by Cdc6 and ORC directs loading of the Mcm2-7 helicase. *Mol. Cell* **21**, 29–39 (2006).
- [75] Jallepalli, P. V., Brown, G. W., Muzi-Falconi, M., Tien, D. & Kelly, T. J. Regulation of the replication initiator protein p65(*cdc18*) by CDK phosphorylation. *Genes Dev.* **11**, 2767–2779 (1997).
- [76] Mailand, N. & Diffley, J. F. X. CDKs promote DNA replication origin licensing in human cells by protecting Cdc6 from APC/C-dependent proteolysis. *Cell* **122**, 915–926 (2005).
- [77] Nishitani, H., Lygerou, Z., Nishimoto, T. & Nurse, P. The Cdt1 protein is required to license DNA for replication in fission yeast. *Nature* **404**, 625–628 (2000).
- [78] Nishitani, H., Taraviras, S., Lygerou, Z. & Nishimoto, T. The human licensing factor for DNA replication Cdt1 accumulates in g1 and is destabilized after initiation of s-phase. *J. Biol. Chem.* **276**, 44905–44911 (2001).
- [79] Li, X., Zhao, Q., Liao, R., Sun, P. & Wu, X. The SCF(Skp2) ubiquitin ligase complex interacts with the human replication licensing factor Cdt1 and regulates Cdt1 degradation. *J. Biol. Chem.* **278**, 30854–30858 (2003).
- [80] Liu, E., Li, X., Yan, F., Zhao, Q. & Wu, X. Cyclin-dependent kinases phosphorylate human Cdt1 and induce its degradation. *J. Biol. Chem.* **279**, 17283–17288 (2004).
- [81] Nishitani, H., Lygerou, Z. & Nishimoto, T. Proteolysis of DNA replication licensing factor Cdt1 in S-phase is performed independently of Geminin through its N-terminal region. *J. Biol. Chem.* **279**, 30807–30816 (2004).

- [82] Dimitrova, D. S., Prokhorova, T. A., Blow, J. J., Todorov, I. T. & Gilbert, D. M. Mammalian nuclei become licensed for DNA replication during late telophase. *J. Cell Sci.* **115**, 51–59 (2002).
- [83] Takahashi, T. S., Yiu, P., Chou, M. F., Gygi, S. & Walter, J. C. Recruitment of *Xenopus* Scc2 and cohesin to chromatin requires the pre-replication complex. *Nature Cell Biology*. **6**, 991–996 (2004).
- [84] Ekholm-Reed, S. *et al.* Deregulation of cyclin E in human cells interferes with pre-replication complex assembly. *J. Cell. Biol.* **165**, 789–800 (2004).
- [85] Ishimi, Y., Komamura-Kohno, Y., Karasawa-Shimizu, K. & Yamada, K. Levels of MCM4 phosphorylation and DNA synthesis in DNA replication block checkpoint control. *J. Struct. Biol.* **146**, 234–241 (2004).
- [86] 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).
- [87] Chapman, J. W. & Johnston, L. H. The yeast gene, DBF4, essential for entry into S phase is cell cycle regulated. *Exp. Cell Res.* **180**, 419–428 (1989).
- [88] de Lichtenberg, U. *et al.* New weakly expressed cell cycle-regulated genes in yeast. *Yeast* **22**, 1191–1201 (2005).
- [89] Pasero, P., Duncker, B. P., Schwob, E. & Gasser, S. M. A role for the Cdc7 kinase regulatory subunit Dbf4p in the formation of initiation-competent origins of replication. *Genes Dev.* **13**, 2159–2176 (1999).
- [90] Tak, Y. S., Tanaka, Y., Endo, S., Kamimura, Y. & Araki, H. A CDK-catalysed regulatory phosphorylation for formation of the DNA replication complex Sld2–Dpb11. *EMBO J.* **25**, 1987–1996 (2006).
- [91] Ricke, R. M. & Bielinsky, A.-K. Mcm10 regulates the stability and chromatin association of DNA polymerase- α . *Mol. Cell Biol.* **16**, 173–185 (2004).
- [92] Kamimura, Y., Tak, Y. S., Sugino, A. & Araki, H. Sld3, which interacts with Cdc45 (Sld4), functions for chromosomal DNA replication in *Saccharomyces cerevisiae*. *EMBO J.* **20**, 2097–2107 (2001).
- [93] Gambus, A. *et al.* GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nature Cell Biology* **8**, 358–366 (2006).
- [94] Kanemaki, M. & Labib, K. Distinct roles for Sld3 and GINS during establishment and progression of eukaryotic DNA replication forks. *EMBO J.* **25**, 1753–1763 (2006).
- [95] Owens, J. C., Detweiler, C. S. & Li, J. J. CDC45 is required in conjunction with CDC7/DBF4 to trigger the initiation of DNA replication. *Proc. Natl. Acad. Sci. U.S.A.* **94**(23), 12521–12526 (1997).
- [96] Krogan, N. J. *et al.* Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**, 637–643 (2006).
- [97] Miles, J. & Formosa, T. Evidence that POB1, a *Saccharomyces cerevisiae* protein that binds to dna polymerase α , acts in DNA metabolism in vivo. *Mol. Cell. Biol.* **12**, 5724–5735 (2006).
- [98] Hanna, J. S., Kroll, E. S., Lundblad, V. & Spencer, F. A. *Saccharomyces cerevisiae* CTF18 and CTF4 are required for sister chromatid cohesion. *Mol. Cell. Biol.* **21**, 3144–3158 (2001).
- [99] Krogh, B. O. & Symington, L. S. Recombination proteins in yeast. *Annu. Rev. Genet.* **38**, 233–271 (2004).
- [100] Araki, H. *et al.* Cloning of *DPB3*, the gene encoding the third subunit of DNA polymerase II of *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **19**, 4867–4872 (1991).
- [101] Mayer, M. L., Gygi, S. P., Aebersold, R. & Hieter, P. Identification of RFC(Ctf18p, Ctf8p, Dcc1p): An alternative RFC complex required for sister chromatid cohesion in *S. cerevisiae*. *Mol. Cell* **7**, 959–970 (2001).
- [102] Ben-Aroya, S., Koren, A., Liefshitz, B., Steinlauf, R. & Kupiec, M. *ELG1*, a yeast gene required for genome stability, forms a complex related to replication factor C. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9906–9911 (2003).
- [103] Yao, R. *et al.* Subcellular localization of yeast ribonucleotide reductase regulated by the DNA replication and damage checkpoint pathways. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 6628–6633 (2003).
- [104] Huang, M. & Elledge, S. J. Identification of RNR4, encoding a second essential small subunit of ribonucleotide reductase in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **17**, 6105–6113 (1997).

- [105] Chabes, A. L., Pfeleger, C. M., Kirschner, M. W. & Thelander, L. Mouse ribonucleotide reductase R2 protein: A new target for anaphase-promoting complex-Cdh1-mediated proteolysis. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3925–3929 (2003).
- [106] Reichard, P., Eliasson, R., Ingemarson, R. & Thelander, L. Cross-talk between the allosteric effector-binding sites in mouse ribonucleotide reductase. *J. Biol. Chem.* **275**, 33021–33026 (2000).
- [107] Chabes, A., Domkin, V. & Thelander, L. Yeast Sml1, a protein inhibitor of ribonucleotide reductase. *J. Biol. Chem.* **274**, 36679–36683 (1999).
- [108] Zhao, X., Chabes, A., Domkin, V., Thelander, L. & Rothstein, R. The ribonucleotide reductase inhibitor Sml1 is a new target of the Mec1/Rad53 kinase cascade during growth and in response to DNA damage. *EMBO J.* **20**, 3544–3553 (2001).
- [109] Chabes, A. *et al.* Survival of DNA damage in yeast directly depends on increased dNTP levels allowed by relaxed feedback inhibition of ribonucleotide reductase. *Cell* **112**, 391–401 (2003).
- [110] Liu, K. *et al.* Cop9/signalosome subunits and Pcu4 regulate ribonucleotide reductase by both checkpoint-dependent and -independent mechanisms. *Genes Dev.* **17**, 1130–1140 (2003).
- [111] Håkansson, P., Dahl, L., Chilkova, O., Domkin, V. & Thelander, L. The *Schizosaccharomyces pombe* replication inhibitor Spd1 regulates ribonucleotide reductase activity and dNTPs by binding to the large Cdc22 subunit. *J. Biol. Chem.* **281**, 1778–1783 (2006).
- [112] Lee, Y. D. & Elledge, S. J. Control of ribonucleotide reductase localization through an anchoring mechanism involving Wtm1. *Genes Dev.* **20**, 334–344 (2006).
- [113] Zhang, Z. *et al.* Nuclear localization of the *Saccharomyces cerevisiae* ribonucleotide reductase small subunit requires a karyopherin and a WD40 repeat protein. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 1422–1427 (2006).
- [114] Hartman, T., Stead, K., Koshland, D. & Guacci, V. Pds5 is an essential chromosomal protein required for both sister chromatid cohesion and condensation in *Saccharomyces cerevisiae*. *J. Cell Biol.* **151**, 613–626 (2000).
- [115] Pannizza, S., Tanaka, T., Hochwagen, A., Eisenhaber, F. & Nasmyth, K. Pds5 cooperates with cohesin in maintaining sister chromatid cohesion. *Curr. Biol.* **10**, 1557–1564 (2000).
- [116] Agarwal, R. & Cohen-Fix, O. Mitotic regulation: The fine tuning of separase activity. *Cell Cycle* **1**, 255–257 (2002).
- [117] Mayer, M. L. *et al.* The centromeric protein Sgo1 is required to sense lack of tension on mitotic chromosomes. *Science* **307**, 130–133 (2005).
- [118] McGuinness, B. E., Hirota, T., Kudo, N. R., Peters, J.-M. & Nasmyth, K. Shugoshin prevents dissociation of cohesin from centromeres during mitosis in vertebrate cells. *PLoS Biology* **3**, e86 (2005).
- [119] Kitajima, T. S. *et al.* Shugoshin collaborates with protein phosphatase 2A to protect cohesin. *Nature* **441**, 46–52 (2006).
- [120] Riedel, C. G. *et al.* Protein phosphatase 2A protects centromeric sister chromatid cohesion during meiosis I. *Nature* **441**, 53–61 (2006).
- [121] Peters, J.-M. The anaphase-promoting complex: Proteolysis in mitosis and beyond. *Mol. Cell* **9**, 931–943 (2002).
- [122] Yoon, H. J. *et al.* Proteomics analysis identifies new components of the fission and budding yeast anaphase-promoting complexes. *Curr. Biol.* **12**, 2048–2054 (2002).
- [123] Passmore, L. A. & Barford, D. Coactivator functions in a stoichiometric complex with anaphase-promoting complex/cyclosome to mediate substrate recognition. *EMBO Reports* **6**, 873–878 (2005).
- [124] Zhou, Y., Ching, Y. P., Chun, A. C. & Jin, D. Y. Nuclear localization of the cell cycle regulator CDH1 and its regulation by phosphorylation. *J. Biol. Chem.* **278**, 12530–12536 (2003).
- [125] Blanco, M. A., Sanchez-Diaz, A., de Prada, J. M. & Moreno, S. APC(ste9/srw1) promotes degradation of mitotic cyclins in G1 and is inhibited by cdc2 phosphorylation. *EMBO J.* **19**, 3945–3955 (2000).
- [126] Yudkovsky, Y., Shteinberg, M., Listovsky, T., Brandeis, M. & Hershko, A. Phosphorylation of Cdc20/Fizzy negatively regulates the mammalian cyclosome/APC in the mitotic checkpoint. *Biochem. Biophys. Res. Commun.* **271**, 299–304 (2000).

- [127] Stergiopoulos, S. G. & Stratakis, C. A. Human tumors associated with Carney complex and germline *PRKARA1* mutations: a protein kinase A disease. *FEBS Lett.* **546**, 59–64 (2003).
- [128] Wilkes, D., Charitakis, K. & Basson, C. T. Inherited disposition to cardiac myxoma development. *Nat. Rev. Cancer* **6**, 157–165 (2006).
- [129] Ferrari, S. Protein kinases controlling the onset of mitosis. *Cell. Mol. Life Sci.* **63**, 781–795 (2006).
- [130] Matyakhina, L., Lenherr, S. M. & Stratakis, C. A. Protein kinase A and chromosomal stability. *Ann. N.Y. Acad. Sci.* **968**, 148–157 (2002).
- [131] Kotani, S. *et al.* PKA and MPF-activated polo-like kinase regulates anaphase-promoting complex activity and mitotic progression. *Mol. Cell* **1**, 371–380 (1998).
- [132] Fung, T. K., Siu, W. Y., Yam, C. H., Lau, A. & Poon, R. Y. C. Cyclin F is degraded during G2-M by mechanisms fundamentally different from other cyclins. *J. Biol. Chem.* **38**, 35140–35149 (2002).
- [133] Miele, L. The biology of cyclins and cyclin-dependent protein kinases. in *Methods in Molecular Biology*, (Giordano, A. & Romano, G., eds), volume 285, 3–21. Humana Press (2004).
- [134] Harper, J. V. & Brooks, G. The mammalian cell cycle. in *Methods in Molecular Biology*, (Humphrey, T. & Brooks, G., eds), volume 296, 113–153. Humana Press (2005).
- [135] Sánchez, I. & Dynlacht, B. D. New insights into cyclins, CDKs, and cell cycle control. *Semin. Cell Dev. Biol.* **16**, 311–321 (2005).
- [136] Nasmyth, K. & Dirick, L. The role of SWI4 and SWI6 in the activity of g1 cyclins in yeast. *Cell* **65**, 995–1014 (1991).
- [137] Dirick, L., Moll, T., Auer, H. & Nasmyth, K. A central role for SWI6 in modulating cell cycle Start-specific transcription in yeast. *Nature* **357**, 508–513 (1992).
- [138] Mendenhall, M. D. & Hodge, A. E. Regulation of Cdc28 cyclin-dependent protein kinase activity during the cell cycle of the yeast *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **62**, 1191–1243 (1998).
- [139] Edgington, N. P. & Futcher, B. Relationship between the function and the localization of G1 cyclins in *S. cerevisiae*. *J. Cell Sci.* **114**, 4599–4611 (2001).
- [140] Breeden, L. Cyclin transcription: Timing is everything. *Curr. Biol.* **10**, R586–R588 (2000).
- [141] Forsburg, S. L. & Nurse, P. Identification of a G1-type cyclin *puc1+* in the fission yeast *Schizosaccharomyces pombe*. *Nature* **351**, 245–248 (1991).
- [142] Mondesert, O., McGowan, C. H. & Russell, P. Cig2, a B-type cyclin, promotes the onset of S in *Schizosaccharomyces pombe*. *Mol. Cell. Biol.* **16**, 1527–1533 (1996).
- [143] Martín-Castellanos, C., Blanco, M. A., de Prada, J. M. & Moreno, S. The *puc1* cyclin regulates the G1 phase of the fission yeast cell cycle in response to cell size. *Mol. Biol. Cell* **11**, 543–554 (2000).
- [144] Yamano, H. *et al.* The spike of S phase cyclin Cig2 expression at the G1–S border in fission yeast requires both APC and SCF ubiquitin ligases. *Mol. Cell* **6**, 1377–1387 (2000).
- [145] Ayté, J., Schweitzer, C., Zarzov, P., Nurse, P. & DeCaprio, J. A. Feedback regulation of the MBF transcription factor by cyclin Cig2. *Nature Cell Biology* **3**, 1043–1050 (2001).
- [146] Xiong, Y., Connolly, T., Futcher, B. & Beach, D. Human D-type cyclin. *Cell* **65**, 691–699 (1991).
- [147] Fitch, I. *et al.* Characterization of four B-type cyclin genes of the budding yeast *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **3**, 805–818 (1992).
- [148] Nurse, P. & Bisset, Y. Gene required in G1 for commitment to the cell cycle and G2 for control of mitosis in fission yeast. *Nature* **292**, 558–560 (1981).
- [149] Christensen, P. U., Bentley, N. J., Martinho, R. G., Nielsen, O. & Carr, A. M. Mik1 levels accumulate in S phase and may mediate an intrinsic link between S phase and mitosis. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 2579–2585 (2000).
- [150] Ng, S. S., Anderson, M., White, S. & McInerney, C. J. *mik1+* G1-S transcription regulates mitotic entry in fission yeast. *FEBS Lett.* **503**, 131–134 (2001).
- [151] Harvey, S. L., Charlet, A., Haas, W., Gygi, S. P. & Kellogg, D. R. Cdk1-dependent regulation of the mitotic inhibitor Wee1. *Cell* **122**, 407–420 (2005).

Acknowledgments

We thank C. von Mering for assistance with detection of orthologs as well as Sir Paul Nurse, Eric Karsenti, Jürg Bähler, and members of the Bork and Brunak groups for comments on the manuscript. Supported by grants from the Danish National Research Foundation, the Danish Technical Research Council, as well as the DIAMONDS Specific Targeted Research Project, contract number LSHB-CT-2004-512143, and the BioSapiens Network of Excellence, contract number LSHG-CT-2003-503265, both funded by the European Commission FP6 Programme.