Supplementary information for: Dynamic complex formation during the yeast cell cycle

Ulrik de Lichtenberg[1*], Lars Juhl Jensen[2*], Søren Brunak[1] and Peer Bork[2,3†]

[1]Center for Biological Sequence Analysis, The Technical University of Denmark, DK-2800 Lyngby, Denmark
[2]European Molecular Biology Laboratory, D-69117 Heidelberg, Germany
[3]Max-Delbrück-Centre for Molecular Medicine, D-13092 Berlin, Germany
*These authors contributed equally
†To whom correspondence should be addressed; Email: bork@embl.de

Materials and methods

We here describe in detail how the dynamic network of the cell cycle was derived. That includes revisions to the curated MIPS complexes, a new topology based scoring scheme for large-scale interaction data sets, and an approach to filter interactions data based on protein subcellular localization data. Finally, we describe a novel procedure for combining data on mRNA expression and protein interactions to extract a cell cycle network.

Protein interaction datasets and their refinement

Three different types of interaction evidence were utilized for constructing the physical interaction network: yeast two-hybrid screens (*S1, S2*), complex pull-down screens (*S3, S4*), and MIPS complexes (*S5*). The manually curated MIPS complexes had to be refined as some annotations did not reflect known dynamic aspects. For example, the Cdc28p complexes are represented in MIPS (*S5*) as a single entity consisting of Cdc28p and all its 9 associated cyclins. However, Cdc28p can only interact with a single cyclin at a time. To correctly reflect this, we instead annotated 9 complexes that each consist of Cdc28p together with a single cyclin. The equivalent correction was made for Pho85p and the 6 cyclins with which it was annotated in MIPS. Furthermore, 4 missing Pho85p-cyclin complexes were added (Pcl5p, Pcl6p, Pcl7p, and Pcl9p).

The myosin and kinesin motor proteins were annotated as two large complexes, but in fact these proteins function individually and do not form complexes with each other. We consequently removed these two complexes. Also, the proteins Aut2p and Aut7p were removed from the "tubulin associated" complex as this published relationship (*S6*) has since then been contested (*S7, S8*).

While SCF was correctly annotated as three entities, namely the core complex together with each of three regulatory subunits, only the core complex was annotated for APC. For consistency, we annotated two APC complexes consisting of the APC core together with each of its two regulatory subunits, Cdc20p and Cdh1p.

The revised version of the MIPS complexes used for our analysis is available from `http://www.cbs.dtu.dk/cellcycle`.

Quality scores for individual binary interactions

For the two different types of high-throughput data sets, scoring schemes were developed that allow the reliability of individual, binary interactions to be compared and integrated across data sources.

For the yeast two-hybrid experiments, the reliability of an interaction was found to correlate well with the number of non-shared interaction partners for each interactor. We summarize this in the following raw quality score:

$$S1 = -\log((N_1 + 1) \cdot (N_2 + 1))$$

where N_1 and N_2 are the numbers of non-shared interaction partners. This score is similar to the IG1 measure suggested by Saito et al. (*S9*). For this scoring, the core interaction set by Ito et al. (*S1*) and the interaction set by Uetz et al. (*S2*) were pooled and treated as one experiment. This was done because the two experiments appeared to be of equal quality as well as to have correlated errors (data not shown). The full set of interactions by Ito et al. (minus interactions in the core set) was treated as a separate set because of its much higher rate of false positives.

In the case of complex pull-down experiments, the reliability of the inferred binary interactions was found to correlate better with the number of times the interactors were co-purified vs. purified individually:

$$S2 = \log\left[\frac{N_{12} \cdot N}{(N_1 + 1) \cdot (N_2 + 1)}\right]$$

Table S1: **Compatible subcellular localizations.** Subcellular localizations between which protein interactions were allowed are marked by dots.

	Bud	Cell wall	Plasma mem.	Cytosol	Nucleus	ER	Vacuole	Mitochondria
Bud	•	•	•	•				
Cell wall	•	•	•					
Plasma mem.	•	•	•	•				
Cytosol	•		•	•	•	•	•	
Nucleus				•	•			
ER				•		•		
Vacuole				•			•	
Mitochondria								•

where N_{12} is the number of purifications containing both proteins, N_1 and N_2 are the numbers of purifications containing either protein 1 or 2, and N is the total number of purifications.

Each of these four interaction sets were benchmarked against co-occurence on KEGG maps. Calibration curves for converting the raw quality scores to probabilistic confidence scores were obtained by fitting sigmoid functions to plots of specificity vs. score. This procedure is identical to that used for other evidence types in the STRING and ArrayProspector servers (*S10, S11*). Because no correlation was observed for false positive interactions from each of the four sets, the individual confidence scores were combined under the assumption of independence:

$$P = 1 - \prod_i (1 - P_i)$$

For this purpose, all interactions between members of a MIPS complex were assigned a confidence value of 0.95, which corresponds to the global agreement between MIPS complexes and KEGG maps.

Interaction filtering by subcellular localization

The final set of binary protein interactions was filtered to remove interactions between proteins with incompatible subcellular localizations. For this purpose, annotated subcellular localizations were downloaded from SGD (*S12*), which includes results of two large scale localization experiments (*S13, S14*).

Table S1 shows the compatibility matrix used for filtering the interactions. As some proteins have multiple subcellular localizations, a pair of proteins was allowed to interact if any of their subcellular localizations were compatible according to Table S1. Proteins of unknown subcellular localization were allowed to interact with all other proteins.

Automated extraction of periodic and non-periodic cell cycle proteins

A set of 600 periodically expressed yeast genes was identified from microarray expression studies (*S15, S16*) as described elsewhere (*S17*). However, as some pairs of periodically expressed genes encode identical proteins (CUP1-1/CUP1-2, HHF1/HHF2, HHT1/HHT2, YRF1-3/YRF1-7, YEL076C-A/YLR464W, and YLR036C/YMR051C), the set of 600 genes corresponds to 595 unique, periodically expressed proteins on which our analysis is based.

This set was augmented by non-periodically expressed ("static") proteins which interact with periodically expressed ("dynamic") ones from the set. For each static, the interaction partners at a confidence of 0.85 or better were examined; if at least 30% of these were dynamic, the static protein was considered cell cycle related. In the network, interactions with confidence scores down to 0.45 are shown among the proteins included. The extraction parameter values were chosen based on manual assessment of the biological correctness of the resulting networks. To allow detailed inspection, an interactive version is available online, as is a tab-delimited file of interactions and their associated confidence scores (`http://www.cbs.dtu.dk/cellcycle`).

We checked all MIPS complexes for overrepresentation of periodic proteins. For each complex, the statistical significance was calculated according to the hypergeometric distribution. Even at a significance level as low as 20%, all members of periodic complexes had already been captured by the approach described above.

Clustering of the cell cycle network

To objectively define the number of modules in the cell cycle network, we clustered the network with the OC implementation of means clustering (*S18*). The interaction confidence scores were used as pairwise similarity scores and a cutoff of 0.01 was applied to define the clusters. This resulted in 31 binary complexes and 29 modules of

three or more proteins.

Statistical significance of correlations

A hypergeometric distribution was used to test the significance of the correlations between transcriptional and post-translational control. 50 putative Cdc28p targets were also dynamic (out of 184 dynamic proteins in total) compared to a total of 59 putative targets among all 300 proteins in the network. This overrepresentation of putative Cdc28p targets among the dynamic proteins is significant at $P < 10^{-4}$.

Among the 300 proteins, 115 were predicted to contain at least one PEST region by the PESTfind program (*S19*). Of these 115 proteins, 34 were putative Cdc28p targets (out of 59 in total); PEST regions are thus overrepresented among Cdc28p targets at a significance level of $P < 10^{-3}$. Similarly, there are 81 dynamic proteins (of 184) among the 115 proteins with PEST regions, meaning that PEST regions are overrepresented among the dynamic proteins at $P < 10^{-2}$.

Error rate, coverage, and robustness

The cell cycle network is based on interactions from the curated MIPS complexes (*S5*) (important for comprehensiveness) as well as interactions from several large-scale experiments (*S1–S4*) (required for novelty). Unfortunately, the latter suffer from about 30–50% false positive interactions (*S20, S21*). For detailed studies of biological systems such as the cell cycle, it is thus of utmost importance to improve specificity. We employ a scheme (see "Materials and methods") that is specifically tailored to extract high-confidence interactions relevant to the cell cycle. In the following, we show that our procedure reduces the error rate by an order of magnitude, i.e. to 3–5%. Moreover, we show that our coverage is four-fold higher than for other integrated data sets with equally low error rates.

Estimating the error rate of our interaction data sets

We benchmark the performance of our extraction procedure by calculating the overlap between an interaction set and the curated MIPS complexes. For this we consider the following four sets:

- de Lichtenberg et al. "all" contains the union of the yeast two-hybrid screens by Uetz et al. (*S2*) and Ito et al. (*S1, full set*) and the matrix representations of the two complex pull-down experiments by Gavin et al. (*S3*) and Ho et al. (*S4*). This interaction set is the raw data for which the error rate has been estimated to be 30–50%.

- de Lichtenberg et al. "45%" consists of the subset of "all" to which we assign a topology-based quality score (see "Materials and methods") of at least 0.45, corresponding to the cutoff for including an interaction between two cell cycle proteins in the final network (Figure 1). Interactions were not filtered based on subcellular localization data.

- de Lichtenberg et al. "85%" consists of the subset of "all" to which we assign a topology-based quality score (see "Materials and methods") of at least 0.85, corresponding to the cutoff for including static proteins in the final network. Interactions were not filtered based on subcellular localization data.

- de Lichtenberg et al. "cell cycle" was constructed by applying the entire scheme described in "Materials and methods" to "all". The MIPS complexes were not included for this analysis.

Figure S1a shows the percentile overlap of each interaction data set with a matrix representation of the MIPS complexes. As expected, filtering the interactions based on our topology score considerably improves the overlap from 3.4% ("all") to 14.2% ("45%") and 21.3% ("85%"). Applying the full extraction scheme to create a cell cycle specific network further increases the overlap to 33.8%, i.e. a ten-fold improvement over the raw data.

Based on the estimate that the error rate in the raw data ("all") is 30–50%, we can estimate the error rate in any interaction data set from its overlap with the MIPS complexes:

$$error = error_{all} \cdot \frac{overlap_{all}}{overlap}$$

The estimated error rate of each interaction data set is shown in Figure S1b, assuming $error_{all}$ of 30% (best case) and 50% (worst case). When calculating these estimates, interactions involving ribosomal proteins were excluded (explained below).

These estimates demonstrate that our scoring and extraction procedure is capable of reducing the error rate by an order of magnitude, such that the resulting interaction network (based on high-throughput data alone) contain only 3–5% false positive interactions. The final network (Figure 1) contains 734 interactions among 300 proteins, of which 198 interactions are supported by high-throughput data alone. Consequently, only 6–10 wrong interactions should be expected in the entire network. As poorly characterized proteins only account for about 15% of these 198 interactions, only 1 or 2 of the assignments should be expected to be wrong. Even in the very unlikely event where all of the false positive interactions are related to poorly characterized proteins, it would still leave us with at least 20 correct novel functional assignments.

Comparison with other published interaction data sets

Numerous other integrative approaches and combined interaction sets have been published over the years based on the same raw data that we analyze:

- Han et al. (*S22*) recently analyzed an interaction data set derived from many of the same sources as those used in our work. Interactions were required to be present in at least two of the following five sets: the union

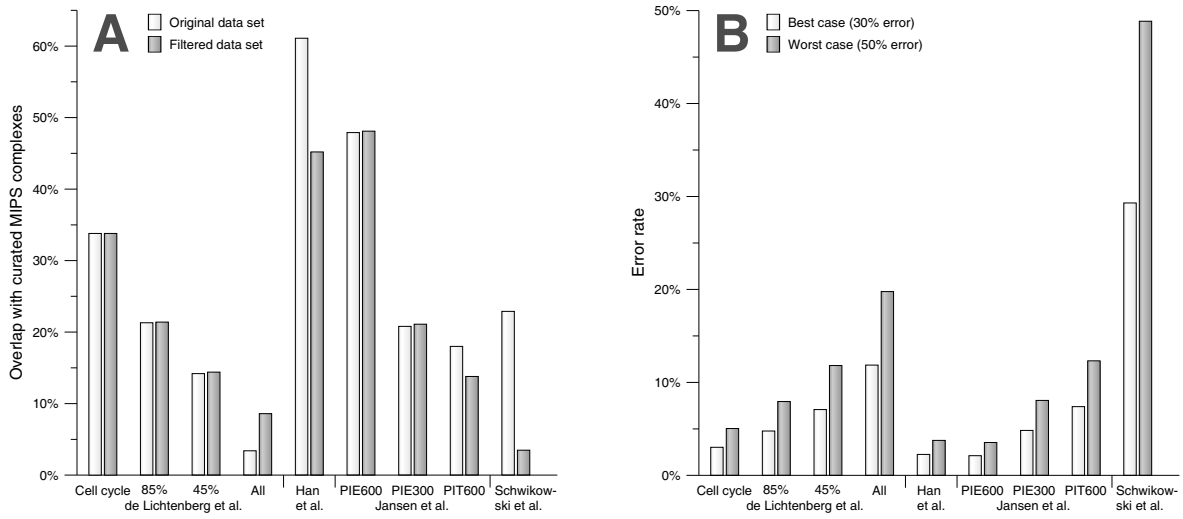

Figure S1: **Agreement with known complexes and error rate estimates.** The accuracy of four data sets from our extraction procedure and five data sets from other groups is compared. **A)** Overlap with the curated MIPS complexes. White bars show the agreement for the original data sets, while gray bars represent versions that have been modified to remove biases. For the Han et al. (*S22*) data set, the gray bar corresponds to not including the small-scale experiments and curated complexes from MIPS. The gray bar for Schwikowski et al. (*S23*) corresponds to removing the MIPS small-scale experiments (*S5*). For all other data sets, the gray bars correspond to excluding interactions involving ribosomal proteins. **B)** Error estimates. From the overlap with the MIPS complexes, the absolute error rate is estimated for each data set under the assumption of 30% (best case, white bars) and 50% (worst case, gray bars) error rate on the raw data.

of Uetz et al. (*S2*) and Ito et al. (*S1, full set*), the union of the spoke representations of Gavin et al. (*S3*) and Ho et al. (*S4*), potential interactions derived from genomic context (*S21*), the MIPS complexes in matrix representation, and the MIPS physical interactions from small-scale experiments (*S5*).

- The Jansen et al. PIE600 set (*S24*) was derived from the same four large-scale data sets (*S1–S4*) included in our "all" data set. Confidence scores were calculated from presence/absence in the four interaction data sets using a full Bayesian scheme.

- The Jansen et al. PIE300 set (*S24*) was constructed exactly like the PIE600 set except from using a less strict confidence cutoff.

- The Jansen et al. PIT600 set (*S24*) relies on a Bayesiean integration of the interaction data with additional data on mRNA expression, gene essentiality, and annotated function from both GeneOntology and MIPS.

- Schwikowski et al. (*S23*) constructed a data set of interactions from Uetz et al. (*S2*), Ito et al. (*S1*) and the MIPS physical interactions from small-scale experiments (*S5*).

Figure S1a shows how the overlap with the curated MIPS complexes compares across different data sets. In the data sets from Han et al. (*S22*) and Schwikowski et al. (*S23*), the small-scale MIPS physical interactions were included. This will likely cause the accuracy of these data sets to be overestimated as these data were probably used when curating the MIPS complexes. We thus benchmarked these two data sets both including and excluding the interactions obtained from MIPS (white and grey bars, respectively).

Ribosomal proteins are known to co-purify with a wide variety of proteins in complex pull-down experiments, leading to many false positive interactions in the Gavin et al. (*S3*) and Ho et al. (*S4*) data sets. Also, interactions among ribosomal proteins accounts for a large fraction of all true interactions. Ribosomal proteins alone can thus greatly influence the apparent quality of an interaction data set. For this reason, we re-benchmarked our own data and the interaction sets of Jansen et al. (*S24*) after removing all interactions involving ribosomal proteins (gray bars). The effect of this is seen most clearly in the raw data ("all").

Figure S1b shows the estimated error rate based on each of the filtered data sets (grey bars in Figure S1a). Only the data set of Han et al. (*S22*) and the two "PIE" data sets by Jansen et al. (*S24*) show an error rate as low or lower than our "85%" set.

Coverage and robustness

The "cell cycle" data set (or network) described above is based on large-scale data sets alone. It consists of 193 interactions among 188 proteins, of which 139 proteins were dynamic and 49 static.

In comparison, the same approach yields 183 dynamic and 98 static proteins when including the curated MIPS complexes (these numbers are lower than those mentioned in the manuscript, because the APC and SCF complexes were added manually). Although including curated complexes obviously improves the network, it is important to note that a large network can be obtained from using high-throughput data sets alone. Also, although there are several tunable parameters in our extraction method, the resulting network changes are minor for even large changes in parameters (data not shown). From these tests we conclude that the approach is highly robust.

For comparison, we also applied the extraction scheme to the Han et al. (*S22*) data set and the two "PIE" data sets by Jansen et al. (*S24*). For the Han et al. (*S22*) and Jansen et al. "PIE600" (*S24*) data sets, the resulting networks consist of only 25 interactions among 42 proteins and 8 interactions among 14 proteins, respectively. The low error rate of these two data sets is thus obtained at the price of a very low coverage. The potential for novel findings would thus be small if these data sets were used.

The Jansen et al. "PIE300" set (*S24*) has about the same global error rate as our "85%" set. When applying the network extraction scheme to the PIE300 set, a cell cycle network of 104 interactions among 121 proteins is obtained, i.e. about 40% smaller than obtained with our own interaction set. Surprisingly, the resulting network only shows 12.5% overlap with the curated MIPS complexes (*S5*) as compared to 33.8% for our "cell cycle" set. Consistent with the higher error rate and lower coverage of the "PIE300" derived network, it provides a cell cycle context for only five proteins of unknown function. Of these, three were also identified in our network, while the remaining two both seem highly unlikely to be correct, as one shows sequence similarity to tRNA-ligases and the other is localized to the mitochondria.

We thus conclude that our scoring and extraction scheme provides the best balance between specificity (low error rate) and sensitivity (coverage) of all the sets investigated.

Details on individual complexes

In this section, we provide additional information on individual proteins and modules included in our interaction network (Figure 1 in the paper). For detailed inspection of the network, we recommend that the reader simultaneously studies either the interactive Java version or the detailed PDF version of the network available from `http://www.cbs.dtu.dk/cellcycle`. They both contain names of individual proteins and allow the user to zoom in and out; the interactive version further allows the layout of the network to be changed.

New cell cycle contexts for uncharacterized proteins

In addition to the few examples mentioned in the paper, the temporal interaction networks allows us to place a handful of other uncharacterized proteins in a cell cycle context.

The temporal network reveals that Ypl014p, a nuclear/cytoplasmic (*S14*) protein of unknown function, is associated with the cyclin dependent kinase Cdc28p complexes: it has the same subcellular localization as its predicted interaction partners, its expression peaks at the M/G_1 transition and it has been co-purified with several other G_1 phase proteins including Sic1p, Cln1p, and Cln2p (*S3*). The association between the latter and Ypl014p was recently confirmed in a targeted proteomics study of the cyclin–Cdc28p module (*S25*). Furthermore, based on various properties of the protein itself, Ypl014p was predicted to play a role in the cell cycle (*S26*).

The two completely uncharacterized proteins Yml119p and Yll032p interact and are expressed close in time in G_2 phase. Their role in the cell cycle is further supported by recent, experimental evidence that both are Cdc28p substrates (*S27*). The uncharacterized Ypl208p is expressed in late G_1 phase and co-immunoprecipitates with Bcp1p in both of the analyzed complex pull-down screens (*S3,S4*). Both Ypl208p and Bcp1p are localized in the cytoplasm as well as the nucleus. In support of a role in the cell cycle, the *Schizosaccharomyces pombe* ortholog of Ypl208p (SPBC1709.13C) is also expressed in M or G_1 phase, which are hardly separable in *S. pombe* (*S28*). Bcp1p is a static protein involved in nuclear export and required for cell cycle progression (*S29*).

The uncharacterized protein Ylr254p is periodically expressed and interacts with Pac1p according to both yeast two-hybrid screens included in our analysis (*S1, S2*). Although Pac1p is not periodically expressed, it is known to be involved in nuclear migration (*S30*). Consistent with this, the expression of Ylr254p peaks in M phase.

In accordance with being involved in biosynthesis of the contractile ring, the chitin synthase Chs2p is expressed during M phase. In a yeast two-hybrid screen (*S1*), it was observed to interact with the putative protein kinase Isr1p, which is also periodically expressed. As phosphorylation events can sometimes be seen in yeast two-hybrid interactions (*S31*), it is tempting to speculate that Chs2p may be a Isr1p substrate.

Ydl156p, a protein of unknown function, has been co-purified with a complex of DNA repair proteins (*S3*), six of which are included in the cell cycle network (Msh2p, Msh3p, Msh6p, Rfa2p, Rfa3p, and Shs1p). Ydl156p is periodically expressed with a timing of peak expression similar to these genes, suggesting a role in DNA repair.

With relatively low confidence, the unchacterized Ycl042p interacts with a DNA replication protein, Rfc3p, based on co-purification evidence (*S4*). In support of the interaction, Ycl042p is expressed in G_1 phase like many known components of the DNA replication machinery.

Additional detail on the nucleosome/bud formation module

The cell cycle network (Figure 1) reveals a novel module that appears to connect processes related to chromosome structure with mitotic events in the bud (Figure 3a). The module is composed of two mitotic signaling kinases (Gin4p and Kcc4p), a histone variant (Htz1p), a nucleosome assembly protein (Nap1p), and two poorly characterized proteins (Nis1p and Yol070p). The two kinases (Gin4p and Kcc4p) are both expressed in G_1 phase and are involved in budding and assembly of the septin ring. Nap1p, which connects the module to the nucleosome complex, is known to shuttle between the nucleus and cytosol (*S32*) where it is required for activation and deactivation of Gin4p during mitosis (*S33*).

The module is supported by numerous independent physical interactions. Using Nap1p as bait, Gin4p, Kcc4p, Htz1p, Nis1p, and Yol070p were co-purified by Gavin et al. (*S3*). Nap1p also has also been observed to interact with Hta1p (*S2, S4*), Htb2p (*S3*), Gin4p (*S4*), Kcc4p (*S2*), and Nis1p (*S1,S2*). The inclusion of the potential Cdc28p substrate Yol070p in the module is further supported by a yeast two-hybrid interactions with Nis1p (*S2*). Both proteins are expressed in mitosis and localize to the bud neck.

The interactions of Nis1p with Nap1p, Gin4p, and Kcc1p have also been identified in small-scale experiments, based on which it was suggested to play a role in mitotic signaling (*S34*). Considering

that Nap1p interacts with Hta1p and Htb2p, its interaction with the histone H2A variant Htz1p is also very plausible. The observed interaction between Htz1p and Mps3p (*S2*) links the module to the sister chromatid assembly, which is consistent with the suggested role of the *S. pombe* ortholog of Htz1p in chromosome segregation (*S35*).

Additional detail on the pre-replication complex

Even though the pre-replication complex has been extensively studied, the network suggests several novel hypothesis related to this complex (Figure 3b). The array data show that the CDC45 gene is periodically transcribed and this may seem to conflict with the earlier observation that the protein level of Cdc45p is constant through the cell cycle (*S36*). However, the expression of this gene is induced right before the time where it is known to function in recruiting the replication machinery to origins of replication. Phosphorylation and other post-translation mechanisms are known to play a key role in regulating the assembly of the pre-replication complexes (*S37, S38*), and we propose that the conflicting observations for Cdc45p could be explained by simultaneous synthesis of new, unmodified Cdc45p and degradation of the modified form.

Furthermore, Mcm2p was co-purified with an uncharacterized putative helicase (Yil177p) and the trehalose complex (Tps1p, Tps2p, Tps3p, and Tsl1p) (*S3*). Yil177p belongs to a family of highly homologous putative helicases, Y' helicases, that all appear to be co-expressed in early G_1 phase (*S16*). Although the Mcm complex is known to exert helicase activity (*S38*) on its own, the interaction suggests that one or more Y' helicases could also be involved in this process. Yil177p and Tsl1p (the only dynamic subunit of the trehalose complex) appear functionally unrelated, however, their interaction is supported by the fact that the expression of the two genes peaks simultaneously during G_1 phase. Indeed, trehalose synthesis has been shown to be influenced by the length of the G_1 phase (*S39*).

Glycogen synthesis

Like the trehalose synthesis, glycogen synthesis is also affected by the length of the G_1 phase (*S39*). This is consistent with the observation that the major isoform of glycogen synthetase, Gsy2p, is expressed in G_1 phase. Two other glycogen biosynthesis genes are also included in the network: the minor isoform of glycogen synthetase (Gsy1p, expressed in late M phase), and a glycogen synthesis initiator protein (Glg2p, expressed in G_2 phase). In spite of the genes being expressed at very different phases of the cell cycle, all possible pairwise interactions have been observed in a yeast two-hybrid screen (*S1*). The functional importance of the expression patterns on complex assembly remains unclear, though.

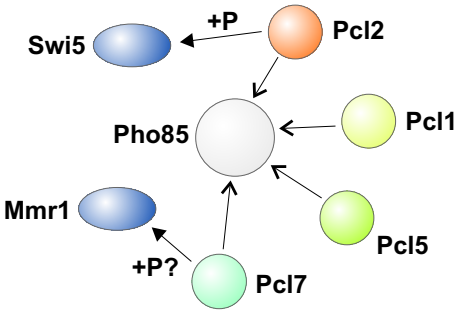

Figure S2: **The Pho85p cyclin module.** This module contains the cyclin dependent kinase Pho85p together with four of its associated cyclin (Pcl1p, Pcl2p, Pcl5p, and Pcl7p). The interactions correctly recover the known phosphorylation of Swi5p by Pho85p–Pcl2p and predict Mmr1p as a novel Pho85p–Pcl7p phosphorylation target.

Pho85p–cyclin complexes

One of the proteins known to regulate glycogen synthesis is the cell cycle related cyclin dependent kinase Pho85p (*S40*). Similar to the Cdc28p module, the network places Pho85p together with four of its associated cyclins. We assign the expression of the cyclins to peak at G_1 phase (Pcl1p and Pcl2p), S phase (Pcl5p), and G_2 phase (Pcl7p). Interestingly, Pcl7p is at the time of writing not listed as one of the cell cycle related Pho85-cyclins in SGD (*S12*). In the context of this module, we also find the cell cycle transcription factor Swi5p and the uncharacterized phosphoprotein Mmr1p. Swi5p interacts with Pcl2p, which is known to target it for phosphorylation by Pho85p (*S40*), possibly to affects its localization to the nucleus. Based on the interaction between Pcl7p and Mmr1p, we therefore propose Mmr1p as a novel Pho85p–Pcl7p target. In support of a cell cycle role, Mmr1p localizes to the bud via an mRNA transport system involving She2p, which is co-expressed with Mmr1p and also peaks in early M phase.

Cell cycle transcription factors

In addition to Swi5p, our network contains several other cell cycle related transcription factors. Swi6p and Mbp1p together form the transcription factor complex called MBF, involved predominantly in transcription of genes related to DNA replication and repair. Swi6p is synthesized in late S phase, whereas Mbp1p is expressed even later, in G_2 phase. This is very surprising, since MBF induces transcription of its targets in late G_1 phase, i.e. before the expression of both its subunits. Recent evidence suggest that Swi6p is exported from the nucleus around the G_2/M phase, somehow modified in the cytoplasm, and imported

back into the nucleus (*S41*). These observations could indicate a mechanisms in which an inactive form of Swi6p is synthesized and only activated after shuttling between the nucleus and the cytoplasm. In either case, the biological importance of the transcriptional regulation of Swi6p is not understood (*S42*).

Mbp1p has also been co-purified with Fkh1p, a forkhead transcription factor expressed in G_2, and a functionally uncharacterized protein expressed in late S phase, Ymr144p (*S4*). The role of Ymr144p in the cell cycle is strongly supported by chromatin-IP experiments as the YMR144W promoter is bound by Mbp1p, Swi4p, Fkh2p, and Ndd1p at $P < 10^{-3}$ in both the Simon et al. (*S43*) and Lee et al. (*S44*) experiments. Although we have found no known connection between Mbp1p and Fkh1p, their interaction is supported by recently published evidence of weak cooperativity between the two factors (*S45*).

Swi4p (which forms the G_1 transcription factor SBF with Swi6p) is found in the network together with Tbf1p (a DNA binding protein essential for mitotic growth) and Rad53p (a kinase involved in cell cycle arrest in response to DNA damage). SBF activates the transcription of another transcription factor, Tos4p, that interacts with the histone protein Hta1p, both of which are expressed around the G_1/S transition. In addition to being transcriptionally regulated, Tos4p is a potential Cdc28p substrate (*S27*).

Septin filaments

Also a transcription factor, Ace2p is known to localize to the bud and later to the nucleus of the daughter cell where it is responsible for expression of e.g. enzymes involved in septum degradation (*S46*). Consistent with this function, Ace2p directly interacts with the septin ring component Cdc11p with which it is also co-expressed in S phase. Cdc11p belongs to the conserved septin family, which was first discovered in yeast and subsequently found in numerous other fungi and animals, including human, mouse, *D. melanogaster*, and *C. elegans* (*S47, S48*). The four yeast septins that have been studied extensively (Cdc3p, Cdc10p, Cdc11p, and Cdc12p) are required for cytokinesis, axial bud site selection, as well as the correct localization of several other proteins involved in cytokinesis, morphogenesis, and bud site selection.

Anaphase promoting complex

The anaphase promoting complex (APC) has two regulatory subunits, Cdh1p and Cdc20p. Only the latter is periodically expressed and peaks in M phase. Cdh1p has an interaction with Clb2p, its well known substrate (*S49*).

One of the core subunits (Apc1p) of APC interacts with Yck3p, which clusters with Yck2p and Yck1p. All three are membrane-bound casein kinase I homologs, but only Yck1p is periodically expressed with peak expression in G_2 phase. In support of the interaction with the APC, at least two of the three kinases (Yck1p and Yck2p) appear to be involved in phosphorylation of PEST sequences required for the ubiquitination of some proteins, e.g. Ste3p and Fur4p (*S50, S51*).

Sister chromatid cohesin complex

In yeast, a complex known as cohesin (Smc1p, Smc3p, Irr1p, Mcd1p) is responsible for holding together the sister chromatids during S, G_2 and beginning of M phase (*S52*). In agreement with this, the network contains a cluster containing these four dynamic and co-expressed proteins; all synthesized in early S phase. The module also contains the periodically expressed kinase Cdc5p, which is synthesized in G_2 phase and plays important parts in regulating exit from mitosis (*S53*). Also linked to the cohesin complex, is an uncharacterized nuclear protein, Pwp1p, which has additional interactions with the histone Hta2p and the nucleolar protein Cbf5p. Recent studies of the human homolog, endonuclein, have found that it is up-regulated in pancreatic cancer and that both expression and localization of the protein varies in a cell cycle dependent manner (*S54*). These data clearly support a relation to the cell cycle, although the precise functional role of the protein is unknown in both yeast and human. The network context, however, may help to generate more precise hypotheses to be tested experimentally.

Mitotic exit

The segregation of the sister chromatids is mediated by the dissociation of the cohesin complex, trigged by the protease Esp1, which specifically targets Mcd1p. This protease is kept inactive by its association with the protein Pds1p, which is expressed in G_1 phase and targeted for degradation by the APC–Cdc20p complex at the onset of anaphase. The interaction between the static Esp1p and its dynamic partner Pds1p is captured in the network. Interestingly, we also observe an interaction between Esp1p and another dynamic protein, Amn1p, expressed at the M/G_1 transition. Amn1p has recently been identified as part of a daughter-specific system that resets the cell cycle after exit from mitosis. It competes with Cdc15p for binding to Tem1p, which is involved in activating the Cdc14p phosphatase that inhibits Cdc28p activity through a number of different mechanisms (*S53*). At the time of writing, however, no functional link between Amn1p and Esp1p appears to have been described in the literature. The interaction is interesting because Amn1p is synthesized shortly after Pds1p is destroyed, suggesting that Amn1p could associate

with the free form of Esp1p, possibly after its cleaveage of Mcd1p.

Esp1p also interacts with Spt16p, which is expressed in G_1 phase and involved in RNA polymerase II transcription. Spt16p interacts with another RNA polymerase II related protein, Paf1p, which is expressed in M phase and required for expression of key regulatory cell cycle genes (*S55*).

Two distinct types of modules and protein hubs

This section reports the existence of two types of modules and shows that these are closely related to two recently discovered classes of protein hubs.

Two distinct types of modules

In the derived cell cycle network (Figure 1) we observe that interacting proteins are typically expressed close in time; this is quantified by the histogram in Figure S3a, showing the distribution of distance in the cell cycle between interacting dynamic proteins. The majority of complexes and modules thus contain dynamic subunits that are expressed at approximately the same time (Figure S3a). Examples of this include the nucleosome complex, the DNA replication machinery, and the spindle pole body (SPB) (see Figure 1 and "Details on individual complexes"). However, there is also a second rarer type of module in which the proteins are expressed at different stages of the cell cycle; these are exemplified by the Cdc28p complexes and the novel nucleosome–bud linker module (see Figure 3a,c and "Details on individual complexes").

Party and date hubs

To better understand the biological importance of these two classes of modules, we concentrated on the most highly connected proteins, also known as "hubs". We defined hubs as proteins with more than five interaction partners and computed the average difference in peak time (distance in the cell cycle) among the interaction partners of each hub. As shown in Figure S3b the distribution of these average distances is bimodal, supporting the recent proposal of two distinct classes of protein hubs, namely "party hubs" (interaction partners are co-expressed) and "date hubs" (interaction partners vary and are condition-dependent) (*S22*). These classes were discovered by Han et al. based on gene expression across different cellular conditions (*S22*); however, our results show that they also exist in the context of a single dynamic process and that they are closely related two the two types of modules observed in the cell cycle interaction network (Figure 1). A prime example of a date hub is the static cyclin-dependent kinase Cdc28p, the activity and specificity of which is regulated by association with different interaction partners at different stages of the cell cycle (see Figure 3c).

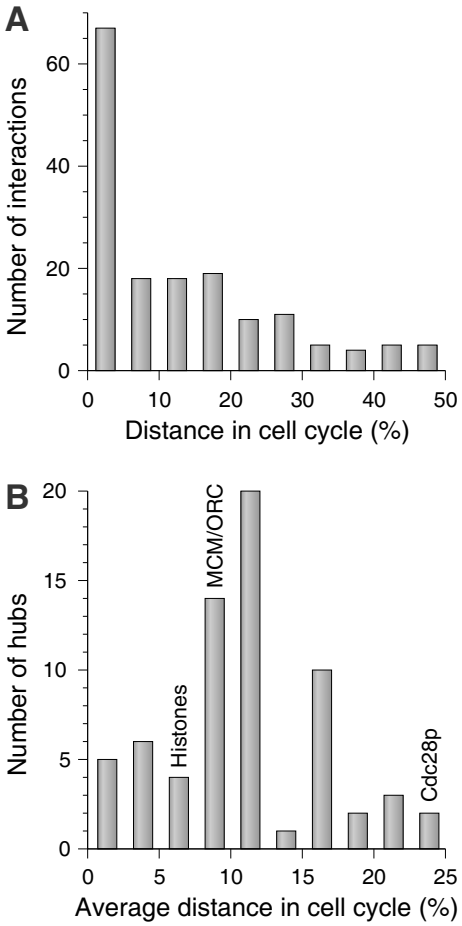

Figure S3: **Timing of expression for interaction partners. A)** The distribution of time differences between dynamic interaction partners reveals a strong preference for interactions between proteins expressed close in time. **B)** For each hub protein, the average difference in peak time was calculated among the dynamic interaction partners. The distribution is bimodal, supporting the proposition of two classes of hubs, namely "party" and "date" hubs (*S22*). Representative hub proteins (histones, MCM/ORC and Cdc28p) are marked in the figure to illustrate the relation between the two types of hubs and modules.

References

S1 T. Ito, *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1143 (2000).

S2 P. Uetz, *et al.*, *Nature* **403**, 623 (2000).

S3 A.-C. Gavin, *et al.*, *Nature* **415**, 141 (2002).

S4 Y. Ho, *et al.*, *Nature* **415**, 180 (2002).

S5 H. W. Mewes, *et al.*, *Nucleic Acids Res.* **32**, D41 (2004).

S6 T. Lang, *et al.*, *EMBO J.* **17**, 3597 (1998).

S7 T. Kirisako, *et al.*, *J. Cell Biol.* **147**, 435 (1999).

S8 W.-P. Huang, S. Scott, D. Kim, J.and Klionsky, *J. Biol. Chem.* **275**, 5845 (2000).

S9 R. Saito, H. Suzuki, Y. Hayashizaki, *Nucleic Acids Res.* **30**, 1163 (2002).

S10 C. von Mering, *et al.*, *Nucleic Acids Res.* **31**, 258 (2003).

S11 L. J. Jensen, J. Lagarde, C. von Mering, P. Bork, *Nucleic Acids Res.* (2004).

S12 K. R. Christie, *et al.*, *Nucleic Acids Res.* **32**, D311 (2004).

S13 A. Kumar, *et al.*, *Genes Dev.* **16**, 707 (2002).

S14 W. K. Huh, *et al.*, *Nature* **425**, 686 (2003).

S15 R. J. Cho, *et al.*, *Molecular Cell* **2**, 65 (1998).

S16 P. T. Spellman, *et al.*, *Mol. Biol. Cell* **9**, 3273 (1998).

S17 U. de Lichtenberg, *et al.*, *Bioinformatics* (2004). Doi:10.1093.

S18 G. J. Barton, Oc a cluster analysis program (1993).

S19 M. Rechsteiner, S. W. Rogers, *Trends Biochem. Sci.* **21**, 267 (1996).

S20 A. M. Edwards, *et al.*, *Trends in Genetics* **18**, 529 (2002).

S21 C. von Mering, *et al.*, *Nature* **417**, 399 (2002).

S22 J. D. Han, *et al.*, *Nature* **430**, 88 (2004).

S23 B. Schwikowski, P. Uetz, S. Fields, *Nature Biotechnology* **18**, 1257 (2000).

S24 R. Jansen, *et al.*, *Science* **302**, 449 (2003).

S25 V. Archambault, *et al.*, *Molecular Cell* **14**, 699 (2004).

S26 U. de Lichtenberg, T. S. Jensen, L. J. Jensen, S. Brunak, *J. Mol. Biol.* **329**, 663 (2003).

S27 J. A. Ubersax, *et al.*, *Nature* **425**, 859 (2003).

S28 G. Rustici, *et al.*, *Nature Genetics* **36**, 809 (2004).

S29 A. Audhya, S. D. Emr, *EMBO J.* **22**, 4223 (2003).

S30 W. L. Lee, J. R. Oberle, J. A. Cooper, *J. Cell. Biol.* **160**, 355 (2003).

S31 L. J. Jensen, P. Bork, *Drug Discovery Today: TARGETS* **3**, 51 (2004).

S32 M. Miyaji-Yamaguchi, *et al.*, *Molecular and Cellular Biology* **23**, 6672 (2003).

S33 R. Altman, D. Kellogg, *Journal of Cell Biology* **138**, 119 (1997).

S34 M. Iwase, A. Toh-e, *Genes Genet. Syst.* **76**, 335 (2001).

S35 A. M. Carr, *et al.*, *Mol. Gen. Genet.* **245**, 628 (1994).

S36 J. C. Owens, C. S. Detweiler, J. J. Li, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12521 (1997).

S37 M. L. DePamphilis, *Gene* **310**, 1 (2003).

S38 S. P. Bell, A. Dutta, *Annual Review of Biochemistry* **71**, 333 (2002).

S39 J. W. Paalman, *et al.*, *FEMS Yeast Res.* **3**, 261 (2003).

S40 V. Measday, H. McBride, J. Moffat, D. Stillman, B. Andrews, *Molecular Microbiology* **35**, 825 (2000).

S41 E. Queralt, J. C. Igal, *Mol. Cell. Biol.* **23**, 3126 (2003).

S42 L. L. Breeden, *Current Biology* **13**, R31 (2003).

S43 I. Simon, *et al.*, *Cell* **106**, 697 (2001).

S44 T. I. Lee, *et al.*, *Science* **298**, 799 (2002).

S45 N. Banerjee, M. Q. Zhang, *Nucleic Acids Res.* **31**, 7024 (2003).

S46 E. L. Weiss, *et al.*, *J. Cell Biol.* **158**, 885 (2002).

S47 M. S. Longtine, *et al.*, *Curr. Opin. Cell Biol.* **8**, 106 (1996).

S48 C. M. Field, D. Kellogg, *Trends Cell Biol.* **9**, 387 (1999).

S49 M. Tyers, P. Jorgensen, *Curr. Opin. Genet. Dev.* **10**, 54 (2000).

S50 Y. Feng, N. G. Davis, *Mol. Cell. Biol.* **20**, 5350 (2000).

S51 C. Marchal, R. Haguenauer-Tsapis, D. Urban-Grimal, *J. Biol. Chem.* **275**, 23608 (2000).

S52 K. Nasmyth, J. M. Peters, F. Uhlmann, *Science* **288**, 1379 (2000).

S53 Y. Wang, T. Shirogane, D. Liu, J. W. Harper, S. J. Elledge, *Cell* **112**, 697 (2003).

S54 B. Honore, U. Baandrup, S. Nielsen, H. Vorum, *Oncogene* **21**, 1123 (2002).

S55 C. Koch, P. Wollmann, M. Dahl, F. Lottspeich, *Nucleic Acids Res.* **27**, 2126 (1999).