# Cyclebase.org: version 2.0, an updated comprehensive, multi-species repository of cell cycle experiments and derived analysis results

**Nicholas Paul Gauthier[1,2], Lars Juhl Jensen[3], Rasmus Wernersson[4], Søren Brunak[3,4],* and Thomas S. Jensen[4]**

[1]Computational Biology Center, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10065, [2]Gerstner Sloan-Kettering Graduate School of Biomedical Sciences, 1275 York Avenue, New York, NY 10065, USA, [3]NNF Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, Blegdamsvej 3b, DK-2200 Copenhagen N, Denmark and [4]Center for Biological Sequence Analysis, BioCentrum, Technical University of Denmark, Building 208, DK-2800 Lyngby, Denmark

## ABSTRACT

Cell division involves a complex series of events orchestrated by thousands of molecules. To study this process, researchers have employed mRNA expression profiling of synchronously growing cell cultures progressing through the cell cycle. These experiments, which have been carried out in several organisms, are not easy to access, combine and evaluate. Complicating factors include variation in interdivision time between experiments and differences in relative duration of each cell-cycle phase across organisms. To address these problems, we created Cyclebase, an online resource of cell-cycle-related experiments. This database provides an easy-to-use web interface that facilitates visualization and download of genome-wide cell-cycle data and analysis results. Data from different experiments are normalized to a common timescale and are complimented with key cell-cycle information and derived analysis results. In Cyclebase version 2.0, we have updated the entire database to reflect changes to genome annotations, included information on cyclin-dependent kinase (CDK) substrates, predicted degradation signals and loss-of-function phenotypes from genome-wide screens. The web interface has been improved and provides a single, gene-centric graph summarizing the available cell-cycle experiments. Finally, key information and links to orthologous and paralogous genes are now included to further facilitate comparison of cell-cycle regulation across species. Cyclebase version 2.0 is available at http://www.cyclebase.org.

## INTRODUCTION

The process by which cells replicate and pass on their genetic information, termed the cell cycle, is fundamental to life and has been intensely studied in the biological sciences. The past decade has witnessed an explosion in data derived from cell-cycle specific and other high-throughput experiments. These data include mRNA expression profiling using microarrays (1–9), overexpression (10,11) and knock-down studies (12), prediction of degradation signals (13), and systematic determination of kinase substrates (14–16). Of particular interest are the mRNA profiling experiments, which are performed on samples aliquoted from synchronously growing cells progressing through the cell cycle. These studies provide a wealth of transcriptome data during the division process, which can be analyzed to deduce the subset of genes that are subjected to transcriptional regulation during the cell cycle. Gathering, comparing and analyzing such a vast amount of data require a significant effort.

In order to address the problems mentioned above, we developed Cyclebase (17), a web resource of cell-cycle microarray data sets and derived analysis results. The database was filled with over 20 time-series microarray experiments. In order to remove experimental condition differences and variation in the speed with which cells progress through the cell cycle, experimental data from each study were first normalized to a common time scale. Data from multiple studies were then plotted on a single chart for each gene. This intuitive visual representation, which depicts hundreds of experimental

measurements in a single image, allows researchers to easily compare expression profiles across studies and gage the reproducibility of the experimental data. Each graph was supplemented by results from state-of-the-art analyses, including measures for periodicity, magnitude of regulation and the point in the division process when the transcription level is highest.

The first version of Cyclebase made it possible to easily assess transcriptional regulation of individual genes in single organisms. However, within the cell-cycle community there is a need for comparing both conservation of transcriptional regulation across species as well as assessing additional cell-cycle relevant information. To address these needs, we have expanded the functionality of Cyclebase, and further updated the database to account for changes in genomic annotations.

## CYCLEBASE VERSION 2.0

In order to provide easier access to more information about each genes' role in the cell cycle, we have performed a major update of Cyclebase. The Gene Details page, which is the centerpiece of the web site, contains many of these updates (Figure 1). This section highlights the major additions and changes to Cyclebase and describes its core components.

### Display of orthologous and paralogous genes

The recent findings that cell-cycle regulation is only rarely conserved at the individual gene level, but appears to be conserved at higher systemic levels (13), highlight the importance of comparing transcriptional regulation across species. To facilitate such comparisons, each gene is now supplemented with a list of orthologous and paralogous genes found in Cyclebase (Figure 1a4). These assignments were taken from the eggNOG database (18). This list contains analysis results, a link to display Reflect information (19), and an icon that, when clicked on, displays a graphic of all available normalized expression profiles for the ortholog or paralog selected (see Figure 1b). Multiple expression profiles can be opened at the same time, further easing comparison between homologous genes across organisms.

### Addition of cell-cycle relevant data

Transcriptional regulation is one of the several regulatory layers used to control the cell cycle. Easy access to additional data relevant to the division process helps to facilitate studies that focus on the interplay between different regulatory mechanisms. Genes in Cyclebase version 2.0 now include a variety of other data related to the cell cycle. We have included cell cycle relevant features such as lists of CDK substrates (14–16), degradation motifs (13) and phenotypic effects of knock-down (12) and overexpression (10,11) experiments. These 'gene features' are presented on the Gene Details page (Figure 1a2).

### Ability to search using BLAST

As with the original version of Cyclebase, the web-interface still queries for genes by name, alias and description. Users can continue to browse all the genes within an organism, select example genes or enter complex queries through the Advanced Search page. In addition, Cyclebase version 2.0 introduces the ability to query for genes using either amino acid or nucleotide sequence, which can be useful when performing detailed searches, e.g. searching for specific genomic sequences in the human data derived from cDNA microarray experiments. Users can either enter the primary sequence directly into the search field or use the Advanced Search feature to input a FASTA entry. Genes are queried with both BLASTP and BLASTX, the results are combined and by default are sorted by *E*-value.

### Update to core Cyclebase components

In addition to the more visible updates, several aspects in the underlying data structure have also changed. For example, the original version of Cyclebase was organized around microarray probesets rather than genes. Multiple probesets often target the same gene and, unfortunately, single probesets may target multiple genes (i.e. there is a many-to-many relationship). Centering the new version of Cyclebase around genes, the new interface is more intuitive and warns users when a many-to-many relationship exists for the gene/probeset they are viewing. In another major change to the backend database, we have updated all data sets to account for changes in genome annotations, which provides up-to-date lists of periodically expressed genes.

Cyclebase continues to provide full documentation of analysis methodology, frequently asked questions and information on each individual experiment. In addition, well-documented downloads are available for all analysis results and, when permission from original authors has been given, normalized expression data for each experiment. All the documentation has been updated to account for the changes introduced in Cyclebase version 2.0 and all downloads have been updated with more recent genome annotations.

## PERSPECTIVES

With the new functional improvements and the updated backend, Cyclebase is well positioned to store and present other temporal cell-cycle-related data sets, e.g. protein and phospho-protein expression profiles. Although only sparsely available right now, experiments that generate these types of data are expected to become more and more common in the future. Such data will help deconvolute the complexity of cell-cycle regulation, allowing researchers to further understand how regulatory mechanisms evolve, how differentiation and the cell cycle are intimately linked and how errors in the process can lead to complicated diseases such as cancer.
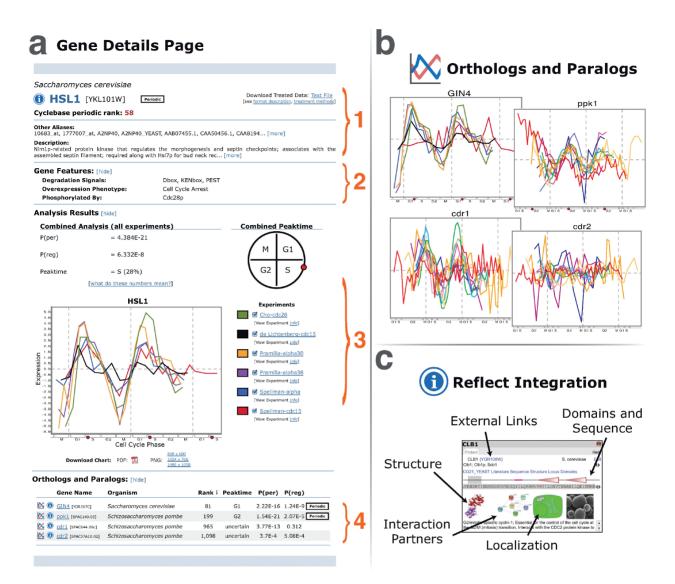
**Figure 1.** (**a**) General overview of the Cyclebase Gene Details page. (**a1**) Header information displays gene name, Cyclebase periodic ranking, aliases, description and links to download raw data. (**a2**) Annotations provide information about predicted degradation signals, kinases that phosphorylate the protein and results of overexpression and knock-down experiments. (**a3**) Analysis results (*P*-value for periodicity, *P*-value for regulation and peaktime value) along with an graphic depicting the peaktime are displayed above the expression chart. This chart shows all the available experiments for a given gene, each normalized to the same time scale, allowing the *x*-axis to be shown as phases of the cell cycle. Researchers can download the graph in both vector graphic (PDF) and image (PNG) format for use in their own publications. (**a4**) Orthologous and paralogous genes are shown in the same table format as the search screen. Users quickly get an overview of the similarity across organisms and can click on each gene name to see the full Cyclebase entry. (**b**) Clicking the chart preview icon for any ortholog or paralog expands a chart for that gene. Multiple charts can be opened simultaneously to further aid cross and inter-species gene comparisons. (**c**) Clicking on most gene names or information icons throughout Cyclebase provides a Reflect pop-up, which presents a variety of information about the gene selected.

## REFERENCES

1. Cho,R.J., Campbell,M.J., Winzeler,E.A., Steinmetz,L., Conway,A., Wodicka,L., Wolfsberg,T.G., Gabrielian,A.E., Landsman,D., Lockhart,D.J. *et al.* (1998) A genome-wide transcriptional analysis of the mitotic cell cycle. *Molecular Cell*, **2**, 65–73.
2. Lichtenberg,U.D., Wernersson,R., Jensen,T.S., Nielsen,H.B., Fausbøll,A., Schmidt,P., Hansen,F.B., Knudsen,S. and Brunak,S. (2005) New weakly expressed cell cycle-regulated genes in yeast. *Yeast*, **22**, 1191–1201.
3. Menges,M., Hennig,L., Gruissem,W. and Murray,J.A.H. (2003) Genome-wide gene expression in an Arabidopsis cell suspension. *Plant Mol. Biol.*, **53**, 423–442.
4. Oliva,A., Rosebrock,A., Ferrezuelo,F., Pyne,S., Chen,H., Skiena,S., Futcher,B. and Leatherwood,J. (2005) The cell cycle-regulated genes of Schizosaccharomyces pombe. *PLoS Biol.*, **3**, e225.

5. Peng,X., Karuturi,R.K.M., Miller,L.D., Lin,K., Jia,Y., Kondu,P., Wang,L., Wong,L.-S., Liu,E.T., Balasubramanian,M.K. *et al.* (2005) Identification of cell cycle-regulated genes in fission yeast. *Mol. Biol. Cell*, **16**, 1026–1042.

6. Pramila,T., Wu,W., Miles,S., Noble,W.S. and Breeden,L.L. (2006) The forkhead transcription factor hcm1 regulates chromosome segregation genes and fills the s-phase gap in the transcriptional circuitry of the cell cycle. *Genes Dev.*, **20**, 2266–2278.

7. Rustici,G., Mata,J., Kivinen,K., Lió,P., Penkett,C.J., Burns,G., Hayles,J., Brazma,A., Nurse,P. and Bähler,J. (2004) Periodic gene expression program of the fission yeast cell cycle. *Nat. Genet.*, **36**, 809–817.

8. Spellman,P.T., Sherlock,G., Zhang,M.Q., Iyer,V.R., Anders,K., Eisen,M.B., Brown,P.O., Botstein,D. and Futcher,B. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Mol. Biol. Cell*, **9**, 3273–3297.

9. Whitfield,M.L., Sherlock,G., Saldanha,A.J., Murray,J.I., Ball,C.A., Alexander,K.E., Matese,J.C., Perou,C.M., Hurt,M.M., Brown,P.O. *et al.* (2002) Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol. Biol. Cell*, **13**, 1977–2000.

10. Niu,W., Li,Z., Zhan,W., Iyer,V.R. and Marcotte,E.M. (2008) Mechanisms of cell cycle control revealed by a systematic and quantitative overexpression screen in S. cerevisiae. *PLoS Genet.*, **4**, e1000120.

11. Sopko,R., Huang,D., Preston,N., Chua,G., Papp,B., Kafadar,K., Snyder,M., Oliver,S.G., Cyert,M., Hughes,T.R. *et al.* (2006) Mapping pathways and phenotypes by systematic gene overexpression. *Mol. Cell*, **21**, 319–330.

12. Mukherji,M., Bell,R., Supekova,L., Wang,Y., Orth,A.P., Batalov,S., Miraglia,L., Huesken,D., Lange,J., Martin,C. *et al.* (2006) Genome-wide functional analysis of human cell-cycle regulators. *Proc. Natl Acad. Sci USA*, **103**, 14819–14824.

13. Jensen,L.J., Jensen,T.S., Lichtenberg,U.D., Brunak,S. and Bork,P. (2006) Co-evolution of transcriptional and post-translational cell-cycle regulation. *Nature*, **443**, 594–597.

14. Diella,F., Gould,C.M., Chica,C., Via,A. and Gibson,T.J. (2008) Phospho.elm: a database of phosphorylation sites–update 2008. *Nucleic Acids Res.*, **36**, D240–D244.

15. Loog,M. and Morgan,D.O. (2005) Cyclin specificity in the phosphorylation of cyclin-dependent kinase substrates. *Nature*, **434**, 104–108.

16. Ubersax,J.A., Woodbury,E.L., Quang,P.N., Paraz,M., Blethrow,J.D., Shah,K., Shokat,K.M. and Morgan,D.O. (2003) Targets of the cyclin-dependent kinase cdk1. *Nature*, **425**, 859–864.

17. Gauthier,N.P., Larsen,M.E., Wernersson,R., Lichtenberg,U.D., Jensen,L.J., Brunak,S. and Jensen,T.S. (2008) Cyclebase.org–a comprehensive multi-organism online database of cell-cycle experiments. *Nucleic Acids Res.*, **36**, D854–D859.

18. Muller,J., Szklarczyk,D., Julien,P., Letunic,I., Roth,A., Kuhn,M., Powell,S., von Mering,C., Doerks,T., Jensen,L.J. *et al.* (2009) eggNOG v2.0: extending the evolutionary genealogy of genes with enhanced non-supervised orthologous groups, species and functional annotations. *Nucleic Acids Res.*, **38**, D190–D195.

19. Pafilis,E., O'Donoghue,S.I., Jensen,L.J., Horn,H., Kuhn,M., Brown,N.P. and Schneider,R. (2009) Reflect: augmented browsing for the life scientist. *Nat. Biotechnol.*, **27**, 508–510.