Cyclebase 3.0: a multi-organism database on cell-cycle regulation and phenotypes

Alberto Santos[1], Rasmus Wernersson[2,3] and Lars Juhl Jensen[1,*]

[1]Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen, Denmark, [2]Center for Biological Sequence Analysis, Technical University of Denmark, 2800 Lyngby, Denmark and [3]Intomics A/S, 2800 Lyngby, Denmark

ABSTRACT

The eukaryotic cell division cycle is a highly regulated process that consists of a complex series of events and involves thousands of proteins. Researchers have studied the regulation of the cell cycle in several organisms, employing a wide range of high-throughput technologies, such as microarray-based mRNA expression profiling and quantitative proteomics. Due to its complexity, the cell cycle can also fail or otherwise change in many different ways if important genes are knocked out, which has been studied in several microscopy-based knockdown screens. The data from these many large-scale efforts are not easily accessed, analyzed and combined due to their inherent heterogeneity. To address this, we have created Cyclebase—available at http://www.cyclebase.org—an online database that allows users to easily visualize and download results from genome-wide cell-cycle-related experiments. In Cyclebase version 3.0, we have updated the content of the database to reflect changes to genome annotation, added new mRNA and protein expression data, and integrated cell-cycle phenotype information from high-content screens and model-organism databases. The new version of Cyclebase also features a new web interface, designed around an overview figure that summarizes all the cell-cycle-related data for a gene.

INTRODUCTION

One of the arguably most fundamental processes to eukaryotic life is the mitotic cell cycle, the process through which a cell replicates its genetic material and divides to become two cells. This process has thus been intensely studied for decades in several model organisms, both at the molecular level and at the phenotypic level.

Today, numerous large-scale datasets related to the mitotic cell cycle exist. These include microarray-based time courses of mRNA expression (1–9), mass-spectrometry-based proteomics on protein expression during the cell cycle (10,11), systematic screens for cyclin-dependent kinase (CDK) substrates (12,13) and high-content screening for knockdown phenotypes (14–22). Together, these datasets provide a wealth of information on the mitotic cell cycle and its many regulatory layers. However, it takes great effort to collect, analyze and combine this amount of heterogeneous data, especially when it is scattered across databases and supplementary files from articles.

The aim of Cyclebase is to address exactly that problem. Earlier versions of Cyclebase primarily addressed the challenge of jointly analyzing and visualizing the many available mRNA expression time courses for a gene and to allow easy comparison across orthologous and paralogous genes. In this new version, we have greatly expanded the scope of the database to include also the results from more recent proteomics and high-content phenotype screening efforts. To accommodate these new types of data into the resource, we have completely redesigned the web interface and the underlying database architecture. The centerpiece of the new interface provides a simple overview of the complex underlying data on the cell-cycle regulation and phenotypes of a gene.

NEW AND UPDATED DATA IN CYCLEBASE 3.0

All data for a given organism in Cyclebase is mapped onto a common set of genes. In version 3.0, we have updated these gene sets to be consistent with the latest version of the eggNOG database (23), from which we also obtain information on orthologs and paralogs.

In addition to remapping all existing microarray studies from the previous version of Cyclebase, we have incorporated one additional study for *Saccharomyces cerevisiae*, which used tiling arrays to globally measure gene expression with high temporal resolution in synchronized cell cultures (9). After inclusion of these additional time courses, we reassessed the periodicity of all *S. cerevisiae* genes and recal-

culated the time of peak expression for all genes deemed periodic.

For human genes, we have complemented the existing microarray expression data with data from two quantitative proteomics studies (10,11). Both studies used mass spectrometry to quantify protein levels in cell cultures from six different time intervals of the cell cycle, which approximately represent G1, G1/S, early S, late S, G2 and M phase. To make the two datasets as comparable to each other as possible, we represent the observed intensity value for each time interval as the intensity ratio relative to unsynchronized cells, which both studies also measured.

Post-translational regulation is at least as important as transcriptional regulation and explains much of the difference observed between transcriptomics and proteomics studies. To capture also this aspect of cell-cycle regulation, we import information on experimentally determined substrates of cell-cycle-related kinases from the latest version of the Phospho.ELM database (24). Unlike earlier versions of Cyclebase, we import information not only for members of the CDK family of kinases, but also for the Polo, Aurora, NEK and DYRK families. For *S. cerevisiae*, we also import the Cdc28p substrates identified in two high-throughput screens (12,13). For CDKs, we complement the known substrates with sequence-based predictions using a regular expression for a CDK consensus site ([S/T]P.[KR]); we opted to use this simple approach because the results were nearly identical to the highest confidence CDK sites predicted by the NetPhorest tool (25). As in earlier versions, we also include sequence-based predictions of PEST regions and D-box and KEN-box motifs, all of which suggest that the protein may be targeted for ubiquitin-mediated proteasomal degradation. As a new feature, we now also include data from two mass-spectrometry studies of a cell-cycle-related Posttranslational Modifications (PTM) closely related to ubiquitination, namely SUMOylation (26,27).

The largest addition of new data to Cyclebase v3.0, however, comes from high-throughput screens and database annotations of cell-cycle phenotypes. Recent years have seen a flood of such datasets made by automated microscopy and siRNA knockdown in human cell lines, such as the Mitocheck Project (19,20) and several other high-content screens (15–18,21). To ensure the quality of the data integrated in Cyclebase, we filtered these screens to include only phenotypes observed with at least half of the siRNAs targeting the gene in question. Large-scale cell-cycle phenotype screens also exist for *S. cerevisiae* (14) and *Schizosaccharomyces pombe* (22). However, as these screens are included in the respective model organism databases (28,29) along with phenotype annotations from many other experiments, we opted to import phenotype associations from these databases instead of the individual screens. To standardize the phenotype terminology, we made use of existing ontologies, namely the Cellular Microscopy Phenotype Ontology for the screens of human cell lines and the Ascomycete Phenotype Ontology and the Fission Yeast Phenotype Ontology (30) for the two yeasts. As these are organism-specific ontologies, we unified them using the entity–quality model to map all their cell-cycle-related terms to combinations of Gene Ontology and Phenotypic Quality Ontology terms.

These common terms are the basis for the visualization of the phenotypes.

REDESIGNED WEB INTERFACE

Until now, the main focus of Cyclebase web interface was on visualizing data from time-course microarray studies. Because of the broader scope of Cyclebase 3.0, we have completely redesigned the web interface to put less emphasis on the showing expression profiles in individual experiments and more focus on providing an overview of the heterogeneous cell-cycle data related to a gene of interest.

The central part of the new web interface is the new overview figure (Figure 1), which aims to provide an at-a-glance overview of the transcriptional and post-translational regulation during the cell cycle as well as the cell-cycle-related phenotypes. To this end, we map the information onto a schematic of the cell-cycle phases. Inside the circle representing the phases, we summarize the available expression data. We show a running average of the data from transcriptomics time courses as a circular blue scale heat map, with a red arrow designating the estimated time of peak expression. When proteomics data are available—presently only the case for human—we show these inside of the transcriptomics data. As for the latter, we show the average of the results from the available experiments; however, because the proteomics time courses are not nearly as finely time resolved, we display the average expression within each of six-phases window instead of as a running average. Outside the schematic of the phases we show icons that represent the presence of certain PTM sites and degradation signals as well as observed cell-cycle phenotypes. The icons are placed at the point in the cell cycle that they are primarily associated with. Hovering over any part of the overview figure will provide additional information in the form of tooltips.

For users who want to see further on the temporal regulation, Cyclebase 3.0 provides plots of the temporal mRNA expression profile of a gene according to each individual time-course experiment. This visualization will be familiar to users of earlier versions of Cyclebase but features improved interactivity and visual quality on modern displays. We have achieved this by reimplementing the functionality using client-side vector graphics instead of server-side bitmap graphics.

Two tables below the two visualizations provide further detail on the PTMs/degradation signals and phenotypes, respectively. In addition to what can also be seen in the overview figure, the tables show the source of the evidence and provides linkouts to the source whenever possible. The phenotype table furthermore lists additional phenotypes that are related to the cell cycle, but cannot be associated with any particular phase or transition.

The last table on the page about a gene shows its orthologs and paralogs in the organisms covered by Cyclebase. Like in the previous version of the database, the table shows the results of the analyses of transcriptomics studies for each gene. We have made this part of the table more compact so that it now only shows for each gene if it was deemed periodically expressed or not, and in which phase or at which transition its transcript level peaks. This allows us
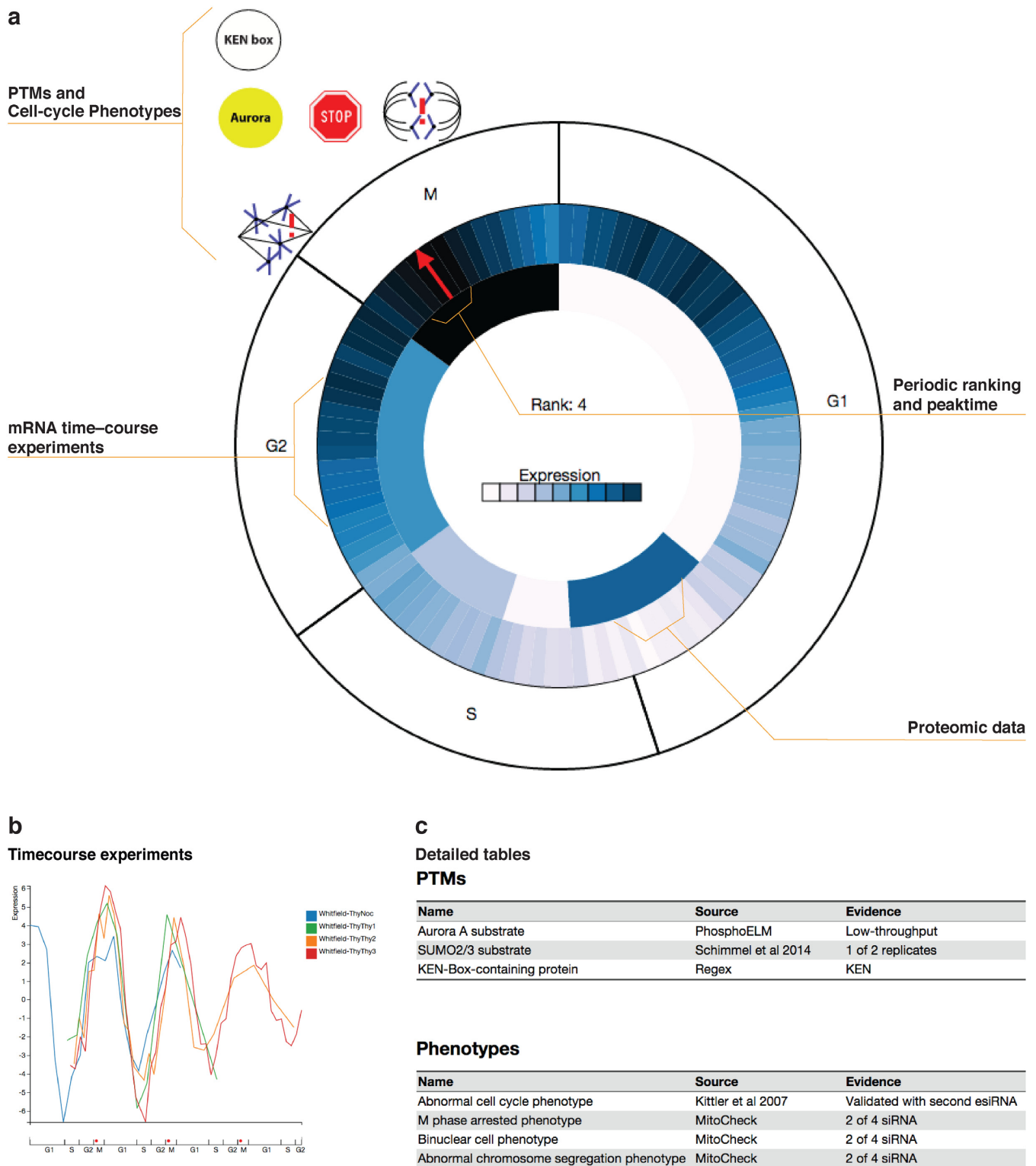
Figure 1. Overview of cell-cycle regulation and phenotypes. (**a**) A key feature of Cyclebase 3.0 is the new visualization, which aims to provide a concise overview of cell-cycle regulation and phenotypes for a gene. (**b**) For a more detailed view of the transcriptomic data, we normalize and align the individual time course studies, to allow all expression data for a gene to be plotted on a common time scale (percentage of cell cycle). (**c**) Further detail on PTMs, degradation signals and organism-specific phenotypes is provided in the form of tables with linkouts to the original sources whenever possible.

to also summarize for which of the genes cell-cycle-related phenotypes were observed and which phases these phenotypes relate to.

PERSPECTIVES

With this major update, Cyclebase is well positioned to continue to serve the scientific community as a one-stop-shop for cell-cycle data. The scope of the database has been expanded from its initial focus on transcriptomics time courses to reflect the developments in experimental technologies, most notably proteomics and high-content screening. Having such data integrated into Cyclebase will help illuminate the complex interplay between the different layers of regulation and understand how regulation relates to the phenotypes observed when knocking down or overexpressing.

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. *Mol. Cell*, **2**, 65–73.
2. 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.
3. 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.
4. 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.
5. 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.
6. Peng,X., Karuturi,R.K., 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.
7. De Lichtenberg,U., 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.
8. 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.
9. Granovskaia,M.V., Jensen,L.J., Ritchie,M.E., Toedling,J., Ning,Y., Bork,P., Huber,W. and Steinmetz,L.M. (2010) High-resolution transcription atlas of the mitotic cell cycle in budding yeast. Genome Biol., **11**, R24.
10. Olsen,J.V., Vermeulen,M., Santamaria,A., Kumar,C., Miller,M.L., Jensen,L.J., Gnad,F., Cox,J., Jensen,T.S., Nigg,E.A. *et al.* (2010) Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. Sci. Signal., **3**, ra3.
11. Ly,T., Ahmad,Y., Shlien,A., Soroka,D., Mills,A., Emanuele,M.J., Stratton,M.R. and Lamond,A.I. (2014) A proteomic chronology of gene expression through the cell cycle in human myeloid leukemia cells. *Elife*, **3**, e01630.
12. 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.
13. Loog,M. and Morgan,D.O. (2005) Cyclin specificity in the phosphorylation of cyclin-dependent kinase substrates. *Nature*, **434**, 104–108.
14. Giaever,G., Chu,A.M., Ni,L., Connelly,C., Riles,L., Véronneau,S., Dow,S., Lucau-Danila,A., Anderson,K., André,B. *et al.* (2002) Functional profiling of the Saccharomyces cerevisiae genome. *Nature*, **418**, 387–391.
15. 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. U.S.A.*, **103**, 14819–14824.
16. Draviam,V.M., Stegmeier,F., Nalepa,G., Sowa,M.E., Chen,J., Liang,A., Hannon,G.J., Sorger,P.K., Harper,J.W. and Elledge,S.J. (2007) A functional genomic screen identifies a role for TAO1 kinase in spindle-checkpoint signalling. *Nat. Cell Biol.*, **9**, 556–564.
17. Kittler,R., Pelletier,L., Heninger,A.-K.K., Slabicki,M., Theis,M., Miroslaw,L., Poser,I., Lawo,S., Grabner,H., Kozak,K. *et al.* (2007) Genome-scale RNAi profiling of cell division in human tissue culture cells. *Nat. Cell Biol.*, **9**, 1401–1412.
18. Rines,D.R., Gomez-Ferreria,M.A., Zhou,Y., DeJesus,P., Grob,S., Batalov,S., Labow,M., Huesken,D., Mickanin,C., Hall,J. *et al.* (2008) Whole genome functional analysis identifies novel components required for mitotic spindle integrity in human cells. *Genome Biol.*, **9**, R44.
19. Hutchins,J.R.A., Toyoda,Y., Hegemann,B., Poser,I., Hériché,J.-K., Sykora,M.M., Augsburg,M., Hudecz,O., Buschhorn,B.A., Bulkescher,J. *et al.* (2010) Systematic analysis of human protein complexes identifies chromosome segregation proteins. *Science*, **328**, 593–599.
20. Neumann,B., Walter,T., Hériché,J.-K.K., Bulkescher,J., Erfle,H., Conrad,C., Rogers,P., Poser,I., Held,M., Liebel,U. *et al.* (2010) Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature*, **464**, 721–727.
21. Fuchs,F., Pau,G., Kranz,D., Sklyar,O., Budjan,C., Steinbrink,S., Horn,T., Pedal,A., Huber,W. and Boutros,M. (2010) Clustering phenotype populations by genome-wide RNAi and multiparametric imaging. *Mol. Syst. Biol.*, **6**, 370.
22. Hayles,J., Wood,V., Jeffery,L., Hoe,K.-L., Kim,D.-U., Park,H.-O., Salas-Pino,S., Heichinger,C. and Nurse,P. (2013) A genome-wide resource of cell cycle and cell shape genes of fission yeast. *Open Biol.*, **3**, 130053.
23. Powell,S., Forslund,K., Szklarczyk,D., Trachana,K., Roth,A., Huerta-Cepas,J., Gabaldón,T., Rattei,T., Creevey,C., Kuhn,M. *et al.* (2014) eggNOG v4.0: nested orthology inference across 3686 organisms. *Nucleic Acids Res.*, **42**, D231–D239.
24. Dinkel,H., Chica,C., Via,A., Gould,C.M., Jensen,L.J., Gibson,T.J. and Diella,F. (2011) Phospho.ELM: a database of phosphorylation sites–update 2011. *Nucleic Acids Res.*, **39**, D261–D267.
25. Horn,H., Schoof,E.M., Kim,J., Robin,X., Miller,M.L., Diella,F., Palma,A., Cesareni,G., Jensen,L.J. and Linding,R. (2014)

KinomeXplorer: an integrated platform for kinome biology studies. *Nat. Methods*, **11**, 603–604.

26. Schimmel,J., Eifler,K., Sigurðsson,J.O., Cuijpers,S.A.G., Hendriks,I.A., Verlaan-de Vries,M., Kelstrup,C.D., Francavilla,C., Medema,R.H., Olsen,J. V *et al.* (2014) Uncovering SUMOylation dynamics during cell-cycle progression reveals FoxM1 as a key mitotic SUMO target protein. *Mol. Cell*, **53**, 1053–1066.
27. Schou,J., Kelstrup,C.D., Hayward,D.G., Olsen,J.V. and Nilsson,J. (2014) Comprehensive identification of SUMO2/3 targets and their dynamics during mitosis. *PLoS One*, **9**, e100692.
28. Costanzo,M.C., Engel,S.R., Wong,E.D., Lloyd,P., Karra,K., Chan,E.T., Weng,S., Paskov,K.M., Roe,G.R., Binkley,G. *et al.* (2014)

Saccharomyces genome database provides new regulation data. *Nucleic Acids Res.*, **42**, D717–D725.
29. Wood,V., Harris,M.A., McDowall,M.D., Rutherford,K., Vaughan,B.W., Staines,D.M., Aslett,M., Lock,A., Bähler,J., Kersey,P.J. *et al.* (2012) PomBase: a comprehensive online resource for fission yeast. *Nucleic Acids Res.*, **40**, D695–D699.
30. Harris,M.A., Lock,A., Bähler,J., Oliver,S.G. and Wood,V. (2013) FYPO: the fission yeast phenotype ontology. *Bioinformatics*, **29**, 1671–1678.