Published online 24 November 2012 *Nucleic Acids Research, 2013, Vol. 41, Database issue* ***D1009–D1013***
doi:10.1093/nar/gks1161

CircaDB: a database of mammalian circadian gene expression profiles

Angel Pizarro[1,*], Katharina Hayer[1], Nicholas F. Lahens[2] and John B. Hogenesch[1,2]

[1]The Institute for Translational Medicine and Therapeutics, University of Pennsylvania, 3400 Civic Center Boulevard, Building 421 and [2]Department of Pharmacology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

Received August 15, 2012; Revised October 4, 2012; Accepted October 28, 2012

ABSTRACT

CircaDB (http://circadb.org) is a new database of circadian transcriptional profiles from time course expression experiments from mice and humans. Each transcript's expression was evaluated by three separate algorithms, JTK_Cycle, Lomb Scargle and DeLichtenberg. Users can query the gene annotations using simple and powerful full text search terms, restrict results to specific data sets and provide probability thresholds for each algorithm. Visualizations of the data are intuitive charts that convey profile information more effectively than a table of probabilities. The CircaDB web application is open source and available at http://github.com/itmat/circadb.

INTRODUCTION

Circadian rhythms are biological rhythms of ~24 h in many physiological and behavioral processes (1,2). These rhythms are generated by a cell autonomous circadian clock, present in most cells in mammals. This circadian clock is composed of interlocked transcriptional, translational feedback loops, where transactivators activate repressors that later feedback on the activators (3). Components of the required E-box loop include Bmal1, Bmal2, Clock and Npas2, bHLH-PAS transactivators, Per1, Per2 and Per3, PAS domain containing repressors and Cry1 and Cry2 (4), transcriptional repressors related to cryptochromes from plants and insects. An important secondary loop also exists, the ROR loop, which comprises Rev-erb-alpha, Rev-erb-beta, transcriptional repressors, as well as Rorα, Rorb and Rorγ, transcriptional activators (5–7). Factors in this loop regulate transcript levels of several of the E-box components including Bmal1, Cry1, Npas2 and Per2. The cAMP Responsive Element Binding Protein (CREB) pathway (8,9) and D-box binding factors, Dbp, Hlf, Tef, Nfil3, also regulate clock function (10,11). Thus, transcription factors play a major role in the functioning of the core clock.

In addition to regulating transcription of each other, clock factors also impart circadian rhythms in expression of many 'output' genes. First order clock control genes are those directly regulated by clock factors (e.g. Clock/Bmal1), while second order output genes could be regulated by a first-order clock-control gene, but not clock components (12–14). Because of this, the research community has spent more than a decade cataloging genes under clock control (12,13,15–17). Historically, these include many disease genes, drug targets and important components of various biological pathways (1,18–20). For example, HMG-CoA reductase, the rate limiting enzyme of cholesterol biosynthesis and target of statins, is under clock control in liver (21). Several factors have catalysed a more complete description of circadian rhythms, including the advent of DNA arrays (16) and now RNA sequencing (22), powerful statistical approaches to find rhythmic genes (23) and appropriate experimental design.

The goal of CircaDB is to systematically collect, analyse and visualize circadian expression profiles for bench researchers in a simple and straightforward fashion. Common queries are supported and include straightforward queries of expression profiles, as well as compound queries searching keywords in the gene annotation, in multiple tissues, with the ability to restrict results by probability of cycling.

MATERIALS AND METHODS

Various publicly available microarray time course studies (23–26) were collected (Table 1). References and links to download the expression data sets are outlined on the website. Data from each study were re-analysed using three circadian rhythm detection algorithms: JTK_CYCLE, Lombe Scargle, de Lichtenberg (23,27,28). Table 2 lists the runtime parameters of the algorithms on each data set. The reported expression values from each study

*To whom correspondence should be addressed. Tel: +1 215 573 3736; Fax: +1 215 573 9736; Email: angel@upenn.edu

Table 1. Expresssion data sets in CircaDB

Name	Time points	Species/tissue
Panda 2002	12	Mouse suprachiasmatic nuclei (SCN) of the hypothalamus, and liver
Hughes 2009	48	Mouse liver, NIH3T3 cells, pituitary gland and human U2OS cells
Miller 2007 and Andrews 2010	12 (WT)	Wild type mouse liver, SCN and skeletal muscle
	7 (KO)	Clock mutant mouse liver, SCN and skeletal muscle
Rudic 2004	12	Mouse aorta, kidney

Table 2. Runtime parameters for each data set and algorithm

Data set	JTK_CYCLE	Lomb Scargle	De Lichtenberg
Panda 2002	Periods: 16–32 h	minFrequency = 1/32, maxFrequncy = 1/18; (periods = 18–32 h; #test frequencies: 4*N	Period = 24 h #Permutations = 10 000
Hughes 2009 (mouse)	Periods: 6–42 h	minFrequency = 1/6, maxFrequncy = 1/42; (periods = 6–42 h; #test frequencies: 4*N	Period = 24 h #Permutations = 10 000
Hughes 2009 (human)	Periods: 6–42 h	minFrequency = 1/6, maxFrequncy = 1/42; (periods = 6–42 h; #test frequencies: 4*N	Period = 24 h #Permutations = 10 000
Miller 2007	Periods: 16–32 h	minFrequency = 1/32, maxFrequncy = 1/18; (periods = 18–32 h; #test frequencies: 4*N	Period = 24 h #Permutations = 10 000
Andrews 2010	Periods: 20–28 h	minFrequency = 1/6, maxFrequncy = 1/42; (periods = 6–42 h; #test frequencies: 4*N	Period = 24 h #Permutations = 10 000
Rudic 2004	Periods: 16–32 h	minFrequency = 1/32, maxFrequncy = 1/18; (periods = 18–32 h; #test frequencies: 4*N	Period = 24 h #Permutations = 10 000

Data sets are located in Table 1.
N = number of time points in the series.

were not filtered, as each algorithm accounts for technical replicates. The significance calls and other results reported by each algorithm were entered into a MySQL database.

Gene annotation data were downloaded from the Affymetrix NetAffx resource (http://www.affymetrix.com/analysis/index.affx). Annotations were then entered into the database alongside the unfiltered experimental values and the results of the circadian rhythm detection algorithms. Transcript information was supplemented with links to the GeneWiki project (29,30) and Homologene (http://www.ncbi.nlm.nih.gov/homologene). The data model for the database is described in Figure 1.

The transcript annotation and the statistical results were indexed with the Sphinx full text search system (http://sphinxsearch.com/). Visualization of data is accomplished by created using pre-formatted URI requests to the Google Charts API (https://developers.google.com/chart/). The web application was coded using the Ruby on Rails framework (http://rubyonrails.org/).

All source code for data loading and the web application is licensed under the GNU General Public License (GPL-2.0) license and available at http://github.com/itmat/circadb.

RESULTS AND DISCUSSION

In creating CircaDB, we have provided the research community a clear, concise and powerful interface for querying genes within the context of circadian expression profile data. Another circadian expression database, Diurnal 2.0 (31), provides a similar resource to CircaDB but focuses on plant data. It also restricts its initial search to transcript accessions, whereas CircaDB allows full query capabilities on gene annotation. CircaDB provides advanced keyword search capabilities of gene annotation. This includes the ability to search by phrases, boolean conditions and combinations thereof. Queries can also be restricted by a given experiment's data set, phase of expression and significance of a particular algorithm (Figure 2).

The Database of Circadian Gene Expression (24), part of the Gene Atlas Project (32), contains a subset of the same data sets in CircaDB, but uses a single circadian expression algorithm. CircaDB contains all of these data and re-analysed them with newer and more robust set of algorithms (23,27,28). Three algorithms were used to allow for the inspection of the differences between each algorithm's results (Figure 3). CircaDB is actively maintained and will continue to add new features and data sets as time they become available. Requests for integration of data sets are handled via submitting a request via the project site at Github. CiraDB also provides integration expression profiles for use within BioGPS (33).

Finally, to facilitate use of this database framework by other researcher groups, we have made the source code for the application freely available under the GPL 2.0 open source license. The project has been recently used to visualize circadian experiments for Anopheles gambiae (34). All of these together make CircaDB a unique and valuable resource for the circadian research community.

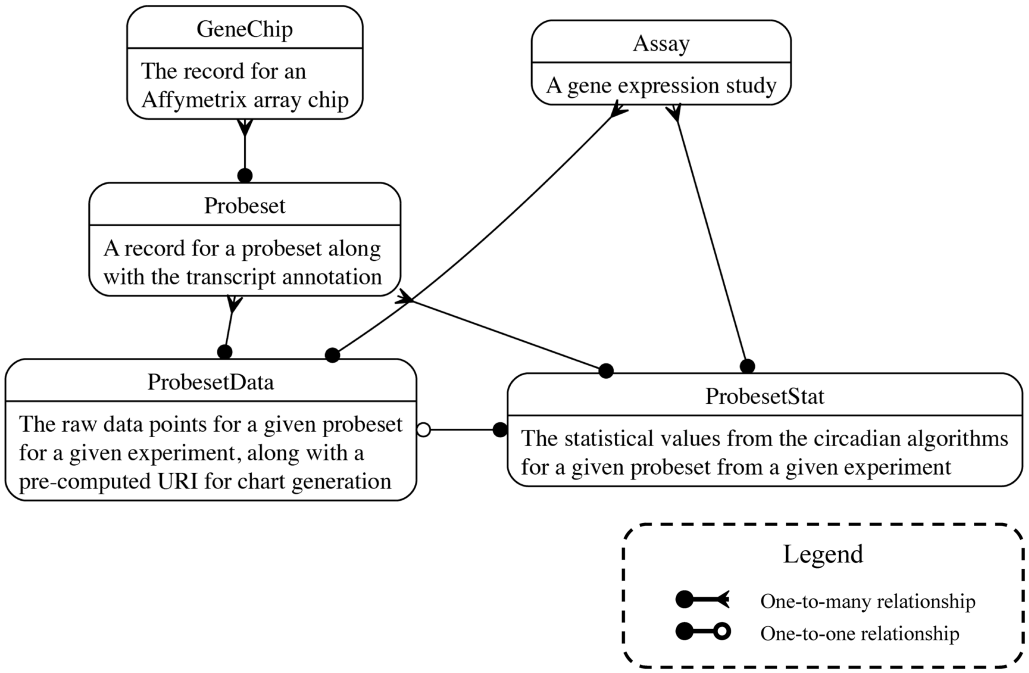


Figure 1. The database schema. Boxes represent table, and edges represent foreign key relationships. Further documentation is available at http://github.com/itmat/circadb.

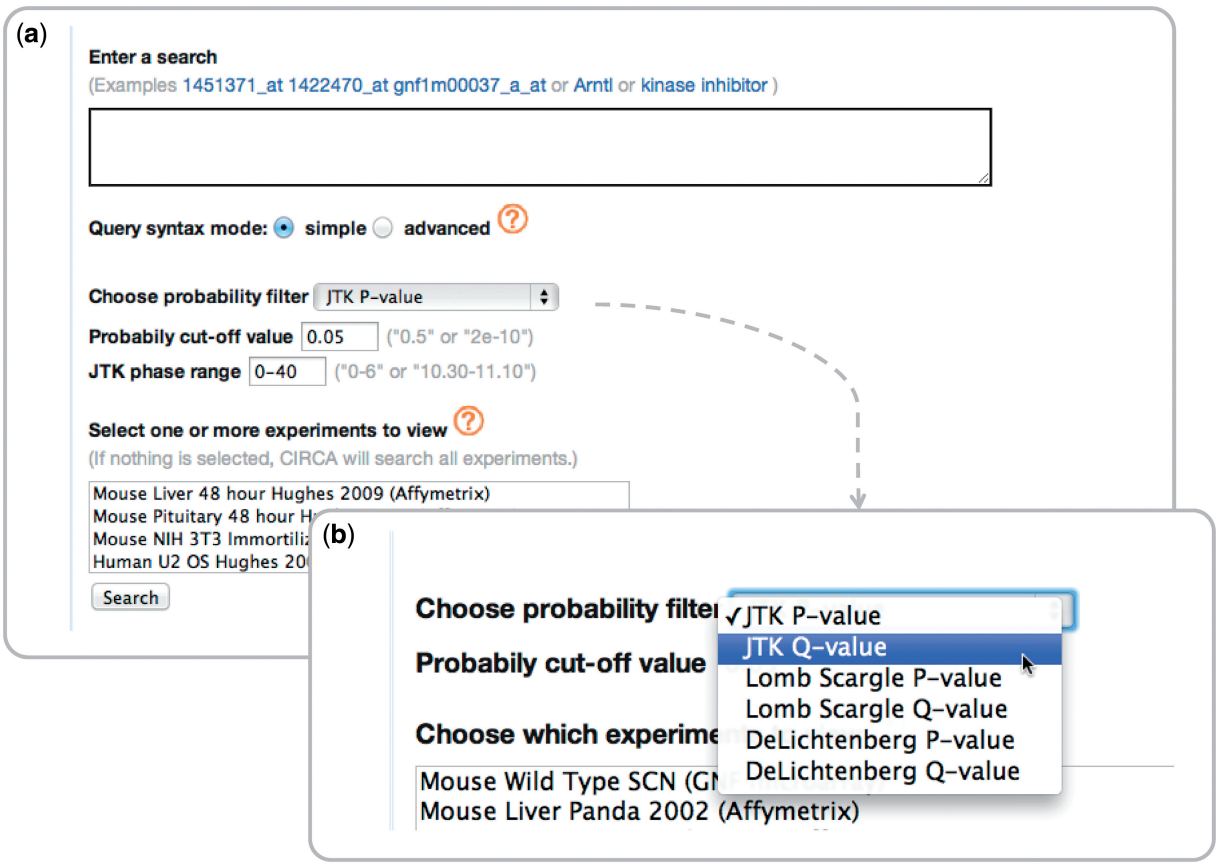
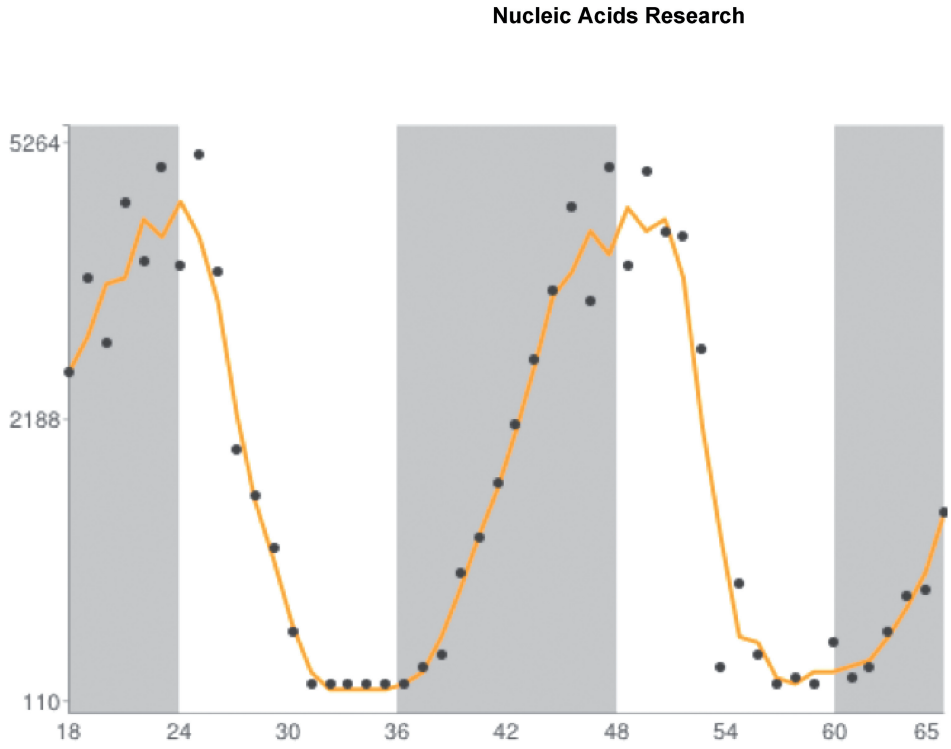


Figure 2. (**a**) The query interface for CircaDB. The interface consists of a simple and powerful full-text search capability, with possible restrictions on the data sets, phase information and a significance threshold for a given algorithm. (**b**) The set of available threshold categories for the circadian classification algorithms.



Tissue liver

UCSC RNAseq
RNAseq_NM_007489

Probeset
1425099_a_at

Links Wikipedia
HomoloGene

Symbol Arntl

Unigene Mm.440371

RefSeq Protein
NP_031515

RefSeq DNA
NM_007489

Description aryl hydrocarbon receptor nuclear translocator-like

	p-Value	q-Value	period	phase
JTK	3.0e-16	2.71e-13	24.5	4.5
Lomb Scargle	2.58e-08	9.13045e-05	24.383	29.97
DeLichtenberg	0.0	0.0	24.0	*NA*

back to the top

Figure 3. Expression profile report. A simple visualization of the data accompanies the main annotation of the gene probe, probability values from various circadian rhythm detection algorithms and other circadian information.

FUNDING

The National Institutes of Health, the National Center for Advancing Translational Sciences [8UL1TR000003] (to Garret FitzGerald, University of Pennsylvania); National Heart, Lung, and Blood Institute [1R01HL097800-04 to J.B.H.]; the Defense Advanced Research Projects Agency [BAA-11-65] (to John Harer, Duke University). Funding for open access charge: Departmental Funds.

Conflict of interest statement. None declared.

REFERENCES

1. Hastings,M.H., Reddy,A.B. and Maywood,E.S. (2003) A clockwork web: circadian timing in brain and periphery, in health and disease. *Nat. Rev. Neurosci.*, **4**, 649–661.
2. Green,C.B., Takahashi,J.S. and Bass,J. (2008) The meter of metabolism. *Cell*, **134**, 728–742.
3. Lowrey,P.L. and Takahashi,J.S. (2004) Mammalian circadian biology: elucidating genome-wide levels of temporal organization. *Annual review of genomics and human genetics*, **5**, 407–4.
4. Ko,C.H. and Takahashi,J.S. (2006) Molecular components of the mammalian circadian clock. *Hum. Mol. Genet.*, **15**, R271–R277.
5. Yin,L. and Lazar,M.A. (2005) The orphan nuclear receptor Rev-erbalpha recruits the N-CoR/histone deacetylase 3 corepressor to regulate the circadian Bmal1 gene. *Mol. Endocrinol.*, **19**, 1452–1459.
6. Guillaumond,F., Dardente,H., Giguère,V. and Cermakian,N. (2005) Differential control of Bmal1 circadian transcription by REV-ERB and ROR nuclear receptors. *J. Biol Rhythms*, **20**, 391–403.
7. Takeda,Y., Jothi,R., Birault,V. and Jetten,A.M. (2012) RORγ directly regulates the circadian expression of clock genes and downstream targets in vivo. *Nucleic Acids Res.*, **40**, 8519–8535.
8. Akashi,M., Hayasaka,N., Yamazaki,S. and Node,K. (2008) Mitogen-activated protein kinase is a functional component of the

autonomous circadian system in the suprachiasmatic nucleus. *J Neurosci.*, **28**, 4619–4623.
9. Sanada,K., Okano,T. and Fukada,Y. (2002) Mitogen-activated protein kinase phosphorylates and negatively regulates basic helix-loop-helix-PAS transcription factor BMAL1. *J. Biol. Chem.*, **277**, 267–271.
10. Ueda,H.R., Hayashi,S., Chen,W., Sano,M., Machida,M., Shigeyoshi,Y., Iino,M. and Hashimoto,S. (2005) System-level identification of transcriptional circuits underlying mammalian circadian clocks. *Nat. Genet.*, **37**, 187–192.
11. Ukai-Tadenuma,M., Yamada,R.G., Xu,H., Ripperger,J.A., Liu,A.C. and Ueda,H.R. (2011) Delay in feedback repression by cryptochrome 1 is required for circadian clock function. *Cell*, **144**, 268–281.
12. Hughes,M.E., DiTacchio,L., Hayes,K.R., Vollmers,C., Pulivarthy,S., Baggs,J.E., Panda,S. and Hogenesch,J.B. (2009) Harmonics of circadian gene transcription in mammals. *PLoS Genet.*, **5**, e1000442.
13. Gachon,F., Olela,F.F., Schaad,O., Descombes,P. and Schibler,U. (2006) The circadian PAR-domain basic leucine zipper transcription factors DBP, TEF, and HLF modulate basal and inducible xenobiotic detoxification. *Cell Metabol.*, **4**, 25–36.
14. Poliandri,A.H.B., Gamsby,J.J., Christian,M., Spinella,M.J., Loros,J.J., Dunlap,J.C. and Parker,M.G. (2011) Modulation of clock gene expression by the transcriptional coregulator receptor interacting protein 140 (RIP140). *J. Biol. Rhythms*, **26**, 187–199.
15. Storch,K.-F., Lipan,O., Leykin,I., Viswanathan,N., Davis,F.C., Wong,W.H. and Weitz,C.J. (2002) Extensive and divergent circadian gene expression in liver and heart. *Nature*, **417**, 78–83.
16. Kornmann,B., Schaad,O., Bujard,H., Takahashi,J.S. and Schibler,U. (2007) System-driven and oscillator-dependent circadian transcription in mice with a conditionally active liver clock. *PLoS Biol.*, **5**, e34.
17. Hughes,M.E., Hong,H.-K., Chong,J.L., Indacochea,A.A., Lee,S.S., Han,M., Takahashi,J.S. and Hogenesch,J.B. (2012) Brain-specific rescue of clock reveals system-driven transcriptional rhythms in peripheral tissue. *PLoS Genet.*, **8**, e1002835.
18. Takahashi,J.S., Hong,H.-K., Ko,C.H. and McDearmon,E.L. (2008) The genetics of mammalian circadian order and disorder: implications for physiology and disease. *Nat. Rev. Genet.*, **9**, 764–75.
19. Curtis,A.M. and Fitzgerald,G.A. (2006) Central and peripheral clocks in cardiovascular and metabolic function. *Ann. Med.*, **38**, 552–9.
20. Sancar,A., Lindsey-Boltz,L.A., Kang,T.-H., Reardon,J.T., Lee,J.H. and Ozturk,N. (2010) Circadian clock control of the cellular response to DNA damage. *FEBS Lett.*, **584**, 2618–2625.
21. Le Martelot,G., Claudel,T., Gatfield,D., Schaad,O., Kornmann,B., Sasso,G.L., Moschetta,A. and Schibler,U. (2009) REV-ERBalpha participates in circadian SREBP signaling and bile acid homeostasis. *PLoS Biol.*, **7**, e1000181.
22. Hughes,M.E., Grant,G.R., Paquin,C., Qian,J. and Nitabach,M.N. (2012) Deep sequencing the circadian and diurnal transcriptome of Drosophila brain. *Genome Res.*, **22**, 1266–81.
23. Hughes,M.E., Hogenesch,J.B. and Kornacker,K. (2010) JTK_CYCLE: an efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets. *J. Biol. Rhythms*, **25**, 372–380.
24. Panda,S., Antoch,M.P., Miller,B.H., Su,A.I., Schook,A.B., Straume,M., Schultz,P.G., Kay,S.A., Takahashi,J.S. and Hogenesch,J.B. (2002) Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell*, **109**, 307–320.
25. Andrews,J.L., Zhang,X., McCarthy,J.J., McDearmon,E.L., Hornberger,T.A., Russell,B., Campbell,K.S., Arbogast,S., Reid,M.B., Walker,J.R. *et al.* (2010) CLOCK and BMAL1 regulate MyoD and are necessary for maintenance of skeletal muscle phenotype and function. *Proc. Natl Acad. Sci. USA*, **107**, 19090–19095.
26. Rudic,R.D., McNamara,P., Curtis,A.-M., Boston,R.C., Panda,S., Hogenesch,J.B. and Fitzgerald,G.A. (2004) BMAL1 and CLOCK, two essential components of the circadian clock, are involved in glucose homeostasis. *PLoS Biol.*, **2**, e377.
27. Glynn,E.F., Chen,J. and Mushegian,A.R. (2006) Detecting periodic patterns in unevenly spaced gene expression time series using Lomb-Scargle periodograms. *Bioinformatics*, **22**, 310–316.
28. de Lichtenberg,U., Jensen,L.J., Fausbøll,A., Jensen,T.S., Bork,P. and Brunak,S. (2005) Comparison of computational methods for the identification of cell cycle-regulated genes. *Bioinformatics*, **21**, 1164–1171.
29. Huss,J.W., Orozco,C., Goodale,J., Wu,C., Batalov,S., Vickers,T.J., Valafar,F. and Su,A.I. (2008) A gene wiki for community annotation of gene function. *PLoS Biol.*, **6**, e175.
30. Huss,J.W., Lindenbaum,P., Martone,M., Roberts,D., Pizarro,A., Valafar,F., Hogenesch,J.B. and Su,A.I. (2010) The Gene Wiki: community intelligence applied to human gene annotation. *Nucleic acids research*, **38**, D633–D639.
31. Mockler,T.C., Michael,T.P., Priest,H.D., Shen,R., Sullivan,C.M., Givan,S.A., McEntee,C., Kay,S.A. and Chory,J. (2007) The DIURNAL project: DIURNAL and circadian expression profiling, model-based pattern matching, and promoter analysis. *Cold Spring Harb. Symp. Quant. Biol.*, **72**, 353–363.
32. Su,A.I., Wiltshire,T., Batalov,S., Lapp,H., Ching,K.A., Block,D., Zhang,J., Soden,R., Hayakawa,M., Kreiman,G. *et al.* (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA*, **101**, 6062–6067.
33. Wu,C., Orozco,C., Boyer,J., Leglise,M., Goodale,J., Batalov,S., Hodge,C.L., Haase,J., Janes,J., Huss,J.W. *et al.* (2009) BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol.*, **10**, R130.
34. Rund,S.S.C., Hou,T.Y., Ward,S.M., Collins,F.H. and Duffield,G.E. (2011) Genome-wide profiling of diel and circadian gene expression in the malaria vector Anopheles gambiae. *Proc. Natl Acad. Sci. USA*, **108**, E421–E430.