

# CGDB: a database of circadian genes in eukaryotes

Shujing Li<sup>1,2,†</sup>, Ke Shui<sup>1,†</sup>, Ying Zhang<sup>1,†</sup>, Yongqiang Lv<sup>1</sup>, Wankun Deng<sup>1</sup>, Shahid Ullah<sup>1</sup>, Luoying Zhang<sup>1,\*</sup> and Yu Xue<sup>1,\*</sup>

<sup>1</sup>Key Laboratory of Molecular Biophysics of Ministry of Education, College of Life Science and Technology and the Collaborative Innovation Center for Brain Science, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China and <sup>2</sup>Department of Life Sciences, Bengbu Medical College, Bengbu, Anhui 233030, China

Received August 08, 2016; Revised October 17, 2016; Editorial Decision October 17, 2016; Accepted October 18, 2016

## ABSTRACT

**We report a database of circadian genes in eukaryotes (CGDB, <http://cgdb.biocuckoo.org>), containing ~73 000 circadian-related genes in 68 animals, 39 plants and 41 fungi. Circadian rhythm is ~24 h rhythm in behavioral and physiological processes that exists in almost all organisms on the earth. Defects in the circadian system are highly associated with a number of diseases such as cancers. Although several databases have been established for rhythmically expressed genes, a comprehensive database of cycling genes across phyla is still lacking. From the literature, we collected 1382 genes of which transcript level oscillations were validated using methods such as RT-PCR, northern blot and *in situ* hybridization. Given that many genes exhibit different oscillatory patterns in different tissues/cells within an organism, we have included information regarding the phase and amplitude of the oscillation, as well as the tissue/cells in which the oscillation was identified. Using these well characterized cycling genes, we have then conducted an orthologous search and identified ~45 000 potential cycling genes from 148 eukaryotes. Given that significant effort has been devoted to identifying cycling genes by transcriptome profiling, we have also incorporated these results, a total of over 26 000 genes, into our database.**

## INTRODUCTION

Circadian rhythms are fundamental phenomena of life that are manifested as daily oscillations in vast biological processes and driven by endogenous clocks that exist in most if not all organisms on the earth (1–4). The molecular clock consists of a number of genes that form transcriptional–translational feedback loops that function with a near 24 h period, resulting in rhythmic expression of its components

which are referred to as core clock genes (3–5). In mammals, these genes include *BMAL1* and *CLOCK*, which form the positive limb of the feedback loops, while *PERIOD1-3* (*PER1-3*) and *CRYPTOCHROME1-2* (*CRY1-2*) form the major negative limb of the feedback loops (3,4). Since most of our physiological processes and behaviors are circadianly regulated, disruptions of circadian rhythms have been implicated in various diseases and disorders, such as sleep disorders, metabolic diseases, psychiatric disorders, neurological diseases and cancers (6–10). Indeed, genetic variations in core clock genes have been shown to be associated with numerous metabolic diseases, cancers, sleep disorders and mental disorders (11). Animal models revealed that core clock genes are involved in sleep and mood regulation, as well as the development of obesity, diabetes and cancer (7,12–15). At the molecular level, ~1% to over 60% of the transcriptome exhibits oscillatory expression, and the majority of the best-selling drugs as well as essential medicines of World Health Organization target cycling genes (1,16,17). As many of these drugs have short half-lives, delivering them at the appropriate time of the day will likely improve their efficacy (17,18). Therefore, the identification of circadian genes and understanding the role of circadian clock in regulating rhythmic global gene expression will benefit the development of treatments and therapies for human diseases.

Numerous efforts have been undertaken to identify oscillating genes (1,19,20). Due to the rapid progresses in the development of high-throughput techniques such as microarray and RNA sequencing (RNA-seq), a number of computational methods, such as COSOPT (21), Fisher's *G*-test (22), HAYSTACK (23), JTK\_CYCLE (24), ARSER (25) and LSPR (26), were developed for identifying potential circadian genes from the temporal transcriptome data. Recently, a regularized supervised learning algorithm was used, and ZeitZeiger was developed as an excellent R package for identifying periodic genes from genome-wide gene expression datasets (27). Also, Agostinelli *et al.* adopted deep learning approaches, and constructed a more accurate predictor of BIO\_CYCLE (28). Because more and more

\*To whom correspondence should be addressed. Tel: +86 27 87793903; Fax: +86 27 87793172; Email: xueyu@hust.edu.cn

Correspondence may also be addressed to Luoying Zhang. Email: zhangluoying@hust.edu.cn

†These authors contributed equally to the paper as first authors.

cycling genes have been identified, the collection and integration of the data is helpful for further experimental consideration. Previously, four major databases were developed for circadian genes, including DIURNAL (23), CircaDB (29), SCNseq (30) and Bioclock (31–33). The DIURNAL database maintains circadian microarray data for three plants, including *Arabidopsis thaliana*, poplar and rice (23). CircaDB contains over 3000 potential circadian genes in human and mouse (29). SCNseq contains 4569 cycling genes, and 3187 intergenic non-coding RNAs in mice, whereas Bioclock contains 1674 potential circadian genes exclusively in *Aedes aegypti* and ~1000 potential circadian genes in *Anopheles gambiae* (31–33). Although a number of computational efforts have been contributed to this area, an integrative resource is still not available.

In this study, we aimed to collect known oscillating genes across phyla. Given that many genes exhibit different temporal expression patterns in different tissues/cells within an organism, and that each tissue/cell type has a distinct collection of cycling genes, we have included information regarding the phase and amplitude of the oscillation, as well as the tissue/cells in which the oscillation was identified (3). From the scientific literature, we first collected 1382 cycling genes that were validated by methods such as RT-PCR, Northern blot and *in situ* hybridization. We then conducted an orthologous search and further identified 44 836 potential oscillating genes. We also collected 26 582 oscillating genes that were identified by microarray and/or RNA-seq. Finally, we developed a comprehensive circadian gene database (CGDB), containing 72 800 non-redundant cycling genes in 148 eukaryotes. We believe CGDB can serve as an integrated and convenient source for genes with oscillatory expression, which will provide useful information for circadian biologists working in any eukaryote. Moreover, since a substantial portion of the transcriptome exhibits daily oscillation, it would be important to take this into consideration when studying the expression and function of a particular gene. In this sense, CGDB will be valuable to a much broader group of researchers, and not limited to those working in the field of circadian rhythms.

## CONSTRUCTION AND CONTENT

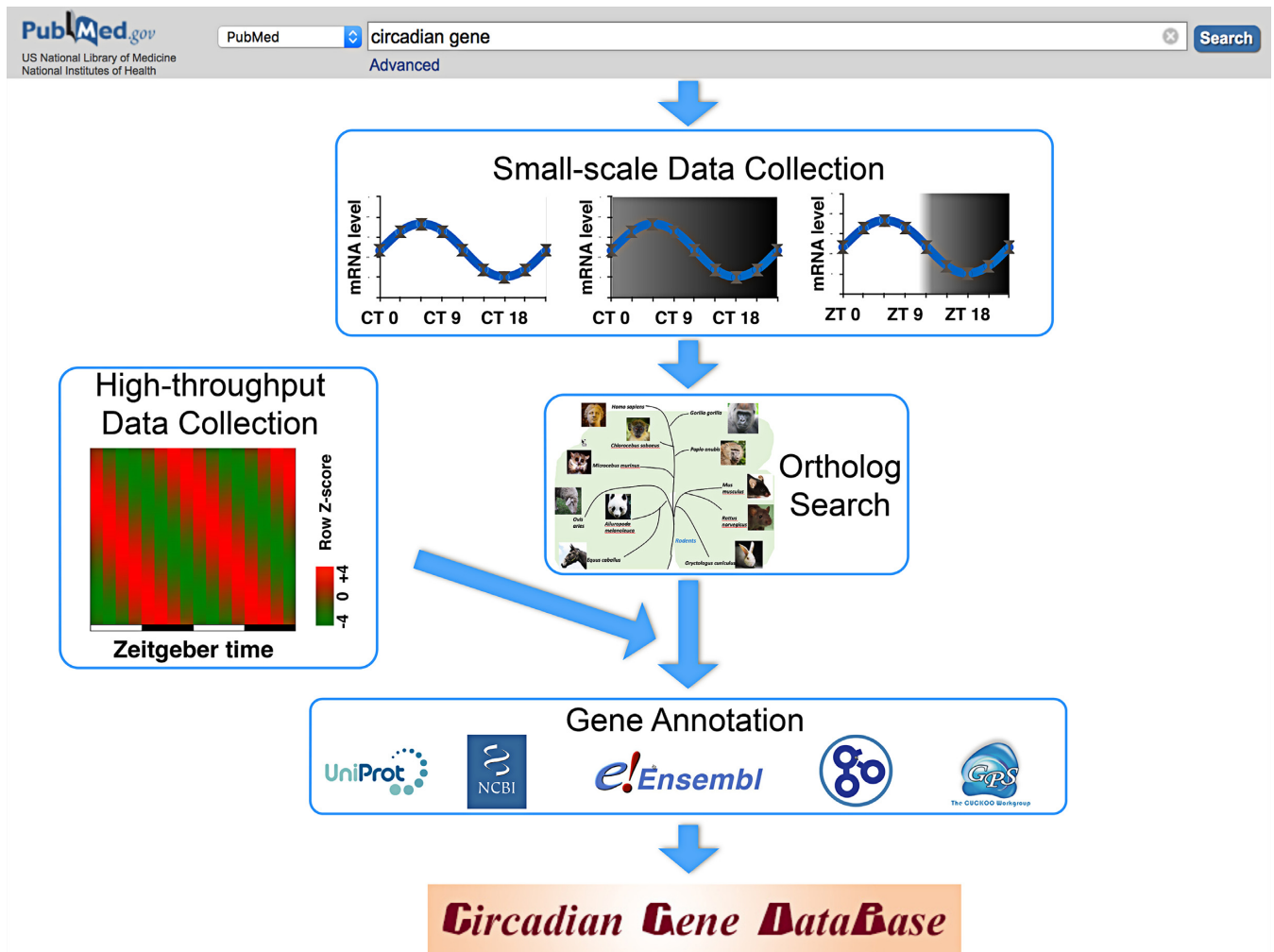
In this study, we defined circadian genes as genes which have daily oscillatory expression patterns. To construct a reliable benchmark data set, we manually curated experimentally identified circadian genes by using the keyword of ‘circadian gene’ to directly search the literature in PubMed (published before 1 March 2016, Figure 1). First, we collected 1382 cycling genes identified using methods such as RT-PCR, *in situ* hybridization and Northern blot. These genes were from 25 eukaryotes, which were 2 fungi, *Saccharomyces cerevisiae* and *Neurospora crassa*, 13 animals including *Macaca mulatta*, *Caenorhabditis elegans*, *Gasterosteus aculeatus*, *Meleagris gallopavo*, *Gadus morhua*, *Ovis aries*, *Equus caballus*, *Gallus gallus*, *Drosophila melanogaster*, *Danio rerio*, *Rattus norvegicus*, *Mus musculus* and *Homo sapiens* and 10 plants including *Arabidopsis lyrata*, *Oryza indica*, *Glycine max*, *Triticum aestivum*, *Solanum tuberosum*, *Physcomitrella patens*, *Zea mays*, *Hordeum vulgare*, *Oryza sativa* and *Arabidopsis thaliana*.

The oscillation of these genes were measured either under entrained condition (zeitgeber time, ZT) with environmental cues such as light/dark cycles or temperature cycles, or under free-running condition (circadian time, CT) which lacks environmental cues. We have included this information along with the peak and trough time points of the oscillation, as well as the amplitude values (amplitude is defined as peak value/trough value). Considering that many genes exhibit distinct temporal expression patterns in different tissues/cells in an organism, we have included information regarding the tissue/cells in which the oscillation was identified and a brief description of the potential function of the oscillation.

To provide information for species beyond the 25 organisms, we performed homologous detection to search for orthologs, which might be potential cycling genes and could be helpful in further studies (Figure 1). The complete reference proteomes of 148 eukaryotes with sequenced genomes were downloaded from Ensembl database (34). These include 68 animals, 39 plants and 41 fungi. As described previously (35), the strategy of reciprocal best hits (RBHs) with BLAST package (36) were applied to detect potential orthologs of cycling genes. In total, 44 836 orthologs were computationally identified as potential cycling genes, which were also integrated into the CGDB database. Lastly, we included 26 582 oscillating genes identified by microarray and RNA-seq studies (Figure 1). The detailed information of 27 964 known cycling genes was provided (Supplementary Table S1), whereas the numbers of oscillating genes from different organisms were summarized (Supplementary Table S2). All the genes in the database were annotated with primary references, cycling information, post-translational modifications (PTMs) and annotations from UniProt and Ensembl databases (34,37). All cycling information and sequences of proteins in CGDB were available for download at <http://cgdb.biocuckoo.org/download.php>.

## USAGE

The CGDB database web interface was constructed in an easy-to-use manner. Four search options were provided, including the simple search option providing an interface to query the CGDB database with one or multiple keywords or database accession numbers such as UniProt ID or CGDB ID (Figure 2A), ‘Advanced search’ based on combined keywords with up to two search terms (Figure 2B), ‘Multiple search’ using multiple keywords or accession numbers in a line-by-line format (Figure 2C), and ‘BLAST search’ based on protein sequence (Figure 2D). For example, if a keyword ‘PER2\_HUMAN’ in ‘UniProt\_Accession’ was submitted for a simple search (Figure 2A), the website will return the circadian gene *PER2* from *H. sapiens* in a tabular format with CGDB ID, UniProt/Ensembl accession, species and gene name/alias (Figure 2A). In advanced search option, two terms specified in two areas are combined with operators ‘and’, ‘or’ and ‘exclude’ to conduct a complex query (Figure 2B). For example, querying the database with ‘Per’ in ‘Gene name’ and ‘Human’ in ‘Organism’ will return five *PER* genes in *H. Sapiens* (Figure 2B). Moreover, users could input a list of keywords to perform a multiple search. For example, three core clock genes could be retrieved by



**Figure 1.** The workflow in constructing the CGDB database. Cycling genes that were identified via non-high throughput methods were collected from published literature and ortholog search was conducted subsequently. This data was then combined with cycling genes identified by high throughput methods. Finally, all cycling genes were annotated by established databases including Uniprot, NCBI and Ensembl.

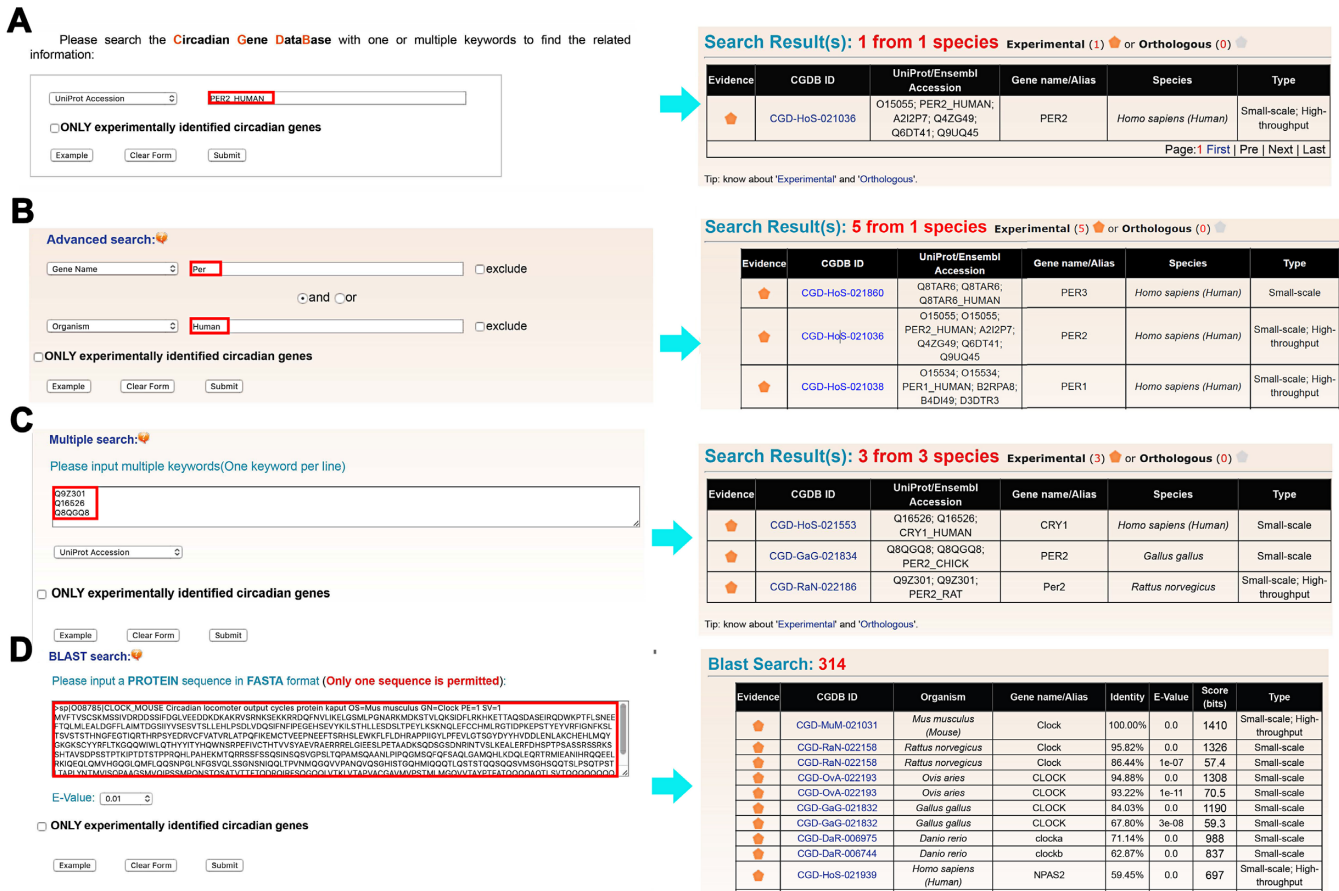
querying a list of their UniProt Accessions (Figure 2C). In addition, users could search identical or homologous proteins by submitting a protein sequence in FASTA format in 'BLAST Search' (Figure 2D). For example, the FASTA sequence of mouse CLOCK protein could be input in the FASTA format to search for homologous proteins in the database. In particular, there is a checkbox of 'ONLY experimentally identified circadian genes' for each search option (Figure 2). Once selected, only experimentally identified cycling genes will be queried and returned.

To conveniently browse CGDB database, two strategies were implemented, including 'by species' and 'by external condition'. In the 'Browse by species' option, the right tree is a representation of phylogenetic relations or classification of eukaryotic species in Ensembl (34), while the left tree lists the Ensembl taxonomy categories, including primates, rodents, laurasiatheria, afrotheria and so on (Figure 3A). By clicking on the '*Homo sapiens*' icon, with the checkbox 'ONLY experimentally identified circadian genes' selected, experimentally validated circadian genes in *H. sapiens* can be shown (Figure 3B). Furthermore, CGDB can be

browsed by external condition (Figure 3C). The numbers on the ZT clock can be clicked to browse genes that have been experimentally verified to peak or trough at a specific time point under entrained conditions. The checkbox 'Peak only' and 'Trough only' can also be selected individually to show genes that peak or trough at a specific time point (Figure 3C). The CGDB ID was adopted to organize the database, while the UniProt/Ensembl ID was used as secondary accession (Figure 3D). The users can click on 'CGD-HoS-021288' to view the detailed information of human *HTR1B* including cycling information, primary references, PTM information and its orthologs (Figure 3D).

## RESULTS AND DISCUSSION

Early physiological experiments have established that circadian clocks exist endogenously in almost every organism and drives oscillatory changes in a myriad of behavioral and physiological processes (1–3,5). Circadian rhythms are believed to arise from endogenous molecular clocks consisting of transcriptional and translational feedback loops that are more or less conserved across phyla (1,4). At the



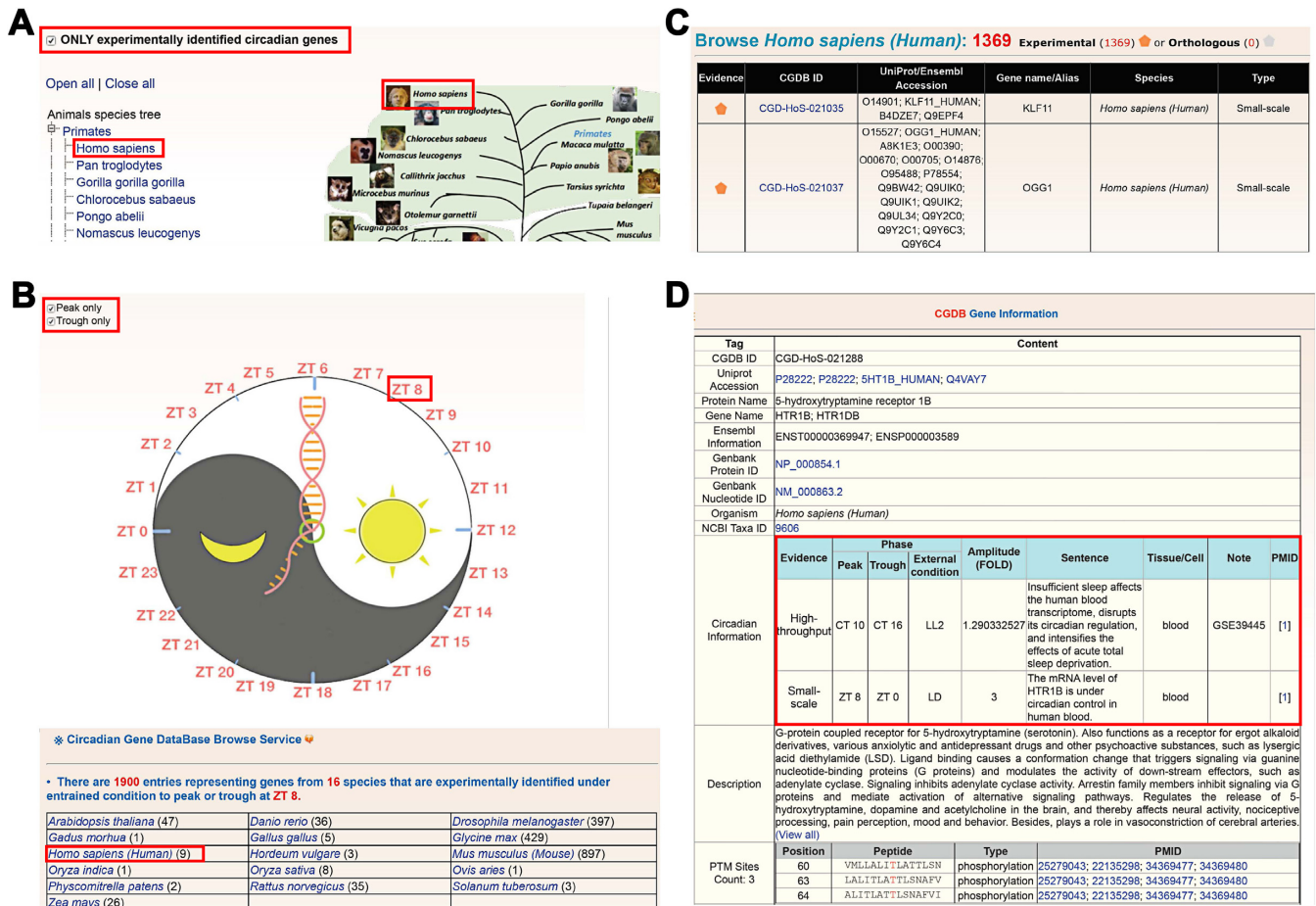
**Figure 2.** The search options of CGDB. (A) ‘Simple search’. The database can be queried with one or multiple keywords. (B) ‘Advanced search’. Users can input up to two terms to perform a precise search. (D) ‘Multiple search’. A list of keywords in a line-by-line format can be input to search multiple genes at once. (C) ‘Blast search’. The option was designed for searching the database with one protein sequence in FASTA format.

heart of these feedback loops are transcriptional activators which promote the expression of transcriptional repressors. The repressors suppress the activity of the activators, thus repressing their own expression. These repressors also undergo a series of translational and PTM regulations, leading to their degradation (1–3,5). As the protein levels of the repressors decrease, the activators can once again initiate transcription, starting a new cycle. Here we establish CGDB as a database of genes with oscillatory expression at the transcript level, accompanied by relevant phase and amplitude information, which are the two key indices of oscillation. We have also included known PTMs which is a critical part of circadian regulation.

To further understand the function of circadian clocks and how they regulate physiological rhythms, we analyzed the functional distribution of human cycling (and potential cycling) genes by mapping the non-redundant human proteome set to two databases of Gene Ontology Annotation (UniProt-GOA, <http://www.ebi.ac.uk/GOA>) (38) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (39). In total, we detected 6352 and 19 194 human proteins annotated with at least one KEGG or GO term, respectively. Then the hypergeometric test was adopted to perform enrichment analyses for 1889 human circadian genes, whereas the statistically over-represented KEGG ( $p$ -value < 0.001)

and GO terms ( $p$ -value <  $10^{-5}$ ) were shown, respectively (Supplementary Figure S1). For simplicity, only the top most significant KEGG terms and GO terms of biological processes, molecular functions, and cellular components were shown (Supplementary Figure S1). In our analyses, the most enriched KEGG term and GO biological pathway are ‘circadian rhythm’ (KEGG ID: hsa04710) and ‘circadian rhythm’ (GO:0007623). The most enriched GO molecular function is ‘RNA-polymerase II transcription factor activity’ (GO:0004879), consistent with the known function of clock proteins (3,4). Interestingly, the most enriched GO cellular compartment is ‘caveola’ (GO:0005901) which has not been known to be oscillatory, revealing the power of our database to identify new potential roles of the circadian clock.

Besides circadian rhythm, several metabolic pathways are significantly enriched (Supplementary Figure S1), consistent with a known role for the clock in regulating glucose homeostasis, lipid metabolism, protein synthesis and proliferation/differentiation (40–43). Among these pathways, insulin signaling has the largest number of cycling (or potential cycling) genes in human (data not shown). Therefore we are using insulin signaling pathway as a detailed example of how cycling genes may carry out time-dependent functions (Figure 4). We have indicated the peak expression



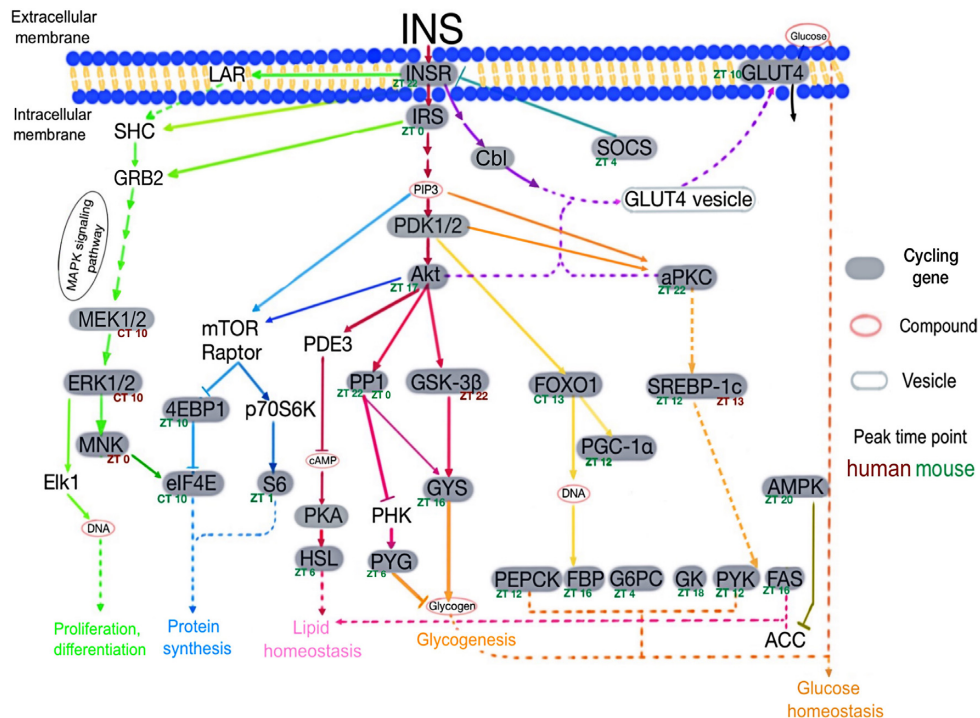
**Figure 3.** The browse options of CGDB. We provide two methods to browse the database: (A) by species and (B) by external condition. (C) The browse results can be displayed in a tabular format with CGDB ID, UniProt/Ensembl accession, species, and protein name. (D) The detailed information of human *HTR1B*.

time point in human and mouse in the figure, which likely reflects the time of day when these genes exert most of their effects. For example, the great majority of genes involved in glucose homeostasis peaks at night (ZT/CT12-ZT/CT0) in the mouse. This is consistent with the mouse being nocturnal and consuming most of the food at night (44).

Because human cycling genes were mostly obtained from blood samples, it is possible that the enriched pathways are biased by the specific physiological properties of blood and do not reflect a more general mechanism. We next conducted KEGG analyses of mouse cycling genes, which were identified from various different tissues/organs in the body (Supplementary Figure S2). ‘Adipocytokine signaling pathway’ (KEGG ID: mmu04920), ‘peroxisome proliferator-activated receptor signaling pathway’ (KEGG ID: mmu03320), ‘peroxisome’ (KEGG ID: mmu04146), and ‘insulin signaling pathway’ (KEGG ID: mmu04910) are once again among the top enriched pathways, indicating a general necessity for time-dependent function of these pathways. This also implies a conserved role of the mammalian circadian clock in regulating these pathways.

To get a better understanding of potential functions of the circadian clock across phyla, we also performed

the KEGG-based enrichment analysis on *Drosophila*, *Neurospora* and *Arabidopsis*, three other commonly used model organisms in circadian research (Supplementary Figure S2). Metabolic pathways appear to be significantly enriched in all organisms analyzed. This evolutionary conservation suggests that daily oscillations in metabolic pathways are critical for an organism to adapt to daily alterations in the environment, and perhaps reflecting a conserved role for circadian clocks across phyla. In both mouse and *Drosophila*, ‘glutathione metabolism’ is significantly enriched, implying a conserved need for cyclic regulation of this pathway in animals. There are also pathways specifically enriched in one organism, such as ‘biosynthesis of unsaturated fatty acids’ (KEGG ID: mmu01040) in mouse, ‘terpenoid backbone biosynthesis’ (KEGG ID: dme00900) in *Drosophila*, ‘ribosome’ (KEGG ID: ncr03010) in *Neurospora*, and photosynthesis-related pathways in *Arabidopsis*. This means cyclic expression of these pathways may be required to meet the demands of specific environmental challenges encountered by each organism, and may also indicate the unique roles of the circadian clock in each organism.



**Figure 4.** The insulin signaling pathway is significantly enriched in human cycling genes (the hypergeometric test,  $p$ -value =  $9.70E-7$ , E-ratio = 2.26). Part of the insulin signaling pathway adapted from KEGG. Cycling genes are filled gray ovals, whereas compounds are open pink ovals and vesicles are open gray ovals. Peak time points for human and mouse cycling genes are indicated in brown and green, respectively. Solid lines indicate direct protein interactions while dashed lines indicate indirect effects (i.e. other factors are involved which are not shown in the figure). PPI [regulatory subunit 3B (ZT20)/regulatory subunit 3C (ZT0)].

Taken together, our study provides a highly useful data resource for further analyzing temporal pattern of gene expression and function. Some of these oscillating genes are part of the clock or directly controlled by the clock, whereas others are driven by rhythmic changes in the environment (i.e. daily changes of light, temperature, food availability, etc.) or indirectly regulated by the clock (e.g. driven by rhythms in behavior or physiological processes such as the sleep/wake cycle or timed meals). Nonetheless, it is the combined actions of all these genes that carry out the rhythmic biological processes in response and adaptation to the daily changes in environment. CGDB will be continuously maintained and updated when new cycling genes are experimentally characterized. Since some of these oscillations are clock-driven, this shall provide implications regarding the function of circadian clocks. However, not all clock genes or clock-controlled genes are rhythmic at the transcript level, as rhythmic regulations at the translational or post-translational level and/or rhythmic activities are also critical steps in circadian regulation (1–3,5). Therefore in the future, genes with oscillating protein levels and PTMs will be integrated for an even more comprehensive understanding of the regulatory mechanism and function of daily rhythms.

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR Online.

**FUNDING**

National Basic Research Program (973 project) [2013CB933900]; Natural Science Foundation of China [31471125, 81272578, 31671360, J1103514]; International Science & Technology Cooperation Program of China [2014DFB30020]; Natural Science Foundation of Anhui Province [1608085QC51]. Funding for open access charge: National Natural Science Foundation of China. *Conflict of interest statement.* None declared.

**REFERENCES**

- Li,S. and Zhang,L. (2015) Circadian control of global transcription. *Biomed. Res. Int.*, **2015**, 187809.
- Lakin-Thomas,P.L., Bell-Pedersen,D. and Brody,S. (2011) The genetics of circadian rhythms in neurospora. *Adv. Genet.*, **74**, 55–103.
- Zhang,L., Jones,C.R., Ptacek,L.J. and Fu,Y.H. (2011) The genetics of the human circadian clock. *Adv. Genet.*, **74**, 231–247.
- Lowrey,P.L. and Takahashi,J.S. (2011) Genetics of circadian rhythms in mammalian model organisms. *Adv. Genet.*, **74**, 175–230.
- Hogenesch,J.B. and Ueda,H.R. (2011) Understanding systems-level properties: timely stories from the study of clocks. *Nat. Rev. Genet.*, **12**, 407–416.
- 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**, 764–775.
- Sahar,S. and Sassone-Corsi,P. (2009) Metabolism and cancer: the circadian clock connection. *Nat. Rev. Cancer*, **9**, 886–896.
- Kondratova,A.A. and Kondratov,R.V. (2012) The circadian clock and pathology of the ageing brain. *Nat. Rev. Neurosci.*, **13**, 325–335.
- Bedrosian,T.A., Fonken,L.K. and Nelson,R.J. (2016) Endocrine effects of circadian disruption. *Annu. Rev. Physiol.*, **78**, 109–131.

10. Masri, S. and Sassone-Corsi, P. (2013) The circadian clock: a framework linking metabolism, epigenetics and neuronal function. *Nat. Rev. Neurosci.*, **14**, 69–75.
11. Zhang, L., Ptacek, L.J. and Fu, Y.H. (2013) Diversity of human clock genotypes and consequences. *Prog. Mol. Biol. Transl. Sci.*, **119**, 51–81.
12. Cirelli, C. (2009) The genetic and molecular regulation of sleep: from fruit flies to humans. *Nat. Rev. Neurosci.*, **10**, 549–560.
13. Sehgal, A. and Mignot, E. (2011) Genetics of sleep and sleep disorders. *Cell*, **146**, 194–207.
14. Marcheva, B., Ramsey, K.M., Buhr, E.D., Kobayashi, Y., Su, H., Ko, C.H., Ivanova, G., Omura, C., Mo, S., Vitaterna, M.H. *et al.* (2010) Disruption of the clock components CLOCK and BMAL1 leads to hypoinsulinaemia and diabetes. *Nature*, **466**, 627–631.
15. Zram, R.V., Kowalczyk, M.S., de Boer, C.G., Schneider, R.K., Miller, P.G., McConkey, M., Tothova, Z., Tejero, H., Heckl, D., Jaras, M. *et al.* (2016) Core circadian clock genes regulate leukemia stem cells in AML. *Cell*, **165**, 303–316.
16. Zhang, R., Lahens, N.F., Ballance, H.I., Hughes, M.E. and Hogenesch, J.B. (2014) A circadian gene expression atlas in mammals: implications for biology and medicine. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 16219–16224.
17. Levi, F., Okyar, A., Dulong, S., Innominato, P.F. and Clairambault, J. (2010) Circadian timing in cancer treatments. *Annu. Rev. Pharmacol. Toxicol.*, **50**, 377–421.
18. Levi, F. and Schibler, U. (2007) Circadian rhythms: mechanisms and therapeutic implications. *Annu. Rev. Pharmacol. Toxicol.*, **47**, 593–628.
19. Zhang, E.E. and Kay, S.A. (2010) Clocks not winding down: unravelling circadian networks. *Nat. Rev. Mol. Cell Biol.*, **11**, 764–776.
20. Wu, G., Zhu, J., Yu, J., Zhou, L., Huang, J.Z. and Zhang, Z. (2014) Evaluation of five methods for genome-wide circadian gene identification. *J. Biol. Rhythms*, **29**, 231–242.
21. Straume, M. (2004) DNA microarray time series analysis: automated statistical assessment of circadian rhythms in gene expression patterning. *Methods Enzymol.*, **383**, 149–166.
22. Wichert, S., Fokianos, K. and Strimmer, K. (2004) Identifying periodically expressed transcripts in microarray time series data. *Bioinformatics*, **20**, 5–20.
23. 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.
24. 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.
25. Yang, R. and Su, Z. (2010) Analyzing circadian expression data by harmonic regression based on autoregressive spectral estimation. *Bioinformatics*, **26**, i168–174.
26. Yang, R., Zhang, C. and Su, Z. (2011) LSPR: an integrated periodicity detection algorithm for unevenly sampled temporal microarray data. *Bioinformatics*, **27**, 1023–1025.
27. Hughey, J.J., Hastie, T. and Butte, A.J. (2016) ZeitZeiger: supervised learning for high-dimensional data from an oscillatory system. *Nucleic Acids Res.*, **44**, e80.
28. Agostinelli, F., Ceglia, N., Shahbaba, B., Sassone-Corsi, P. and Baldi, P. (2016) What time is it? Deep learning approaches for circadian rhythms. *Bioinformatics*, **32**, i8–i17.
29. Pizarro, A., Hayer, K., Lahens, N.F. and Hogenesch, J.B. (2013) CircaDB: a database of mammalian circadian gene expression profiles. *Nucleic Acids Res.*, **41**, D1009–D1013.
30. Pembroke, W.G., Babbs, A., Davies, K.E., Ponting, C.P. and Oliver, P.L. (2015) Temporal transcriptomics suggest that twin-peaking genes reset the clock. *eLife*, **4**, e10518.
31. Leming, M.T., Rund, S.S., Behura, S.K., Duffield, G.E. and O'Tousa, J.E. (2014) A database of circadian and diel rhythmic gene expression in the yellow fever mosquito *Aedes aegypti*. *BMC Genomics*, **15**, 1128.
32. Rund, S.S., 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. U.S.A.*, **108**, E421–E430.
33. Rund, S.S., Gentile, J.E. and Duffield, G.E. (2013) Extensive circadian and light regulation of the transcriptome in the malaria mosquito *Anopheles gambiae*. *BMC Genomics*, **14**, 218.
34. Cunningham, F., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S. *et al.* (2015) Ensembl 2015. *Nucleic Acids Res.*, **43**, D662–D669.
35. Huang, Z., Ma, L., Wang, Y., Pan, Z., Ren, J., Liu, Z. and Xue, Y. (2015) MiCroKITS 4.0: a database of midbody, centrosome, kinetochore, telomere and spindle. *Nucleic Acids Res.*, **43**, D328–D334.
36. Boratyn, G.M., Camacho, C., Cooper, P.S., Coulouris, G., Fong, A., Ma, N., Madden, T.L., Matten, W.T., McGinnis, S.D., Merezuk, Y. *et al.* (2013) BLAST: a more efficient report with usability improvements. *Nucleic Acids Res.*, **41**, W29–W33.
37. The UniProt Consortium. (2015) UniProt: a hub for protein information. *Nucleic Acids Res.*, **43**, D204–D212.
38. Huntley, R.P., Sawford, T., Mutowo-Muullenet, P., Shypitsyna, A., Milla, C., Martin, M.J. and O'Donovan, C. (2015) The GOA database: gene ontology annotation updates for 2015. *Nucleic Acids Res.*, **43**, D1057–D1063.
39. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. and Tanabe, M. (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.*, **44**, D457–D462.
40. Aviram, R., Manella, G., Kopelman, N., Neufeld-Cohen, A., Zwighaft, Z., Elimelech, M., Adamovich, Y., Golik, M., Wang, C., Han, X. *et al.* (2016) Lipidomics analyses reveal temporal and spatial lipid organization and uncover daily oscillations in intracellular organelles. *Mol. Cell*, **62**, 636–648.
41. Borgs, L., Beukelaers, P., Vandenbosch, R., Belachew, S., Nguyen, L. and Malgrange, B. (2009) Cell "circadian" cycle: new role for mammalian core clock genes. *Cell Cycle*, **8**, 832–837.
42. Janich, P., Arpat, A.B., Castelo-Szekely, V., Lopes, M. and Gatfield, D. (2015) Ribosome profiling reveals the rhythmic liver transcriptome and circadian clock regulation by upstream open reading frames. *Genome Res.*, **25**, 1848–1859.
43. Perelis, M., Ramsey, K.M. and Bass, J. (2015) The molecular clock as a metabolic rheostat. *Diabetes Obes. Metab.*, **17**(Suppl. 1), 99–105.
44. Patton, D.F. and Mistlberger, R.E. (2013) Circadian adaptations to meal timing: neuroendocrine mechanisms. *Front. Neurosci.*, **7**, 185.