communications biology

ARTICLE

https://doi.org/10.1038/s42003-022-03722-0 OPEN

Gene networks under circadian control exhibit diurnal organization in primate organs

Jie Li [1,3], Pengxing Nie [1,3], Christoph W. Turck[2] & Guang-Zhong Wang [1✉]

Mammalian organs are individually controlled by autonomous circadian clocks. At the molecular level, this process is defined by the cyclical co-expression of both core transcription factors and their downstream targets across time. While interactions between these molecular clocks are necessary for proper homeostasis, these features remain undefined. Here, we utilize integrative analysis of a baboon diurnal transcriptome atlas to characterize the properties of gene networks under circadian control. We found that 53.4% (8120) of baboon genes are oscillating body-wide. Additionally, two basic network modes were observed at the systems level: daytime and nighttime mode. Daytime networks were enriched for genes involved in metabolism, while nighttime networks were enriched for genes associated with growth and cellular signaling. A substantial number of diseases only form significant disease modules at either daytime or nighttime. In addition, a majority of SARS-CoV-2-related genes and modules are rhythmically expressed, which have significant network proximities with circadian regulators. Our data suggest that synchronization amongst circadian gene networks is necessary for proper homeostatic functions and circadian regulators have close interactions with SARS-CoV-2 infection.

[1] CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China. [2] Max Planck Institute of Psychiatry, Proteomics and Biomarkers, Munich 80804, Germany. [3] These authors contributed equally: Jie Li, Pengxing Nie. ✉email: guangzhong.wang@picb.ac.cn

The mammalian circadian system is hierarchical in structure, with the brain's suprachiasmatic nucleus (SCN) acting as a master pacemaker to orchestrate clocks in other organs[1–3]. This system is replicated in individual cells, which need to be synchronized with each other to properly perform their organ's functions[3–5]. The central clock regulates peripheral clocks slightly each day to adapt to the 24 h day-night cycle, as the intrinsic circadian period is longer than a day[6]. At the molecular level, circadian oscillation in mammals is regulated by the transcriptional/translational autoregulatory feedback loop (TTFL). The core circadian network consists of CLOCK (along with its paralog NPAS2) and its heterodimeric partner BMAL1 (ARNTL), which bind to their downstream targets including *PER1*, *PER2*, *CRY1*, and *CRY2*[7–9]. These core clock genes regulate the expression of other molecules, thus generating the rhythmicity of various biological physiology. Although it is well known that a substantial fraction (3–16%) of transcribed mRNAs show rhythmic expression[10–13], the effects of global circadian synchronization on the transcriptomes of individual organs remains elusive.

To help understand the influence of synchronization between subordinate oscillators, we propose the concepts of "global cycling genes" and "rhythmic interactions" for analysis of the transcriptome. Global cycling genes are those transcripts that exhibit a rhythmic expression profile across multiple organs/tissues. Rhythmic interactions refer to gene pairs that are rhythmic across time. Testing these concepts has been difficult so far because only one or a few tissues are analyzed by high-throughput sequencing technologies in any given study[14–16]. Still, several large-scale circadian transcriptomes have been recently reported[13,17,18], making examination of this conceptual framework feasible. Through quantification of the circadian gene expression of 12 mouse organs, ~43% of protein-coding genes and a large number of non-coding genes show oscillations in their expression in at least one organ[13]. Likewise, 64 baboon tissues were collected every 2 h and utilized in RNA-seq and the data was then constructed into a high-resolution atlas of the circadian transcriptome[17]. Use of such comprehensive datasets could allow examination of the molecular design principles within the circadian clock from a systemic perspective[19].

Here, by integrative analysis of the high-resolution baboon diurnal transcriptome, we identified a large number of global cycling genes and cycling interactions, which implicate the fundamental circadian network organization of the primate body. We then characterized the co-expression properties of networks across different time points and discovered that networks can exhibit either "daytime" or "nighttime" status. Finally, we found that SARS-CoV-2 related proteins tend to be encoded by global cycling genes whose module is linked to circadian rhythms in the protein-protein interaction network. Together, these results demonstrate the effect of circadian rhythm on the whole-body primate transcriptome.

Results

More than half of all baboon genes are globally expressed in a rhythmic pattern. We identified global cycling genes by assessing the circadian transcriptome of more than 60 baboon (*Papio anubis hamadryas*) organs[17]. No obvious outliers for any organs or circadian time points were detected based on PCA of these data (Supplementary Fig. 1). Thus, no additional normalization on FPKM value was performed to improve the detection of global cycling genes. For the 15,219 expressed genes considered, JTK_CYCLE was performed to identify the significant rhythmic signals[20,21]. Notably, 53.4% (8120) exhibited a significant circadian expression signal after a multiple comparison correction test (Benjamini-Hochberg adjusted $P < 0.05$) (Supplementary Data 1), suggesting that body-wide gene expression is of a robust, rhythmic nature (Fig. 1a). To examine the robustness of the approach, we performed leave-one-out cross validation. In each experiment, we removed one organ at a time and used the remaining data to identify the global cycling genes. By removing one organ at a time, we found that >96% global cycling genes can be reproduced in more than half of the experiments (33 of 63 experiments) and >91% global cycling genes can be detected in all experiments, demonstrating the robustness of our methodology (Supplementary Fig. 2). By using mouse circadian atlas[13], we conducted global cycling gene identification analysis on mice as well. We found that 50.5% of the genes (6,613 genes) in mice can be detected as global cycling genes, with a significant overlap with baboon (3,954) (Supplementary Fig. 3a, odds ratio (OR) = 1.58, $P = 1.32 \times 10^{-44}$). With the exception of *RORA* and *RORB*, the majority of the core circadian clock genes exhibited a significant global oscillation signal (*CLOCK, BMAL1, NPAS2, PER1, PER2, PER3, CRY1, CRY2, RORC, NR1D1, NR1D2* and *DBP*, Supplementary Fig. 4). *CLOCK* showed a weak but significant rhythmic expression profile (Benjamini-Hochberg adjusted $P = 0.022$) at body-wide level, while all other 11 core circadian genes exhibited a strong and significant oscillation signal (Benjamini-Hochberg adjusted $P < 10^{-6}$) (Supplementary Data 1).

Consistent with previous work[17], two peak phases were observed for global cycling genes, with the first peak around noon and the second at midnight (ZT06 and ZT18, Fig. 1b). Additionally, we found that the phase of the global cycling genes represents the median phase of cycling genes in each organ (Fig. 1c, d), and thus there is a strong positive correlation between these two parameters (Pearson correlation coefficient $r = 0.93$, $P < 2.2 \times 10^{-16}$) (Fig. 1d). Subsequent functional analysis of these global cycling genes revealed that they are largely involved in basic cellular functions such as metabolic process, RNA biosynthesis, cellular localization, intracellular transport, DNA repair and others (Fig. 1e), a finding that is consistent with analysis of individual organs[17]. Similar results were observed for the mouse (Supplementary Fig. 3b, c). Interestingly, cell type enrichment analysis revealed that global cycling genes are strongly associated with immune cell type and muscle cells (Supplementary Fig. 5). To further explore the effect of global cycling genes on disease, GWAS gene-set analysis was performed using MAGMA[22]. Global cycling genes, especially the ones that have phase at ZT06, are significantly enriched in schizophrenia and type 2 diabetes risk genes ($P = 0.00256$ and 0.0008, Supplementary Fig. 6). These results indicate that the identification of global cycling genes is a useful concept that helps delineate circadian synchronization between the transcriptomes of peripheral oscillators.

One third of gene-gene interactions exhibit rhythmic co-expression. Next, we examined to what extent gene-gene interactions exhibit rhythmic patterns. For this purpose, 10,000 gene pairs were randomly sampled from all protein-coding genes. For every circadian time point, Spearman correlation coefficients were estimated by integrating organ-based expression profiles (Fig. 2a). Thus, "gene-gene interaction" specifically indicates an association between the expression of gene pairs among diverse organs and high coefficients suggest similar expression profiles across organs. The significance of rhythmic co-expression for these gene pairs was determined by JTK_CYCLE[20,21] and the procedure was repeated 100 times in order to obtain an unbiased assessment. 3180 gene pairs displayed a strong signal of rhythmic co-expression (Benjamini-Hochberg adjusted $P < 0.05$) (Fig. 2b). The phase distribution of these rhythmic interactions was similar

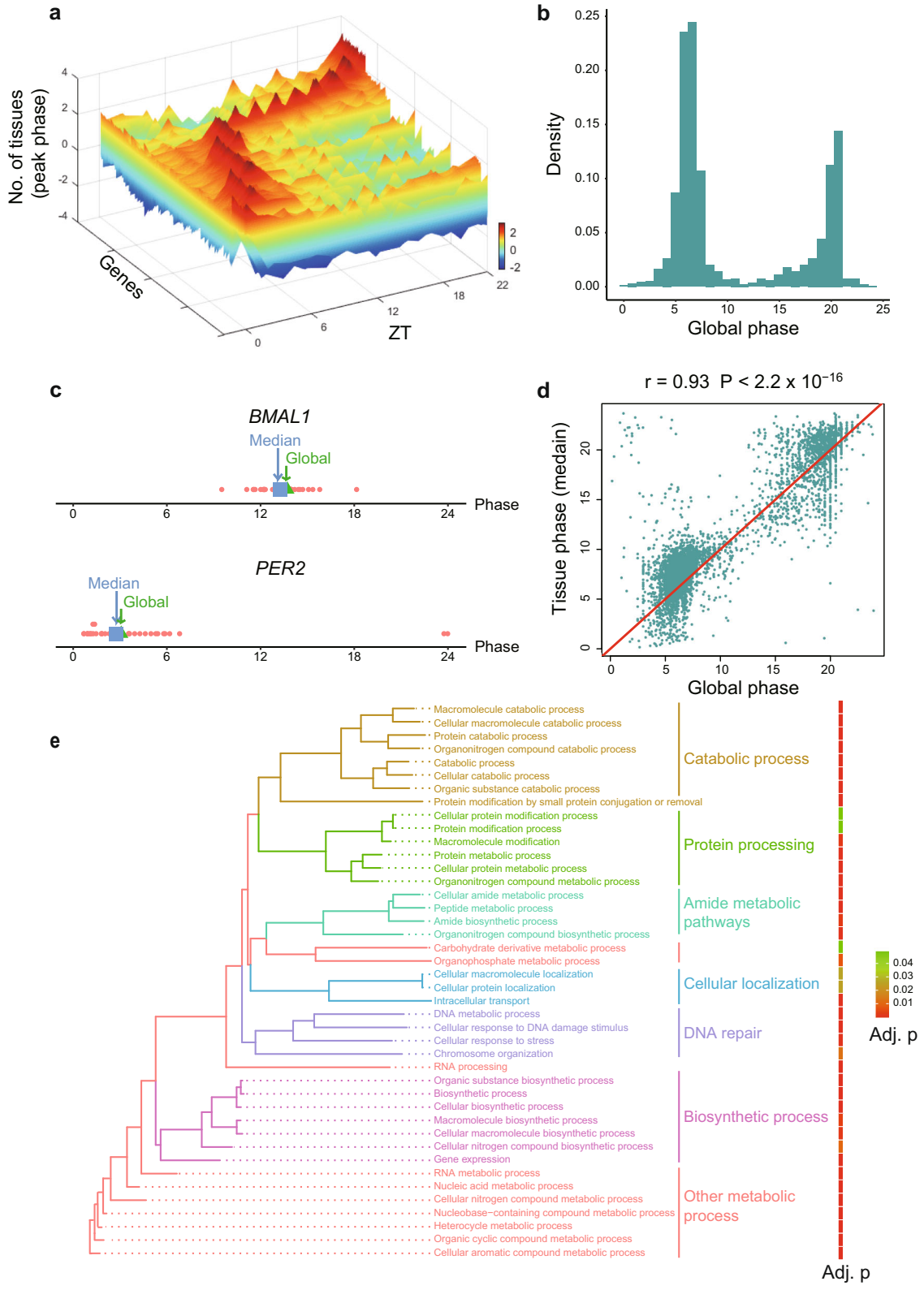

Fig. 1 Global cycling genes and their functions. a 3D plot of called global cycling genes and the number of organs with peak expression. Scale bar indicates the number of organs with peak expression, with a greater number indicated in red and a lower number indicated in blue. **b** Phase distribution of global cycling genes. **c** Global phase (blue) of *BMAL1* and *PER2* are close to the median phase (green) of individual organs. **d** The correlation between global phase and median phase in individual organs for all global cycling genes. **e** The enrichment of biological processes for global cycling genes.

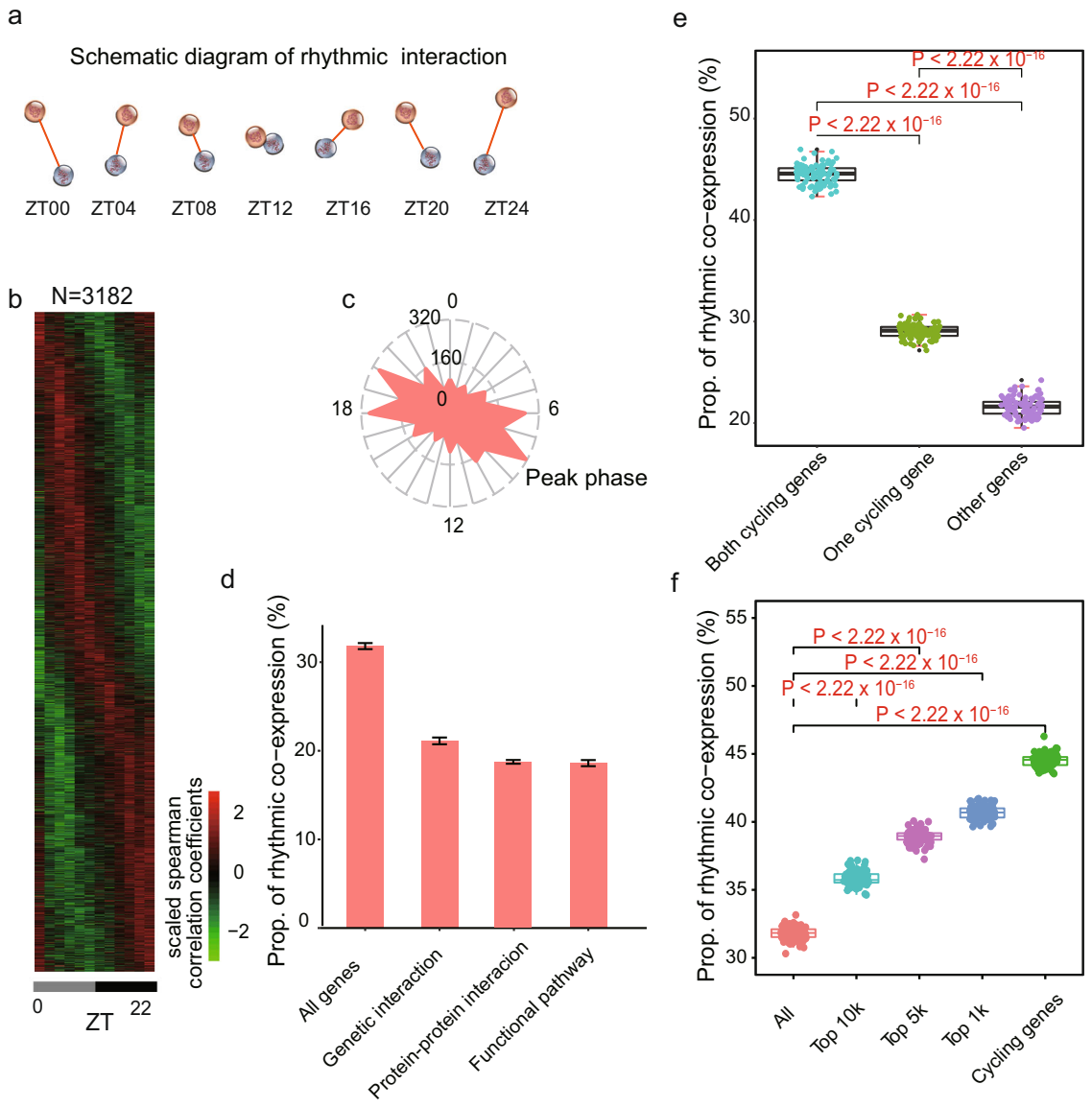

Fig. 2 Properties of rhythmic co−expression interactions. a Schematic diagram of rhythmic co-expression interactions. Spearman correlation coefficients were used to measure co-expression relationships between gene pairs at given time points. Greater coefficients indicate stronger interactions between gene pairs. **b** Heatmaps of 3,182 rhythmic co-expression interactions from 10,000 randomly sampled gene pairs. Stronger interactions are marked in red while weaker interactions are marked in green. **c** Radial plot of the peak phase distribution of rhythmic co-expression links. **d** Proportion of rhythmic co-expression interactions in different networks, including protein-protein interactions (STRING), genetic interactions (HIPPIE), and functional relevant interactions (KEGG). **e** Relationships between rhythmic co-expression links and global cycling genes. Gene pairs containing both global cycling genes show the highest percentage of rhythmic interactions while gene pairs containing no global cycling genes show the lowest percentage of rhythmic interactions. **f** Relationships between rhythmic co-expression links and expression levels. Genes were ranked from highest to lowest, according to their expression levels and top 1,000, 5000, and 10,000 were grouped in the classification.

with that of global cycling genes (Fig. 2c). As the variation of individual phase increases, both the proportion of global cycling genes and rhythmic interactions decrease rather than increase (Supplementary Fig. 7a, 7b), which suggests that the rhythmicity in rhythmic co-expression cannot be explained by the variation in the cycling phase of individual tissues. For a comparison with other types of networks, a similar approach was also applied to protein-protein interactions, genetic interactions, and functional relevant interactions (genes located in the same KEGG pathway)[23–26]. These data illustrated that approximately 20% of the interactions in these network types are rhythmically co-expressed, which is notably lower than that observed for the co-expression network ($P < 2.2 \times 10^{-16}$) (Fig. 2d).

To investigate the contribution of global cycling genes on rhythmic interactions, three situations were considered: (i) when both are global cycling genes, (ii) when only one is a global cycling gene, and (iii) when neither one is global cycling gene. For each category we examined the proportion of rhythmic interaction. For 10,000 randomly selected gene pairs in case (i), 4,450 interactions were rhythmically co-expressed, while in cases (ii) and (iii) only 2903 and 2158 gene pairs exhibited rhythmic links, respectively ($P < 2.2 \times 10^{-16}$) (Fig. 2e). As cycling genes are highly expressed[27,28], we then sampled the same number of gene pairs from the top 1000, 5000 and 10,000 expressed genes. We found that a considerably greater percentage of rhythmic co-expression links were detected in these groups than expected by

chance ($P < 2.2 \times 10^{-16}$) (Fig. 2f). Such results demonstrate that global cycling genes substantially contribute to body-wide rhythmic interactions.

Next, we assessed whether the interactions of the members of the core circadian regulatory network exhibit a rhythmic co-expression pattern. Strikingly, *CLOCK* and *BMAL1* are robustly and rhythmically co-expressed, with a peak phase at ZT13 (Benjamini-Hochberg adjusted $P = 4.22 \times 10^{-5}$), indicating that the two core circadian partners bind to each other in a time dependent manner. For the potential interactions amongst the 14 core circadian genes, 62.22% (56/91) were observed to exhibit a robust rhythmic interaction, which was significantly higher than the proportion of other detected interaction pairs (Supplementary Data 2; Supplementary Fig. 8). The high proportion of rhythmic co-expression among core circadian genes implies that the periodicity of these interactions may be a fundamental feature of core circadian oscillators[29].

Rhythmic expression of body-wide network modules. Because the co-expression of gene pairs results in distinct gene clusters that form network modules[30,31], we asked whether the latter also exhibit rhythmic characteristics. Weighted Gene Co-expression Network Analysis (WGCNA) was constructed by using all the samples from the circadian expression atlas[32,33], with 20 modules detected (Fig. 3a, b; Supplementary Data 3). 9 modules exhibited significant rhythmic signals based on either their module eigengenes or the average expression of each module (brown, yellow, cyan, tan, midnight blue, black, red, turquoise, salmon modules, Benjamini-Hochberg adjusted $P < 0.05$) (Fig. 3c; Supplementary Data 3). The peak phases of 5 of these rhythmic modules were around noon (brown, yellow, cyan, midnight blue, black modules), while the peak phases of other 4 rhythmic modules occurred around midnight (tan, red, turquoise, salmon modules), suggesting that phase peak times also impact network modules. We also found that 7 of the 9 rhythmic modules are enriched for global cycling genes (Fig. 3d). Additionally, 3 core circadian genes (*PER1, NR1D1*, and *DBP*) were found in the "brown" cycling module, whereas 6 core circadian genes (*NR1D2, CLOCK, RORA, PER3, RORC*, and *BMAL1*) were located in the "turquoise" cycling module (Fig. 3e). MAGMA analysis revealed that there is significant enrichment of disease risk genes in cycling modules (Supplementary Fig. 9). For example, the yellow module is highly related to obesity and T2D risk genes ($P = 0.03$ and 0.04, respectively) and the SCZ risk genes are overrepresented in turquoise module ($P = 0.00002$). Remarkably, 8 out of 9 cycling modules (turquoise, brown, yellow, tan, red, salmon, cyan, black) are conserved between baboon and mouse, while only one module (midnight blue) is not conserved (Supplementary Fig. 10). The results suggest that cycling modules, together with other modules, are preferably preserved during evolution These results indicate that circadian oscillation is a basic property of body-wide gene networks.

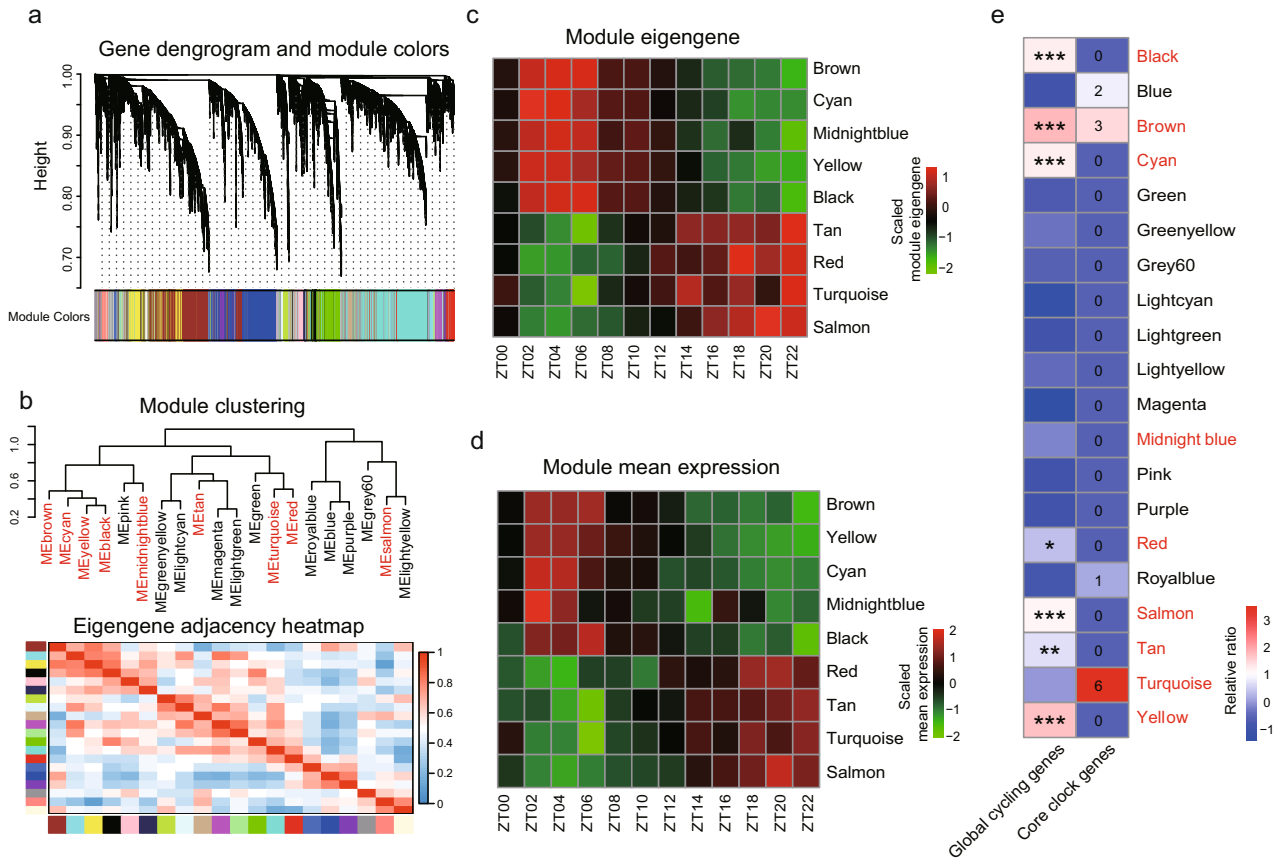

Fig. 3 Exhibition of network rhythmicity at modular level. a Gene cluster dendrograms and module dictation using all expressed genes in the circadian transcriptome atlas. **b** Hierarchical clustering dendrogram and eigengene similarity heatmap of network modules. Rhythmic modules are indicated in red. The color bar of heatmap represents correlation level among modules, with blue for low values and red for high values. Heatmap of the eight rhythmic modules indicated by eigengene or mean expression. Module eigengene (**c**) and mean expression (**d**) were scaled. Modules are ordered by circadian phase. **e** The relationship between rhythmic modules and global cycling genes. Blocks with stars indicate the significance of the enrichment (*$P < 0.01$, **$P < 0.0001$, ***$P <$ 1e−10; Fisher's exact test), and blocks with numbers indicate the number of core circadian genes included in each module.

Co-expression networks are more internally connected at nighttime than during daytime. To investigate the dynamics of body-wide networks across time, we then built detailed WGCNA modules for each circadian time point and computed the network topology accordingly[34]. In order to make different networks comparable, similar parameters were used (soft threshold = 16) for each network's construction (Supplementary Code). We found that these networks exhibited distinct properties (Supplementary Data 4). The median network connectivity at night is significantly higher than during daytime (Average degree = 38 vs. 35, respectively, $P = 0.013$) (Fig. 4a), suggesting that the expression profiles of different organs are more similar to each other at night. The network density fluctuated about 54% between daytime and night, and the network is denser at night than during daytime (0.00515 vs. 0.003), respectively, ($P = 0.0076$) (Fig. 4b). The differences found between nighttime- and daytime-assigned networks were preserved when using other network parameters, such as centrality ($P = 0.0076$) (Fig. 4c), cluster coefficients ($P = 0.021$) (Fig. 4d), heterogeneity ($P = 0.021$) (Fig. 4e), and Maximum Adjacency Ratio ($P = 0.0022$) (Fig. 4f). Overall, the topological differences suggested that the body-wide co-expression networks are flexible during daytime and robust at night (Fig. 4g–l).

Networks at the two peak phases impact distinct functions. To further characterize the differences between daytime and nighttime networks, we focused on the phase peaks at ZT06 and ZT18 (Fig. 1b), which correspond to noon and midnight, respectively. The midnight (ZT18) networks possessed both a higher connectivity and clustering coefficient than the daytime networks (ZT06) (Fig. 5a, b). 43 and 42 network modules were classified in each network, respectively (Fig. 5c, d; Supplementary Data 5). Although the two network types exhibited distinct network topologies (Fig. 5a, b), approximately half of the modules detected were shared between them (Supplementary Fig. 11). Many of these "consensus" modules were related to a specific organ (Supplementary Fig. 12; Supplementary Data 6), indicating the conservation properties of these two network types. Brain, muscle, and immune system tissues were associated with five, four, and four consensus modules, respectively (Supplementary Data 6), and included 5223 global cycling genes in total. The connectivity of global cycling genes is significantly higher than other genes in both networks ($P = 5.31 \times 10^{-13}$ for daytime network and $P = 2.55 \times 10^{-233}$ for nighttime network, Wilcoxon rank sum test, Supplementary Fig. 13). We next examined the differentially connected nodes of these two network types by focusing on genes with a network connectivity that is at least 2-fold higher in one network than the other. Of the 5755 differentially connected genes, 3767 (65%) were global cycling genes (Fig. 5e), which were over-represented ($P = 2.05 \times 10^{-122}$, Fisher's exact test). In addition, we also determined the Euclidean distance for each gene between ZT06 and ZT18 by taking the co-expression weights into consideration. We found that global cycling genes have greater Euclidean distance compared to other genes, indicating that global cycling genes tended to have different network partners than other genes in the two networks ($P < 2.2 \times 10^{-16}$) (Fig. 5f). Subsequently, we asked whether global cycling genes can explain the distinct strength of network connections. We found that global cycling gene pairs displayed greater differences in co-expression coefficients compared with cases where only one or none of the two interacting partners was a global cycling gene ($P < 2.2 \times 10^{-16}$).

Disease module is an important concept in network medicine, implicating the pathology of complex diseases and their related drug partitioning[35,36]. The daytime and nighttime networks in baboon implicate that those disease modules are not static. Indeed, we found that many diseases only form significant disease modules at either daytime or nighttime, i.e., disease modules formation depends on the network dimorphism (Fig. 5g). We have plotted 20 representative disease modules according to whether they are daytime or nighttime oriented. Interestingly, it appears that most of the diseases are nighttime oriented (rest period or with low locomotor activity in baboon) (Fig. 5h), such as glucose metabolic disorders (Supplementary Fig. 14). This may have important implications for treatment and disease pathology. Finally, we found that of the 2024 approved drug target genes in the U.S. Food and Drug Administration Drug Bank[37](https://go.drugbank.com/), >30% (634) are differentially connected between daytime and nighttime networks (Supplementary Figs. 15 and 16), of which 11% (70) represent network hubs, which further implicates the biomedical significance of the two circadian network statuses.

Besides the differences in network topology, we also found that the daytime and nighttime networks are associated with distinct functions and organs may cycle through functions throughout the day. We then focused on the genes that have less connectivity with other genes, as such these genes could indicate dissimilarities in organ activities across circadian time. Genes with fewer network partners at ZT06 (daytime) were associated with various metabolic processes (including *organic acid catabolic process*, adjusted $P = 1.86 \times 10;^{-06}$ *carboxylic acid catabolic process*, adjusted $P = 1.86 \times 10;^{-06}$ *small molecule metabolic process*, adjusted $P = 2.84 \times 10^{-05}$ and *organic substance metabolic process*, adjusted $P = 0.0002$) while genes with small amounts of network connectivity at ZT18 (nighttime) were implicated in growth-related functions (*bone mineralization*, adjusted $P = 0.035$, *response to growth factor*, adjusted $P = 0.040$ and *biomineral tissue development*, adjusted $P = 0.040$) or cellular signaling pathways (*enzyme linked receptor protein signaling pathway*, adjusted $P = 0.0059$) (Fig. 5i, j; Supplementary Fig. 17; Supplementary Data 7). Thus, the metabolic functions of specific gene modules were more dynamic during daytime, seemingly fitting the energetic demands of diurnal animals. Interestingly, five core circadian genes (*BMAL1*, *PER1*, *PER2*, *DBP*, and *RORB*), most of which cycled in a body-wide pattern, displayed at least two-times higher connectivity at night (ZT18) than at daytime (ZT06), implicating the core circadian clock in this is transcriptomic architecture.

Global cycling genes and rhythmic interactions tend to be COVID-19 related. The COVID-19 pandemic has resulted in hundreds of millions of infections with millions of those being fatal[38,39]. The relationship between SARS-CoV-2 infection and circadian regulation has not been fully explored. We thus examined the recently identified 332 human proteins that are reported to interact with SARS-CoV-2[40]. Surprisingly, we found that 69.01% (216) of these genes were global cycling genes (OR = 1.97, $P = 1.57 \times 10^{-8}$, Fisher's exact test). This enrichment cannot be explained by the high expression level of these genes ($P = 3.79 \times 10^{-8}$, logistic regression analysis). Further, among these 216 cycling genes, 92 exhibited oscillating expression in 13 human tissues[18] and 61 of them showed periodic expression in at least one organ in both mouse and human[13,18] (Supplementary Data 8). Enrichment analysis using the Human Protein Atlas[41] suggested that the 61 conserved global cycling genes were highly expressed in respiratory epithelial cells, bronchi, the placenta, the epididymis, glandular cells, the kidneys, and cells in tubules (adjusted $P < 0.001$). In addition, 52.06% of co-expression interactions among those 216 baboon global cycling genes were highly rhythmic, which was greater than for gene pairs from the entire set of 313 SARS-CoV-2 interacting genes (42.68%) or from all the global cycling genes (44.52%) in the co-expression network.

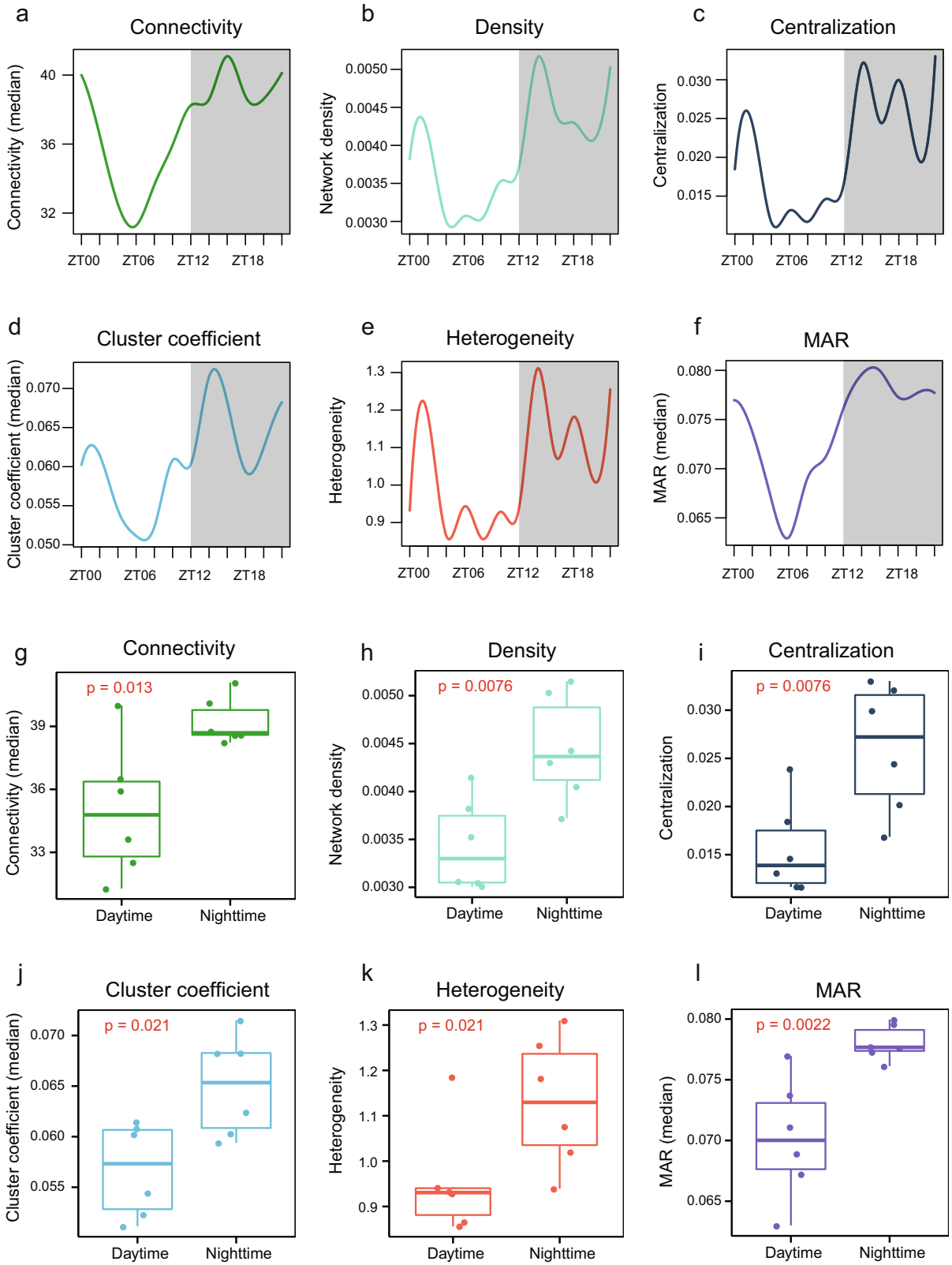

Fig. 4 Comparison of network topology at daytime and nighttime. a–f Change in network density, centralization, heterogeneity, connectivity, cluster coefficient, and maximum adjacency ratio (MAR) across different time points. Daytime phase is indicated in white and nighttime phase is indicated in gray. **g–l** Comparison of network topological parameters between daytime and nighttime networks (Wilcoxon's rank-sum test).

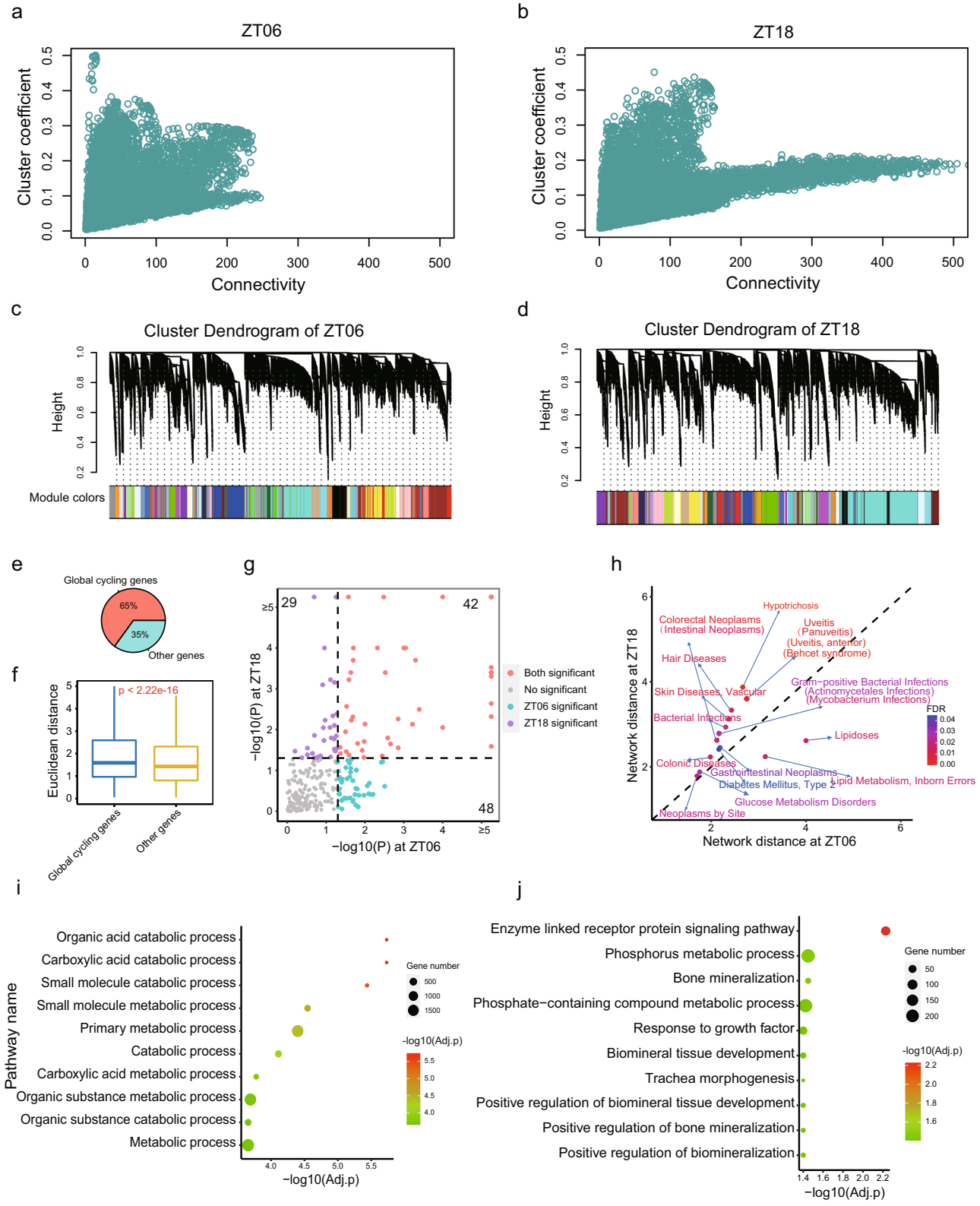

Fig. 5 Network comparisons between phase peaks at noon (ZT06) and midnight (ZT18). a, **b** The relationship between connectivity and cluster coefficients at ZT06 and ZT18. **c**, **d** Gene clustering dendrogram and module distributions at ZT06 and ZT18, of which 43 and 42 modules were assigned, respectively. **e** The proportion of global cycling genes and non-cycling genes of all differentially connected genes. **f** Euclidean distance between global cycling genes and non-cycling genes within the networks. **g** Significance of the disease modules at ZT06 and ZT18. **h** The network-based distance of representative disease modules at ZT06 and ZT18 and the statistical significance of the difference. **i** Top 10 enriched biological pathways of genes with less connectivity at ZT06. **j** Top 10 enriched biological pathways for genes with less connectivity at ZT18.

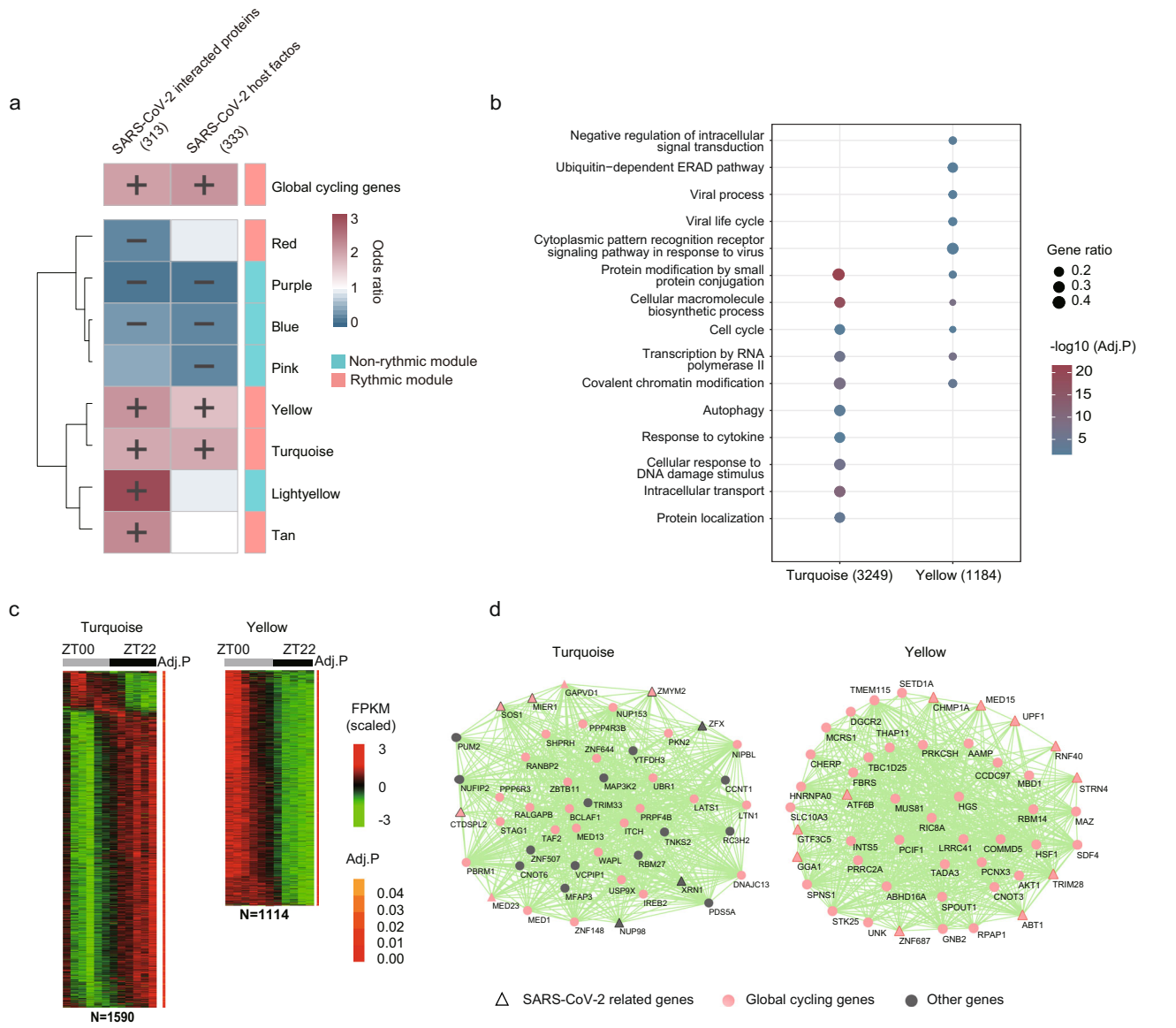

Fig. 6 Relationship between SARS-CoV-2 related genes and diurnal modules. a Heatmap shows the over-representation of SARS-CoV-2 related genes in global cycling genes and diurnal modules. **b** Functional enrichment of rhythmic modules that are enriched in SARS-CoV-2 related genes. **c** Heatmap representation of the global cycling genes in module yellow and turquoise. **d** Network view of the top 50 genes defined by the highest intramodular connectivity in module yellow and turquoise. Pink nodes are global cycling genes, dark grey nodes are non-global cycling genes, while SARS-CoV-2 related genes are shown as triangles.

To validate the enrichment between baboon global cycling genes and SARS-CoV-2 related proteins, we explored an independent dataset, i.e. SARS-CoV-2 host factors[42–45]. It appears that this enrichment still exists (OR = 2.10, $P = 2.41 \times 10^{-10}$, Fisher's exact test) and cannot be explained by expression level ($P = 9.55 \times 10^{-10}$, logistic regression analysis, Fig. 6a). We also found that >80% of the SARS-CoV-2 related genes enriched network modules are rhythmic modules (Fig. 6a), whose functions are tightly associated with autophagy ($P = 0.0017$), viral process ($P = 0.0087$) and intracellular transport ($P = 6.43 \times 10^{-12}$) (Fig. 6b–d). Furthermore, we found that for the differential connected SARS-CoV-2 interacting proteins between ZT18 and ZT06, > 90% of them are global cycling genes (33 out of 36 genes, 44 out of 48 genes, respectively).

In addition, we also estimate whether global cycling genes are enriched in other virus-related pathways. To do this, we used the 323 host factors that co-immunoprecipitated with influenza viral proteins and affected influenza virus replication (influenza host factors)[46]. It seems that influenza virus-host proteins are overrepresented in baboon global cycling genes (OR = 1.93, $P = 9.6 \times 10^{-8}$, Fisher's exact test). These results demonstrated that global cycling genes are important for various virus pathways.

Most importantly, the module of circadian regulation and SARS-CoV-2 interacting genes (either host factors or SARS-CoV-2 interacting proteins) are significantly closer in the protein-protein interaction network ($P = 0.011$ and 0.020, respectively, Supplementary Data 9), suggesting that SARS-CoV-2 infection may disrupt the circadian regulatory pathway as well. Indeed, many proteins that surround the two network modules are related to viral infection process ($P = 2.9 \times 10^{-119}$ and 4.2×10^{-123}, Fig. 7a–c) and immune response ($P = 0.0079$ and 0.0154, Fig. 7b, d, Supplementary Data 9). We thus posit that the treatment of circadian disruption may also be beneficial for the recovery of SARS-CoV-2 infection and the relief of the long COVID symptoms.

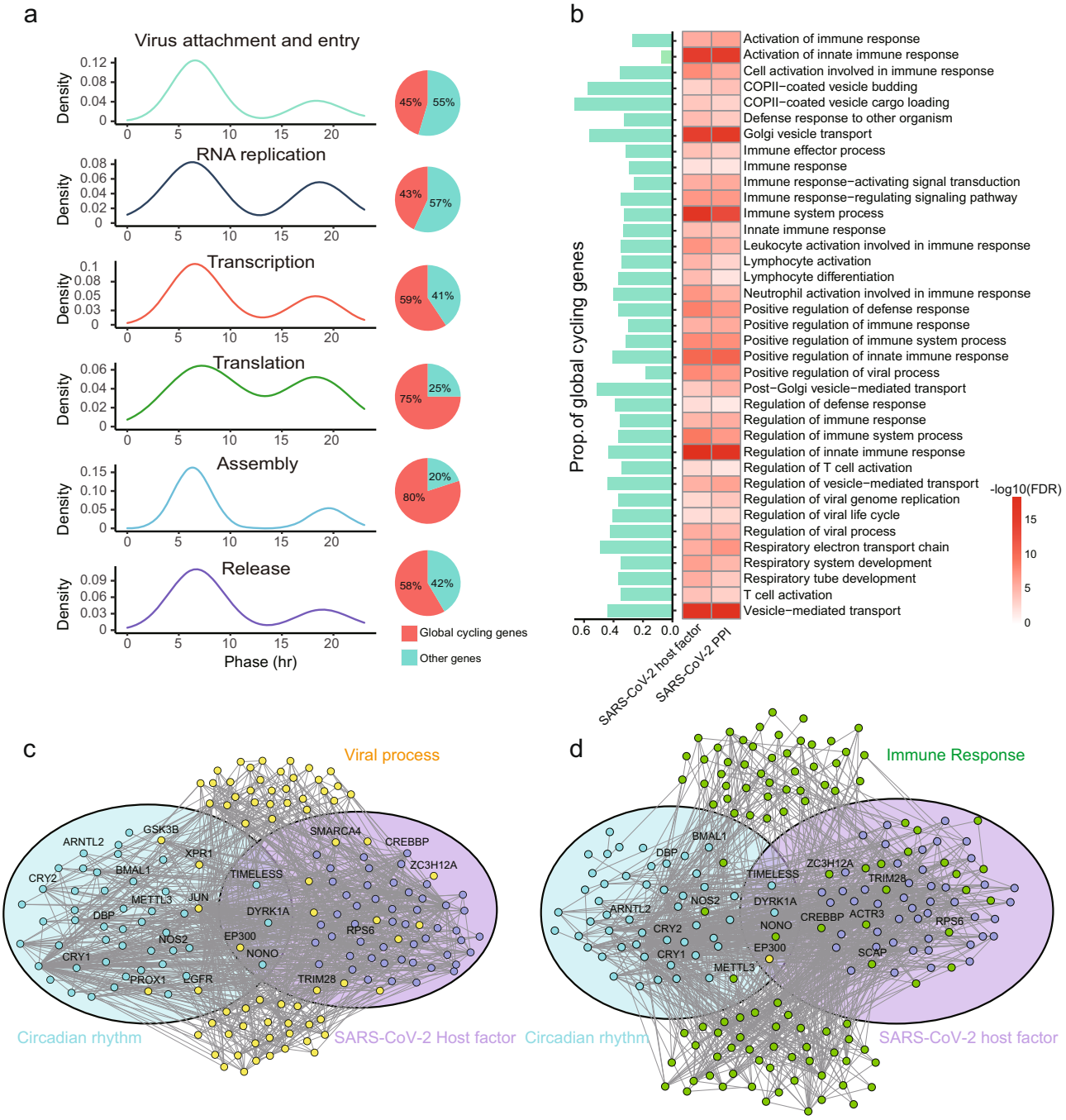

Fig. 7 Network-based model of SARS-CoV-2 related genes and circadian regulation. a Phase distributions and proportion of global cycling genes involved in different processes of viral infection. **b** Functional annotation of internodes between "circadian rhythm" and "SARS-CoV-2 host factors" or "SARS-CoV-2 interacted proteins". A network landscape of circadian rhythm related genes and SARS-CoV-2 host factors, with genes involved in viral process (**c**) and immune response (**d**) surrounding them. Light blue circle represents the genes whose function is directly related to circadian rhythm, and genes in light purple circle are SARS-CoV-2 host factors. Yellow nodes are "viral process" related genes and green nodes are "immune response" related genes.

Discussion

Individual organs require coordination with other organs to perform unified functions and ensure homeostasis. Similarly, individual cells in an organ need to be synchronized in order to implement the function of that organ, and this relies on gene co-expression. Hence, two genes that are related in function need to be properly co-expressed across time. By analyzing body-wide diurnal transcriptome data from >60 baboon organs, we found that more than half (53%) of transcripts is rhythmically expressed on a global scale. The phase of these genes represents the median phase of the cycling genes in each organ. More importantly, two modes of network status were discovered, with daytime networks associated with metabolic functions and nighttime networks associated with growth-related processes. At least one third of interactions and half of all network modules were rhythmic, revealing a cyclical nature to organ-specific output. Interestingly, this rhythmicity is widely enriched in genes encoding SARS-CoV-2 interacting proteins.

Several novel aspects of circadian rhythms were revealed by our approach of evaluating global cycling genes and rhythmic

interactions. Firstly, by integrating the circadian transcriptomes of multiple organs, robust detection of cycling genes was achieved with high-confidence (Benjamini-Hochberg adjusted $P < 0.05$), even though the sequencing depth of samples was relatively low (median total of 18.7 million sequence reads per sample). Secondly, this approach allowed detection of coordination between multiple oscillators, resulting in dynamic regulation of thousands of genes, which has been demonstrated in both baboon and mouse. Thirdly, most of the core circadian genes including *CLOCK* exhibited oscillation at the body-wide level, which is distinct from their behavior in individual tissues[17]; lastly, the interactions among human SARS-CoV-2 interacting genes were also rhythmic and have close links with circadian regulation.

The nature of gene networks has been extensively studied and several important features have been revealed. For instance, 'small-world' and 'scale-free' features characterize the structural organization of biological networks[47,48]. Hub genes in the network frequently indicate portend the functional importance of a particular gene[49,50]. Our data show that rhythmicity is another fundamental feature of biological networks. This property might be extended to protein-protein interaction networks, genetic interaction networks, functional association networks and regulatory networks, although the proportion of the rhythmic interactions each network could differ. The rhythmicity of biological networks reveals that gene networks may function in a temporally-organized manner, allowing different parts of the network to work sequentially.

Overall, co-expression networks were significantly more internally connected at night suggesting that there are two basic network modes: daytime and nighttime mode. This network dimorphism implicates that circadian clock is regulating metabolism to assist feeding/digestion during the day and tissue repair/growth at night which may have biomedical implications and is consistent with the fact that the baboon has higher locomotor activity and body temperature (enhanced metabolic functions) during daytime compared with nighttime. It has been proposed that circadian time should be considered as a factor during drug development, as it may affect drug efficiency, especially for drugs with short half-lives[51–53]. Indeed, we found that many diseases only form disease modules either at daytime or nighttime. Since network dimorphism indicates the similarity of drug target activities in these organs, we submit that for circadian medicine, these two network modes need to be validated and then tested for their usefulness in drug development and usage.

Although the cycling genes display two phase peaks, the time point of these phase peaks can differ between genes[13,17]. Moreover, the diurnal nature of such peaks has remained elusive. By constructing peak phase co-expression networks, we have found that daytime- and nighttime-assigned networks have differential topologies and associated functions, with greater regulation of metabolic pathways during daytime and more regulation of growth and cell-signaling functions at night. The function of these peak phases is consistent with the daily activity of baboon.

The inflammatory course of viral infections are closely linked to the circadian system[54–57], especially for influenza. The rhythmic expression of genes encoding SARS-CoV-2 interacting proteins and host factors together with their enrichment in rhythmic modules shows that SARS-CoV-2 may be associated with circadian rhythms, which has potential implications for health care efforts[58,59]. The organs with the most circadian-based regulation of SARS-CoV-2 interacting genes may be the ones where the response to SARS-CoV-2 infection is time dependent. More importantly, SARS-CoV-2 related proteins have closer network proximity with circadian rhythm than expected by chance. These links may imply that SARS-CoV-2 patients should increase their "circadian robustness" by avoiding behaviors that disrupt the circadian system, such as insufficient sleep and excessive fatigue. Time restricted eating may be benefit to the recovery as well. The design of the COVID-19 vaccination may also need to take this chronobiology factor into consideration[60] and thus this represents a rich area for future investigation.

In conclusion, we find that both global rhythmic expression patterns and interaction signatures are robust features of gene networks; thus, reflecting circadian synchronization amongst different organs and tissues. We show that such models can be successfully applied to the analysis of single-cell oscillations, as the transcriptome of individual cells should be aligned together with the cycling functions of their particular organs. Forthcoming research that applies single-cell sequencing technologies to the molecular architecture of the circadian clock could help validate our findings herein[61].

Methods

Primate gene expression data. FPKM expression values from 756 samples were downloaded from NCBI's Gene Expression Omnibus, with GEO accession number: GSE98965[17]. For each organ, only genes with expression levels above 0 for at least half of all the time points were considered to be expressed. In the final dataset, the expression profiles of 15,219 genes were included for further analyses. The multi-organ single-cell data from *Tabula Sapiens* were obtained from public resource[62] and subjected to findmarker analysis[63] to search for cell type markers. Only genes with at least 4-fold higher expression level in a specific cell type were considered as a potential cell type marker.

Baboon global cycling genes identification. To detect global cycling genes, all the expression information from 63 organs was considered at a particular time point. JTK_CYCLE[20,21] was employed to detect the global rhythmic parameters, such as amplitude, phase period, and the significance level. Only genes with a Benjamini-Hochberg adjusted P value <0.05 were considered to be global cycling genes. In addition, leave-one-out cross validation was performed to verify the robustness of identification of global cycling genes, by removing one organ at a time and using the remaining 62 organs to calculate global cycling genes.

Mouse global cycling gene identification. We downloaded the circadian microarray data of 12 mouse tissues from NCBI (GSE54650)[13]. Then the expression levels of 13,087 mouse genes that are homologous with baboon genes were extracted. Next, ComBat was used to eliminate the variations among organs[64] and the JTK_CYCLE[20,21] was used to detect global cycling genes. Only genes with Benjamini-Hochberg adjusted $P < 0.05$ were considered to be global cycling genes.

Rhythmic interaction identification. 10,000 gene pairs were randomly sampled from the expressed genes. For each gene pair, the Spearman correlation coefficient was estimated based on expression among 63 organs. Then, JTK_CYCLE was used to estimate the rhythmic parameters and only gene pairs with Benjamini-Hochberg adjusted P values < 0.05 were retained as rhythmic interactions. The above procedures were repeated 100 times to obtain an unbiased assessment. Protein-protein interactions were obtained from the STRING[26] and HIPPIE[25] databases; genetic interactions were downloaded from the BioGRID[24] database, and functional relevant interactions were downloaded from the KEGG pathway database[23]. Then the proportion of rhythmic co-expression in these types of interactions was analyzed via JTK_CYCLE[20,21]. R function *pheatmap* with the parameter *scale* = *row* was used to display the rhythmic changes of interaction strength between gene pairs. z-score was used to transform the raw spearman correlation coefficients (ρ) by this formula: $\rho_i' = \frac{\rho_i - \bar{\rho}}{\sigma_\rho}$, where $\bar{\rho}$ is the mean of $\rho_1, \rho_2, \ldots, \rho_{12}$ and σ_ρ represents the standard deviation of $\rho_1, \rho_2, \ldots, \rho_{12}$.

Co-expression network analysis. To investigate rhythmic co-expression characteristics at the module level, co-expression networks were constructed based on the expression data among diverse organs. In brief, Pearson correlation coefficients were used to calculate expression similarities across tissues. Then, the power parameter was selected as 16. In total, 15,219 genes were assigned into 20 modules. To estimate whether one particular module is rhythmic, both the module eigengene and the mean expression of this module were analyzed by JTK_CYCLE[20,21]. Modules with Benjamini-Hochberg adjusted P values < 0.05 were considered to be rhythmically expressed.

Then we applied the same soft threshold (16) to build the co-expression network at each time point. To compare the network topologies at different time points, six network indices were calculated by the building function *fundamentalNetworkConcepts* in WGCNA[33]; including network connectivity, cluster coefficient, maximum adjacency ratio (MAR), network density, centralization and heterogeneity.

Co-expression network comparison at noon and midnight. To compare the peak phase networks, which occur at noon and midnight, the above expression networks built at ZT06 and ZT18 were chosen for downstream analyses, respectively. Differentially connected genes were defined as genes with >2-fold changes in their connection degree in the two networks. Then the Euclidean distance of each node in the two co-expression networks and the absolute number of differences of correlation coefficients between gene pairs was estimated. These three parameters were used for the network comparisons. To compare network modules, we calculated the module similarity based on the *jaccard index* among modules[65].

Diurnal variation analysis of disease modules. The 299 diseases were obtained as previously reported[35]. Firstly, we constructed unweighted gene co-expression networks at ZT06 and ZT18 respectively, with Pearson Correlation Coefficient greater than 0.8. Then for each of the disease gene set, we determine the shortest distance d_s based on previously reported approach[35]. To test whether disease genes tend to form a module in a network, the same number of genes was randomly selected to calculate the average distance. To detect whether the distance of disease gene sets is differential in the ZT06 and ZT18 networks, 10,000 permutation tests were performed for each disease with the same number of disease genes sampled. *P* values were corrected for multiple tests using FDR. In the disease gene enrichment analysis, the GWAS annotation of 8 diseases (obesity; insomnia; schizophrenia, SCZ; bipolar disorder, BD; heart failure, HF; nonischemic cardiomyopathy, NICM; type 2 diabetes, T2D; autistic-spectrum disorder, ASD) were collected from previously published studies[66–72].

Functional enrichment analysis. The function annotation of global cycling genes was performed by g:Profiler[73]. Each term was ranked according to the false discovery rate (FDR), and the significance threshold was set to 0.05. To assess the main function of global cycling genes, we restricted the term size to between 500 and 5000. As many enriched terms have intersections, term similarity based on the *jaccard index* was calculated and then clustered together by using a neighbor-joining method[74].

Rhythmic interactions analysis of SARS-CoV-2 linked protein and host factors. To investigate whether rhythmic interactions exist among human SARS-CoV-2 interacting proteins, all information of 332 SARS-CoV-2 interacting proteins were downloaded as recently reported[40]. In addition, 374 SARS-Cov-2 infection host factors identified by three large-scale genome-wide CRISPR screens were collected as well. 313 interacting proteins and 333 host factors were found to be expressed in the baboon circadian transcriptome and were included in the downstream analysis. Finally, rhythmic interactions among these genes were estimated accordingly based on the baboon circadian transcriptome. In addition, 323 host factors that coimmunoprecipitated with influenza viral proteins (influenza host factors) were collected from a previous reported study. The enrichment of SARS-CoV-2 interacting proteins and host factors in global cycling genes was performed using a two-sided Fisher's exact test with odds ratio >1 and an FDR-adjusted $P < 0.05$.

Human interactome. The human protein–protein interactome was assembled by Zhou et al.[75], which includes five types of protein–protein interactions (PPI): protein complexes data identified by AP-MS, binary protein-protein interactions tested by high-throughput yeast-two-hybrid (Y2H) systems, kinase–substrate interactions, signaling networks and literature-curated protein-protein interactions or protein 3D structures from public databases. The final dataset contains 17,706 proteins with 351,444 interactions.

Gene sets for circadian rhythm, viral infection process and immune related functions. We downloaded viral infection process and immune related terms gene sets from AmiGO[76]. Only the human proteins are filtered. For genes that are involved in circadian rhythm, we used the "GO class (direct)" to limit them to the ones annotated directly to this function.

Calculation of network proximity. We calculated the proximity of the SARS-CoV-2 host factors and SARS-CoV-2 interacting proteins to genes directly related to "circadian rhythm" based on previously reported methodology[75]. For the closest distance d_{AB} between group A and group B:

$$d_{AB} = \frac{1}{||A|| + ||B||}\left(\sum_{a\in A} \min_{b\in B} d(a,b) + \sum_{b\in B} \min_{a\in A} d(a,b)\right) \quad (1)$$

where $d(a,b)$ is the shortest distance of gene a and gene b in the interactome, $||A||$ represents the size of A, $||B||$ represents the size of B. To evaluate the statistical significance, 1000 permutation tests were performed in which two randomly gene sets were chosen with the same gene number and degree distribution as A and B.

Statistics and reproducibility. All the analyses were performed in R (v4.0.2). Statistical significance was determined at Benjamini-Hochberg adjusted *P* value < 0.05. Gene set enrichment analysis was performed using a two-sided Fisher's exact test.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All datasets analyzed in this paper are publicly available. The RNA sequencing data of baboon and mouse are available in NCBI's Gene Expression Omnibus, with GEO accession number: GSE98965 and GSE54650. The multi-organ single-cell data from Tabula Sapiens is available at https://tabula-sapiens-portal.ds.czbiohub.org/.

Code availability

The codes for co-expression network construction are included in the supplementary code and all other codes necessary to replicate the analyses are available upon request.

Received: 23 November 2021; Accepted: 14 July 2022;
Published online: 29 July 2022

References

1. Hastings, M. H., Reddy, A. B. & Maywood, E. S. A clockwork web: circadian timing in brain and periphery, in health and disease. *Nat. Rev. Neurosci.* **4**, 649–661 (2003).
2. Welsh, D. K., Takahashi, J. S. & Kay, S. A. Suprachiasmatic nucleus: cell autonomy and network properties. *Annu Rev. Physiol.* **72**, 551–577 (2010).
3. Albrecht, U. Timing to perfection: the biology of central and peripheral circadian clocks. *Neuron* **74**, 246–260 (2012).
4. Mohawk, J. A., Green, C. B. & Takahashi, J. S. Central and peripheral circadian clocks in mammals. *Annu Rev. Neurosci.* **35**, 445–462 (2012).
5. Dibner, C., Schibler, U. & Albrecht, U. The mammalian circadian timing system: organization and coordination of central and peripheral clocks. *Annu Rev. Physiol.* **72**, 517–549 (2010).
6. Czeisler, C. A. et al. Stability, precision, and near-24-hour period of the human circadian pacemaker. *Science* **284**, 2177–2181 (1999).
7. Partch, C. L., Green, C. B. & Takahashi, J. S. Molecular architecture of the mammalian circadian clock. *Trends Cell Biol.* **24**, 90–99 (2014).
8. Rijo-Ferreira, F. & Takahashi, J. S. Genomics of circadian rhythms in health and disease. *Genome Med.* **11**, 82 (2019).
9. Takahashi, J. S. Transcriptional architecture of the mammalian circadian clock. *Nat. Rev. Genet.* **18**, 164–179 (2017).
10. Storch, K. F. et al. Extensive and divergent circadian gene expression in liver and heart. *Nature* **417**, 78–83 (2002).
11. Panda, S. et al. Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell* **109**, 307–320 (2002).
12. Hughes, M. E. et al. Harmonics of circadian gene transcription in mammals. *PLoS Genet.* **5**, e1000442 (2009).
13. Zhang, R., Lahens, N. F., Ballance, H. I., Hughes, M. E. & Hogenesch, J. B. A circadian gene expression atlas in mammals: implications for biology and medicine. *Proc. Natl Acad. Sci. USA* **111**, 16219–16224 (2014).
14. Sun, L., Ma, J., Turck, C. W., Xu, P. & Wang, G. Z. Genome-wide circadian regulation: A unique system for computational biology. *Comput. Struct. Biotechnol. J.* **18**, 1914–1924 (2020).
15. Doherty, C. J. & Kay, S. A. Circadian control of global gene expression patterns. *Annu. Rev. Genet* **44**, 419–444 (2010).
16. Hughes, M. E. et al. Guidelines for genome-scale analysis of biological rhythms. *J. Biol. Rhythms* **32**, 380–393 (2017).
17. Mure, L. S. et al. Diurnal transcriptome atlas of a primate across major neural and peripheral tissues. *Science* **359**, eaao0318 (2018).
18. Ruben, M. D. et al. A database of tissue-specific rhythmically expressed human genes has potential applications in circadian medicine. *Sci Transl. Med.* **10**, eaat8806 (2018).
19. Hogenesch, J. B. & Ueda, H. R. Understanding systems-level properties: timely stories from the study of clocks. *Nat. Rev. Genet.* **12**, 407–416 (2011).
20. Wu, G., Anafi, R. C., Hughes, M. E., Kornacker, K. & Hogenesch, J. B. MetaCycle: an integrated R package to evaluate periodicity in large scale data. *Bioinformatics* **32**, 3351–3353 (2016).
21. Hughes, M. E., Hogenesch, J. B. & Kornacker, K. JTK_CYCLE: an efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets. *J. Biol. Rhythms* **25**, 372–380 (2010).
22. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
23. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).

24. Oughtred, R. et al. The BioGRID interaction database: 2019 update. *Nucleic Acids Res.* **47**, D529–D541 (2019).
25. Alanis-Lobato, G., Andrade-Navarro, M. A. & Schaefer, M. H. HIPPIE v2.0: enhancing meaningfulness and reliability of protein-protein interaction networks. *Nucleic Acids Res.* **45**, D408–D414 (2017).
26. Szklarczyk, D. et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
27. Wang, G. Z. et al. Cycling transcriptional networks optimize energy utilization on a genome scale. *Cell Rep.* **13**, 1868–1880 (2015).
28. Cheng, Y., Chi, Y., Zhang, L. & Wang, G. Z. A single factor dominates the behavior of rhythmic genes in mouse organs. *BMC Genom.* **20**, 879 (2019).
29. Wallach, T. et al. Dynamic circadian protein-protein interaction networks predict temporal organization of cellular functions. *PLoS Genet.* **9**, e1003398 (2013).
30. Barabasi, A. L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101–113 (2004).
31. Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. From molecular to modular cell biology. *Nature* **402**, C47–C52 (1999).
32. Langfelder, P. & Horvath, S. Eigengene networks for studying the relationships between co-expression modules. *BMC Syst. Biol.* **1**, 54 (2007).
33. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinforma.* **9**, 559 (2008).
34. Horvath, S. & Dong, J. Geometric interpretation of gene coexpression network analysis. *PLoS Comput. Biol.* **4**, e1000117 (2008).
35. Menche, J. et al. Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science* **347**, 1257601 (2015).
36. Barabasi, A. L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* **12**, 56–68 (2011).
37. Wishart, D. S. et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074–D1082 (2018).
38. Zhu, N. et al. A Novel Coronavirus from patients with Pneumonia in China, 2019. *N. Engl. J. Med* **382**, 727–733 (2020).
39. World Health Organization. Coronavirus disease (COVID-19) Weekly Epidemiological Update and Weekly Operational Update. (2020).
40. Gordon, D. E. et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020).
41. Uhlen, M. et al. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
42. Wei, J. et al. Genome-wide CRISPR screens reveal host factors critical for SARS-CoV-2 infection. *Cell* **184**, 76–91.e13 (2021).
43. Wang, R. et al. Genetic screens identify host factors for SARS-CoV-2 and common cold Coronaviruses. *Cell* **184**, 106–119.e114 (2021).
44. Schneider, W. M. et al. Genome-scale identification of SARS-CoV-2 and pan-coronavirus host factor networks. *Cell* **184**, 120–132.e114 (2021).
45. Daniloski, Z. et al. Identification of required host factors for SARS-CoV-2 infection in human cells. *Cell* **184**, 92–105.e116 (2021).
46. Watanabe, T. et al. Influenza virus-host interactome screen as a platform for antiviral drug development. *Cell Host Microbe* **16**, 795–805 (2014).
47. Barabasi, A. L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
48. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).
49. Konopka, G. et al. Human-specific transcriptional networks in the brain. *Neuron* **75**, 601–617 (2012).
50. Yang, Y. et al. Gene co-expression network analysis reveals common system-level properties of prognostic genes across cancer types. *Nat. Commun.* **5**, 3231 (2014).
51. Ruben, M. D., Smith, D. F., FitzGerald, G. A. & Hogenesch, J. B. Dosing time matters. *Science* **365**, 547–549 (2019).
52. Panda, S. The arrival of circadian medicine. *Nat. Rev. Endocrinol.* **15**, 67–69 (2019).
53. Cederroth, C. R. et al. Medicine in the fourth dimension. *Cell Metab.* **30**, 238–250 (2019).
54. Gachon, F., Yeung, J. & Naef, F. Cross-regulatory circuits linking inflammation, high-fat diet, and the circadian clock. *Genes Dev.* **32**, 1359–1360 (2018).
55. Hong, H. K. et al. Requirement for NF-kappaB in maintenance of molecular and behavioral circadian rhythms in mice. *Genes Dev.* **32**, 1367–1379 (2018).
56. Sengupta, S. et al. Circadian control of lung inflammation in influenza infection. *Nat. Commun.* **10**, 4107 (2019).
57. Edgar, R. S. et al. Cell autonomous regulation of herpes and influenza virus infection by the circadian clock. *Proc. Natl Acad. Sci. USA* **113**, 10085–10090 (2016).
58. Liu, K. et al. Effects of progressive muscle relaxation on anxiety and sleep quality in patients with COVID-19. *Complement Ther. Clin. Pr.* **39**, 101132 (2020).
59. Zhang, R. et al. COVID-19: Melatonin as a potential adjuvant treatment. *Life Sci.* **250**, 117583 (2020).
60. Long, J. E. et al. Morning vaccination enhances antibody response over afternoon vaccination: A cluster-randomised trial. *Vaccine* **34**, 2679–2685 (2016).
61. Wen, S. et al. Spatiotemporal single-cell analysis of gene expression in the mouse suprachiasmatic nucleus. *Nat. Neurosci.* **23**, 456–467 (2020).
62. Tabula Sapiens, C. et al. The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science* **376**, eabl4896 (2022).
63. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e3529 (2021).
64. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
65. Jaccard, P. Nouvelles recherches sur la distribution florale. *Bull. Soc. Vaud. Sci. Nat.* **44**, 223–270 (1908).
66. Trubetskoy, V. et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).
67. Grove, J. et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet* **51**, 431–444 (2019).
68. Stahl, E. A. et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat. Genet* **51**, 793–803 (2019).
69. Jansen, P. R. et al. Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nat. Genet* **51**, 394–403 (2019).
70. Xue, A. et al. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat. Commun.* **9**, 2941 (2018).
71. Aragam, K. G. et al. Phenotypic Refinement of Heart Failure in a National Biobank Facilitates Genetic Discovery. *Circulation* **139**, 489–501 (2018).
72. Berndt, S. I. et al. Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat. Genet.* **45**, 501–512 (2013).
73. Raudvere, U. et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191–W198 (2019).
74. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).
75. Zhou, Y. et al. A network medicine approach to investigation and population-based validation of disease manifestations and drug repurposing for COVID-19. *PLoS Biol.* **18**, e3000970 (2020).
76. Carbon, S. et al. AmiGO: online access to ontology and annotation data. *Bioinformatics* **25**, 288–289 (2009).

Acknowledgements

We thank J.B.H., Hiroki R. Ueda, Luoying Zhang, Hung-Chun Chang, Daniel J. Araujo and Yi Liu for helpful comments on the draft of the manuscript. This work was supported by the National Natural Science Foundation of China (Grant nos. 81827901, 31600960, and 31871333) and the National Key R&D Program of China (Grant nos. 2016YFC0901700 and 2016YFC1303100).

Author contributions

Conceiving the project, G-Z.W.; performing the analyses, J.L. and P.N.; writing, reviewing, and editing the manuscript, J.L., P.N., C.W.T., and G-Z.W.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-022-03722-0.

Correspondence and requests for materials should be addressed to Guang-Zhong Wang.

Peer review information *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Simona Chera and Karli Montague-Cardoso.