Supplementary Material for

Diurnal transcriptome atlas of a primate across major neural and peripheral tissues

Ludovic S. Mure, Hiep D. Le, Giorgia Benegiamo, Max W. Chang, Luis Rios, Ngalla Jillani, Maina Ngotho, Thomas Kariuki, Ouria Dkhissi-Benyahya, Howard M. Cooper,* Satchidananda Panda*

*Corresponding author. Email: howard.cooper@inserm.fr (H.M.C.); satchin@salk.edu (S.P.)

Published 8 February 2018 as *Science* First Release
DOI: 10.1126/science.aao0318

This PDF file includes:

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencemag.org/content/science.aao0318/DC1)

Tables S1 to S12 as a separate Excel file
Databases 1 and 2 as separate .zip files

Materials and Methods

Sample Collection and Processing.

Animal recordings and tissue collection was performed at the Institute of Primate Research (IPR, National Museums of Kenya, Nairobi). The IPR is a WHO Collaborating Center and is part of the EU Primate network EUPRIM.

The protocols and ethics were approved by the Kenya IACUC (Institute Animal Care and Use Committee, Approval 160811) and (ISERC, (Institution Scientific and Ethical Review Committee). The primate species used was the baboon, *Papio anubis hamadrya,* listed by the IUCN as a species of *Least Concern*, stable to increasing population in Kenya (http://www.iucnredlist.org/details/40647/0). Permits for export of biological specimens were obtained from the KENYA WILDLIFE SERVICE (No. 93-85129), CITES (No. 0380542), Import Permit to France from CITES (FR1306901234-I), re-export permit to the United States from CITES (FR1406902037-R) and import permit by the Department of Health and Human Services (2014-06-130). Although it is customary to sample tissues from multiple animals (>3) per time point in every 2 hours for at least 48 hours, due to ethical considerations to minimize the number of animals, yet obtain enough data across one full day, we decided to sample tissue from only one animal every 2 h over 24 h. Furthermore, to maximize the use of this tissue resource, we have set up two tissue banks (one in INSERM, Lyon, and another at Salk Institute, La Jolla) to enable future studies and access to remaining tissues and brain regions by our group and by any interested academic researcher.

For one month before the experiment, twelve male baboons (age 5-6 yrs, 7-11 kg) were maintained in a 12hr light-12 hr dark cycle (artificial fluorescent light, 300 lux). Animals were fed twice a day at ZT 3 (9:00) and ZT 9 (15:00) with fruits and primate meal pellets, with access to water ad libitum. Daily rest-activity rhythms were recorded using infrared movement sensitive detectors installed above each cage (IPR Chronobiology facility, CAMS activity system™) and body temperature recorded with temperature loggers (i-button, Texas Inst) implanted between the scapula. The facility was not insulated regarding temperature and sound.

Timed biosamples were collected every 2 hrs from ZT0 to ZT22, one primate at each sample time. The Institute veterinary services administered animals a lethal dose of sodium pentobarbital, after which blood (centrifuged for plasma), urine and CSF samples collected. Animals were perfused through the left ventricle with 4 liters of ice cold (2°C) buffered, oxygenated Ame's solution. Collection of brain sections used standardized procedures for NHPs (*20*). The brain and eyes were rapidly removed (less than 5 min) and kept at 4°C on ice. Tissue collections were made simultaneously in three stations. In Station 1, the brain was cut into 4 mm thick serial sections using a specifically built baboon brain matrix (J.A. Cooper). Sections were placed on individual 0.25 inch thick aluminium ice cold plates, each section photographed and punches obtained from selected brain structures (see Fig.1A and Table S1). The Primate Brain map (corrected with the *Papio* linear ratio (*21*)) was used as the reference guide for brain tissue punches. Sections were packaged in individual sterilized plastic pouches (Nasco™) then frozen in dry ice and stored at -80°C. At the same time (Station 2), samples were taken from the eyes (retina, RPE, vitreous, lens,

cornea, iris, optic nerve), divided into aliquots and frozen in dry ice. During the night-time periods, the head and eyes were covered with opaque plastic prior to perfusion and dissections of the eye and retina were performed in the dark using infrared viewers. Station 3 collected various body tissues and organs, listed in Fig.1A and Table S1. All tissues were distributed into individual aliquots with certain larger tissues stored in bulk samples (liver, muscle, etc.), frozen in dry ice, and then stored at -80°C. The whole procedure, from collection to storage lasted no more than 120 min.

The INSERM-SALK NHP tissue bank collection consisted of 284 brain sections (22 to 25 per animal), and 42 peripheral tissues divided into 1008 aliquots.

RNA extraction

Total RNA was extracted from samples using various methods optimized for the type of tissue. For many samples, RNA was isolated and purified by QIAGEN's RNeasy Mini kit according to the manufacturer's instructions. For fatty samples, Trizol (Thermo Fisher) was used following manufacturer's instructions to obtain RNA. Proteinase K was added to samples that contained high amounts of connective tissue. Homogenization methods such as rotor/stator, bead-based, pestle and Qiagen's QIAshredder were used to improve RNA extraction.

Tissue	Method		Tissue	Method		Method	Description
ADC	1		OES	3		1	QIAshredder, RNeasy Mini
ADM	1		OLB	3		2	Rotor/stator homogenization, Proteinase K, RNeasy Mini
AMY	1		OMF	3		3	Rotor/stator homogenization, Trizol
ANT	3		ONH	3		4	Bead-based homogenization, Proteinase K, RNeasy Mini
AOR	3		PAN	3		5	Pestle Homogenization, Rneasy Mini
ARC	3		PIN	1			
ASC	3		PIT	1			
AXL	3		PON	3			
BLA	3		PRA	5			
BOM	3		PRC	3			
CEC	3		PRO	3			
CER	1		PUT	3			
COR	4		PVN	5			
DEC	3		RET	3			
DMH	5		RPE	2			
DUO	3		SCN	5			
HAB	3		SKI	2			
HEA	3		SMM	3			
HIP	1		SON	3			
ILE	3		SPL	1			
IRI	2		STF	3			
KIC	1		SUN	3			
KIM	1		TES	3			
LGP	3		THA	3			
LH	3		THR	3			
LIV	1		VIC	3			
LUN	3		VMH	3			
MEL	3		WAM	3			
MGP	3		WAP	3			
MMB	3		WAR	3			
MUA	3		WAS	3			
MUG	3		WAT	3			

Libraries preparation and high-throughput sequencing

Libraries were prepared using Illumina's TruSeq Stranded mRNA HT kit according to manufacturer's instructions. In brief, total RNA starting with 1ug was poly-A selected, fragmented by metal-ion hydrolysis and then converted to cDNA using SuperScript II. For four tissues - habenula, lateral globus pallidus, medial globus pallidus, and substantia negra only 400ng of total RNA was used for library production due to the limiting nature of the source tissue. The cDNA was then end-repaired, adenylated and ligated with Illumina sequencing adapters. Finally, the libraries were enriched by 15 cycles of PCR amplification. Libraries were pooled and sequenced using an Illumina HiSeq 2500 with 50-bp single-read chemistry.

Read mapping and annotation

Sequencing reads were aligned to the olive baboon PapAnu2.0 genome using STAR aligner (*22*). Gene-level read counts were obtained using featureCounts (*23*) with gene annotation from Ensembl (http://www.ebi.ac.uk/ena/data/view/GCA_000264685.1). DESeq2 (*24*) was used to generate normalized read counts and FPKM values.

Statistical analysis of rhythmic gene expression

Only genes with the average of FPKM values of the 12 time points >0.1 were considered expressed (*9*).We then use meta2d, a function of the R package Metacycle, to evaluate periodicity in the RNA seq data. Briefly, meta2d incorporates ARSER, JTK_CYCLE and Lomb-Scargle and implements N-version programming concepts using a suite of algorithms and integrating their results (p and q values, period, phase and amplitude). Genes were considered to be rhythmically expressed when the integrated p value was <0.05. Other analysis, comparisons and statistical tests were performed using Matlab (The MathWorks, Inc) and Graph Pad Prism 5.

Gene ontology and pathway over-representation analysis (ORA)

For gene ontology and pathway ORA analysis *Papio Anubis* ensembl gene IDs were converted to the orthologous human ensemble IDs and analysed using GO-Elite homo sapiens databases (*16*). Each analyzed term was ranked according to a Z-score, calculated with a normal approximation to the hypergeometric distribution along with a permutation or a Fisher's exact test P-value. False-discovery rate adjusted P-values are calculated using a Benjamini–Hochberg correction. The ontology ORA results from this step were further evaluated by a pruning method. Pruning occurs by importing the ORA statistics (Z-score, P-values and gene counts), matching default filtering options and building all unique branch paths of these results based on the ontology tree structure. Branch paths are pruned to obtain the nodes with the largest Z-score relative to all corresponding child and parent nodes, to report the most informative, highest scoring term for a network of related terms. We used 2000 permutations, Z-score cut-off of 2, permuted p-value cut-off of 0.05. Ontology terms were sorted by Z-score.

Phase Set Enrichment Analysis (PSEA)

To analyse rhythmicity of annotated groups of genes, we used Phase Set Enrichment Analysis software (*17*). Files containing the list of cycling genes relative to a given gene ontology annotation and their peak phases of expression were inputted in PSEA software. PSEA evaluated each selected gene set for evidence of temporally coordinated transcription and provided summary statistics.

Mouse data sets

For comparison between Baboon and mouse rhythmic gene expression, gene expression data from twelve different mouse organs were obtained from recent studies (*1*, *18*, *25*). ADR, AOR, BST, CER, HEA, KID, LIV, LUN, MUS, WAT datasets are from RNA seq experiments (*1*), PIT and SCN (*18*) from microarray experiments. Mouse ADR, BST and KID were compared to baboon ADC, PON and KIC respectively.

The mouse data were re-analyzed using meta2d, a function of the R package Metacycle. As described for the baboon data, genes were considered to be rhythmically expressed when the integrated p value was <0.05.

Data availability

The data discussed in this study have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number: GSE98965. Rhythmicity parameters and statistics calculated via metacycle are available in supplementary databases S1 for the baboon and databases S2 for the mouse.

Supplementary figures.

FigureS1

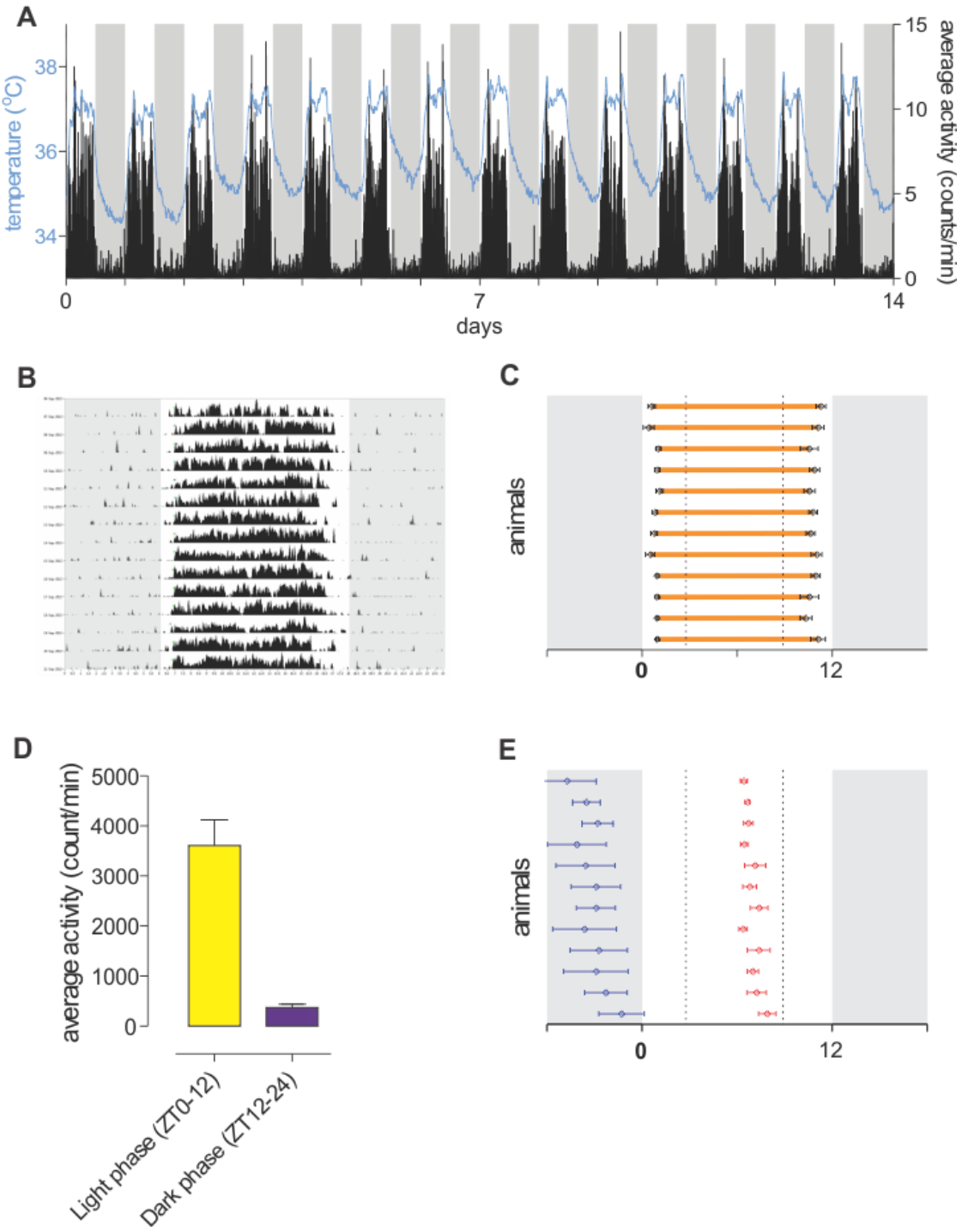

Fig. S1: (**A**) Average locomotor activity and body temperature of the 12 animals over the 2 weeks immediately preceding the tissue collection (n=12). (**B**) Example of actogram of locomotor activity of a baboon for two weeks before the experiment. Each horizontal line represents one day of activity. Shaded portion represents nighttime. (**C**) mean onsets and offsets of locomotor activity and active period (interval shown in orange bar) of the 12 baboons for two weeks preceding the experiment. (**D**) Average locomotor activity that occurred daily during the light and dark phases respectively (average =/-SEM, n=12). (**E**) Average times of minimum temperature (blue) and mid-maximum body temperatures phase (red).

FigureS2

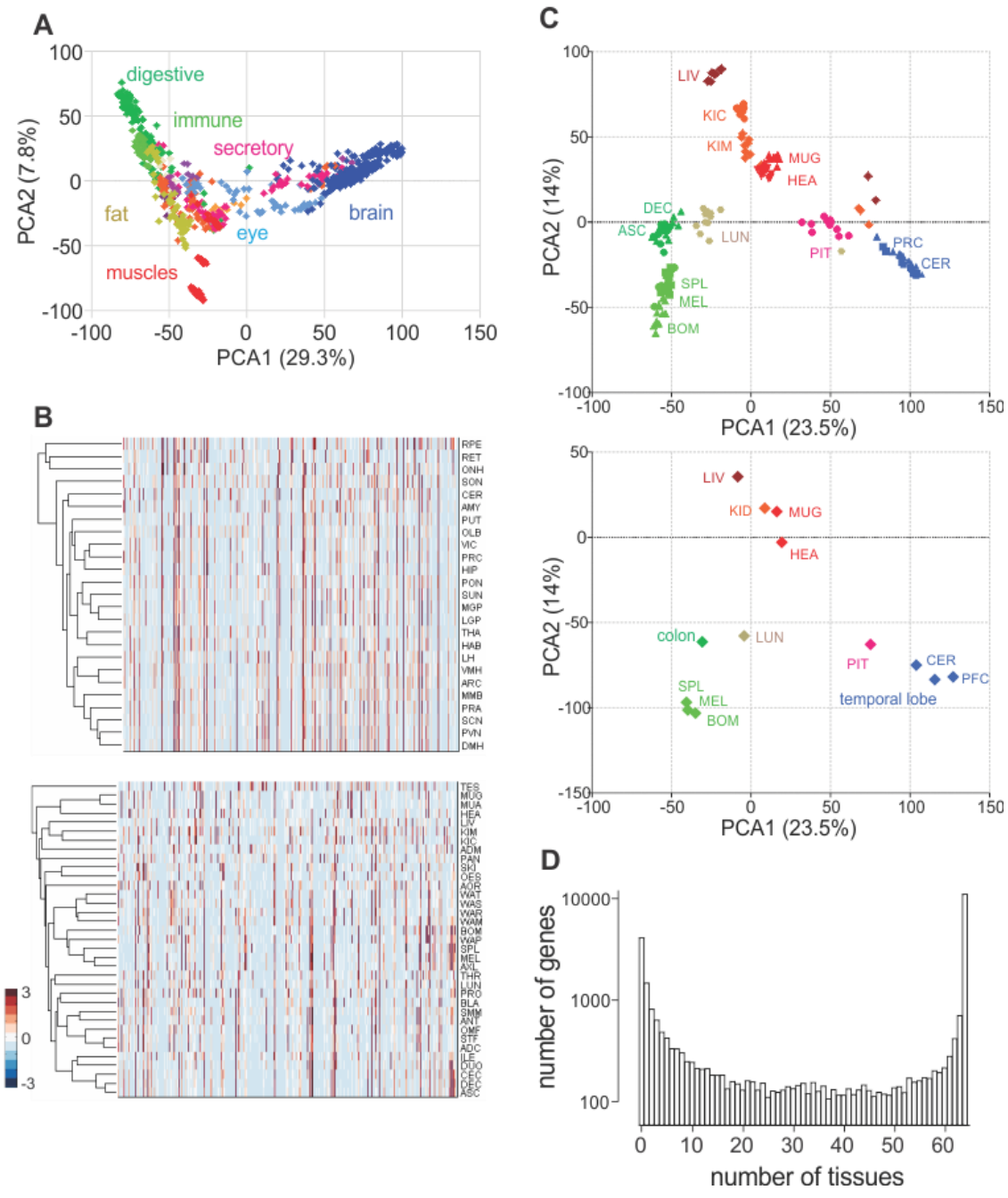
A
digestive
immune
secretory
fat
brain
eye
muscles
PCA2 (7.8%)
PCA1 (29.3%)
B
C
LIV
KIC
KIM
MUG
HEA
DEC
ASC
LUN
PIT
PRC
CER
SPL
MEL
BOM
PCA2 (14%)
PCA1 (23.5%)
KID
colon
PFC
temporal lobe
D
number of genes
number of tissues

Fig. S2: (**A**) Principal component analysis performed on the 12 time points of the 64 tissues of the present study. (**B**) Dendrograms obtained from the clustering of the central nervous system (above) and the peripheral organs (below) transcriptomes (log2(expressed genes-UEGs)). (**C**) Principal component analysis performed on the 12 time points of 14 tissues of the present study (above) and 13 corresponding tissues (below) from the NHPRTR database (*12*) shows similar clustering of tissues groups based on gene expression profiles. Data from the Pipes et al study are from an RNA seq experiment on one female *Papio anubis* of 6 years 2 months. (**D**) Distribution of the genes as a function of the number of tissues they are expressed in (genes are considered expressed if their average FPKM over the 12 time points is >0.1).

FigureS3

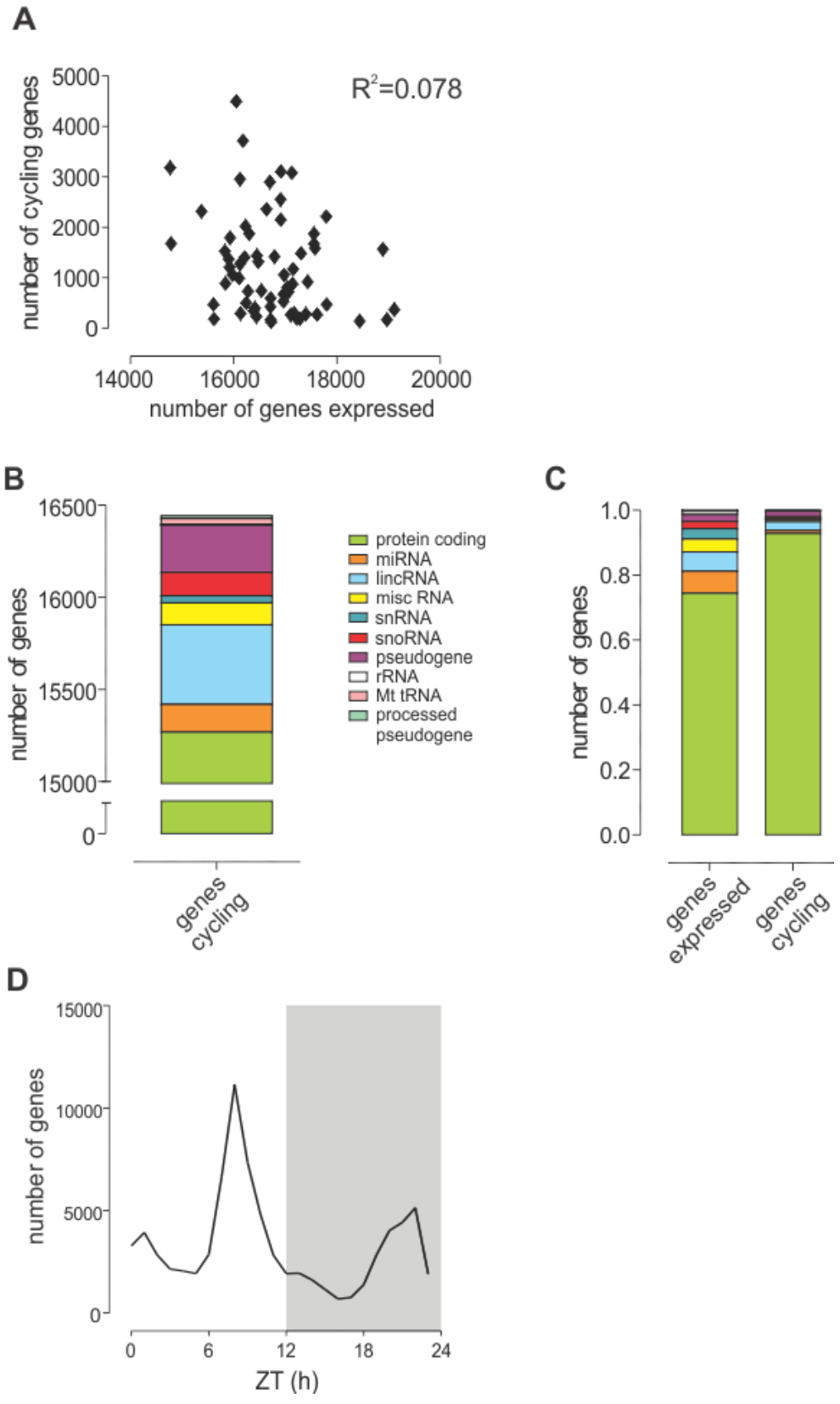
A
number of cycling genes
5000
4000
3000
2000
1000
0
R²=0.078
14000
16000
18000
20000
number of genes expressed
B
number of genes
16500
16000
15500
15000
0
protein coding
miRNA
lincRNA
misc RNA
snRNA
snoRNA
pseudogene
rRNA
Mt tRNA
processed pseudogene
genes cycling
C
number of genes
1.0
0.8
0.6
0.4
0.2
0.0
genes expressed
genes cycling
D
number of genes
15000
10000
5000
0
0
6
12
18
24
ZT (h)

Fig. S3: (**A**) A tissue by tissue comparison showing the lack of correlation of the number of genes expressed and the number of genes cycling in the same tissue. (**B**) Number of genes cycling (meta2d, p value >0.05) in at least one tissue classified by gene type (**C**). Relative distribution of genes found expressed (left) and cycling (right) in at least one of the 64 tissues according to their gene types. (**D**) Distribution across 24h and all the tissues sampled of the total number of genes phase of peak expression.

FigureS4

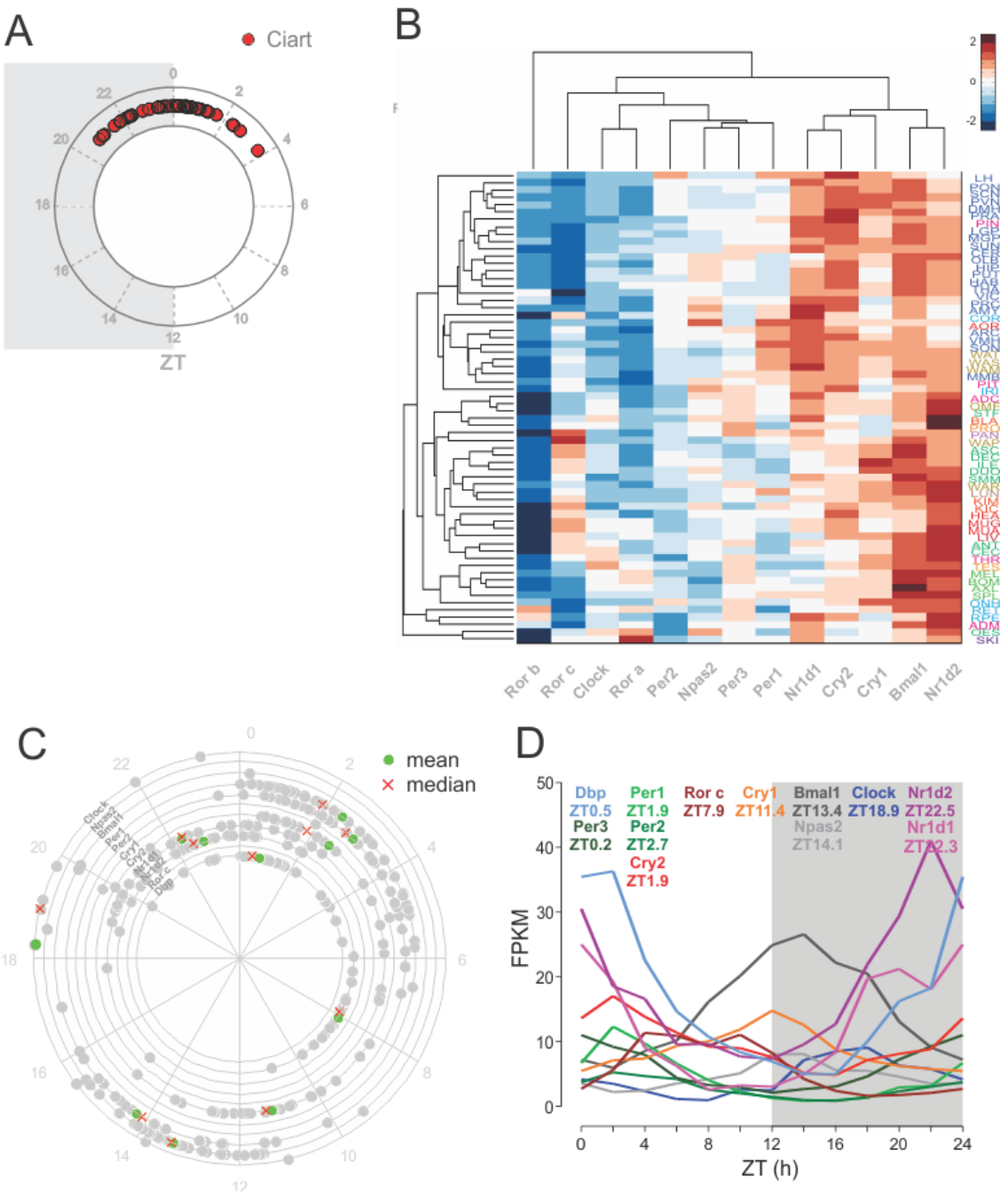
A
Ciart
ZT
B
Ror b
Ror c
Clock
Ror a
Per2
Npas2
Per3
Per1
Nr1d1
Cry2
Cry1
Bmal1
Nr1d2
C
mean
median
D
Dbp ZT0.5
Per3 ZT0.2
Per1 ZT1.9
Per2 ZT2.7
Cry2 ZT1.9
Ror c ZT7.9
Cry1 ZT11.4
Bmal1 ZT13.4
Npas2 ZT14.1
Clock ZT18.9
Nr1d2 ZT22.5
Nr1d1 ZT22.3
FPKM
ZT (h)

Fig. S4: (**A**) Circular distribution of the peak phases of *Ciart*. (**B**) Clustering of 10 clock genes expressed in 64 tissues based on their average level of expression over 24h. (**C**) Average (green circle) and median (red arrow) distributions of the peak phases of core clock components in the tissues where they are detected as cycling (grey circles). (**D**) Distribution of the median peak phases of core clock components.

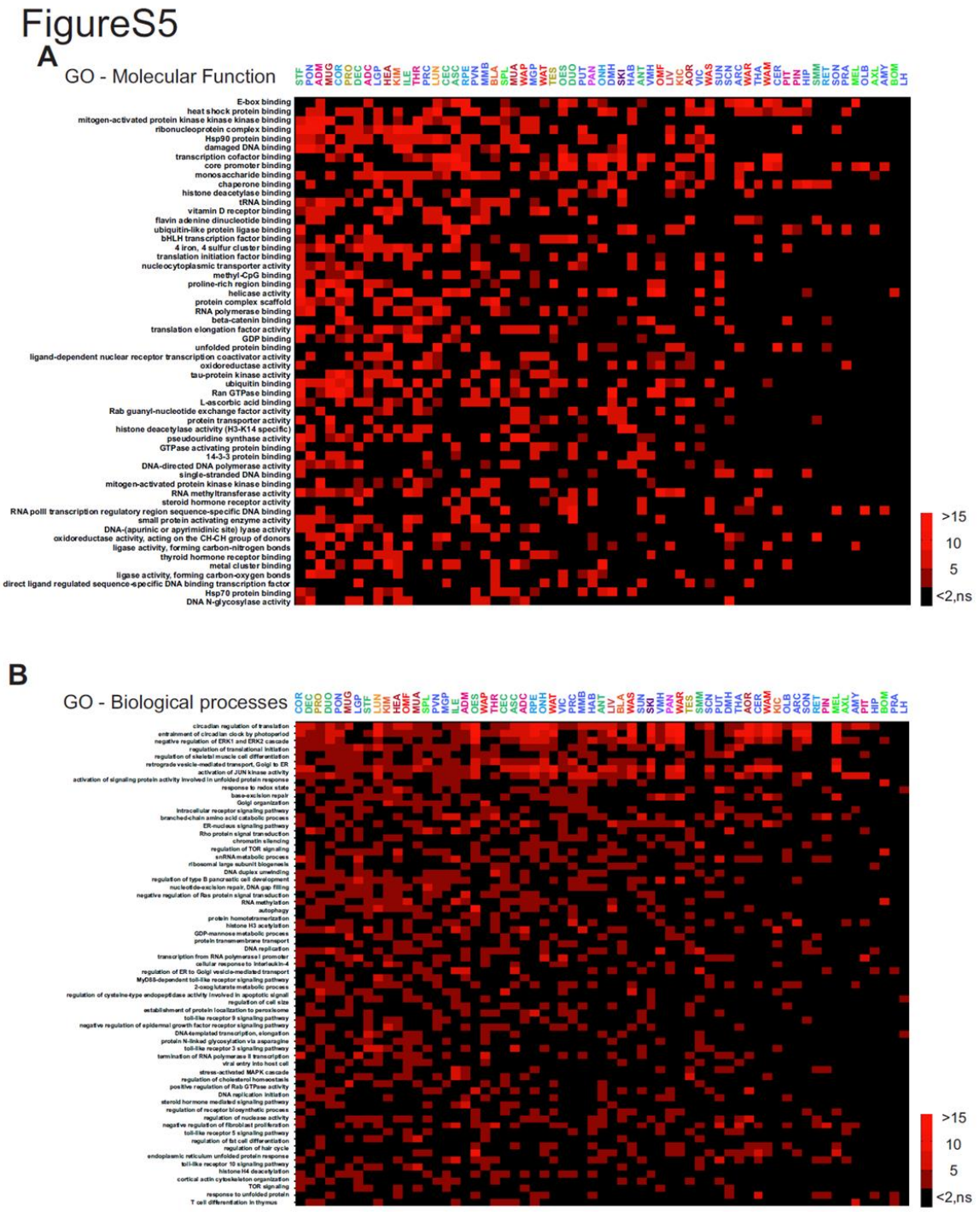

Fig. S5: (**A**) Heatmap of the GO-molecular functions terms enriched for genes cycling in more than 10 tissues. In red are represented the pathways satisfying the criteria of over-representation analysis (ORA, gene number>2, z-score>2 and permuted p<0.05). The color intensity codes for the z-score. (**B**) Heatmap of the GO-biological processes terms enriched for cycling genes in more than 10 tissues.

FigureS6

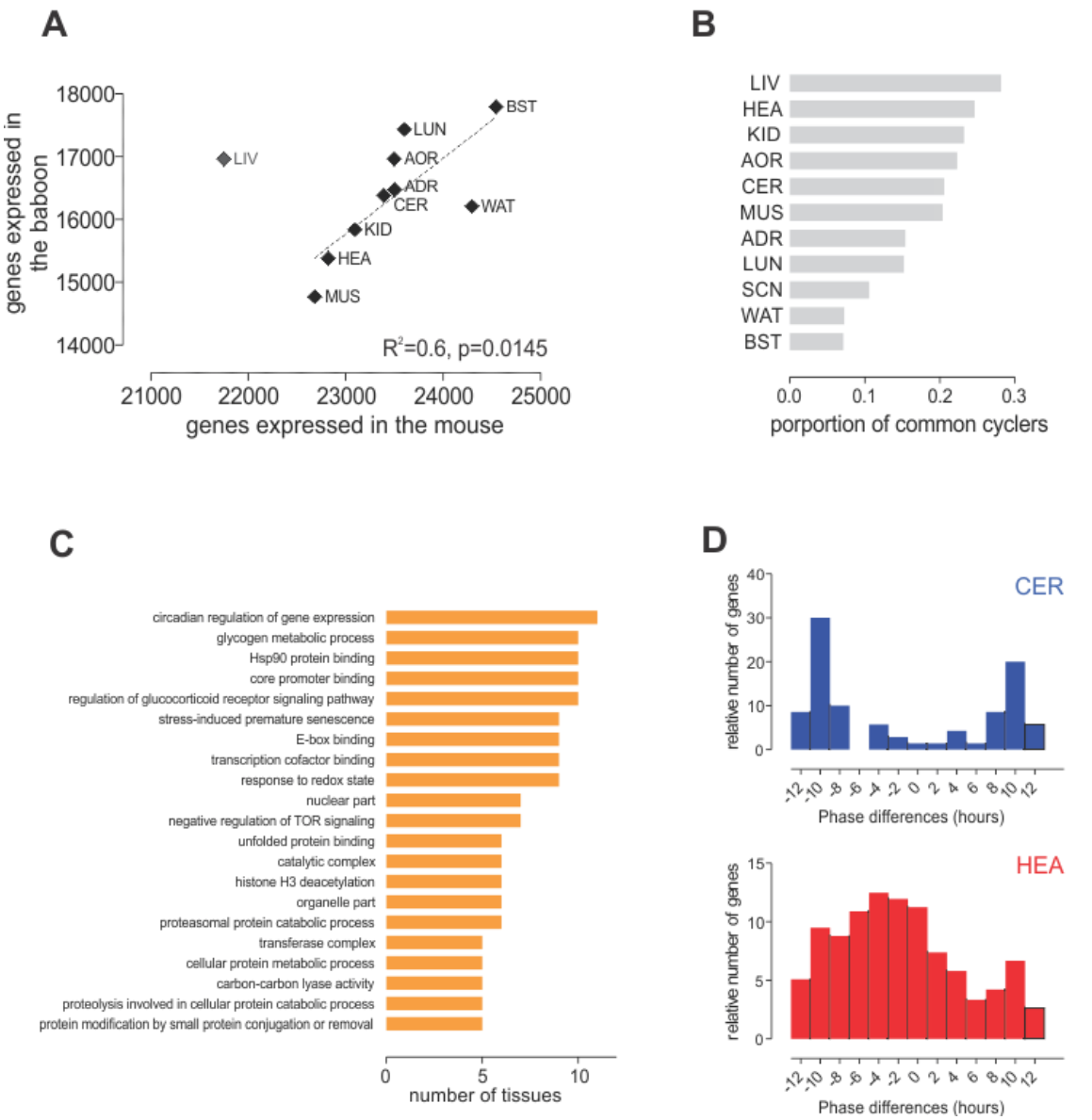

Fig. S6: (**A**) Correlation between the numbers of genes expressed in mouse and baboon tissues. (**B**) Fraction of genes cycling in the baboon also cycling in the mouse. (**C**) histogram of the KEGG pathways enriched for genes found cycling in both mouse and baboon (criteria of over-representation analysis: gene number>2, z-score>2 and permuted p<0.05). (**D**) Phase differences of the genes cycling in both mouse and baboon in cerebellum (CER) and heart (HEA).

FigureS7

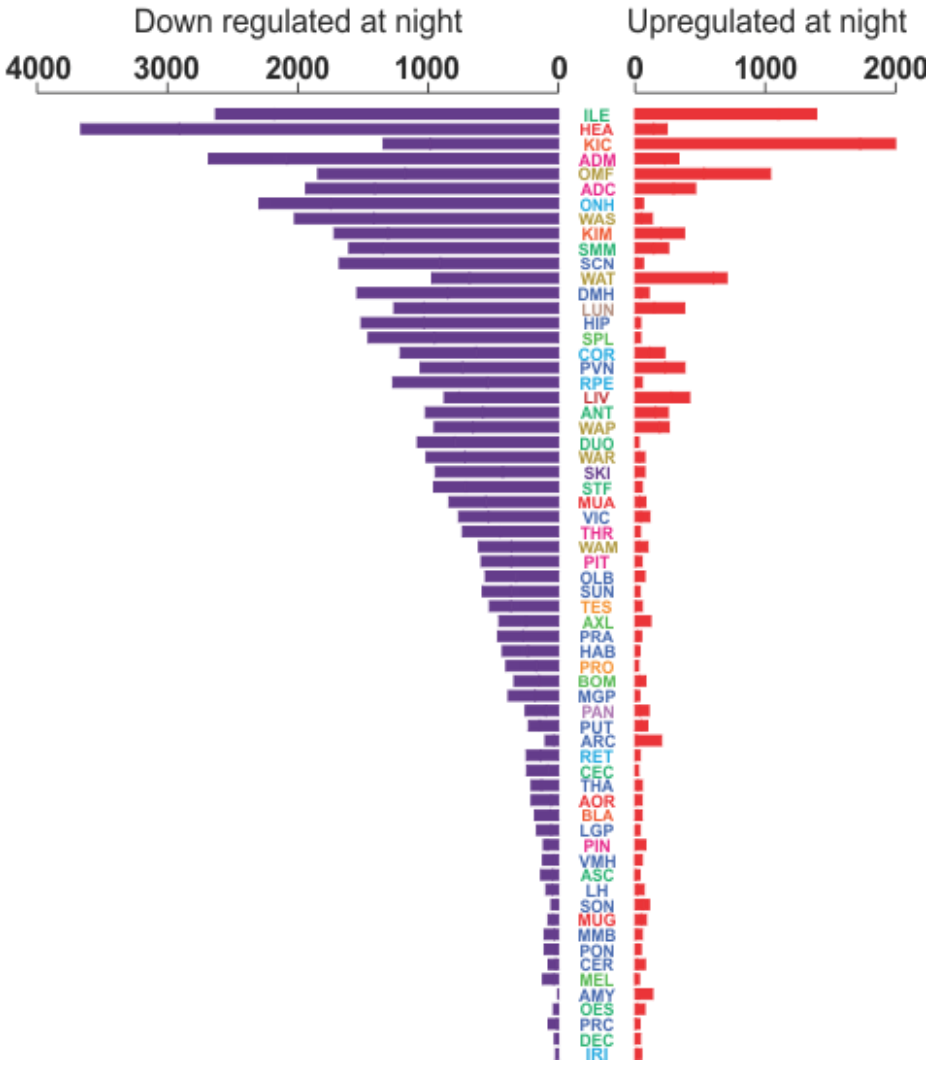

Fig. S7- Up (>2fold) and down regulation (<2fold) of gene expression between the light/activity/feeding phase (ZT0-12) and the dark/sleep/fasting phase (ZT12-24) in different tissues.

Additional Data tables (separate file)

Table S1: Tissues nomenclature and number of expressed genes
Table S2: Mapping Statistics
Table S3: KEGG annotations of UEGs
Table S4: Lists of genes expressed at least in one tissue and in all tissues (UEGs)
Table S5: 100 Top expressed genes by tissue
Table S6: List of cycling genes in each tissue
Table S7: Expressed and cycling genes overlaps tissue by tissue
Table S8: List of all cycling genes
Table S9: List of non-cycling UEGs
Table S10: List of genes cycling in more than 20 tissues
Table S11: Phases of the Core Clock Genes across tissues
Table S12: List of FDA approved drug targets cycling

Additional Databases (separate file)

Database S1: Baboon meta2d statistics
Database S2: Mouse meta2d statistics

References and Notes

1. R. Zhang, N. F. Lahens, H. I. Ballance, M. E. Hughes, J. B. Hogenesch, A circadian gene expression atlas in mammals: Implications for biology and medicine. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 16219–16224 (2014). doi:10.1073/pnas.1408886111 Medline

2. M. F. Ceriani, J. B. Hogenesch, M. Yanovsky, S. Panda, M. Straume, S. A. Kay, Genome-wide expression analysis in Drosophila reveals genes controlling circadian behavior. *J. Neurosci.* **22**, 9305–9319 (2002). Medline

3. S. L. Harmer, J. B. Hogenesch, M. Straume, H.-S. Chang, B. Han, T. Zhu, X. Wang, J. A. Kreps, S. A. Kay, Orchestrated transcription of key pathways in Arabidopsis by the circadian clock. *Science* **290**, 2110–2113 (2000). doi:10.1126/science.290.5499.2110 Medline

4. H. Oster, E. Challet, V. Ott, E. Arvat, E. R. de Kloet, D.-J. Dijk, S. Lightman, A. Vgontzas, E. Van Cauter, The functional and clinical significance of the 24-h rhythm of circulating glucocorticoids. *Endocr. Rev.* **38**, er20151080 (2016). doi:10.1210/er.2015-1080

5. W. J. Schwartz, S. M. Reppert, S. M. Eagan, M. C. Moore-Ede, In vivo metabolic activity of the suprachiasmatic nuclei: A comparative study. *Brain Res.* **274**, 184–187 (1983). doi:10.1016/0006-8993(83)90538-3 Medline

6. C. S. Möller-Levet, S. N. Archer, G. Bucca, E. E. Laing, A. Slak, R. Kabiljo, J. C. Y. Lo, N. Santhi, M. von Schantz, C. P. Smith, D.-J. Dijk, Effects of insufficient sleep on circadian rhythmicity and expression amplitude of the human blood transcriptome. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E1132–E1141 (2013). doi:10.1073/pnas.1217154110 Medline

7. J. Z. Li, B. G. Bunney, F. Meng, M. H. Hagenauer, D. M. Walsh, M. P. Vawter, S. J. Evans, P. V. Choudary, P. Cartagena, J. D. Barchas, A. F. Schatzberg, E. G. Jones, R. M. Myers, S. J. Watson Jr., H. Akil, W. E. Bunney, Circadian patterns of gene expression in the human brain and disruption in major depressive disorder. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 9950–9955 (2013). doi:10.1073/pnas.1305814110 Medline

8. C. Vollmers, R. J. Schmitz, J. Nathanson, G. Yeo, J. R. Ecker, S. Panda, Circadian oscillations of protein-coding and regulatory RNAs in a highly dynamic mammalian liver epigenome. *Cell Metab.* **16**, 833–845 (2012). doi:10.1016/j.cmet.2012.11.004 Medline

9. M. Melé, P. G. Ferreira, F. Reverter, D. S. DeLuca, J. Monlong, M. Sammeth, T. R. Young, J. M. Goldmann, D. D. Pervouchine, T. J. Sullivan, R. Johnson, A. V. Segrè, S. Djebali, A. Niarchou, F. A. Wright, T. Lappalainen, M. Calvo, G. Getz, E. T. Dermitzakis, K. G. Ardlie, R. Guigó, R. Guigo, GTEx Consortium, The human transcriptome across tissues and individuals. *Science* **348**, 660–665 (2015). doi:10.1126/science.aaa0355 Medline

10. L. Pipes, S. Li, M. Bozinoski, R. Palermo, X. Peng, P. Blood, S. Kelly, J. M. Weiss, J. Thierry-Mieg, D. Thierry-Mieg, P. Zumbo, R. Chen, G. P. Schroth, C. E. Mason, M. G. Katze, The non-human primate reference transcriptome resource (NHPRTR) for comparative functional genomics. *Nucleic Acids Res.* **41** (D1), D906–D914 (2013). doi:10.1093/nar/gks1268 Medline

11. GTEx Consortium, The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648–660 (2015). doi:10.1126/science.1262110 Medline

12. G. Wu, R. C. Anafi, M. E. Hughes, K. Kornacker, J. B. Hogenesch, MetaCycle: An integrated R package to evaluate periodicity in large scale data. *Bioinformatics* **32**, 3351–3353 (2016). doi:10.1093/bioinformatics/btw405 Medline

13. H. R. Ueda, W. Chen, A. Adachi, H. Wakamatsu, S. Hayashi, T. Takasugi, M. Nagano, K. Nakahama, Y. Suzuki, S. Sugano, M. Iino, Y. Shigeyoshi, S. Hashimoto, A transcription factor response element for gene expression during circadian night. *Nature* **418**, 534–539 (2002). doi:10.1038/nature00906 Medline

14. K. F. Storch, O. Lipan, I. Leykin, N. Viswanathan, F. C. Davis, W. H. Wong, C. J. Weitz, Extensive and divergent circadian gene expression in liver and heart. *Nature* **417**, 78–83 (2002). doi:10.1038/nature744 Medline

15. G. Z. Wang, S. L. Hickey, L. Shi, H.-C. Huang, P. Nakashe, N. Koike, B. P. Tu, J. S. Takahashi, G. Konopka, Cycling Transcriptional Networks Optimize Energy Utilization on a Genome Scale. *Cell Reports* **13**, 1868–1880 (2015). doi:10.1016/j.celrep.2015.10.043 Medline

16. A. C. Zambon, S. Gaj, I. Ho, K. Hanspers, K. Vranizan, C. T. Evelo, B. R. Conklin, A. R. Pico, N. Salomonis, GO-Elite: A flexible solution for pathway and ontology over-representation. *Bioinformatics* **28**, 2209–2210 (2012). doi:10.1093/bioinformatics/bts366 Medline

17. R. Zhang, A. A. Podtelezhnikov, J. B. Hogenesch, R. C. Anafi, Discovering Biology in Periodic Data through Phase Set Enrichment Analysis (PSEA). *J. Biol. Rhythms* **31**, 244–257 (2016). doi:10.1177/0748730416631895 Medline

18. M. Hatori, S. Gill, L. S. Mure, M. Goulding, D. D. M. O'Leary, S. Panda, Lhx1 maintains synchrony among circadian oscillator neurons of the SCN. *eLife* **3**, e03357 (2014). doi:10.7554/eLife.03357 Medline

19. M. Uhlén, L. Fagerberg, B. M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson, C. Kampf, E. Sjöstedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C. A.-K. Szigyarto, J. Odeberg, D. Djureinovic, J. O. Takanen, S. Hober, T. Alm, P.-H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, F. Pontén, Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015). doi:10.1126/science.1260419 Medline

20. A. T. Davenport, K. A. Grant, K. T. Szeliga, D. P. Friedman, J. B. Daunais, Standardized method for the harvest of nonhuman primate tissue optimized for multiple modes of analyses. *Cell Tissue Bank.* **15**, 99–110 (2014). doi:10.1007/s10561-013-9380-2 Medline

21. J. Wu, J. Robertson, (Elsevier Science Amsterdam, the Netherlands, 2000).

22. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013). doi:10.1093/bioinformatics/bts635 Medline

23. Y. Liao, G. K. Smyth, W. Shi, featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014). doi:10.1093/bioinformatics/btt656 Medline

24. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014). doi:10.1186/s13059-014-0550-8 Medline

25. M. E. Hughes, L. DiTacchio, K. R. Hayes, C. Vollmers, S. Pulivarthy, J. E. Baggs, S. Panda, J. B. Hogenesch, Harmonics of circadian gene transcription in mammals. *PLOS Genet.* **5**, e1000442 (2009). doi:10.1371/journal.pgen.1000442 Medline