Cell Metabolism, Volume *25*

Supplemental Information

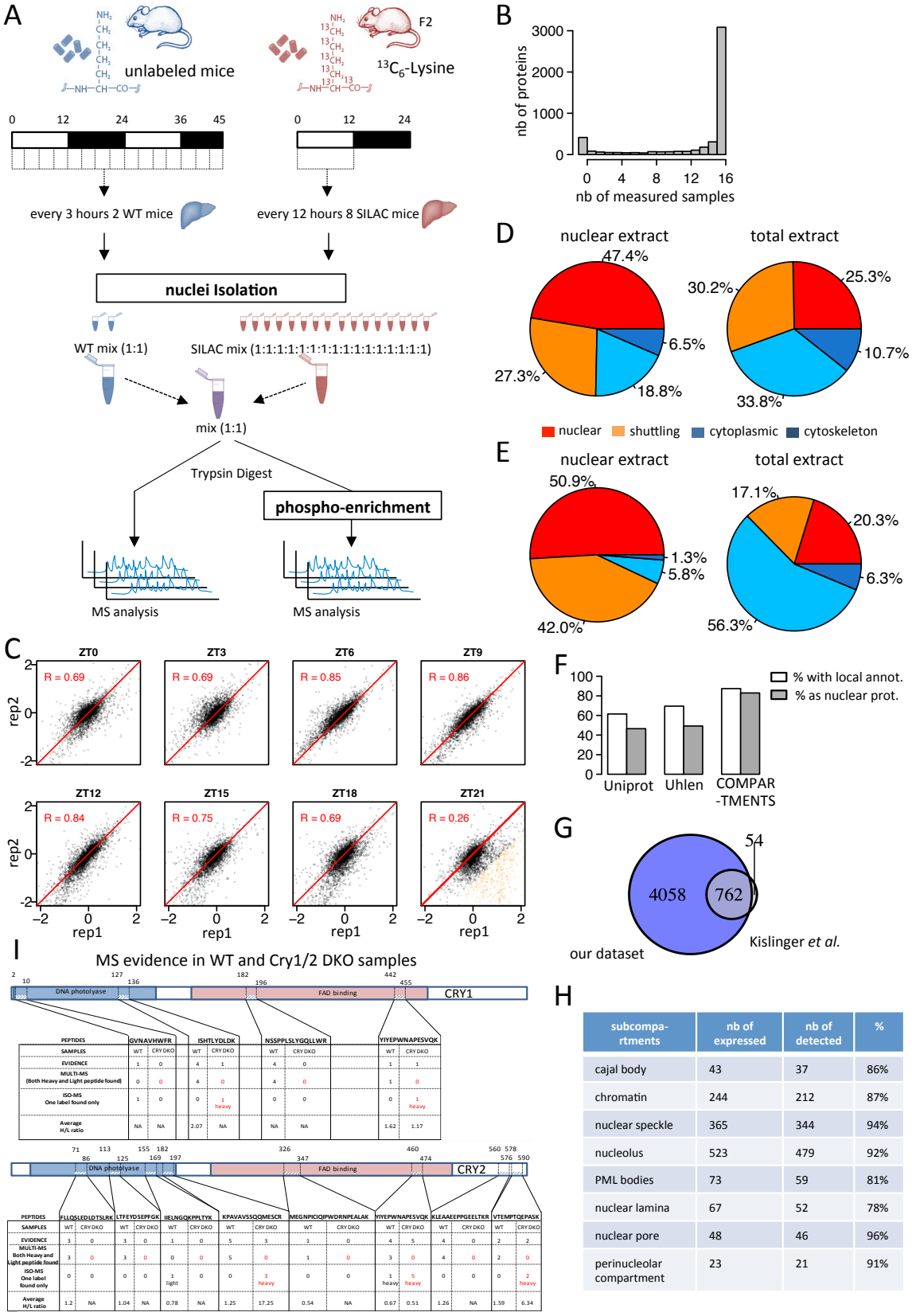
Nuclear Proteomics Uncovers Diurnal Regulatory Landscapes in Mouse Liver

Jingkui Wang, Daniel Mauvoisin, Eva Martin, Florian Atger, Antonio Núñez Galindo, Loïc Dayon, Federico Sizzano, Alessio Palini, Martin Kussmann, Patrice Waridel, Manfredo Quadroni, Vjekoslav Dulić, Felix Naef, and Frédéric Gachon

Supplemental Figures and Table Legends

Supplemental Figures Legends

Figure S1. Experimental design and quality control for SILAC-based quantitative nuclear proteomics, related to Fig 1

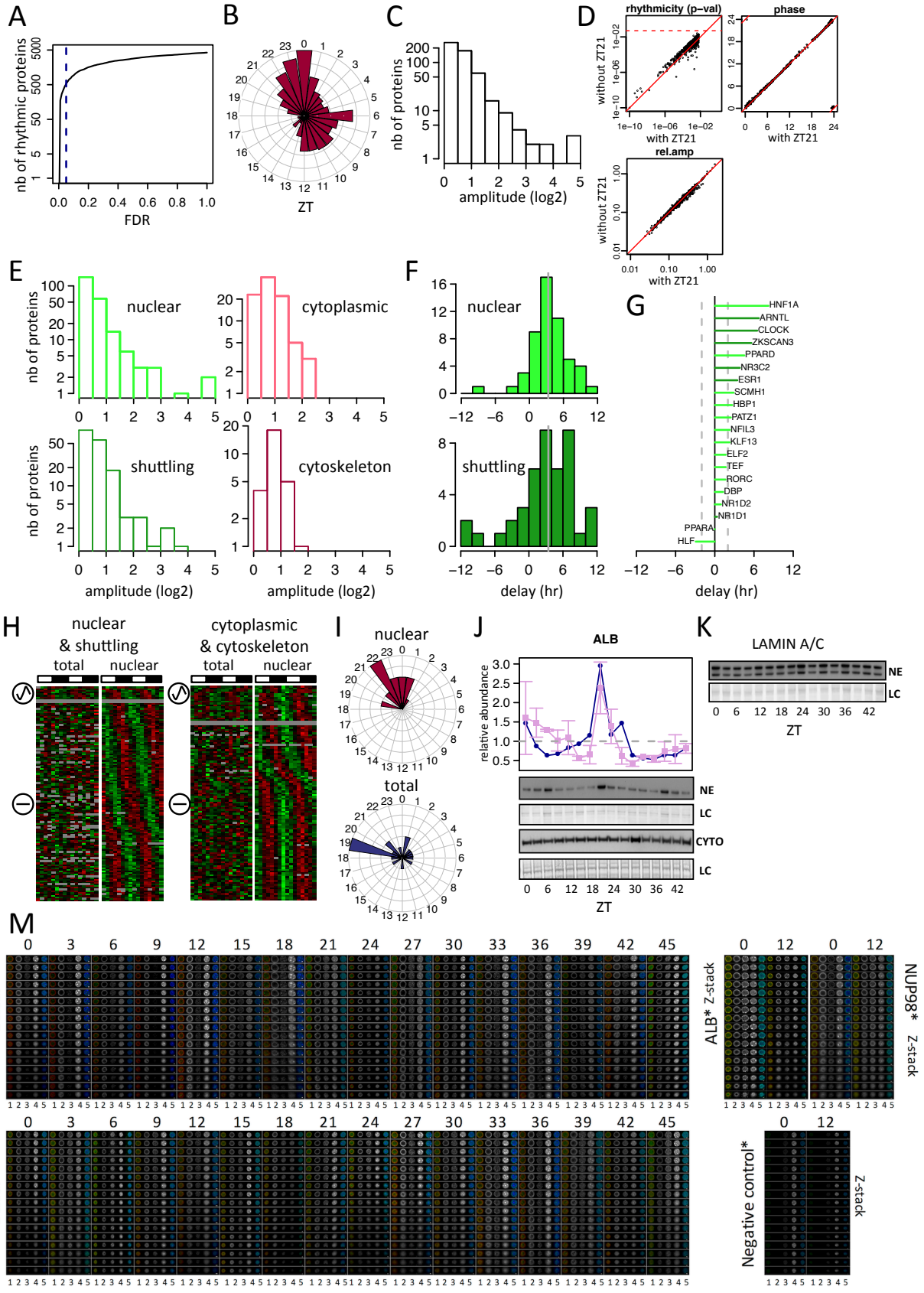


H

subcompa-rtments	nb of expressed	nb of detected	%
cajal body	43	37	86%
chromatin	244	212	87%
nuclear speckle	365	344	94%
nucleolus	523	479	92%
PML bodies	73	59	81%
nuclear lamina	67	52	78%
nuclear pore	48	46	96%
perinucleolar compartment	23	21	91%

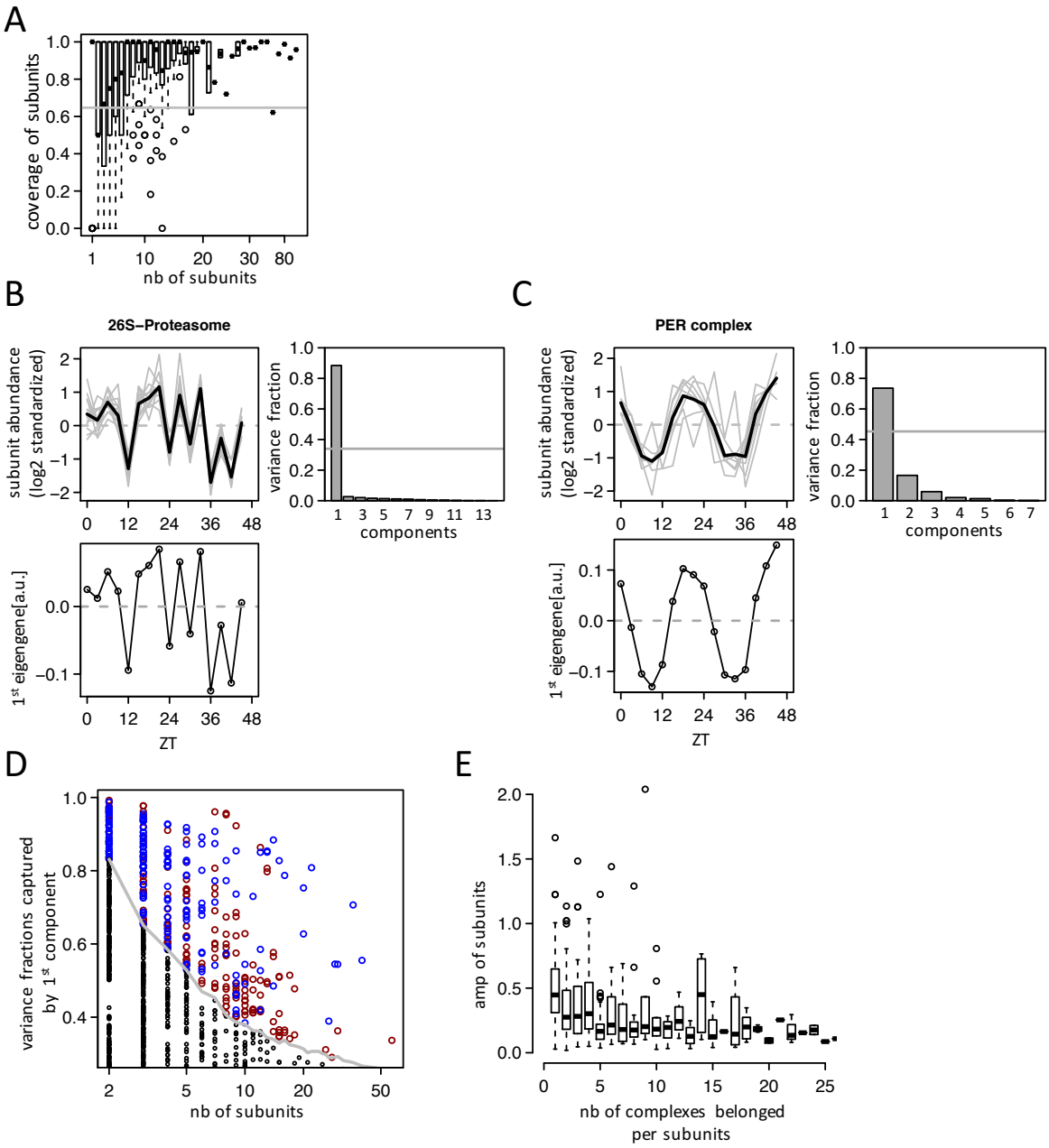
A. Workflow of the SILAC-based MS analysis for proteins and phosphoproteins extracted from purified mouse liver nuclei. Equal amounts of nuclear proteins extracted from two non-SILAC mice were pooled every 3 h during two consecutive days (WT mix). A reference SILAC nuclear protein mix was prepared from 8 SILAC mice collected at ZT0 and 8 collected at ZT12 (SILAC mix). Then, 16 mixes were obtained by adding the same amount of SILAC mix to the WT mix. As a result, the 16 mixes samples, 8 for day 1 and 8 for day 2 (or n=2 biological replicates at each of eight time points), were used for nuclear proteomics and nuclear phosphoproteomics.
B. Distribution of sample numbers quantified for 4820 detected nuclear proteins.
C. Pearson correlation analysis of biological replicates at the eight time points. All values are log2 ratios to the common reference samples. The replicates are overall well correlated (70 % average Pearson correlation), and showing similar spread, indicating that the quality of the nuclear purifications was fairly homogenous. Only one time point (ZT21) showed a lower correlation. We could trace this back to a lower H/L ratio for 327 proteins in sample ZT21-Rep1 (orange dots). These proteins are mostly coming from peroxisomes, potentially due to a higher degree of carry-over from the cytoplasmic fraction in this sample.
D. Relative abundance of identified proteins annotated as nuclear, shuttling between cytoplasm and nucleus, cytoplasmic, and cytoskeleton (Uniprot) for nuclear extract and previously published total extract (Mauvoisin et al., 2014).
E. Same distribution as in C but for raw intensity signals quantified by MS.
F. Fraction of 4820 identified proteins with known localization and annotated as nuclear in Uniprot (The UniProt, 2015), Uhlèn et al. (Uhlén et al., 2015), and COMPARTMENTS (Binder et al., 2014) datasets.
G. Comparison of mouse liver nuclear proteome identified in our dataset and the previous study (Kislinger et al., 2006) in which 762 out of 816 characterized nuclear proteins were overlapped by our analysis. This comparison is done using corresponding gene names.
H. Detection coverage of liver-expressed proteins in nuclear sub-compartments.
I. MS-evidence in protein mixes prepared with SILAC/WT and SILAC/*Cry1/2* DKO nuclear protein extracts shows CRY1 and CRY2 proteins identification only originates from the heavy part (WT protein extract) of the SILAC/*Cry1/2* DKO protein mixes (ISO-MS evidence only). The table shows sums of MS evidence for ZT0, 6, 12 and 18.

Figure S2. Diurnal accumulation of nuclear proteins and confirmation of nuclear localization for individual examples by confocal microscopy, related to Fig 2



A. Dependency of the number of rhythmic nuclear proteins in the function of false discovery rate (FDR). With a value of FDR < 0.05, 522 rhythmic proteins are identified (blue dashed line).
B, C. Phase and peak-to-trough amplitude (log2) distributions for rhythmic nuclear proteins (FDR < 0.05).
D. The potential contamination of the ZT21 sample 1 only had a minor influence on the global rhythmic analysis. Indeed, the identical rhythmicity analysis without this sample presented very little differences on the identity of the rhythmic proteins as well as negligible differences on the estimated phases and amplitudes.
E. Peak-to-trough amplitude distributions of rhythmic proteins in C1 (nuclear and shuttling) and C2 (cytoplasmic and cytoskeleton groups).
F, G. Distribution of phase delay between nuclear protein and mRNA accumulation peaks for C1 (nuclear and shuttling groups). Specific delays are shown in (F) for TFs, which are enriched in the rhythmic nuclear proteins with corresponding rhythmic mRNAs ($p < 10^{-5}$).
H. Heat maps comparing proteins with rhythmic accumulations in nuclear extracts (FDR<0.05) with corresponding proteins in total extracts (TE) (Mauvoisin et al., 2014) for both classes C1 (left panel) and C2 groups (right panel). Sinusoidal sign means rhythmic protein in TE, whereas straight line means constant protein in TE.
I. Phase distributions of C2 proteins in nuclear extracts (upper panel) compared with corresponding rhythmic proteins in total extracts (bottom panel).
J. Example of rhythmic nuclear proteins in C2 (ALB) confirmed by Western blot (WB) analysis in nuclear extracts (NE; upper blots) and cytoplasmic extracts (CYTO; lower blots). The two biological replicates are shown as ZT0-ZT21 (replica 1) and ZT24-ZT45 (replica 2). Quantifications of nuclear extracts by MS (blue lines) and WB (pink lines) are shown in the upper graphs. Data are normalized to the temporal mean and the sixteen time points show the mean and standard error (SEM) from two independent biological samples. Naphtol blue black staining of the membranes was used as a loading control (LC) and serves as references for normalisation of the quantified values.
K. LAMINA/C display flat nuclear accumulation measured by WB.
L. Nuclear detection of rhythmic proteins in C2 (FASN, ACACA and ALB) using confocal microscopy. Secondary antibodies alone were used as negative control and the nuclear pore complex NUP98 as positive control. Nuclei were stained and confocal z-stack images were acquired at 0.20 µm intervals. Each image from top to bottom represents a z-stack from the center to the periphery of a nucleus. Horizontal white scale bars represent 1 µm.

Figure S3. Identification of rhythmic nuclear protein complexes by singular value decomposition (SVD), related to Fig 3



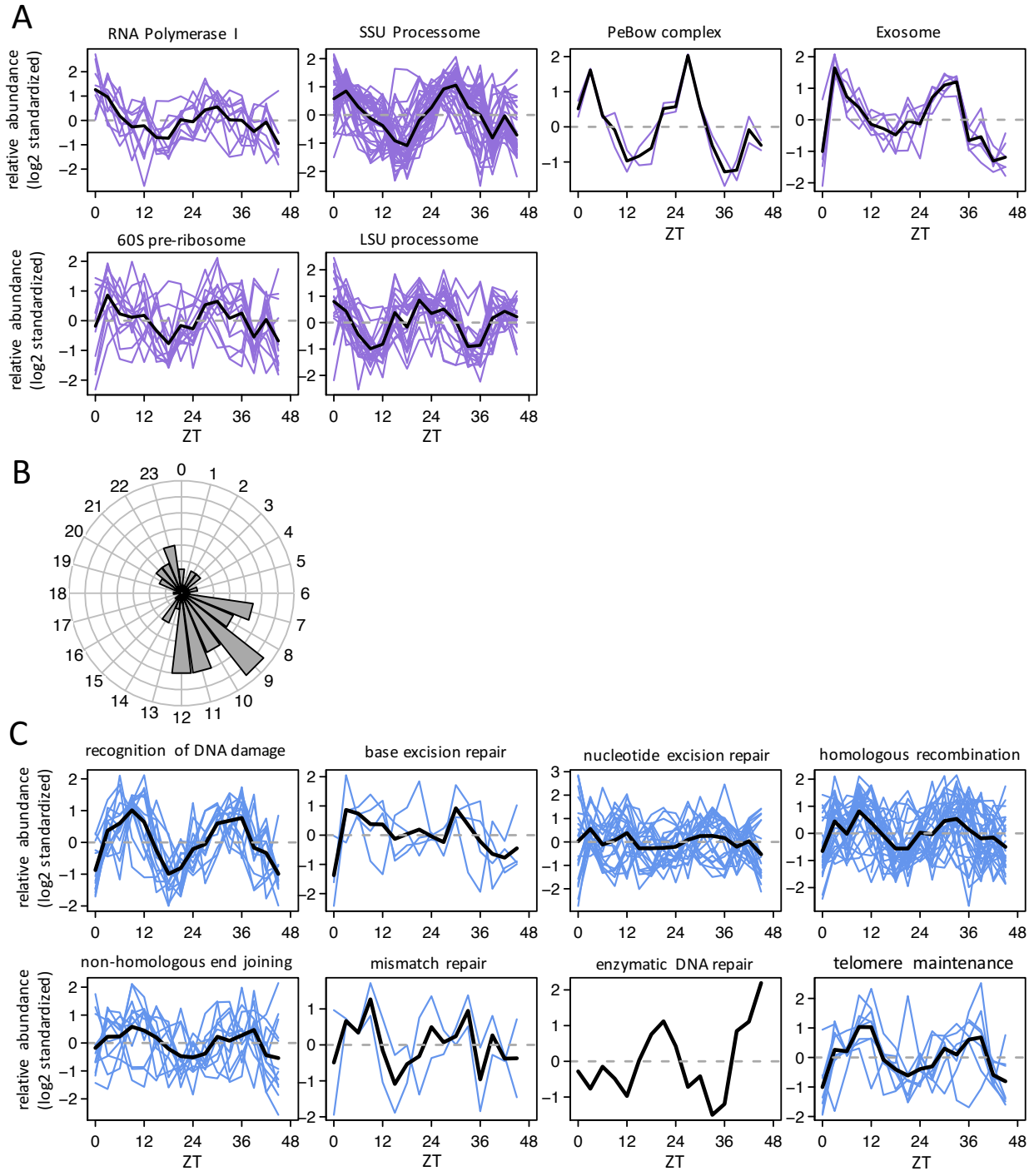
A. Fraction of subunits detected (coverage) in the function of the number of subunits among annotated nuclear protein complexes. Gray line indicates the coverage of proteins not annotated in complexes.

B, C. Temporal accumulation profiles and SVD analysis for two examples of nuclear protein complexes: 26S-protasome (B) and PER complex (C). SVD analysis is applied to the matrix, in which rows represent standardized temporal profiles of all detected subunits within the complex across time points (upper left panel), and the fraction of variance captured by the first component of SVD is tested to be significantly different from the background (upper right panel). The background distribution is calculated by randomly choosing detected nuclear proteins not annotated in the database of complexes and generating complexes of the same number of subunits. The solid gray lines represent the threshold p-value of 0.05. When the first component capture a significant portion of variance, the rhythmicity and phase is assessed on the first eigengene (bottom left panels).

D. Variance fractions captured by the 1st component of SVD for all detected protein complexes. The solid gray line is the threshold of background with p-value = 0.05. Rhythmic nuclear protein complexes are indicated by red cycles and non-rhythmic ones by blue cycles.

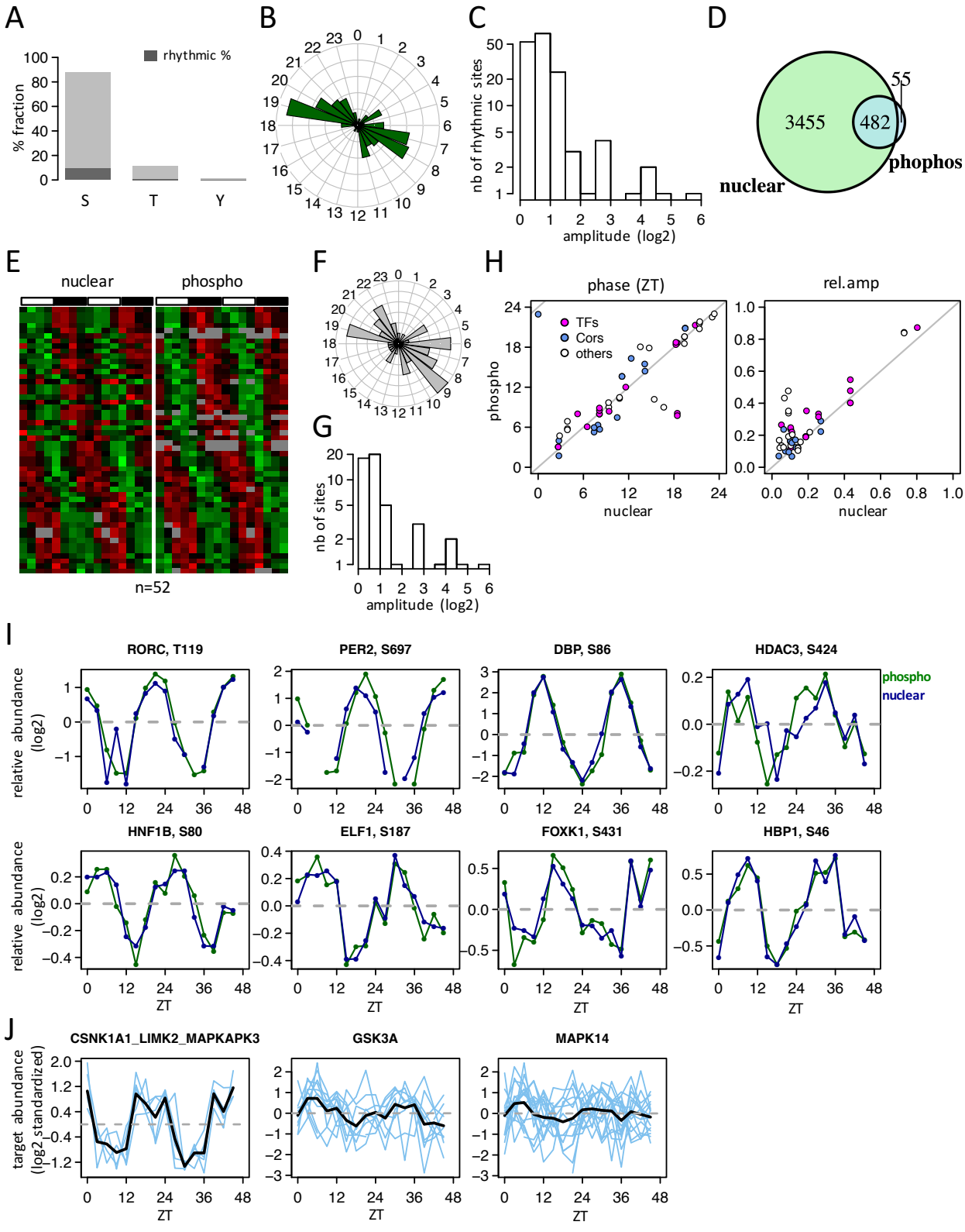
E. Peak-to-trough amplitudes (log2) of subunits show a damping trend with number of complexes they belong.

Figure S4. Diurnal rhythms of nuclear proteins involved in ribosome biogenesis and DNA repair, related to Fig 4



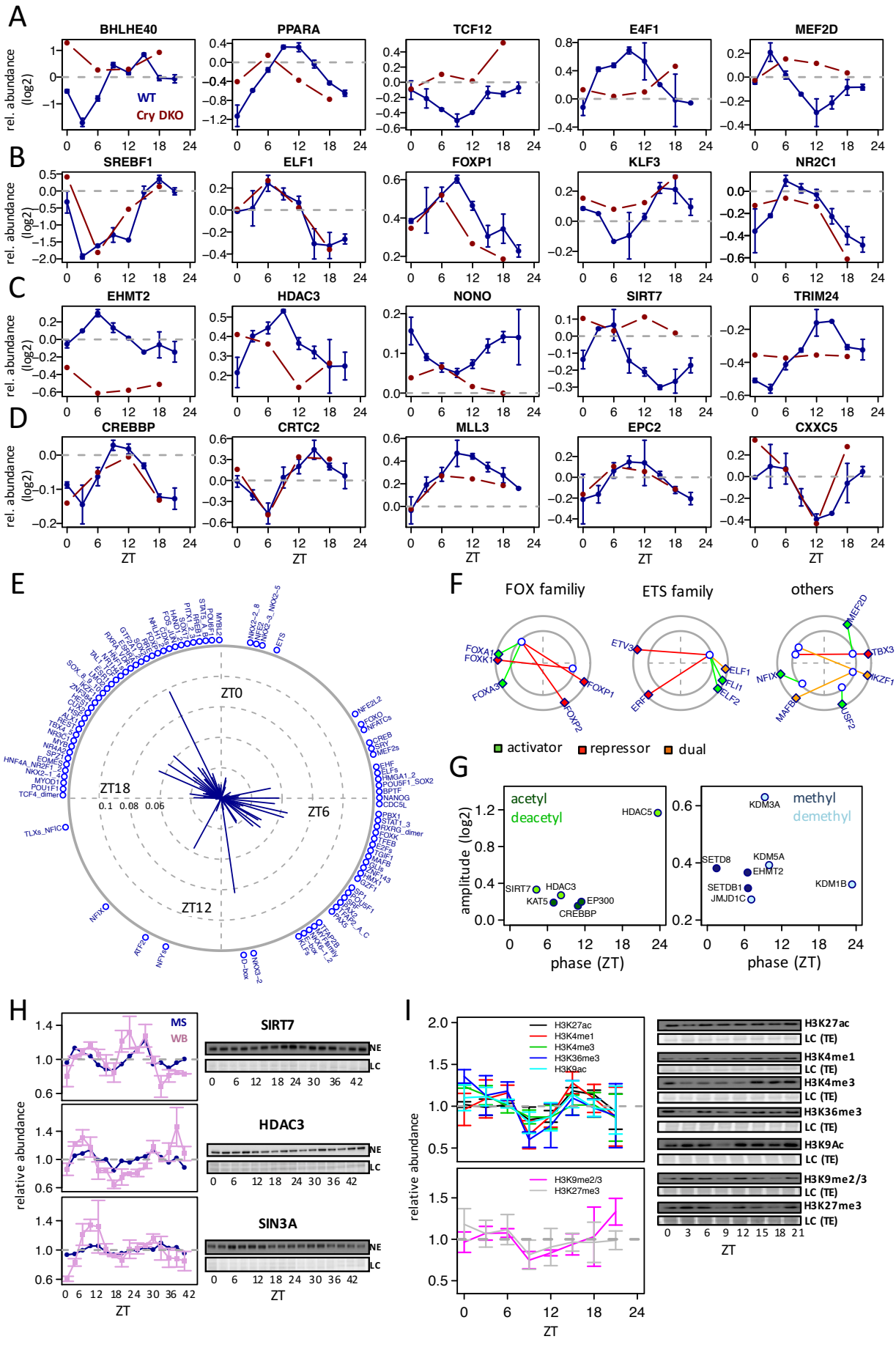
A. Rhythmic accumulation profiles of nuclear proteins involved in ribosome biogenesis, namely RNA polymerase I, SSU processome, PeBow complex, Exosome, 60 pre-ribosome, and LSU processome.
B. Peak time distribution of rhythmic nuclear proteins involved in DNA repair processes.
C. Temporal profiles of rhythmic nuclear proteins involved in recognition of DNA damage, base excision repair, nucleotide excision repair, homologous recombination, non-homologous end joining, mismatch repair, enzymatic DNA repair, and telomere maintenance.

Figure S5. Nuclear phosphoproteome and its relationship with the nuclear proteome, related to Fig 5



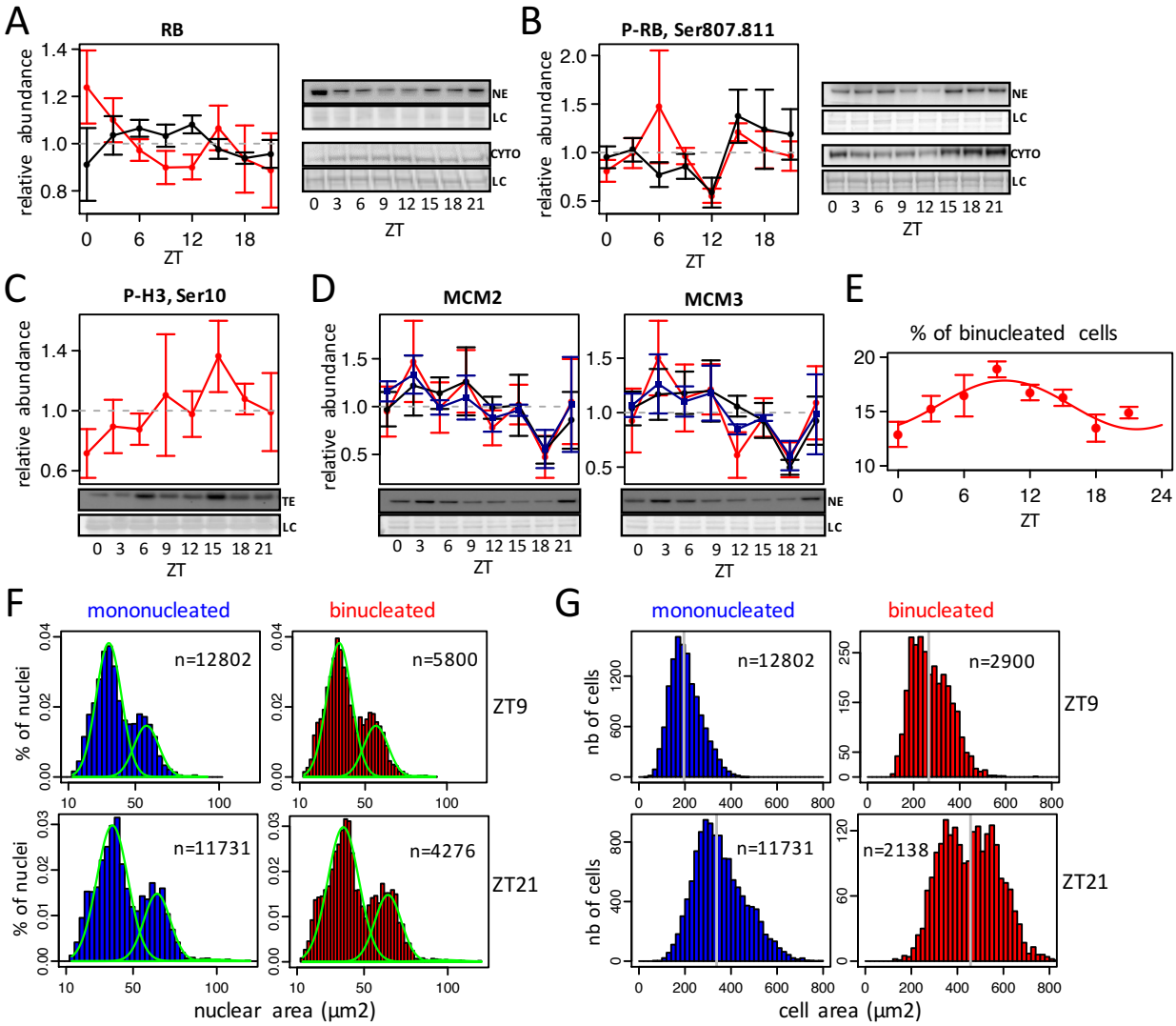
A. Proportions of nuclear phosphorylation sites, Serine (S), Threonine (T) and Tyrosine (Y) quantified in at least 8 out of 16 samples, as well as fractions of sites showing diurnal accumulations (FDR < 0.05).
B, C. Phase and peak-to-trough amplitude distributions for all rhythmic nuclear phosphosites.
D. 482 proteins (corresponding to 859 phosphosites) of 537 proteins (corresponding to 980 phosphosites) with detected phosphorylation sites have corresponding proteins in the nuclear protein dataset.
E. Heat maps representation of rhythmic phosphorylated sites (n = 52, right panel) with rhythmic nuclear proteins (left panel).
F, G. Peak time and peak-to-trough amplitude distributions for the rhythmic nuclear phosphoproteins with rhythmic nuclear proteins in (E).
H. Peak time and relative amplitude comparison between the nuclear phosphoproteins and corresponding nuclear proteins from (E).
I. Temporal accumulation of individual examples from (E) (Blue line: nuclear proteins; Green line: nuclear phosphoproteins). Data are normalized to the temporal mean.
J. Temporal accumulation of individual phosphorylated sites targeted by kinases CSK1A1, GSK3A, and MAPK14.

Figure S6. Analysis of transcriptional and epigenetic regulation related to detected rhythmic TFs and co-regulators, related to Fig 6



A-D. Individual examples of rhythmic TFs (A, B) or co-regulators (C, D) losing (A and C) or retaining (B and D) their rhythms in *Cry1/2* DKO mice.
E. Rhythmic motif activities inferred with DHS and RNA-seq data (pre-mRNA). Direction of lines indicate peak activity times and distances to the center of solid lines indicate activity amplitudes per motif.
F. Peak times of nuclear TF accumulations (outer circle) compared with peak activity times of corresponding DNA binding motif (inner circle) for TF of the FOX and ETS families, and other TFs. Color code indicated the commonly accepted transcriptional function of each TF, namely activator (green), repressor (red) or dual regulation (orange).
G. Amplitude (log2) and phases of rhythmic (FDR < 0.05) transcriptional co-regulators involved in histone acetylation (HATs and HDACs) and methylation (HMTs and HDMs).
H. Nuclear accumulations of rhythmic regulators, SIRT7, HDAC3 and SIN3A, confirmed by Western Blot analysis on nuclear protein extracts (NE). Data are normalized to the temporal mean. The two biological replicates are shown as ZT0-ZT21 (replica 1) and ZT24-ZT45 (replica 2). For the Western blots, the sixteen time points show the mean and standard error (SEM) from two independent biological samples. Naphtol blue black staining of the membranes was used as a loading control (LC) and serves as references for normalisation of the quantified values.
I. Active (H3K27ac, H3K4me1, H3K4me3, H3K36me3 and H3K9ac) and repressive (H3K9me2/3 and H3K27me3) histone modifications are measured and quantified by WB analysis in total protein extracts (TE). Data are normalized to the temporal mean and the eight time points show the mean and standard error (SEM) from four independent biological samples.

Figure S7. Additional evidence for cell-cycle regulation and analysis of IHC images, related to Fig 7



For A to D, data are normalized to the temporal mean and each time point shows the mean and standard error (SEM) from four independent biological samples. Naphtol blue black staining of the membranes was used as a loading control (LC) and also as references for normalisation of the quantified values.

A-D. Temporal accumulations for RB protein (A) and its phosphosites P-RB Serine 807 and 811 (B) measured by WB on both nuclear (NE) and cytoplasmic (CYTO) protein extracts. (C) P-H3 Serine 10 measured by WB using total protein extract (TE). (D) MCM2 and MCM3, two subunit of MCM complex, measured by WB analysis on NE.

E. Diurnal proportion of bi-nucleated hepatocytes extracted from IHC images analysis. Each time point shows the mean and standard error (SEM) of four independent biological samples.

F. Nuclear area distribution for nuclei within mono and bi-nucleated cells at ZT9 and ZT21. Best fits to these distributions using the mixture of Gaussian distribution are also shown. Note that the sizes of the nuclei at the observed peaks exactly match earlier measurements (Danielsen et al., 1986), thus clearly identifying the peak near 35 μm^2 with 2N and that near 54 μm^2 with 4N nuclei.

G. Cell area distribution for mono- and bi-nucleated hepatocytes at ZT9 and ZT21.

Supplemental Tables

Table S1. Complete dataset of the nuclear proteome in mouse liver identified by SILAC-base MS, related to Fig 1

Sheet1: Relative abundance (normalized L/H ratios from the MaxQuant analysis) for all distinct 4820 proteins identified by SILAC MS in the nucleus of mouse liver at different time points and in all replicates under Light/Dark conditions.
Sheet2: Raw data with peptides counts from MaxQuant analysis at different time points in all replicates.
Sheet3: Description of the column headers for Sheet1
Sheet4: Description of the column headers for Sheet2

Table S2. List of 522 rhythmic nuclear proteins (FDR<0.05), related to Fig 2

Sheet 1. Biochemical functions and cellular localization of these rhythmic nuclear proteins
Sheet 2. Relative abundance of rhythmic nuclear proteins and their corresponding phases and amplitudes in WT mice and their correlation in *Cry1/2* DKO
Sheet 3. Corresponding mRNA expression measured by total RNA-seq
Sheet 4. Corresponding previously published proteomic data in whole liver extracts extracts (Mauvoisin et al., 2014)
Sheet 5. Description of the column headers for Sheets1-4

Table S3. List of rhythmic nuclear complexes with SVD analysis, related to Fig 3

Sheet 1. Phase, amplitudes and related functions of the identified rhythmic complexes
Sheet 2. All known complexes with at least two detected subunits in nuclear proteome and corresponding SVD analysis
Sheet 3. Description of the column headers for Sheets1-2

Table S4. Rhythmic nuclear proteins involved in ribosome biogenesis and DNA repair, related to Fig 4

Sheet 1. Rhythmic nuclear proteins with identified phases and amplitudes involved in different steps of ribosome biogenesis
Sheet 2. Rhythmic nuclear proteins with identified phases and amplitudes involved in different steps of DNA repair process
Sheet 3. Description of the column headers for Sheets1-2

Table S5. Complete dataset of phosphorylated sites of nuclear proteins in mouse liver, related to Fig 5

Sheet 1. Complete diurnal phosphoproteome dataset and corresponding rhythmicity analysis
Sheet 2. Kinase motifs with rhythmic activities inferred using linear model with regularization
Sheet 3. Description of the column headers for Sheets1-2

Table S6. Rhythmic transcription factors (TF), co-regulators, and TF motifs activity inference, related to Fig 6

Sheet 1. Rhythmic TF quantified in nuclear proteome and/or nuclear phosphoproteome
Sheet 2. Rhythmic co-regulators quantified in nuclear proteome
Sheet 3. TF motifs with rhythmic activities inferred from DNase hypersensitive sites (DHS) and pre-mRNAs datasets
Sheet 4. Comparison between rhythmic TF and their corresponding motif activity
Sheet 5. Description of the column headers for Sheets1-4

Table S7. Antibodies used in the different experiments, related to Experimental Procedures

Supplemental Experimental Procedures

Preparation of whole cell protein extracts

Livers were homogenized in a lysis buffer containing 8M Urea and protease inhibitors (cOmplete ULTRA®, Roche), a phosphatase inhibitor cocktail (PhosphoSTOP®, Roche), and deacetylase inhibitors (AGK7, salermide, and trichostatin A, all from SantaCruz® Biochemicals). After 20 min incubation on ice, extracts were centrifuged 10 min at 20,000 g. The resulting supernatants constitute total protein extracts (TE).

Preparation of nuclei, and cytoplasmic and nuclear protein extracts

For nuclei purification, livers were homogenized in sucrose homogenization buffer containing 2.2 M sucrose, 15 mM KCl, 2 mM EDTA, 10 mM HEPES (pH 7.6), 0.15 mM spermin, 0.5 mM spermidin, 1 mM DTT, and protease inhibitors (0.5 mM PMSF, 10 µg/ml Aprotinin, 0.7 µg/ml Pepstatin A, and 0.7 µg/ml Leupeptin). Lysates were deposited on a sucrose cushion containing 2.05 M sucrose, 10% glycerol, 15 mM KCl, 2 mM EDTA, 10 mM HEPES (pH 7.6), 0.15 mM spermin, 0.5 mM spermidin, 1 mM DTT, and protease inhibitors. After 45 min of centrifugation at 105,000 g at 4°C, the resulting supernatants were harvested and constitute the cytoplasmic extracts (CYTO).

On the other hand, the nucleus pellets were suspended in a nucleus buffer composed of 10 mM HEPES (pH 7.6), 100 mM KCl, 0.1 mM EDTA, 10% Glycerol, 0.15 mM spermine, 0.5 mM spermidine, 0.1 mM NaF, 0.1 mM sodium orthovanadate, 0.1 mM ZnSO4, 1 mM DTT, and protease inhibitors (0.5 mM PMSF, 10 µg/ml Aprotinin, 0.7 µg/ml Pepstatin A, and 0.7 µg/ml Leupeptin). This constituted the purified nuclei fractions used for Immunofluorescence and FACS analysis.

Nuclear protein extracts were obtained by adding an equal volume of NUN buffer (2 M urea, 600 mM NaCl, 50 mM HEPES (pH 7.6), 1 mM DTT, protease inhibitors (cOmplete ULTRA®, Roche), a phosphatase inhibitor cocktail (PhosphoSTOP®, Roche), and deacetylase inhibitors AGK7, salermide, and trichostatin A (all from SantaCruz® Biochemicals), followed by a 20 min incubation on ice. The supernatants resulting from a 10 min centrifugation at 21,000g at 4°C constituted the nuclear extracts (NE) and were used to perform proteins and phosphoproteins MS-analysis.

Protein extracts were quantified using a BCA protein assay kit (Thermo Fisher Scientific) and were also resolved by SDS-PAGE following western blots using standard procedures. References for the antibodies are given in Table S7.

SILAC-based MS analysis of nuclear proteomics and phosphoproteomics

Sample preparation: Equal amount of nuclear proteins extracted from two non-SILAC mice were pooled for the 16 samples (two biological replicates for eight time points during two consecutive days; WT mix). A complex and common reference SILAC protein mix (SILAC mix) was prepared from sixteen SILAC protein samples (6 SILAC male and 10 SILAC female livers) collected at ZT0 and ZT12 (Figure S1). Thereafter, the 16 mixes were obtained by adding the same amount of SILAC mix to the WT mixes. An equivalent procedure was applied for the 4 *Cry1/2* DKO, protein samples collected every 6 hours for 24 hours and the same SILAC mix was used as a reference. Mixed heavy/light extracts were processed in parallel.

Nuclear protein SILAC-MS and data analysis: Mixed samples were reduced with 5mM DTT and alkylated with 18.75 mM iodoacetamide for 30 min at room temperature in the dark. After an ethanol-acetate precipitation, they were resuspended in 250 mM triethylammonium bicarbonate pH 8.0 containing 4M urea and digested overnight at 37°C with sequencing grade modified trypsin (Promega) at a 1:50 (w/w) trypsin:protein ratio. The obtained peptide mixtures (250 µg total material) were desalted on SepPak C18 cartridges (Waters Corp.), dried, dissolved in 4M Urea with 0.1 % Ampholytes pH 3-10 (GE Healthcare) and fractionated by off-gel focusing as described (Geiser et al., 2011). The 10 fractions obtained were desalted on a microC18 96-well plate (Waters Corp.), dried and resuspended in 0.1 % formic acid, 3 % (v/v) acetonitrile for LC-MS/MS analysis. Samples were analyzed on a hybrid linear trap LTQ-Orbitrap Velos mass spectrometer (Thermo Fisher Scientific) interfaced via a nanospray source to a Dionex RSLC 3000 nanoHPLC system (Dionex). Peptides were separated on a reversed-phase Acclaim Pepmap nanocolumn (75 µm ID x 25 cm, 2.0 µm, 100 Å, (Dionex)) with a gradient from 5 to 85 % acetonitrile in 0.1 % formic acid (total time: 200 min) and a flow rate of 300 nl/min. Full MS survey scans were performed at 60'000 resolution. In data-dependent acquisition controlled by Xcalibur 2.1 software (Thermo Fisher Scientific), the twenty most intense multiply charged precursor ions detected in the full MS survey scan were selected for CID fragmentation in the LTQ linear trap with an isolation window of 4.0 m/z and then dynamically excluded from further selection during 35s.

MS data were analyzed and quantified with MaxQuant version 1.3.0.5 (Cox and Mann, 2008), using Andromeda as search software (Cox et al., 2011) against UniProt (release 2012_02) database restricted to mouse (*Mus musculus*) taxonomy and a custom database containing usual contaminants (digestion enzymes, keratins, etc). Cleavage specificity was trypsin/P (cleavage after K, R, including KP and RP) with two missed cleavages. Mass tolerances were of 6 ppm for the precursor and 0.5 Da for CID tandem mass spectra. The iodoacetamide derivative of cysteine was specified as a fixed modification, and oxidation of methionine and protein N-terminal acetylation were specified as variable modifications. Protein identifications were filtered at 1 % FDR established by MaxQuant against a reversed sequence database. A minimum of one unique peptide was necessary to discriminate sequences which shared peptides. Sets of protein sequences which could not be discriminated based on identified peptides were listed together as protein groups and are fully reported in Table S1. Details of peak quantitation and protein ratio computation by MaxQuant are described elsewhere (Cox and Mann, 2008). Raw mass spectrometry data and search engine outputs have been

deposited to the ProteomeXchange Consortium (proteomexchange.org) via the PRIDE partner repository (Vizcaíno et al., 2016) with the dataset identifier PXD003818.

***Nuclear phosphoprotein SILAC MS and data analysis*:** Protein disulphide bridges were reduced with 10 mM tris(2-carboxyethyl)phosphine hydrochloride for 1 hour at 55 °C. Alkylation was performed with 18.75 mM iodoacetamide for 30 min at room temperature in the dark. To remove lipids and salts, proteins were precipitated using chloroform/methanol. Briefly, methanol, chloroform and water were added sequentially. Mixtures were centrifuged at 13'000 rpm for 5 min at 4 °C. Upper and lower phases were discarded. The white disk precipitates were additionally washed with methanol prior to be dried for 5 min under vacuum. Obtained pellets were resuspended in 100 mM triethylammonium hydrogen carbonate buffer and proteins were digested with premixed Lys-C/trypsin (Promega) 2:100 (w/w) at 37 °C overnight. Samples were cleaned up using Oasis HLB cartridges (Waters) and finally dried. Purified peptides (350 µg of heavy/light material) were then enriched using TiO_2 Mag Sepharose Magnetic beads (GE Healthcare) following manufacturer instructions for enrichment of phosphopeptides. The enriched samples were dissolved in 25 µL water/acetonitrile/formic acid 96.9/3/0.1 for RP-LC MS/MS. RP-LC MS/MS was performed on a hybrid linear ion trap-Orbitrap (LTQ-OT) Elite equipped with an Ultimate 3000 RSLC nano system (Thermo Fisher Scientific). Proteolytic peptides (injection of 5 µl of sample) were trapped on an Acclaim PepMap 75 µm × 2 cm (C18, 3 µm, 100 Å) pre-column and separated on an Acclaim PepMap RSLC 75 µm × 50 cm (C18, 2 µm, 100 Å) column (Thermo Scientific) coupled to a stainless steel nanobore emitter (40 mm, OD 1/32") mounted on a Nanospray Flex Ion Source (Thermo Fisher Scientific). The analytical separation was run for 150 min using a gradient that reached 30% of acetonitrile after 140 min and 80% of acetonitrile after 150 min at a flow rate of 220 nl/min. For MS survey scans, the OT resolution was 120'000 (ion population of 1×10^6) with an m/z window from 300 to 1'500. For MS/MS with collision-induced dissociation at 30% of the normalized collision energy, ion population was set to 1×10^4 (isolation width of 2), and a maximum injection time of 150 ms in the LTQ. A maximum of 20 precursor ions (most intense) were selected for MS/MS. Ions with 1+ and unassigned charge-states were rejected from MS/MS analysis. Dynamic exclusion was set for 60 s within a ± 5 ppm window. A lock mass of m/z = 445.1200 was used. Each sample was analyzed in triplicate.

MaxQuant (Cox and Mann, 2008) (version 1.4.1.2) was used for data processing. Identification was performed using Andromeda (Cox et al., 2011) as search engine against the mouse UniProtKB database (26/06/2013 release; 50818 entries). Variable amino acid modifications were acetyl (Protein N-term), phospho (STY), and oxidation (M). Carbamidometyl (C) was set as fixed modification. Trypsin/P (i.e., cleaves after lysine and arginine also if a proline follows) was selected as the proteolytic enzyme, with a maximum of two potential missed cleavages. Peptide and fragment ion tolerance were set to, respectively, 6 ppm and 0.5 Da. Peptide-spectrum match, protein and site FDRs were fixed at 1% against a reversed sequence database. Quantification was performed with stable isotope with a multiplicity of 2 using lysine $^{13}C_6$ as heavy labels. A maximum of 3 labeled amino acids per peptide was specified. Site quantification used least modified peptide.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (proteomexchange.org) via the PRIDE partner repository (Vizcaíno et al., 2016) with the dataset identifier PXD004191.

Rhythmicity analysis for nuclear proteins and phosphoproteins

We assessed the rhythmicity in temporal nuclear accumulation of proteins and phosphoproteins using harmonic regression, as previously (Mauvoisin et al., 2014). Briefly, focusing on only 24-h periodicity (data are generated under 24-h LD cycles), we used a multiple linear regression for each relative protein time trace $y(t)$ (transformed to log2 units). For this analysis, we used the relation:

$$y(t) = \mu + a \sin\left(2\pi \frac{t}{24}\right) + b \cos(2\pi \frac{t}{24}) + \text{noise}$$

where μ is the mean, whereas a and b are the coefficients of cosine and sine functions with period of 24 h, respectively. The resulting p-values of all proteins are used to estimate false discovery rate (FDR) by the Benjamini–Hochberg method (Benjamini and Hochberg, 1995). All test statistics are provided in Table S1 and S2.

To test whether rhythm of nuclear proteins in WT persists in *Cry1/2* DKO mice, we applied linear regressions combined with model selection, similar to (Atger et al., 2015). Briefly the same multiple linear regression to temporal profiles of nuclear proteins under these two conditions by considering three different cases: nuclear protein is rhythmic in KO condition with the same parameters (μ, a and b) as in WT (M1), or rhythmic with different parameters (M2); the third case is that it is non-rhythmic in KO condition (M3). A model selection approach with Bayesian information criterion (BIC) is used to control the model complexity and select the optimal model given the data. Specifically, for each model (M1, M2 and M3), multiple linear regression is performed and corresponding BIC is computed with the formula:

$$BIC = n log\left(\frac{RSS}{n}\right) + k\ log(n)$$

where RSS the sum of residuals square of the multiple linear regression, n the number of data and k the number of parameters. Schwarz weights (or probability of being the optimal model) are assigned to each model with the following expression:

$$w_j = e^{-\frac{1}{2}BIC_{Mj}} / \sum_j e^{-\frac{1}{2}BIC_{Mj}}$$

The rhythm of nuclear proteins in KO condition is considered to persist with $w_1 > 0.5$.

Identification of rhythmic protein complexes with singular value decomposition (SVD)

To annotate known protein complexes, we used protein complexes of mouse, rat and human from the CORUM database (Ruepp et al., 2010), and we manually added other known proteins complexes implicated in the circadian clock, transcription and histone modification, but missing in the CORUM database, for instance, the PER complex, BMAL1-CLOCK, REV-ERBα-NCOR1-HDAC3, RNA polymerase I, II, and III, H3K4 methyltransferase and HIRA histone chaperon complexes. Annotated known complexes in which at least two subunits detected with >8 out of 16 samples in our nuclear protein dataset were considered. In addition, missing data in some subunits, if any, were interpolated with cosine-curve fitting. To identify protein complexes showing diurnal accumulation in the nucleus, we applied the singular value (SVD) analysis to the matrix E_{gt}, in which each row represents standardized temporal profile of each subunit belonging to the same protein complex (Alter et al., 2000). SVD gives the following expression:

$$E_{gt} = \sum_{r} U_{gr} S_r V_{rt}^T$$

Where g represents the subunit; t the time point. Each column of U_{gr} is the eigengene (eigensubunit n this case) and each column of V_{tr} is the eigenarray (eigentime here) vector. S is a weight corresponding to the r'th component. If a protein complex is rhythmic, temporal profiles of its subunits are expected to be highly correlated. To measure such correlation between subunits, we calculated the fraction of variance captured by the first component $S_1 / \sum_k S_k$. To test statistical significance, background distributions for complexes with different number of subunits were calculated by randomly choosing proteins not annotated as subunits of complexes. If the fraction of variance captured by the first component is significant (p-val < 0.05), the rhythmicity and phase of protein complex were computed with the first eigen-subunit and eigentime vectors.

Inference of rhythmic motif activity of TFs with DNase hypersensitive sites (DHS) and pre-mRNAs

TF motif library curation: To obtain a relatively non-redundant motif library, we used the database of SwissRegulon (Pachkov et al., 2013), in which each TF is associated with only one unique motif. To cover as many as possible rhythmic TFs identified in our dataset of nuclear proteins, we added motifs previously implicated in circadian rhythms for ZNF187, CREBRF, HMBOX1, ETV3 and FOXK1 from Transfac (Matys et al., 2006) and HNF1B, FOXP1 and USF2 from Jaspar (Mathelier et al., 2014).

Linear model with Elastic-net regularization: To infer rhythmic motif activity of TFs, we used DHS hotspots of adult mouse liver from ENCODE (Yue et al., 2014) and temporal profiles of pre-mRNA (intronic signals as proxies of transcription) from total RNA-seq data measured in the same light dark and night-restricted feeding conditions in mouse liver (Atger et al., 2015) (GSE73552) as our nuclear protein dataset. We modeled rhythms in transcription (measured with rhythmic pre-mRNAs with FDR < 0.1 and relative amplitude > 0.2) as a linear combination of diurnal activities of TF motifs in the DHS regions ± 10 kb around transcriptional start sites (TSSs) annotated by Ensemble mm10. In particular, if multiple TSSs were annotated for one gene, only those within DHS regions were considered to be active. The model is expressed as follows:

$$E_g(t) = N_{gm} A_m(t) + \text{noise}$$

where $E_g(t)$ is the centered rhythmic transcription for gene g at time t; N_{gm} is the number of ocurrence for motif m scanned by FIMO (Grant et al., 2011) in all DHS regions ± 10 kb around the active TSSs of gene g; $A_m(t)$ represents the temporal activity of motif m at time t. To reduce the number of parameters we performed the regression in the subspace of 24 24hr periodic functions, leaving only two free parameters per motif m. To control for over-fitting and also redundancy of motifs, we employed Elastic-net penalty in the linear regression (implemented in R package glmnet (Friedman et al., 2010)). In particular, the mixing parameter $\alpha = 0.1$ was empirically chosen and the parameter λ was optimized by cross-validation. After obtaining the rhythmic motif activities, redundant motifs were removed and some motif names were manually modified to match the common terminologies (Table S6).

Inference of kinases with rhythmic activities

To predict kinases with rhythmic activities, we used all rhythmic phosphorylation sites (p-value < 0.05) with non-rhythmic nuclear proteins (p-value > 0.05) and the kinase motif library from *PhosphoNetworks* (Hu et al., 2014). Windows of 15 amino acids centered at the phosphorylation site are scanned by FIMO (Grant et al., 2011) to find hits to kinase motifs, which are potentially responsible for the rhythmic phosphorylation. Thus, we have a motif presence/absence matrix, in which each row is presence/absence of kinase motifs for each rhythmic phosphorylation site and each column indicates targets of each kinase motif. In addition, to avoid considering redundant motifs, motifs with high correlations in this matrix are manually clustered (Table S5). Here, we used two different methods to identify kinases with rhythmic activities.

Phase enrichment analysis: We used phase windows of 3 hrs shifting every 0.2 hr from 0 to 24 hrs. For each position of phase windows, rhythmic phosphorylation sites with phases within the windows are considered; enrichment of each kinase motif is calculated with hypergeometric distribution, in which the background is the total number of targets in the motif presence/absence matrix. Enriched motifs (p-value < 0.05) are selected as predicted candidates.

Linear model with regularization: An alternative method accounting for the amplitudes of rhythmic phosphorylation sites is the liner model with elastic-net regularization, which is very similar to the previous inference of rhythmic motif activities for TFs. Here the mixing parameter alpha = 0.2 was used.

FACS experiment
Purified nuclei were incubated at room temperature for 30 min with a staining solution containing 50 µg/ml of propidium iodide (Sigma Aldrich), 10 µg/ml RNAse A (Sigma Aldrich) and 0.1 % v/v Igepal CA-630 (Sigma Aldrich) in PBS. Samples were then acquired by means of a Canto II analyzer (Becton-Dickinson) equipped with a 488 nm laser line used for PI excitation. Fifty-thousand events for each sample were acquired in the forward/side scatter gate. After debris and doublet exclusion, the PI fluorescence was collected using a 585 ± 42 nm band-pass filter and visualized in a monovariate histogram. To accurately define the 2N peak, some nuclei samples were spiked with whole blood cells from the same animal at a 1:20 ratio (nuclei/blood). Data were stored in Flow Cytometry Standard files (.fcs, version 3.1) and analyzed using FCS Express 4.0 (De Novo Software). Doublets were excluded by gating on area vs height signals and data were expressed as the percentage of nuclei having different ploidy (2N, 4N, 8N) in the gate of single events.

Immunohistochemistry (IHC) experiment
Mouse livers were fixed using Formalin (Sigma Aldrich) for 24 hrs and rinse twice with PBS before being embedded in paraffin and cut in 4µm thick-slices. Tissue sections were deparaffinized and rehydrated in an ethanol series. After PBS washing, sections were treated 10 minutes with 3 % hydrogen peroxide in PBS to quench endogenous peroxidase. Heat induced epitope retrieval was then performed with 10 mM Tri-Na citrate 20 min at 95°C. After washing, sections were blocked in 1% BSA in PBS for 30 min and incubated with β-catenin or Ki-67 antibodies overnight at 4°C. After washing, secondary antibody was incubated 40 min at RT. Cell nuclei were additionally stained with Mayer's hematoxylin. Entire livers slices were imaged using an Olympus slide scanner at 20X magnification.

IHC image segmentation and estimation of different polyploid cell populations
50% (Ki-67 marker) and 25% (for β-catenin) of whole-liver IHC images from the centers of the slices were automatically segmented using a custom script in MATLAB, and using the same parameters across all time points. In particular, for the slices with plasma membrane stained by β-catenin and nuclei by hematoxylin, cells and nuclei were first segmented separately using standard functions from Image Processing Toolbox in MATLAB (e.g. imopen, bwareaopen and watershed). Only cells with both detected nuclei and plasma membrane were then kept. Moreover, cells with more than two nuclei were not analyzed. At the end, 10,000-20,000 of cells were identified with all measured features (e.g. number of nuclei per cell, nuclear and cell areas) using the function 'regionprops' for each time point.
After IHC image segmentation, we found that nuclear areas consistently show bimodal/trimodal distributions in both mono- and bi-nucleated cells across all time point (Fig S7F). It is known that nuclear areas are approximately proportional to their DNA contents (Danielsen et al., 1986; Martin et al., 2002) and we thus modeled those distributions as mixtures of Gaussians to estimate the population sizes of nuclei with a certain ploidy. In addition, to estimate the parameters of the mixture model, we assumed that two nuclei in bi-nucleated cells have the same DNA content and we did not consider liver aneuploidy (Duncan, 2013). Moreover, we constrained the total fractions of nuclear ploidy to those measured by FACS (step 1). Finally we refit (step 2) the proportions of nuclear areas of mono- and bi-nucleated cells in each sample, using the obtained parameters form step 1 (e.g. mean and variance of nuclear areas). The estimated proportions of nuclei with different DNA contents for mono- and bi-nucleated cells, and the percentages of mono- and bi-nucleated cells, allowed us to estimate fractions of all different polyploid cell populations (Fig. 7E).

Glycogen content
Glycogen amount in liver was estimated using the method from (Lo et al., 1970). In brief, frozen liver were incubated 20 min in KOH at 100°C. Then, 95% ethanol was added and after an 840 g centrifugation at 4°C, the glycogen pellet was suspended in water. Then 10 µl of this suspension was diluted fifteen times and absorbance was measured at 490 nm after addition of 150µl of Phenol 5% and 750 µl of sulfuric acid 98%.

Supplemental References

Alter, O., Brown, P.O., and Botstein, D. (2000). Singular value decomposition for genome-wide expression data processing and modeling. Proc Natl Acad Sci U S A 97, 10101-10106.

Atger, F., Gobet, C., Marquis, J., Martin, E., Wang, J., Weger, B., Lefebvre, G., Descombes, P., Naef, F., and Gachon, F. (2015). Circadian and feeding rhythms differentially affect rhythmic mRNA transcription and translation in mouse liver. Proc Natl Acad Sci U S A 112, E6579-E6588.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Statist Soc B 57, 289-300.

Binder, J.X., Pletscher-Frankild, S., Tsafou, K., Stolte, C., O'Donoghue, S.I., Schneider, R., and Jensen, L.J. (2014). COMPARTMENTS: unification and visualization of protein subcellular localization evidence. Database 2014.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26, 1367-1372.

Cox, J.r., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. J Proteome Res 10, 1794-1805.

Danielsen, H., Lindmo, T., and Reith, A. (1986). A method for determining ploidy distributions in liver tissue by stereological analysis of nuclear size calibrated by flow cytometric DNA analysis. Cytometry 7, 475-480.

Duncan, A.W. (2013). Aneuploidy, polyploidy and ploidy reversal in the liver. Semin Cell Dev Biol 24, 347-356.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw 33, 1-22.

Geiser, L., Dayon, L., Vaezzadeh, A.R., and Hochstrasser, D.F. (2011). Shotgun proteomics: a relative quantitative approach using Off-Gel electrophoresis and LC-MS/MS. Methods Mol Biol 681, 459-472.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics 27, 1017-1018.

Hu, J., Rho, H.-S., Newman, R.H., Zhang, J., Zhu, H., and Qian, J. (2014). PhosphoNetworks: a database for human phosphorylation networks. Bioinformatics 30, 141-142.

Kislinger, T., Cox, B., Kannan, A., Chung, C., Hu, P., Ignatchenko, A., Scott, M.S., Gramolini, A.O., Morris, Q., Hallett, M.T., Rossant, J., Hughes, T.R., Frey, B., and Emili, A. (2006). Global Survey of Organ and Organelle Protein Expression in Mouse: Combined Proteomic and Transcriptomic Profiling. Cell 125, 173-186.

Lo, S., Russell, J.C., and Taylor, A.W. (1970). Determination of glycogen in small tissue samples. J Appl Physiol 28, 234-236.

Martin, N.C., McCullough, C.T., Bush, P.G., Sharp, L., Hall, A.C., and Harrison, D.J. (2002). Functional analysis of mouse hepatocytes differing in DNA content: volume, receptor expression, and effect of IFNγ. J Cell Physiol 191, 138-144.

Mathelier, A., Zhao, X., Zhang, A.W., Parcy, F., Worsley-Hunt, R., Arenillas, D.J., Buchman, S., Chen, C.-y., Chou, A., Ienasescu, H., Lim, J., Shyr, C., Tan, G., Zhou, M., Lenhard, B., Sandelin, A., and Wasserman, W.W. (2014). JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucleic Acids Res 42, D142-D147.

Matys, V., Kel-Margoulis, O.V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K., Voss, N., Stegmaier, P., Lewicki-Potapov, B., Saxel, H., Kel, A.E., and Wingender, E. (2006). TRANSFAC® and its module TRANSCompel®: transcriptional gene regulation in eukaryotes. Nucleic Acids Res 34, D108-D110.

Mauvoisin, D., Wang, J., Jouffe, C., Martin, E., Atger, F., Waridel, P., Quadroni, M., Gachon, F., and Naef, F. (2014). Circadian clock-dependent and -independent rhythmic proteomes implement distinct diurnal functions in mouse liver. Proc Natl Acad Sci U S A 111, 167-172.

Pachkov, M., Balwierz, P.J., Arnold, P., Ozonov, E., and van Nimwegen, E. (2013). SwissRegulon, a database of genome-wide annotations of regulatory sites: recent updates. Nucleic Acids Res 41, D214-D220.

Ruepp, A., Waegele, B., Lechner, M., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C., and Mewes, H.W. (2010). CORUM: the comprehensive resource of mammalian protein complexes-2009. Nucleic Acids Res 38, D497-D501.

The UniProt, C. (2015). UniProt: a hub for protein information. Nucleic Acids Res 43, D204-D212.

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

Vizcaíno, J.A., Csordas, A., del-Toro, N., Dianes, J.A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., Xu, Q.-W., Wang, R., and Hermjakob, H. (2016). 2016 update of the PRIDE database and its related tools. Nucleic Acids Res 44, D447-D456.

Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D., Shen, Y., Pervouchine, D.D., Djebali, S., Thurman, R.E., Kaul, R., Rynes, E., Kirilusha, A., Marinov, G.K., Williams, B.A., Trout, D., Amrhein, H., Fisher-Aylor, K., Antoshechkin, I., DeSalvo, G., See, L.-H., Fastuca, M., Drenkow, J., Zaleski, C., Dobin, A., Prieto, P., Lagarde, J., Bussotti, G., Tanzer, A., Denas, O., Li, K., Bender, M.A., Zhang, M.,

Byron, R., Groudine, M.T., McCleary, D., Pham, L., Ye, Z., Kuan, S., Edsall, L., Wu, Y.-C., Rasmussen, M.D., Bansal, M.S., Kellis, M., Keller, C.A., Morrissey, C.S., Mishra, T., Jain, D., Dogan, N., Harris, R.S., Cayting, P., Kawli, T., Boyle, A.P., Euskirchen, G., Kundaje, A., Lin, S., Lin, Y., Jansen, C., Malladi, V.S., Cline, M.S., Erickson, D.T., Kirkup, V.M., Learned, K., Sloan, C.A., Rosenbloom, K.R., Lacerda de Sousa, B., Beal, K., Pignatelli, M., Flicek, P., Lian, J., Kahveci, T., Lee, D., James Kent, W., Ramalho Santos, M., Herrero, J., Notredame, C., Johnson, A., Vong, S., Lee, K., Bates, D., Neri, F., Diegel, M., Canfield, T., Sabo, P.J., Wilken, M.S., Reh, T.A., Giste, E., Shafer, A., Kutyavin, T., Haugen, E., Dunn, D., Reynolds, A.P., Neph, S., Humbert, R., Scott Hansen, R., De Bruijn, M., Selleri, L., Rudensky, A., Josefowicz, S., Samstein, R., Eichler, E.E., Orkin, S.H., Levasseur, D., Papayannopoulou, T., Chang, K.-H., Skoultchi, A., Gosh, S., Disteche, C., Treuting, P., Wang, Y., Weiss, M.J., Blobel, G.A., Cao, X., Zhong, S., Wang, T., Good, P.J., Lowdon, R.F., Adams, L.B., Zhou, X.-Q., Pazin, M.J., Feingold, E.A., Wold, B., Taylor, J., Mortazavi, A., Weissman, S.M., Stamatoyannopoulos, J.A., Snyder, M.P., Guigo, R., Gingeras, T.R., Gilbert, D.M., Hardison, R.C., Beer, M.A., Ren, B., and The Mouse, E.C. (2014). A comparative encyclopedia of DNA elements in the mouse genome. Nature 515, 355-364.