PNAS
www.pnas.org

Supplementary Information for:

Diurnal rhythms across the human dorsal and ventral striatum

Kyle D. Ketchesin[a,1], Wei Zong[b,1], Mariah A. Hildebrand[a], Marianne L. Seney[a], Kelly M. Cahill[b], Madeline R. Scott[a], Vaishnavi G. Shankar[a], Jill R. Glausier[a], David A. Lewis[a], George C. Tseng[b,*], and Colleen A. McClung[a,*]

Colleen A. McClung, Ph.D.
Email: mcclungca@upmc.edu

George C. Tseng, Sc.D.
Email: ctseng@pitt.edu

This PDF file includes:

Other supplementary materials for this manuscript include the following:

Datasets S1 to S17

SI Methods

Human Postmortem Brain Samples

Brain tissue was obtained, following consent from the next of kin, during autopsies conducted at the Allegheny County (Pittsburgh, PA) or Davidson County (Nashville, TN) Medical Examiner's Office. All procedures were approved by the University of Pittsburgh Institutional Review Board for Biomedical Research and Committee for Oversight of Research and Clinical Training Involving Decedents.

The right hemisphere of each brain was blocked coronally, immediately frozen, and stored at -80° C. Tissue blocks containing the body of the caudate, putamen, and NAc were selected for analysis (Atlas of the Human Brain, 4th edition). The medial-lateral border between the caudate and putamen was scored to specifically exclude the internal capsule. The ventral border of the caudate and putamen, and accordingly the dorsal border of the NAc, was defined by the anterior thalamic radiation of the internal capsule or the anterior commissure. Cryostat sections were cut at 40 µm thickness, and each striatal subregion from an individual section was placed into its respective collection tube until a total of volume of 35 mm^3 was collected from each region. This quantity of tissue was determined to contain appropriate levels of RNA in RNA-seq pilot studies."

RNA-Sequencing and Data Preprocessing

FastQC v0.11.3 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was performed to assess the quality of the data. Per base sequence quality was high (Quality score generally>30), indicating good data quality. HISAT2 (1) (HISAT2v2.1.0) was used to align reads to the reference (Homo sapiens Ensembl GRCh38) using default parameters. The Picard Tools module CollectRNASeqMetrics

(https://gatk.broadinstitute.org/hc/en-us/articles/360037057492-CollectRnaSeqMetrics-Picard-) was used to calculate mapping rates to coding vs non-coding regions of the genome (Table S3). The resulting bam files from HISAT2 were converted to expression count data using HTSeq (2) (HTSeq v0.10.0) with default union mode. For the NAc and dorsal striatum sequencing runs (caudate and putamen samples run simultaneously), the average total paired end reads were approximately 46.5 and 46.0 million reads, respectively. The average mapped reads for the NAc and dorsal striatum were 30.5 and 33.6 million reads, respectively. RNA-seq count data were transformed to $\log_2$ continuous counts per million (cpm) data using the cpm function of the Bioconductor edgeR package (3, 4). Transcripts were retained for analysis if $\log_2$(cpm) was greater than 1 in 50% or more of subjects. All Y-chromosome gene were also eliminated from analysis. After filtering, 15,300 (NAc), 15,041 (caudate), and 14,866 (putamen) transcripts remained.

Rhythmicity Analyses

All individual samples were ordered by their TOD and expression for each transcript was fitted to a sinusoidal curve, defined as follows: $y = A\sin(f(t + p)) + b$, where y is the gene expression level, A is an amplitude factor, $f = \pi/12$ is fixed frequency such that 24 hours is a period, t is time of death, p is a phase factor, and b is the offset. Levenberg-Marquardt algorithm was applied to solve the sinusoidal curve fitting. The A, p, and b were estimated, and the peak hour of the circadian pattern and goodness-of-fit coefficient R^2 were calculated. Here, $R^2 = 1 - RSS_m/RSS_0$, where RSS_m is the residual sum of squares of the fitted model and RSS_0 is the residual sum of squares of the null model. The null hypothesis is that there is no rhythmic pattern and R^2 was used to assess

the significance level. The null distribution of R^2 was generated by pooling R^2 of all transcripts in 1,000 different TOD-permuted expression datasets for each region separately, generating ~15,000 (transcripts) x 1,000 (permutations) = 1.5 x 10^7 null values to infer the rhythmic p-values. The null distribution of R^2 did not differ between rhythmic vs non-rhythmic or protein-coding vs non-coding transcripts for each striatal region (Fig. S14). Q-values of R^2 were calculated by Benjamini-Hochberg procedure. Confidence intervals of peak estimates were derived by generating 1000 bootstrap samples. Peaks are estimated for each bootstrap sample and the 90% confidence interval was constructed from the 5% and 95% quantiles of bootstrap peaks with consideration of circular time effect (e.g., peak at 19 hr is equivalent to -5 hr). Confidence interval boundaries were adjusted by +/- 24 hours to lie within [-6, 18]. The peak and phase estimate confidence intervals for all transcripts in each region can be found in Dataset S2.

The Gaussian mixture model was parametrized by the mixture component weights and the component means and common variance. The number of Gaussian components was selected by BIC. With parameters estimated by Expectation-Maximization algorithm, the posterior probability of a given phase difference belonging to each component can be calculated to estimate the decision boundary between clusters. In both of the NAc-caudate and caudate-putamen comparisons, the number of clusters were selected to be 3 with one component centered around zero representing the concordant cluster, one with mean greater than zero representing the advanced phase cluster and one with mean smaller than zero representing the delayed phase clusters. The decision boundaries between the advanced cluster and concordant cluster and the concordant

cluster and delayed cluster are around +4 and -4, respectively, in both comparisons (Fig. S11).

Peak times in each striatal region were also calculated and plotted for transcripts that are rhythmic across regions (Fig. S12). Similar to the phase concordance plots, AW-Fisher method was used to combine p-values across regions followed by Benjamini-Hochberg correction to derive meta-analyzed q-values. Transcripts with meta-analyzed $q<0.05$ and AW-Fisher weight (1,1,1) were considered rhythmic across regions. Transcripts were plotted by their peak time in the NAc (Fig. S12). For transcripts in the caudate and putamen whose peak time difference was greater than 12, relative to the NAc, peak times were adjusted by +/- 24 hours so that all peak time differences were smaller than 12 (Fig. S12).

Pathway Enrichment and Upstream Regulator Analysis

For the pathway and upstream regulator analyses for each individual region, a significance threshold of $p<0.01$ (NAc: 1344 transcripts; caudate: 1053 transcripts; putamen: 3097 transcripts) was used to determine rhythmicity for the transcript input list. The transcripts expressed in each respective region were used as the reference set for pathway and upstream regulator enrichment. For analyses comparing rhythmic overlap between regions, a significance threshold of $p<0.05$ was used to determine rhythmic overlap for the transcript input list. This threshold was chosen so that the transcript input list would be large enough to determine pathway and upstream regulator enrichment. For all analyses involving comparisons between regions, transcripts expressed in any of the three regions (n=15672) were used as the reference set for pathway and upstream regulator enrichment. For all analyses, pathways and upstream regulators were

considered enriched if they met a threshold of $p<0.05$ ($-\log_{10}$(p-value)>1.3 in figures). IPA pathways with <15 or >300 genes were not included in the analyses. For the upstream regulator analysis, only direct relationships were considered. The top 5 significant pathways and upstream regulators are represented for each analysis. Complete lists of all significant pathways and upstream regulators for each analysis can be found in Datasets S4, S7, S9, S12, and S15.

Biological Process Enrichment

For the process enrichment comparisons across the three striatal regions (Fig. 3C), the top 1000 rhythmic transcripts in each region were used instead of $p<0.01$, to avoid bias of enrichment analysis from different input gene numbers in each region. A user-supplied list of the 15672 transcripts expressed in any of the three striatal regions was used as the enrichment background. Following conversion to Entrez gene IDs, 13043 background transcripts remained, as some of the transcripts (e.g., some non-coding RNAs) were not mapped or functionally annotated (see Datasets S5, S8, S10, S13, and S16 for the complete lists of annotated transcripts following Entrez gene ID conversion for each analysis). Terms with $p<0.01$, a minimum count of 3, and an enrichment factor of >1.5 were grouped into clusters based on their membership similarities (kappa score >0.3). P-values were calculated based on the cumulative hypergeometric distribution and q-values were calculated based on the Benjamini-Hochberg procedure (5). The most statistically significant term within a cluster was chosen to represent the cluster. The top 20 significant clusters of processes are depicted in the Cytoscape networks. If more than 10 enriched terms within a cluster were identified, the top 10 significant terms were chosen for visualization in the Cytoscape networks. Complete lists of all the enriched

processes within each cluster can be found in Datasets S5, S8, S10, S13, and S16. In the network plots, the nodes are represented as pie charts, where the size of the pie is proportional to the total number of gene hits for that specific term. The pie charts are color-coded based on the identity of the gene list, where the size of the slice represents the percentage of genes for the term that originated from the corresponding list (5). Terms that are similar (kappa score >0.3) are connected by edges.

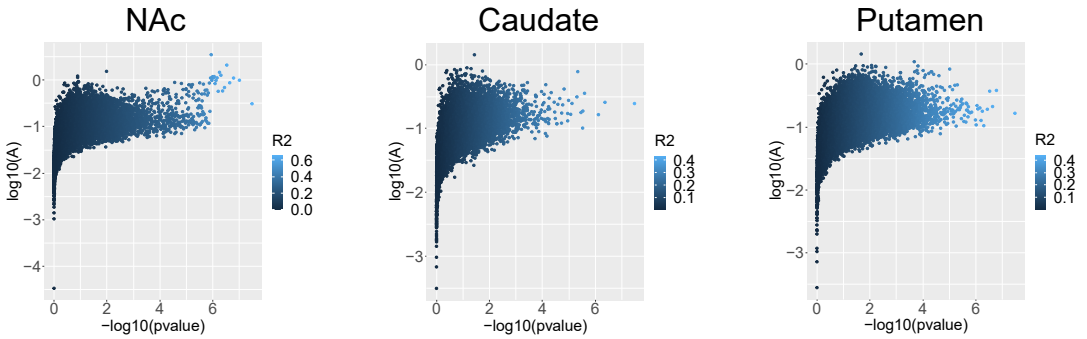
Phase Set Enrichment Analysis

Analysis was limited to the top 1000 rhythmic genes in each brain region. Peak time of transcripts adjusted to [ZT0, ZT24] were used as input. The gene set file "C5 GO biological process" was downloaded from MSigDB. The minimum number of items per gene set was 15. Since the distribution of peak times in gene sets among the top 1000 transcripts were already phase clustered in each region, enrichment was tested against the empirical background distribution (using Kuiper $q<0.05$) to test if each gene set behaves differently from the majority of others within a region.

SI References

1. Kim D, Langmead B, & Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. *Nature methods* 12(4):357-360.
2. Anders S, Pyl PT, & Huber W (2015) HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31(2):166-169.
3. Law CW, Chen Y, Shi W, & Smyth GK (2014) voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome biology* 15(2):R29.
4. Ritchie ME, *et al.* (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* 43(7):e47.
5. Zhou Y, *et al.* (2019) Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature communications* 10(1):1523.

A



B

Average R^2			
	NAc	**Caudate**	**Putamen**
q < 0.01	0.35	0.37	0.29
q < 0.05	0.27	0.29	0.21
p < 0.01	0.22	0.19	0.21
p < 0.05	0.16	0.14	0.17

C

Average Amplitude			
	NAc	**Caudate**	**Putamen**
q < 0.01	0.30	0.29	0.21
q < 0.05	0.22	0.20	0.20
p < 0.01	0.19	0.18	0.20
p < 0.05	0.17	0.17	0.20

Fig. S1. Amplitude (log10(A); y-axis) vs p-value (-log10(p-value); x-axis) plots (A) for NAc (left), caudate (middle), and putamen (right). The strength of the R^2 value is depicted in color. The lower the p-value (higher R^2), the more likely it was that the transcript showed a higher amplitude. Tables showing the average R^2 (A) and amplitude (B) values at various significance thresholds.

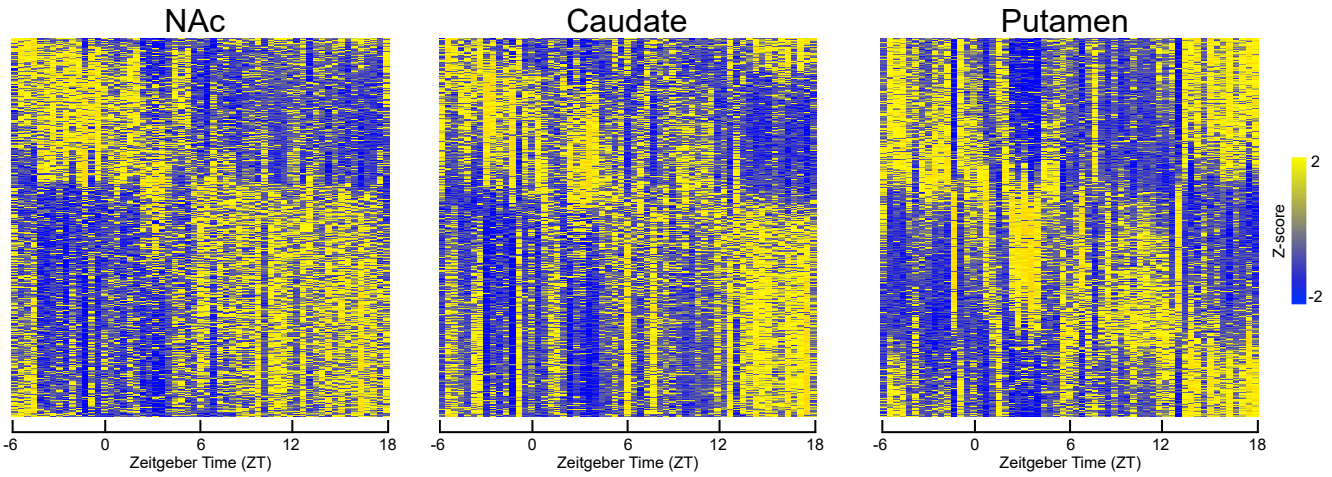


Fig. S2. Heatmaps showing global rhythms in transcript expression ($p<0.01$) across the 3 striatal regions. Expression levels were Z-transformed for each transcript and the transcripts were ordered by the time at which they peak. Each column represents a subject and the subjects were ordered by their time of death.

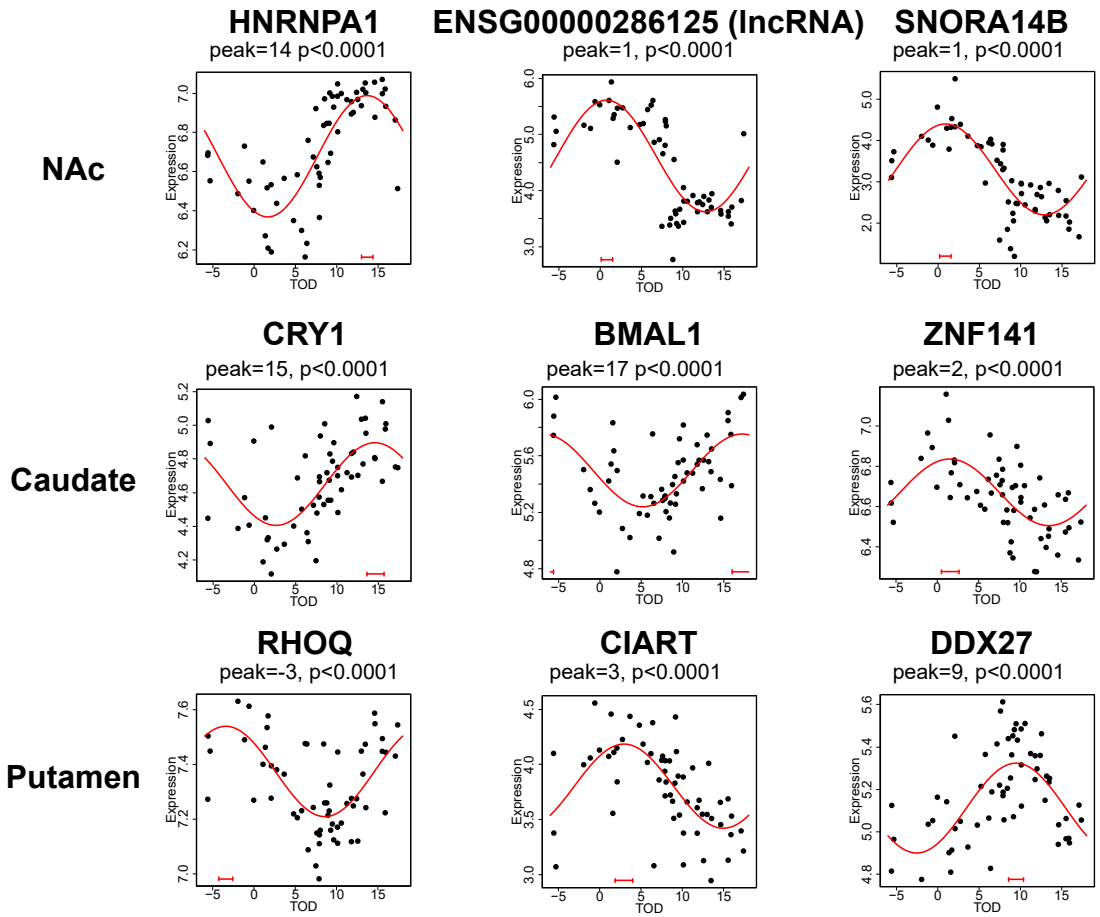


Fig. S3. Top 3 rhythmic transcripts in the NAc (top), caudate (middle), and putamen (bottom). In the scatterplots, each dot represents a subject with the x-axis indicating TOD on a ZT scale (-6 – 18h) and the y-axis indicating level of transcript expression. The red line is the fitted sinusoidal curve. Confidence intervals (90%) for peak estimates are depicted by the red bar on each plot. Peak time and p-values are located above each scatterplot.

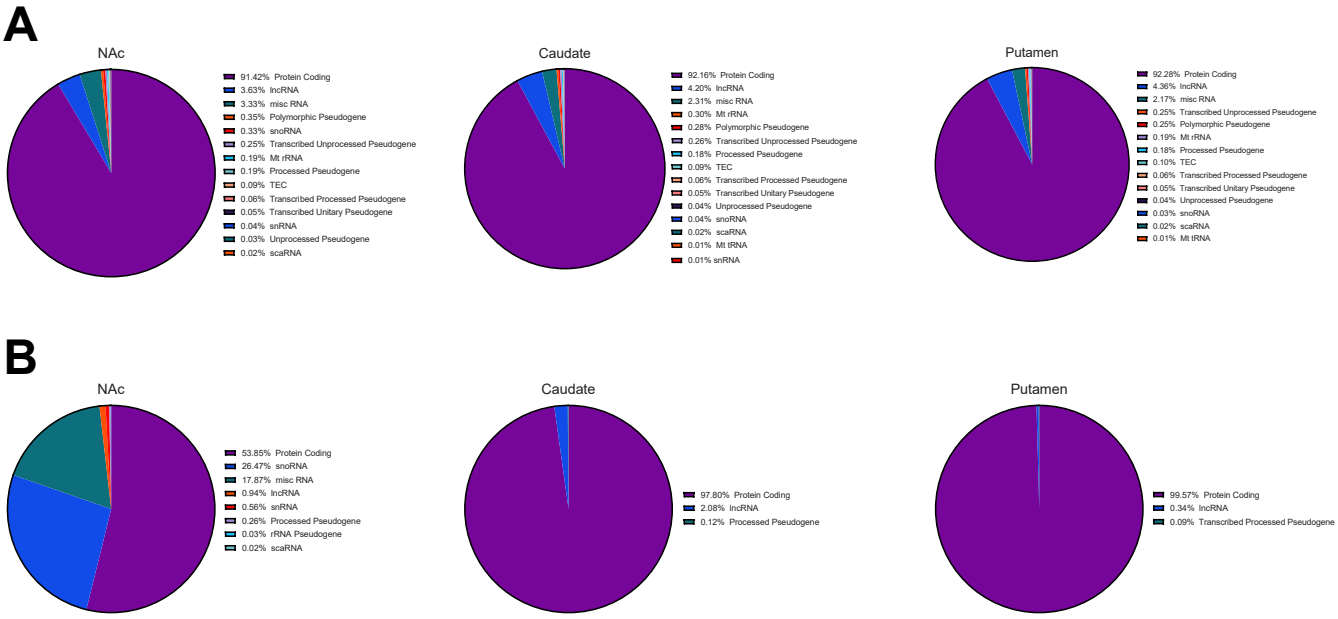


Fig. S4. Biotype charts of the percentage of reads mapped to all expressed transcripts (A) and the top 100 rhythmic transcripts (B) for each striatal region.

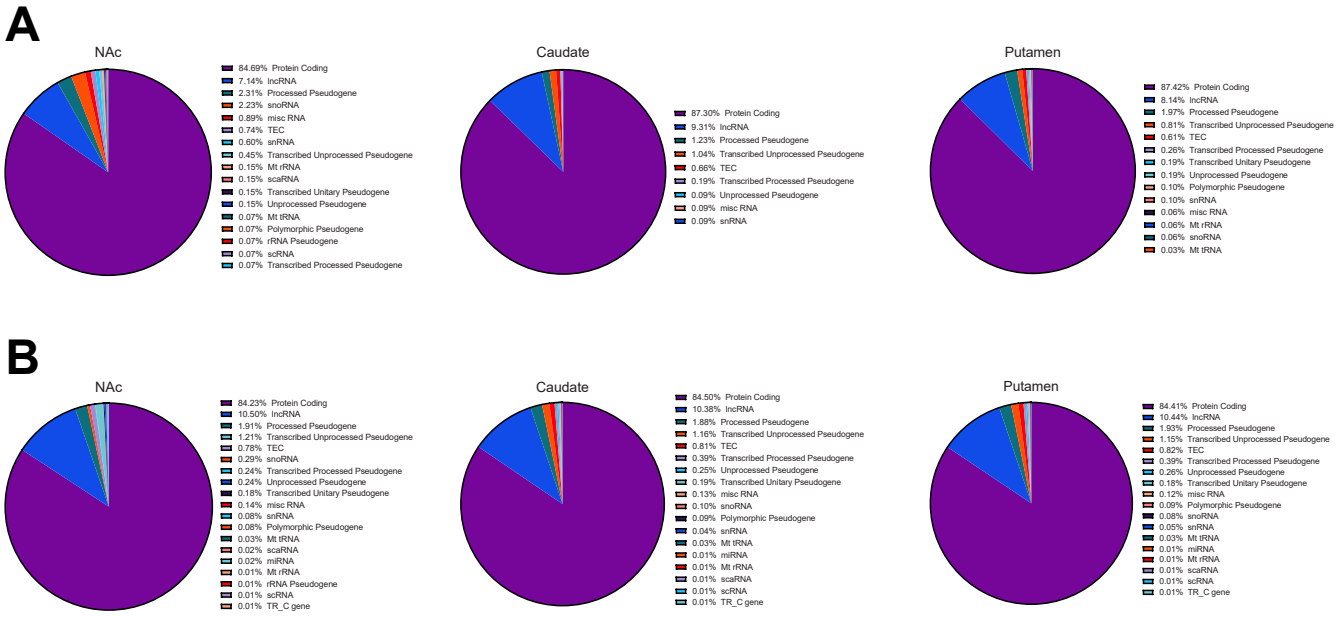
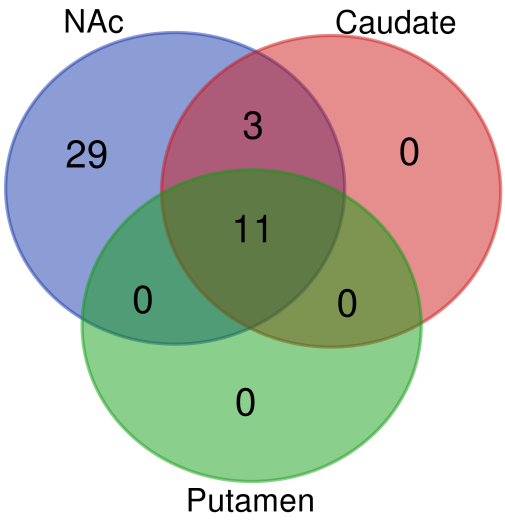


Fig. S5. Biotype charts of rhythmic transcripts at a significant threshold of $p<0.01$ (A) and all expressed transcripts (B) for each striatal region.

A



B

snoRNA Rhythmicity			
	NAc	Caudate	Putamen
Total # of snoRNAs	43	14	11
# of Rhythmic Transcripts (% of Total)	33 (76.7%)	0 (0%)	6 (54.5%)

C

NAc			Putamen		
ENSEMBL	Symbol	p-value	ENSEMBL	Symbol	p-value
ENSG00000207181	SNORA14B	1.63E-07	**ENSG00000263934**	SNORD3A	0.0025
ENSG00000272533	SNORA28	3.58E-07	**ENSG00000208892**	SNORA49	0.0084
ENSG00000201831	SNORD115-1	6.19E-07	**ENSG00000212384**	SNORD113-2	0.0105
ENSG00000207088	SNORA7B	6.84E-07	**ENSG00000212232**	SNORD17	0.0242
ENSG00000235408	SNORA71B	9.45E-07	ENSG00000201700	SNORD113-3	0.0265
ENSG00000207067	SNORA72	1.08E-06	**ENSG00000200087**	SNORA73B	0.0424
ENSG00000271907	SNORA35B	1.34E-06			
ENSG00000263934	SNORD3A	1.60E-06			
ENSG00000206838	SNORA5A	1.92E-06			
ENSG00000208839	SNORA35	2.25E-06			
ENSG00000239039	SNORD13	5.83E-06			
ENSG00000207493	SNORA46	6.42E-06			
ENSG00000208772	SNORD94	1.01E-05			
ENSG00000207233	SNORA37	1.04E-05			
ENSG00000212283	SNORD89	1.32E-05			
ENSG00000264940	SNORD3C	1.53E-05			
ENSG00000201448	SNORA63C	2.41E-05			
ENSG00000207445	SNORD15B	4.73E-05			
ENSG00000208892	SNORA49	5.88E-05			
ENSG00000238961	SNORA47	6.95E-05			
ENSG00000262074	SNORD3B-2	7.33E-05			
ENSG00000200087	SNORA73B	0.0001			
ENSG00000200792	SNORA80A	0.0001			
ENSG00000201229	SNORA63D	0.0002			
ENSG00000201772	SNORA5C	0.0002			
ENSG00000265185	SNORD3B-1	0.0003			
ENSG00000212402	SNORA74B	0.0009			
ENSG00000212232	SNORD17	0.0009			
ENSG00000212384	SNORD113-2	0.0011			
ENSG00000201998	SNORA23	0.0061			
ENSG00000222489	SNORA79B	0.0281			
ENSG00000200959	SNORA74A	0.0409			
ENSG00000272344	SNORD114-21	0.0485			

Bold = Rhythmic in NAc

Fig. S6. Characterization of snoRNA expression in each striatal region. (A) Venn diagram showing the overlap of detected snoRNAs in each striatal region. There are 29 unique snoRNAs expressed in the NAc. (B) Table showing the total number of snoRNAs in each region and the number/percentage that are rhythmic. A large percentage (76.7%) of snoRNAs are rhythmic in the NAc. None of the 14 snoRNAs in the caudate are rhythmic. (C) Lists of the rhythmic transcripts in the NAc and putamen, with corresponding p-values. Bold transcripts in the putamen list are also rhythmic in the NAc.

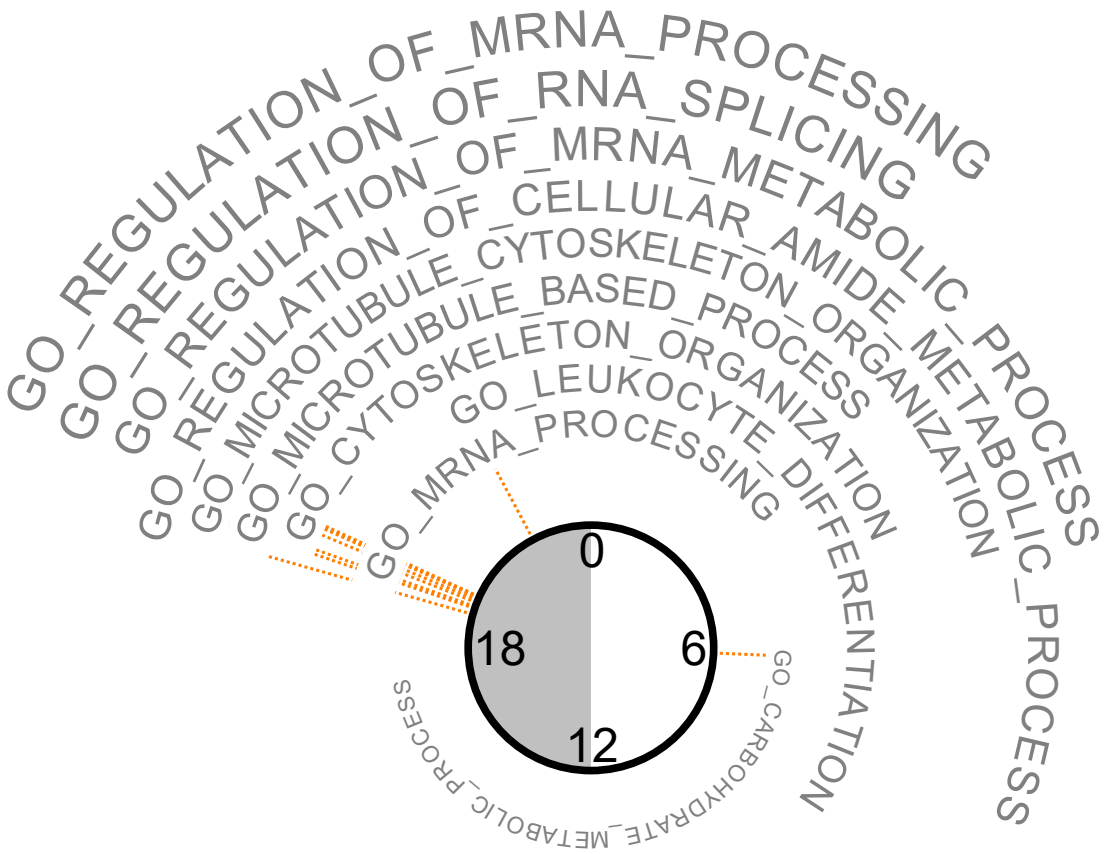


Fig. S7. Summary of significantly phase-clustered biological processes in the NAc. Phase set enrichment analysis was performed on the top 1000 rhythmic transcripts in the NAc. Peak time of transcripts adjusted to [ZT0, ZT24] were used as input. Enrichment was tested against the empirical background distribution (using Kuiper $q<0.05$). Gray and white shading represents night and day, respectively, and 0, 6, 12, and 18 indicate ZT time, with dawn corresponding to ZT0. The circular axis represents the vector-average phase value of all transcripts in a biological process. Biological processes that are further away from the center of the circle (larger font size) have a larger magnitude of the vector-average with more unimodal phase clustering. Biological processes related to mRNA processing and splicing are phase-clustered at night around ZT19-20.

Fig. S8. Summary of significantly phase-clustered biological processes in the caudate. Phase set enrichment analysis was performed on the top 1000 rhythmic transcripts in the caudate. Peak time of transcripts adjusted to [ZT0, ZT24] were used as input. Enrichment was tested against the empirical background distribution (using Kuiper $q<0.05$). Gray and white shading represents night and day, respectively, and 0, 6, 12, and 18 indicate ZT time, with dawn corresponding to ZT0. The circular axis represents the vector-average phase value of all transcripts in a biological process. Biological processes that are further away from the center of the circle (larger font size) have a larger magnitude of the vector-average with more unimodal phase clustering. Biological processes related to translation and mitochondrial function are phase-clustered around ZT21-23, prior to sunrise.

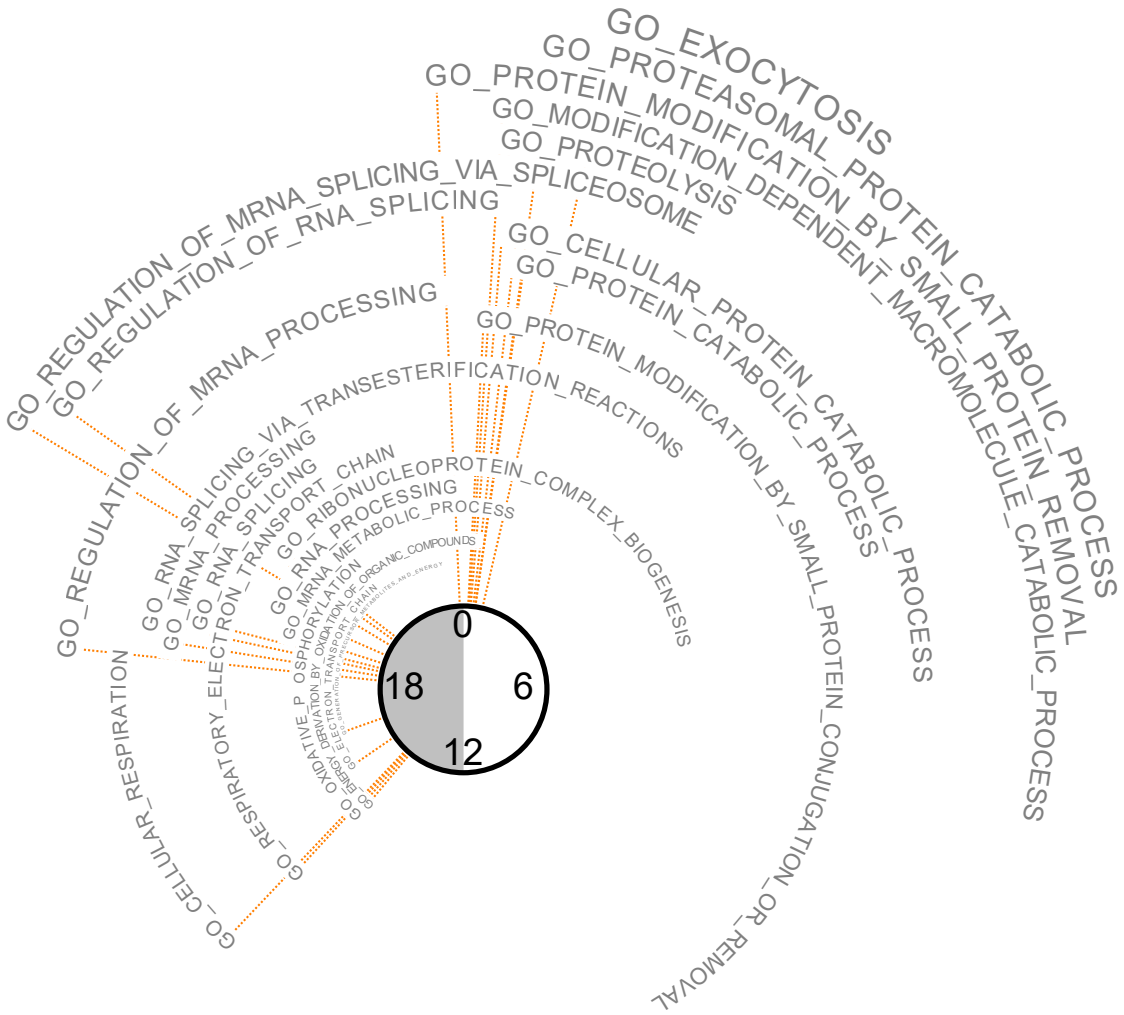


Fig. S9. Summary of significantly phase-clustered biological processes in the putamen. Phase set enrichment analysis was performed on the top 1000 rhythmic transcripts in the putamen. Peak time of transcripts adjusted to [ZT0, ZT24] were used as input. Enrichment was tested against the empirical background distribution (using Kuiper $q<0.05$). Gray and white shading represents night and day, respectively, and 0, 6, 12, and 18 indicate ZT time, with dawn corresponding to ZT0. The circular axis represents the vector-average phase value of all transcripts in a biological process. Biological processes that are further away from the center of the circle (larger font size) have a larger magnitude of the vector-average with more unimodal phase clustering. Biological processes related to protein catabolism and cellular stress are phase-clustered around subjective dawn and processes related to mRNA processing and splicing are phase-clustered at night.

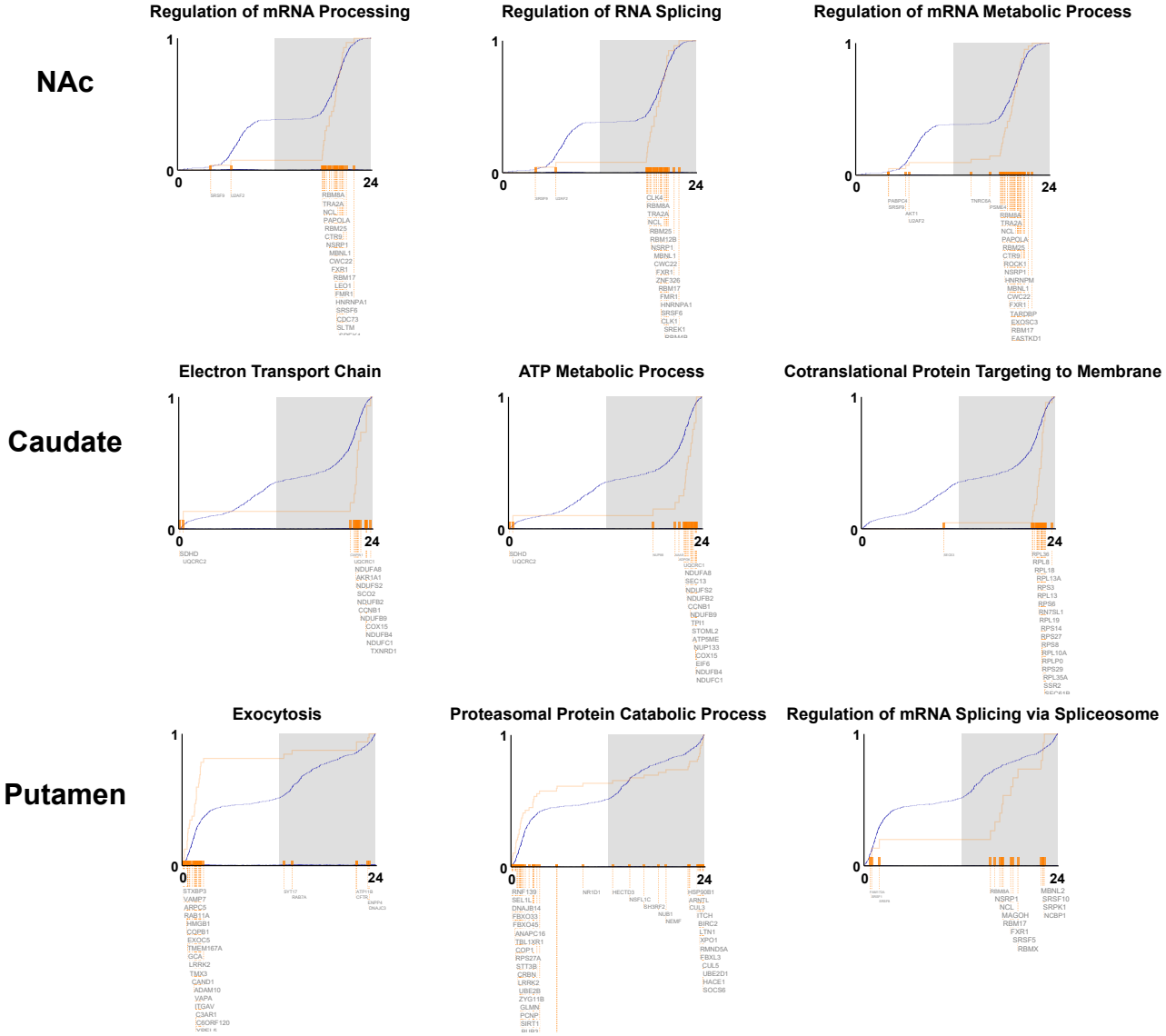


Fig. S10. Phase distributions of rhythmic transcripts for example biological processes in NAc (top), caudate (middle), and putamen (bottom). Phase distributions for the top 3 processes with the largest vector-average magnitudes are shown in the NAc (top) and caudate (middle). Phase distributions for the top 2 processes with the largest vector-average magnitude ("Exocytosis" and "Proteasomal Protein Catabolic Process") and the top process related to splicing ("Regulation of mRNA Splicing via Spliceosome") are shown for the putamen (bottom). The blue line in each plot depicts the empirical background phase distribution and the orange line depicts the empirical cumulative distribution of phases belonging to transcripts in each biological process. Larger font sizes of the transcript name indicate greater contribution by that transcript to the overall phase clustering of a biological process.

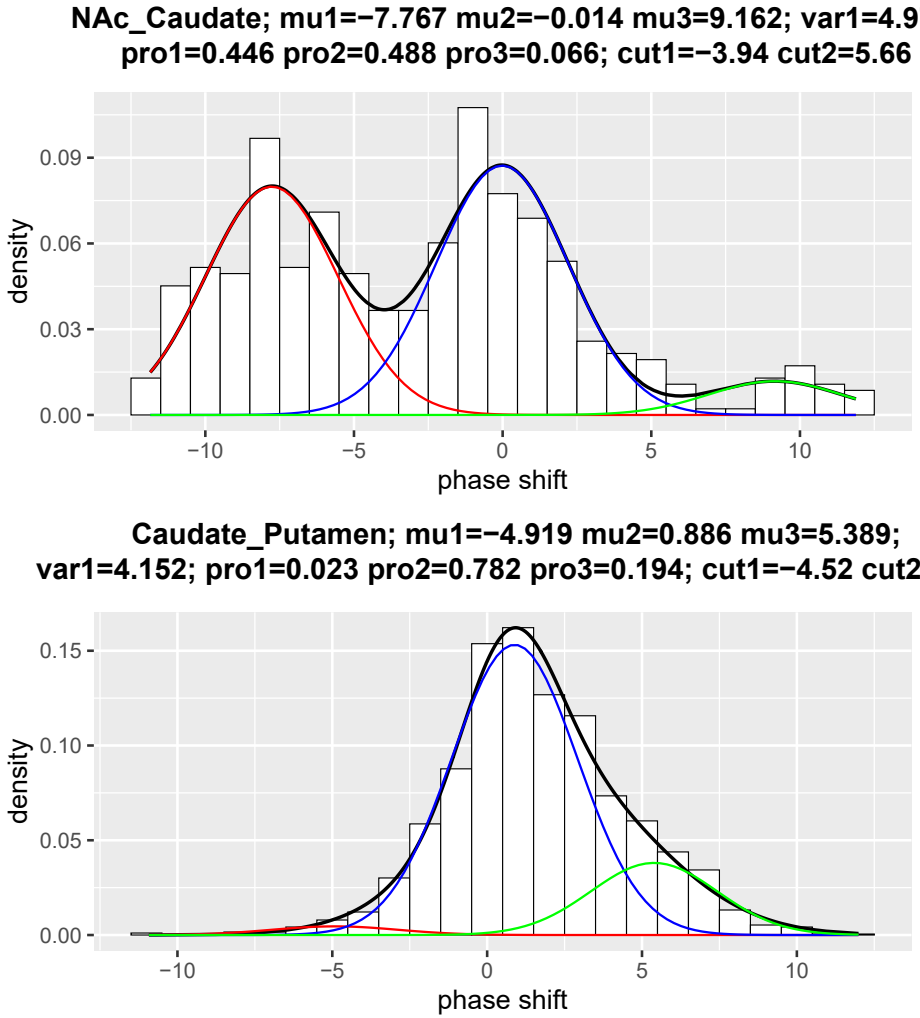


Fig. S11. Univariate Gaussian mixture model used for selecting phase concordance time window (+/- 4 hours). The univariate Gaussian mixture model was fitted to the phase difference of the two brain region comparisons with distinct discordant clusters (NAc and caudate (top) and caudate and putamen (bottom)). The Gaussian mixture model was parametrized by the mixture component weights and the component means and common variance. The number of Gaussian components was selected by BIC. With parameters estimated by Expectation-Maximization algorithm, the posterior probability of a given phase difference belonging to each component was calculated to estimate the decision boundary between clusters. In both of the NAc-caudate (top) and caudate-putamen comparisons (bottom), the number of clusters were selected to be 3 with one component centered around zero representing the concordant cluster, one with mean greater than zero representing the advanced phase cluster and one with mean smaller than zero representing the delayed phase clusters. The decision boundaries between the advanced cluster and concordant cluster and the concordant cluster and delayed cluster are around +4 and -4, respectively, in both comparisons.

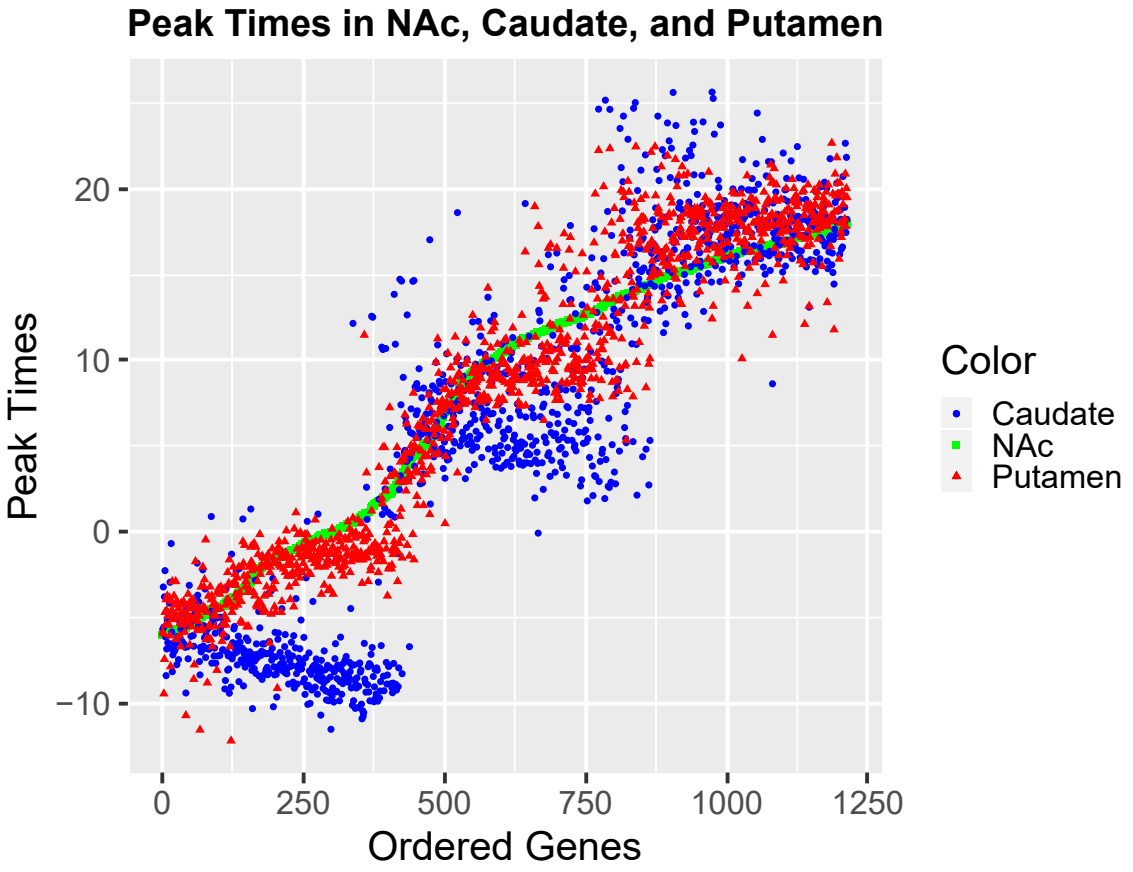


Fig. S12. Peak times in each striatal region for transcripts that are rhythmic across regions. Transcripts with meta-analyzed q-value<0.05 and AW-Fisher weight (1,1,1) were considered rhythmic across regions. Transcripts were first ordered by their peak time in the NAc. For transcripts in the caudate and putamen whose peak time difference was greater than 12, relative to the NAc, peak times were adjusted by +/- 24 hours so that all peak time differences were smaller than 12. The x-axis represents transcripts that were ordered by the NAc peak time from ZT -6 – 18. The y-axis represents peak times of each gene in the NAc (green), caudate (blue), and putamen (red). There are distinct clusters of transcripts in the caudate related to translation and mitochondrial function that peak at distinct times (-12 to -5 and 0 to 10, respectively) compared to NAc and putamen.

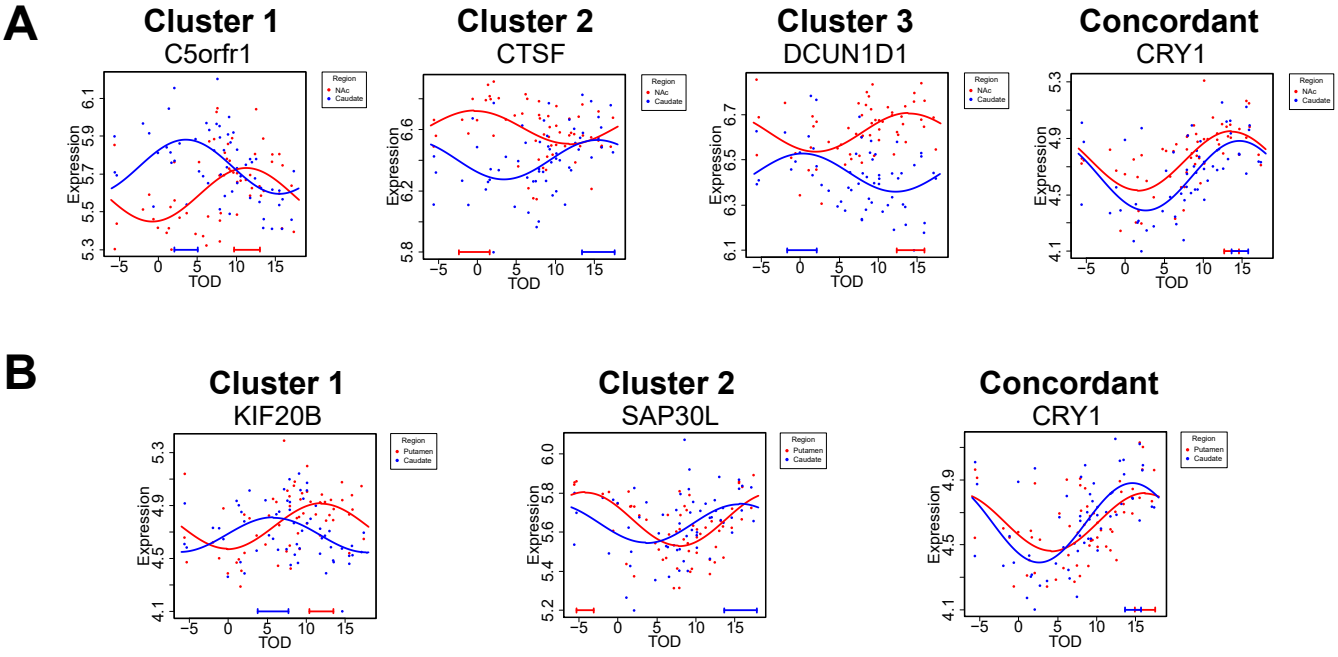


Fig. S13. Example double-plotted scatterplots for NAc and caudate (A) and caudate and putamen (B) with bars depicting 90% confidence intervals of peak estimates. Discordant transcripts in each cluster have peak times with confidence intervals that do not overlap. The core clock gene, CRY1, shows a similar peak time in each region comparison with overlapping confidence intervals.

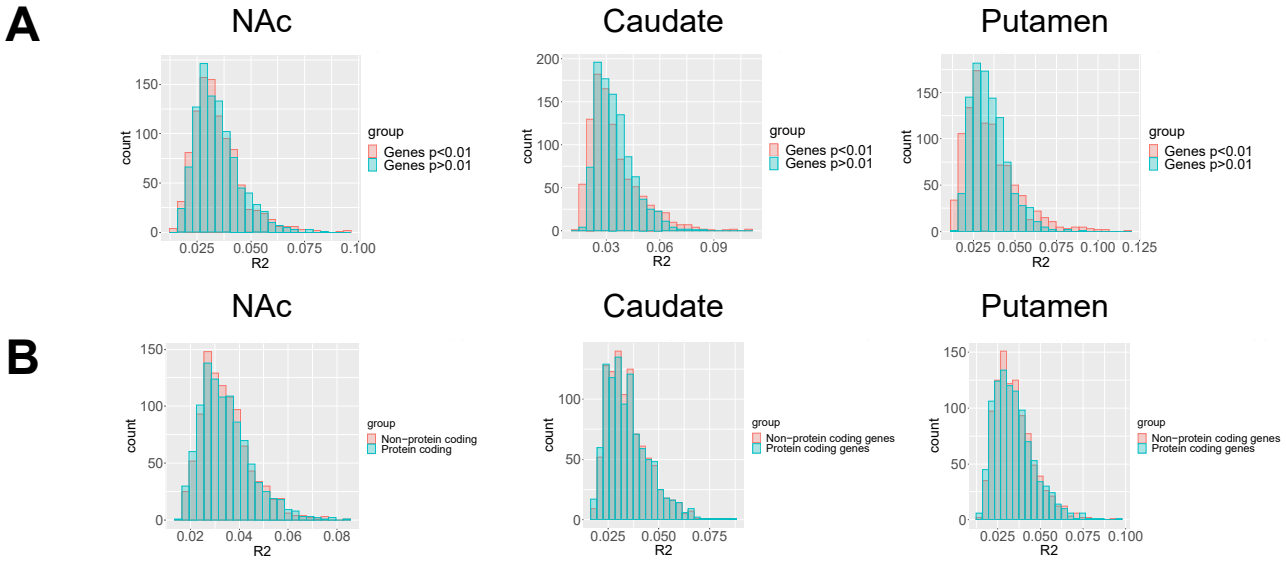


Fig. S14. Histograms of the null distributions of average R^2 comparing rhythmic vs non-rhythmic transcripts (A) and protein-coding vs non-coding transcripts (B) for NAc (left), caudate (middle), and putamen (right). The null distributions between each comparison were very similar across regions.

Table S1. Subject Characteristics	
Sex, No. (%)	
Male	48 (80)
Female	12 (20)
Race, No. (%)	
White/Hispanic	53 (88)
Black	7 (12)
Age, mean ± SD, y	47.0 ± 9.8
PMI, mean ± SD, h	17.7 ± 5.8
Brain pH, mean ± SD	6.6 ± 0.2
TOD, mean ± SD	7.6 ± 5.8

Abbreviations: h, hours; PMI, postmortem interval; SD, standard deviation; TOD, time of death; y, years

Table S2. Rhythm characteristics for concordant and discordant transcripts for NAc vs caudate (top), caudate vs putamen (middle), and NAc vs putamen (bottom) phase comparisons. The number of transcripts, average R^2 values, and percentage of non-overlapping peak estimate confidence intervals are shown for concordant transcripts and discordant clusters.

NAc vs Caudate				
	No. of Transcripts	**Avg. R^2 NAc**	**Avg. R^2 Caudate**	**% Non-Overlapping CI**
Cluster 1	67	0.21	0.15	98.5
Cluster 2	124	0.19	0.15	99.2
Cluster 3	29	0.19	0.17	96.6
Concordant	215	0.18	0.18	1.4

Caudate vs Putamen				
	No. of Transcripts	**Avg. R^2 Caudate**	**Avg. R^2 Putamen**	**% Non-Overlapping CI**
Cluster 1	107	0.13	0.17	87.9
Cluster 2	261	0.15	0.15	85.4
Concordant	1482	0.13	0.20	1.9

NAc vs Putamen				
	No. of Transcripts	**Avg. R^2 NAc**	**Avg. R^2 Putamen**	**% Non-Overlapping CI**
Discordant	211	0.16	0.17	74.4
Concordant	1715	0.16	0.18	2.6

Table S3. The Picard Tools module CollectRNASeqMetrics was used to calculate mapping rates to coding vs non-coding regions of the genome. The values in the table are mean ± SD. There are similar percentages of mapping to intronic, intergenic, and mRNA bases between striatal regions.

	% Intronic Bases	% Intergenic Bases	% mRNA Bases
NAc	46.5 ± 3.2	9.6 ± 3.8	43.9 ± 4.1
Caudate	48.1 ± 3.5	7.4 ± 0.8	44.6 ± 4.0
Putamen	48.4 ± 4.1	6.9 ± 0.6	44.7 ± 4.6

Dataset S1. This spreadsheet contains the subject and tissue characteristics for each of the 60 subjects.

Dataset S2. This spreadsheet contains the best-fitted circadian curve parameters with corresponding p- and q-values for all transcripts in the NAc, caudate, and putamen.

Dataset S3. This spreadsheet contains the snoRNA target analysis of the 33 rhythmic snoRNAs in the NAc using the database tool, snoDB.

Dataset S4. This spreadsheet contains the complete lists of pathways and upstream regulators for the NAc, caudate, and putamen from Figs. 3A and 3B.

Dataset S5. This spreadsheet contains the complete list of all enriched processes within each cluster for the GO Biological Process enrichment in Fig. 3C.

Dataset S6. This spreadsheet contains the list of all phase-enriched biological processes in each region from the PSEA analyses in Figs. S7-S9.

Dataset S7. This spreadsheet contains the complete lists of pathways and upstream regulators for the region overlap in Fig. 4C.

Dataset S8. This spreadsheet contains the complete list of all enriched processes within each cluster for the GO Biological Process enrichment in Fig. 4D.

Dataset S9. This spreadsheet contains the complete lists of pathways and upstream regulators for the concordant and discordant transcripts between NAc and caudate in Fig. 5A.

Dataset S10. This spreadsheet contains the complete list of all enriched processes within each cluster for the GO Biological Process enrichment in Fig. 5A.

Dataset S11. This spreadsheet contains all of the concordant and discordant transcripts with peak estimate confidence intervals for NAc and caudate comparisons in Fig. 5A.

Dataset S12. This spreadsheet contains the complete lists of pathways and upstream regulators for the concordant and discordant transcripts between the NAc and putamen in Fig. 5B.

Dataset S13. This spreadsheet contains the complete list of all enriched processes within each cluster for the GO Biological Process enrichment in Fig. 5B.

Dataset S14. This spreadsheet contains all of the concordant and discordant transcripts with peak estimate confidence intervals for NAc and putamen comparisons in Fig. 5B.

Dataset S15. This spreadsheet contains the complete lists of pathways and upstream regulators for the concordant and discordant transcripts between the caudate and putamen in Fig. 5C.

Dataset S16. This spreadsheet contains the complete list of all enriched processes within each cluster for the GO Biological Process enrichment in Fig. 5C.

Dataset S17. This spreadsheet contains all of the concordant and discordant transcripts with peak estimate confidence intervals for caudate and putamen comparisons in Fig. 5C.