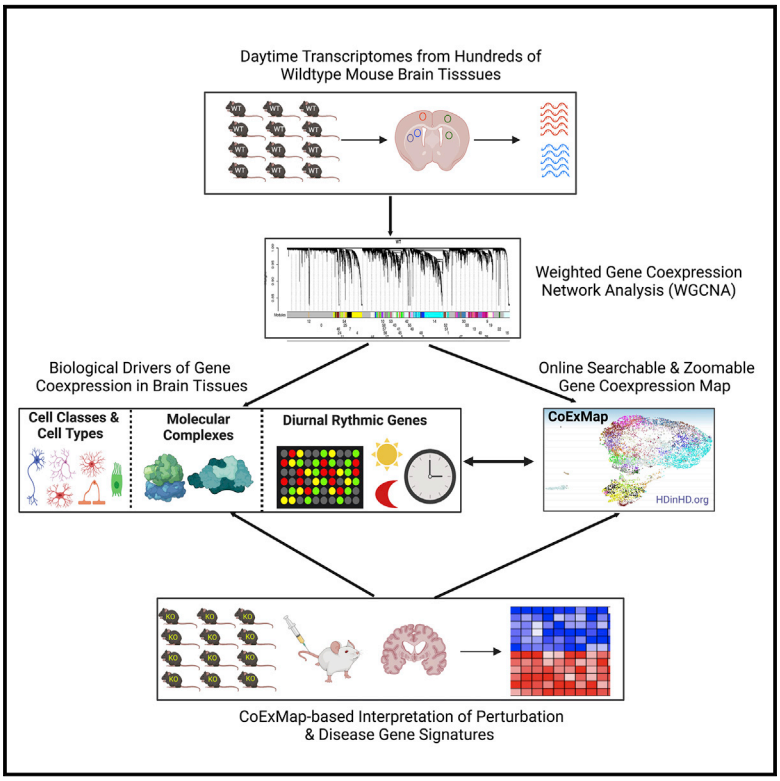
Neuron

Article

Mapping brain gene coexpression in daytime transcriptomes unveils diurnal molecular networks and deciphers perturbation gene signatures

Graphical abstract



Highlights

- A large-scale study defines reproducible gene coexpression modules in the mouse brain
- Diurnal time is a novel driver of gene coexpression in the mouse striatum and cortex
- An online searchable map can help visualize and search the gene coexpression networks
- The CoExMap is a powerful tool to decipher perturbation and disease gene signatures

Authors

Nan Wang, Peter Langfelder, Matthew Stricos, ..., Jim Rosinski, Steve Horvath, X. William Yang

Correspondence

xwyang@mednet.ucla.edu

In brief

We employed a large-scale study to identify reproducible striatal and cortical gene coexpression modules. Besides modules enriched in cell type and molecular complex genes, most modules are enriched in diurnally expressed genes. These modules and a searchable online map (CoExMap) help decipher gene signatures from genetic perturbations and brain diseases.

Wang et al., 2022, Neuron *110*, 3318–3338
October 19, 2022
https://doi.org/10.1016/j.neuron.2022.09.028

CellPress

CellPress
OPEN ACCESS

Neuron

Article

Mapping brain gene coexpression in daytime transcriptomes unveils diurnal molecular networks and deciphers perturbation gene signatures

Nan Wang,[1,2,3,9] Peter Langfelder,[1,2,3,9] Matthew Stricos,[1,2,3] Lalini Ramanathan,[1,2,3] Jeffrey B. Richman,[1,2,3] Raymond Vaca,[1,2,3] Mary Plascencia,[1,2,3] Xiaofeng Gu,[1,2,3] Shasha Zhang,[1,2,3] T. Katherine Tamai,[2,3] Liguo Zhang,[4] Fuying Gao,[1,2] Koliane Ouk,[5] Xiang Lu,[1,2] Leonid V. Ivanov,[6] Thomas F. Vogt,[7] Qing Richard Lu,[4] A. Jennifer Morton,[5] Christopher S. Colwell,[2,3] Jeffrey S. Aaronson,[7] Jim Rosinski,[7] Steve Horvath,[8] and X. William Yang[1,2,10,*]

[1]Center for Neurobehavioral Genetics, Semel Institute for Neuroscience & Human Behavior, University of California, Los Angeles, Los Angeles, CA, USA

[2]Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA

[3]UCLA Brain Research Institute, University of California, Los Angeles, Los Angeles, CA, USA

[4]Department of Pediatrics, Division of Experimental Hematology and Cancer Biology, Brain Tumor Center, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

[5]Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, UK

[6]Rancho Biosciences LLC, San Diego, CA, USA

[7]CHDI Management /CHDI Foundation, Princeton, NJ, USA

[8]Department of Human Genetics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, Los Angeles, CA, USA

[9]These authors contributed equally

[10]Lead contact

*Correspondence: xwyang@mednet.ucla.edu

https://doi.org/10.1016/j.neuron.2022.09.028

SUMMARY

Brain tissue transcriptomes may be organized into gene coexpression networks, but their underlying biological drivers remain incompletely understood. Here, we undertook a large-scale transcriptomic study using 508 wild-type mouse striatal tissue samples dissected exclusively in the afternoons to define 38 highly reproducible gene coexpression modules. We found that 13 and 11 modules are enriched in cell-type and molecular complex markers, respectively. Importantly, 18 modules are highly enriched in daily rhythmically expressed genes that peak or trough with distinct temporal kinetics, revealing the underlying biology of striatal diurnal gene networks. Moreover, the diurnal coexpression networks are a dominant feature of daytime transcriptomes in the mouse cortex. We next employed the striatal coexpression modules to decipher the striatal transcriptomic signatures from Huntington's disease models and heterozygous null mice for 52 genes, uncovering novel functions for Prkcq and Kdm4b in oligodendrocyte differentiation and bipolar disorder-associated Trank1 in regulating anxiety-like behaviors and nocturnal locomotion.

INTRODUCTION

A fundamental process in biology is the flow of genetic information from DNA to RNA and then to proteins to mediate physiological functions. At the cellular level, the expression of functionally related genes is often tightly co-regulated to enable temporal control of molecular-complex formation and avoid conflicting molecular processes (Geschwind and Konopka, 2009; Stuart et al., 2003). One widely used systems biology approach to examine tissue-level transcriptomic programs uses gene coexpression to identify gene modules with strongly correlated expression patterns across an entire dataset (Zhang and Horvath, 2005; Langfelder and Horvath, 2008; Parikshak et al., 2015). Weighted gene correlation network analysis (WGCNA) permits the construction of robust gene coexpression modules from transcriptomic datasets (Langfelder and Horvath, 2008). In such networks, the expression values of genes in the entire module can be summarized by its eigengene (i.e., first principal component), and hub genes in each module can be ranked based on their fuzzy module membership. The biological meaning of each module can be inferred based on gene set enrichment analysis or the known function of its hub genes (Langfelder and Horvath, 2008).

Despite the extensive use of gene coexpression networks to study the mammalian brain, some basic questions remain: what are the types of coexpressed gene modules that could be delineated in wild-type (WT) brain tissues? And what are their underlying biological drivers? Transcriptomic studies of

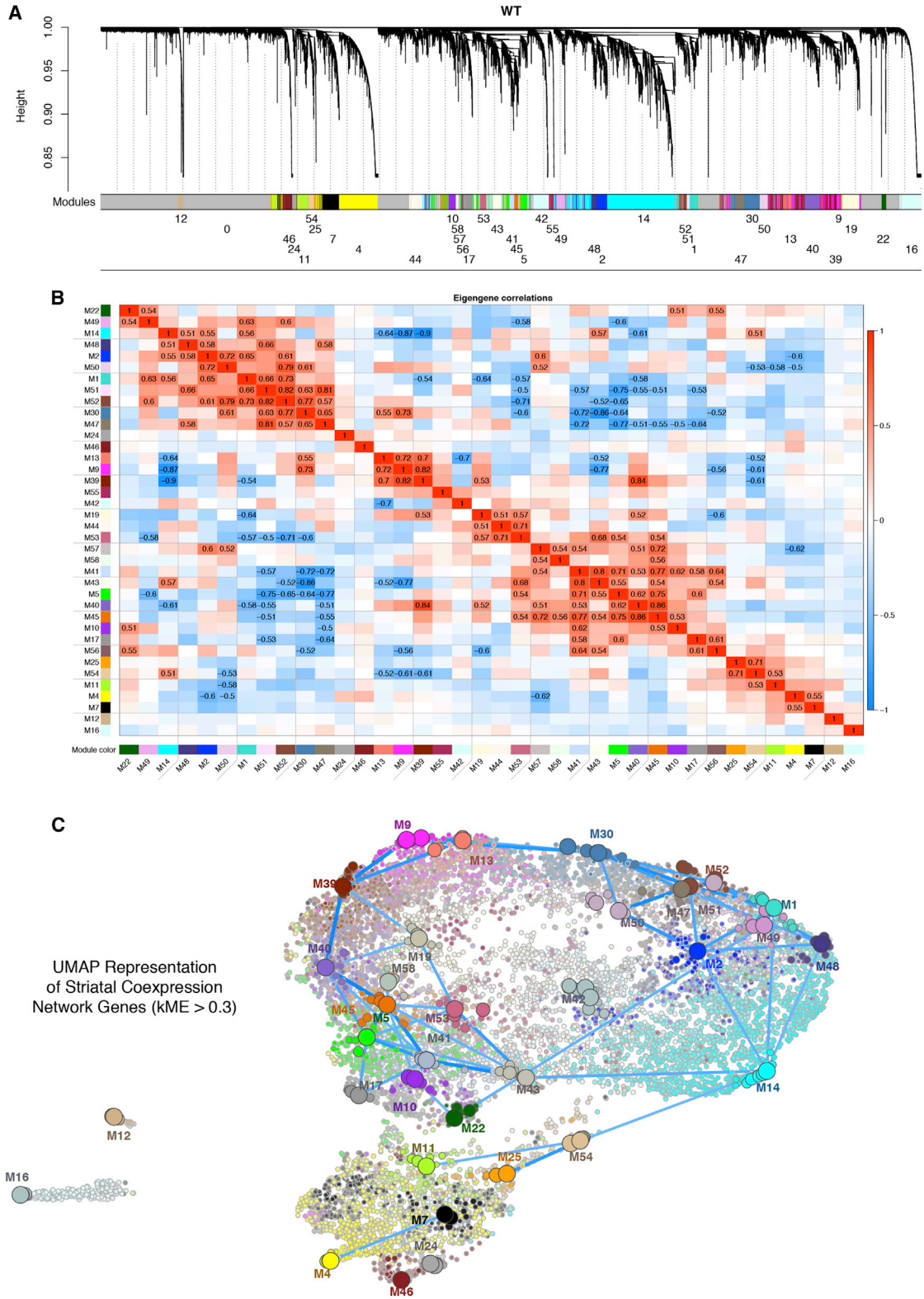


Figure 1. Large-scale striatal gene coexpression networks in adult wild-type mice and its visualization with UMAP

(A) Hierarchical clustering tree (dendrogram) of genes based on the striatum co-expression network. Modules are indicated by colors and numeric labels below the dendrogram.

(legend continued on next page)

mammalian brains to date suggest gene-coexpression modules mostly represent brain regions, and within each brain region, modules correspond to different cell classes or cell types (Kelley et al., 2018; McKenzie et al., 2018; Oldham et al., 2008). It is also possible to detect modules enriched in genes specific to cellular organelles such as ribosomes and mitochondria (Winden et al., 2009). These findings suggest that gene coexpression analysis can detect subtle variations in cellular composition due to natural variations across individuals or tissue dissection (Oldham et al., 2008). However, the full compendium of gene coexpression modules in WT brain tissues is yet to be discerned.

Here, we identified striatal and cortical coexpression networks using daytime transcriptomic data from adult mouse tissues. We annotated coexpression modules based on their enrichment in genes related to cell-type markers and molecular complexes as well as diurnally expressed genes. Moreover, we applied the annotated striatal gene coexpression networks to decipher gene signatures related to Huntington's disease (HD) and those from 52 knockout mouse lines, leading to new biological insights for multiple genes.

RESULTS

Overview of the study design

In this study, we asked what the biological drivers of gene coexpression network modules in adult mouse brain tissues are. We took advantage of an ongoing large-scale transcriptomic study of heterozygous knockout (KO-Het) and WT littermates for genes selected from mutant Huntingtin (mHtt) CAG-length-dependent coexpression networks (Langfelder et al., 2016; Figure S1; Table S1). We used a total of 510 WT striatal and 100 WT cortical RNA-sequencing (RNA-seq) samples in our WGCNA analysis. To interpret the resulting gene coexpression network, we employed an additional 134 striatal and cortical diurnal time-dependent RNA-seq samples for WT mice as well as 404 transcriptomic samples from 52 KO-Het mice. Overall, this study employed >1,100 novel mouse striatal and cortical tissue RNA-seq samples (Figure S1).

Mapping gene coexpression networks in the adult mouse striatum

The striatum is a basal ganglia nucleus that is critical for motor and cognitive function, and it is affected in many neurological and psychiatric disorders (Graybiel and Grafton, 2015; Kreitzer and Malenka, 2008). Here, we set out to thoroughly map gene coexpression modules that occur naturally in the daytime striatal tissue transcriptomes (Figure 1A). Using 407 striatal transcriptomic samples, we applied WGCNA to define 38 gene coexpression modules ranging in size from 47 to 1,695 genes and containing a total of 10,958 genes (Figures 1A and 1B; Table S2). We next used 103 independent WT striatal RNA-seq samples to show that all modules identified with the 407 samples are strongly preserved in the 103 samples (Figure S2; Langfelder et al., 2011). To quantify module preservation, we evaluated eigengene-based intramodular connectivity (*kME*) in all modules between the original discovery and replication datasets and found high preservation (correlations > 0.7) for all but 2 modules (Figure S2); the highest *kME* correlations were found in M16 (0.93) and M12 (0.92). Together, we concluded that the 38 WGCNA modules (referred to as striatal CoExMap hereafter) are highly reproducible in the transcriptomes of adult mouse striatal tissue.

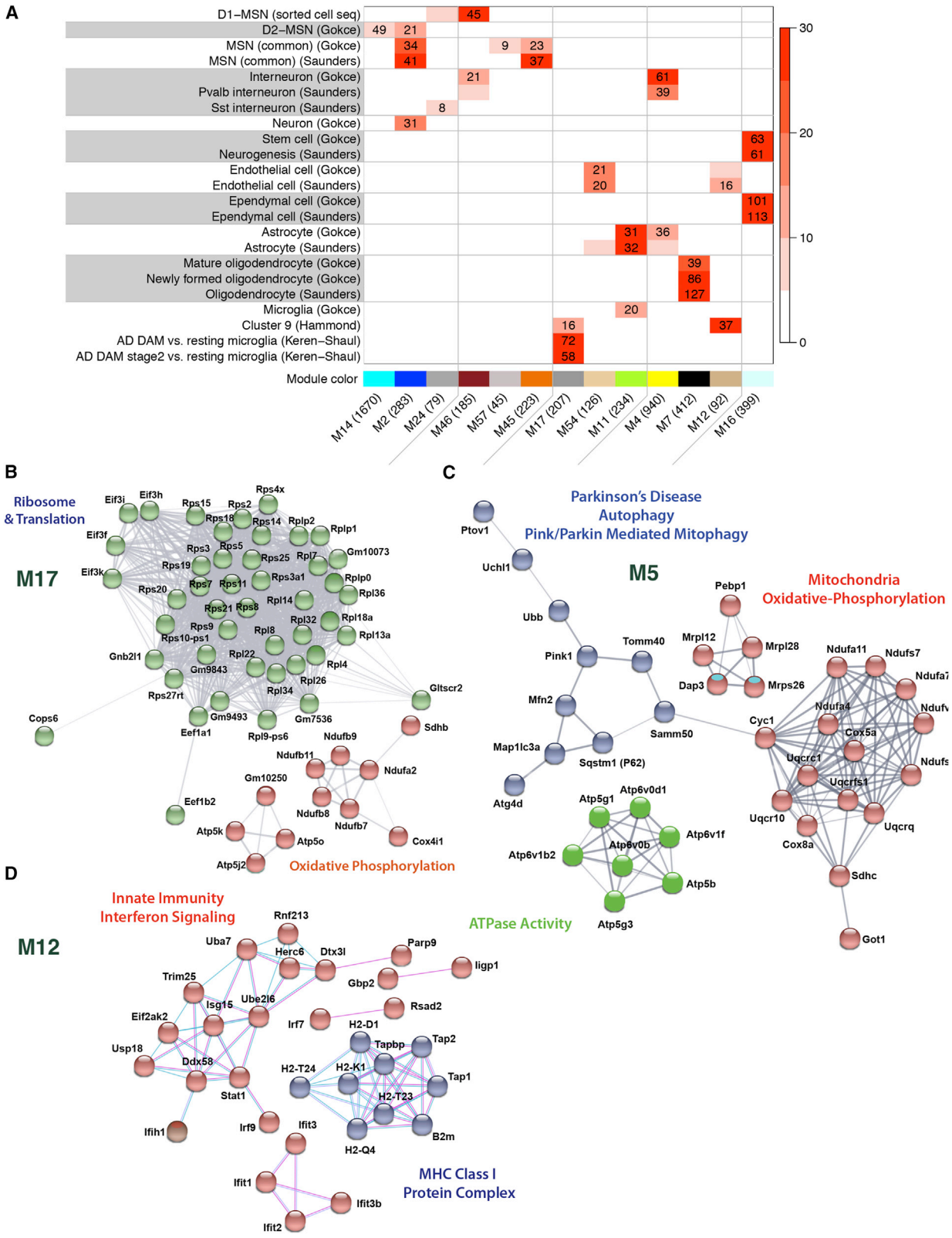
We next asked how many transcriptomic samples are needed to reliably detect these modules. We first selected random subsets of the 407 samples ranging from n = 15 to 220 samples and calculated module preservation (Figure S3). We found that 60 samples are sufficient for most modules (32/38) to reach *kME* correlations of >0.6, and 110 samples allow the majority of modules (31/38) to reach *kME* correlation of >0.8. Second, we asked how many samples would be needed in a *de novo* WGCNA study to identify modules that recapitulate one of the original 38 modules. We found that higher replication sample numbers resulted in more replication modules significantly overlapping with or containing at least 50% of the top hub genes for one of the 38 original modules (Figures S4A and S4B). With the top hub gene-based criteria, we showed that WGCNA analysis with about 60 samples can reach the level of module overlap as the 103-sample replication study.

We also performed separate WGCNA analyses of striatal male and female transcriptomic samples and found that all identified modules were highly concordant with each other, and modules in this sex-specific analysis were highly preserved in the transcriptomic datasets of the other sex (Figures S5A–S5C). Thus, the striatal gene coexpression modules remained mostly unchanged using either sex-specific or sex-adjusted analyses.

Visualization of the adult mouse striatal gene coexpression network

We next asked whether we could develop visualization tools based on a 2-dimensional (2D) embedding of genes in the CoExMap modules with either uniform manifold approximation and projection (UMAP) (McInnes et al., 2018) or t-distributed stochastic neighbor embedding (t-SNE) (Van der Maaten and Hinton, 2008). We used either the gene expression values in each sample or the *kME* of each gene in these modules as inputs (STAR Methods). All four data visualization schemes showed positive correlations (between 0.3 and 0.7) between distance in the visualization and correlations among genes and modules (Figures S6A and S6B). The UMAPs showed slightly higher correlation between map distance and intermodular or gene-gene correlations than the t-SNE maps (Figure S6). We empirically chose the UMAP with *kME* as inputs to be our interactive 2D visualization tool and termed this tool the striatal CoExMap UMAP (Figure 1C). The CoExMap UMAP shows four distinct gene clusters we numbered 1–4 in decreasing order of the number of genes in the cluster (Figure 1C). Clusters 3 and 4 consist of

(B) Correlation heatmap of module eigengenes of the 38 modules identified by WGCNA. Heatmap color and numbers indicate correlations of module eigengenes. (C) UMAP representation of the co-expression network of individual genes with top hub genes in each module indicated by larger circles colored by the module color used in (A) and (B). Positive module eigengene correlations >0.5 are indicated by blue lines connecting the hub genes. See also Figures S1–S7 and Table S1 and S2.



(legend on next page)

M16 and M12, respectively, and are well separated from clusters 1 (25 modules) and 2 (7 modules). Overall, CoExMap UMAP reflects the intramodular connectivity of the vast majority of the modules (31 of the 38 modules), and the hub genes (*kME* > 0.5) for each module are tightly clustered in specific locations on the UMAP (Table S2 and Figure S7).

Interpretation of the striatal CoExMap—modules with significant enrichment of cell class and cell-type marker genes

We next annotated our 38 striatal modules using striatal cell-type-specific marker genes (Figure 2A and Table S3; STAR Methods) (Saunders et al., 2018; Gokce et al., 2016; Doyle et al., 2008; Heiman et al., 2008) and other glial cell marker genes (Hammond et al., 2019; Keren-Shaul et al., 2017). Six modules are enriched in markers of striatal medium spiny neurons (MSNs), which account for about 95% of the striatal neurons. MSNs can be divided into direct- and indirect-pathway MSNs (or D1-MSNs and D2-MSNs, respectively) based on their differential axonal projections and gene expression (Kreitzer and Malenka, 2008; Lobo et al., 2006; Heiman et al., 2008; Gokce et al., 2016; Saunders et al., 2018). Overall, we found modules enriched in marker genes for D2-MSN, D1-MSN, interneurons, oligodendrocytes, astrocytes, microglia, endothelial cells, ependymal cells, and neural stem cells. Importantly, our study reveals multiple modules containing marker genes from more than one cell type (Figure 2A). For example, M4 is significantly enriched in markers of parvalbumin interneurons and astrocytes (Figure S8B), M11 in signatures of astrocytes and microglia, M12 in marker genes for microglia and endothelial cells, and M16 in genes for ependymal and neurogenic niche cells (Figure 2A). Together, our study reveals a set of modules that may represent coordinated marker gene expression across different striatal cell types.

Striatal gene coexpression modules with enrichment in distinct protein complexes

We hypothesized that some of the striatal coexpression modules may be enriched in known protein complexes. Using stringent criteria of $p < 1.0E-10$ and at least 30 interaction edges, we found 11 modules that showed significant enrichment in known protein-protein interactions and molecular pathways in the STRING database (Szklarczyk et al., 2019) and identified their putative upstream regulators using ingenuity pathway analysis (IPA) (Figure S8C and Table S3). Interestingly, four modules are enriched in terms related to mitochondria. M17 contains 33 ribosomal proteins as well as 12 mitochondrial inner membrane proteins (gene ontology or GO), and its predicted upstream regulator is rictor (mTOR pathway) (Figures 2B and S8C). Another module, M16, is enriched in genes related to cilium and microtubule-based movement (Figure S8C). M12 is enriched in innate immunity and interferon signaling genes (Figure 2D), and its predicted upstream regulators, interferon alpha, Stat1, and Irf7, are also M12 module genes (Figure S8C). M5 is particularly interesting in the context of Parkinson's disease (PD) (Figures 2C and S8C). First, it is significantly enriched in familial PD genes (i.e., *Pink1*, *Park7/DJ-1*, *Atp13a2*). Second, it is enriched in genes in Pink1/Parkin-mediated mitophagy (e.g., *Mfn2*, *Sqstm1*, *Tomm40*) and other mitophagy/autophagy genes (e.g., *Samm50*, *Fis1*, *Amfr*, *Fkbp8*, *Atp6v0b*, *Usp20*, *Atg101*, *Atg4d*, *Map1lc3a*). Third, it is highly enriched in GO terms mitochondrion, oxidative phosphorylation, and ubiquitin-mediated proteolysis, which are relevant to PD pathophysiology. The predicted IPA upstream regulator of M5 is the mTOR pathway (i.e., Rictor and Stk11/Lkb1) (Figure S8C). Thus, our unbiased striatal transcriptomic mapping reveals a novel striatal coexpression module that connects the Pink1-Parkin mitophagy pathway to mitochondrial oxidative phosphorylation and ubiquitin-mediated proteolysis and provides novel candidate genes to study these PD-relevant pathways.

Striatal gene coexpression modules enriched in diurnally expressed genes

Only 17 out of 38 striatal modules have robust enrichment in cell-type markers or protein complex genes. We next sought to identify new biological drivers for the remaining unannotated striatal coexpression modules. Two clues suggest that these modules may be related to specific cellular states. First, M42 and M55 modules have numerous immediate-early genes as hub genes (e.g., *Egr1*, *Nr4a3*, *Fosl2*, and *Egr4* in M42 and *Junb*, *Fos*, *Fosb*, *Arc* in M55) and are enriched in neuronal- and synaptic-activity-regulated genes (Figure S9) (Kim et al., 2010; Zhang et al., 2007). Moreover, M58 is significantly enriched in circadian genes (Table S3) and contains several core circadian transcription factors (i.e., *Cry2*, *Dbp*, *Bhlhe40*, *Bhlhe41*, and *Nr1d1*) (Table S3; Takahashi, 2017).

To examine a possible relationship between the striatal CoExMap modules and daily rhythmically transcribed genes, we generated an RNA-seq dataset of adult mouse at 6 m of age and defined between 971 and 3,272 daily rhythmically expressed genes that are significantly upregulated at each of the 8 ZT timepoints around the 24-h time window (FDR < 0.05; Figures 3A and 3B; Table S4; STAR Methods). We validated our analysis by confirming the expected circadian phases of core clock-related genes (Figure S10A) (Takahashi, 2017) and showed our striatal diurnally expressed genes are highly correlated with those reported for the hypothalamus (Figure S10B) (Zhang et al., 2014a). Finally, we compared our analysis of daily rhythmic genes using linear models with the JTK_cyle method (STAR Methods) (Hughes et al., 2010) and found a highly positive correlation for time of maximum significance ($r = 0.98$; Figure S10C) and significance of association with diurnal time ($r = 0.86$; Figure S10D).

Enrichment analyses showed that 18 striatal CoExMap modules are strongly ($p < 10^{-8}$) enriched in experimentally defined

Figure 2. Enrichment of cell class and cell-type marker genes in the striatal coexpression modules

(A) Enrichment of CoExMap modules in selected cell-type marker sets. Heatmap color represents enrichment significance ($-\log_{10}$ of the hypergeometric test p value). Numbers of overlapping genes are shown in those cells where the overlap p value is $< 10^{-8}$.

(B–D) STRING database protein-protein interaction networks of genes in modules M17 (B), M5 (C), and M12 (D) together with their main functional annotations. See also Figures S8 and Table S3.

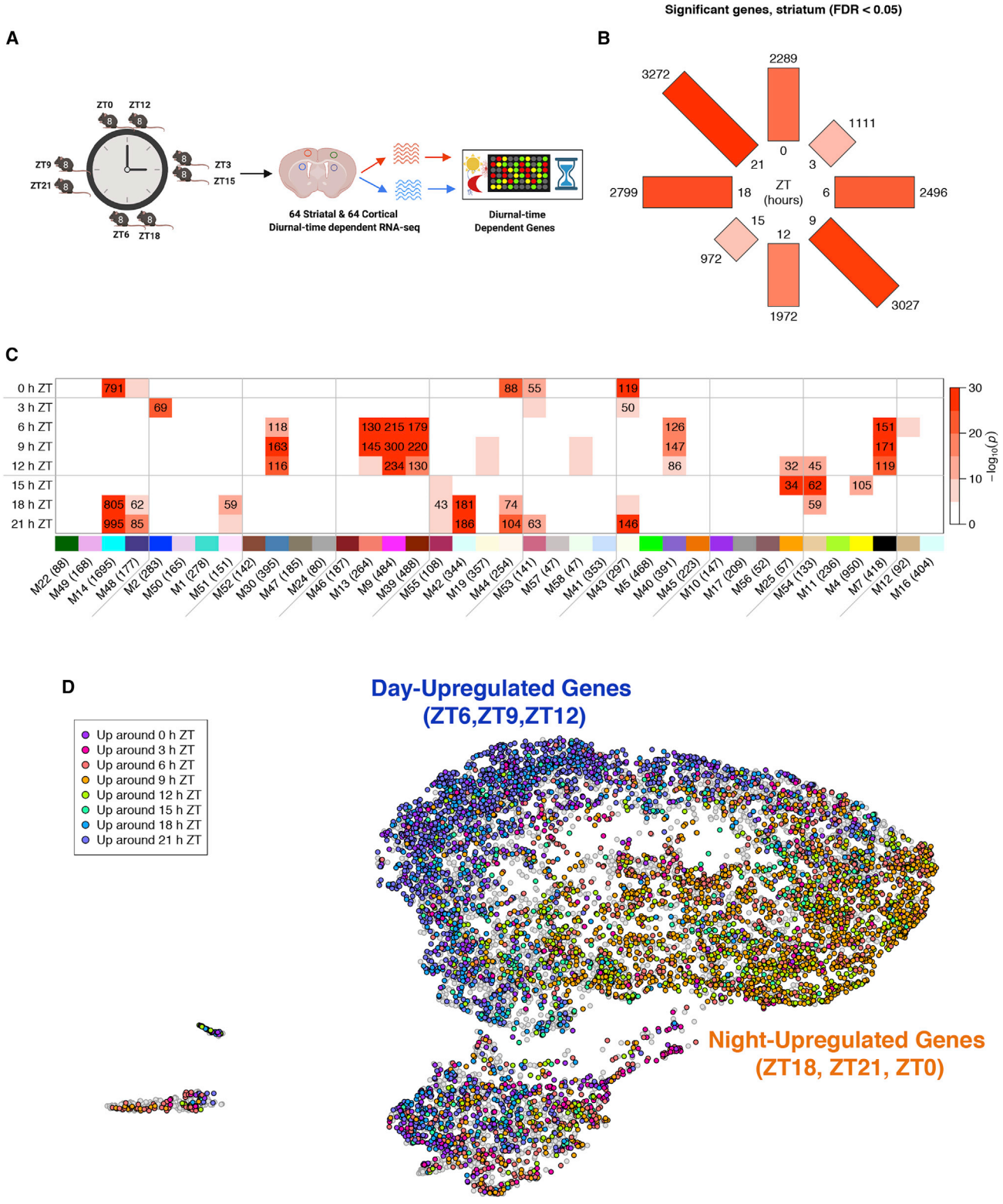


Figure 3. Striatal gene coexpression modules are highly enriched in diurnally expressed genes

(A) Schematic diagram of the diurnal transcriptome variation study.

(B) Numbers of genes significantly (FDR < 0.05) upregulated around each of the 8 time points (STAR Methods).

(legend continued on next page)

diurnal genes; of these, 11 and 7 modules are enriched in genes upregulated during the nighttime and daytime, respectively (Figure 3C). Among the 18 modules, six modules also have prior cell-class/cell-type or molecular complex annotations (Figures 2 and S8). These 18 modules (referred to as ''diurnal modules'') contain a total of 4,651 genes that diurnally vary at FDR < 0.05. Importantly, M14, the largest striatal CoExMap module with 1,695 genes, has strikingly high enrichment in genes that peak toward the end of the night (965 genes at ZT21, FDR = 1.90E−310). Similarly, M9 (300 genes at ZT9, FDR = 6.70E−112) and M39 (220 genes at ZT9, FDR = 1.70E−42) have robust enrichment in genes that peak during the day. We next mapped the diurnal genes onto the striatal CoExMap UMAP and found a striking dichotomous distribution of genes that peak during the daytime versus those that peak during the nighttime in opposite domains of the UMAP (Figures 3D and S11).

Striatal CoExMap diurnal modules show distinct daily temporal kinetics of gene expression levels and biological enrichment

We next asked what the distinct daily temporal kinetics of the CoExMap diurnal modules are and what biological functions are suggested by their enrichment. We focused on 16 of the 18 strongly diurnally enriched CoExMap modules whose module representative in the diurnal data are also strongly varying (STAR Methods). We clustered these modules based on temporal patterns of their module representatives to define five distinct diurnal kinetic patterns, which included two daytime patterns (DP1 and DP2) with 1 and 6 modules each, and three nighttime patterns (NP1–NP3) with 2–4 modules (Figure 4A). As expected, the top hub genes of the diurnal modules exhibited the shared diurnal expression patterns (Figure S12; Table S3).

We next examined the functional enrichment of the diurnal modules in each day or night pattern (Figure 4B; Table S4). DP1 contains only the M2 module that peaks at ZT3 and is enriched in D2-MSN and common MSN genes (Figure 2), dopamine-DARPP32 signaling, and synapses and synaptogenesis (Figure 4B), revealing a coordinated upregulation of genes involved in MSN-specific function and synaptogenesis at the onset of daytime/sleep phase.

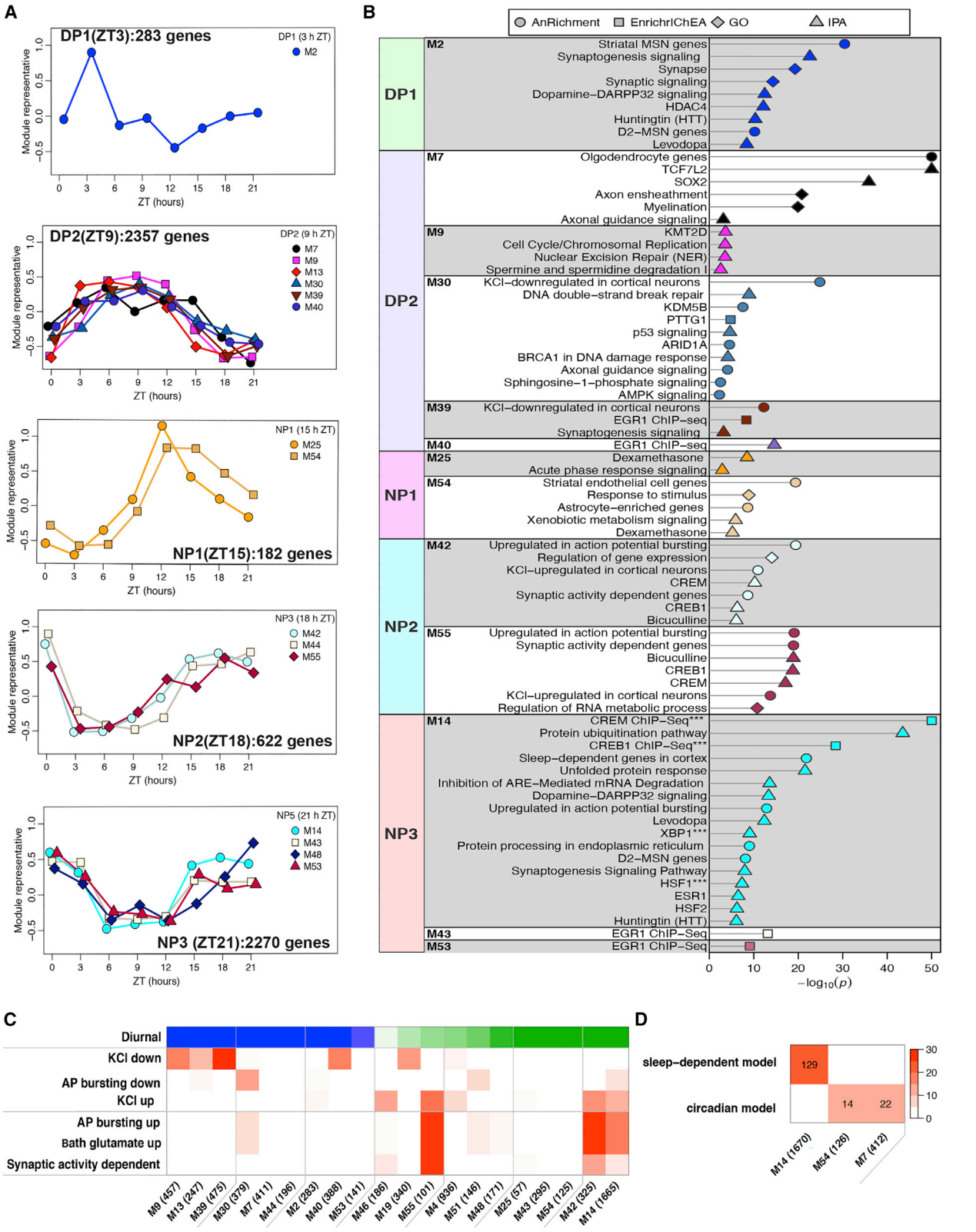
DP2 encompasses six modules with 2,357 genes and has a broad peak during the daytime that plateaus around ZT9 (Figure 4A). The M7 is highly enriched in oligodendrocyte genes and myelination and has TCF7L2 and SOX2 (known regulator of oligodendrocyte differentiation) as its IPA upstream regulators (Figure 4B). This result is consistent with evidence that oligodendrocyte genes are regulated by the circadian clock or sleep/wake cycle (Artiushin and Sehgal, 2020; Bellesi et al., 2013). Interestingly, M9 and M30 in DP2 are significantly enriched in DNA repair genes (Figure 4B; Table S3). M9 is enriched in genes involved in nuclear excision repair (*Haf1a*, *Ercc3*, *Gtf2h5*, *Pola2*, *Rpa2*, *Ube2i*, *Uvssa*, and *Xpc*) and other DNA repair genes (*Brca1*, *Brca2*, *Xrcc6*) (Figure S12 and Tables S2 and S3). It also has an enrichment in cell-cycle control of chromosomal replication (Figure 4B), indicating that it may be related to cell proliferation. M30 is significantly enriched in genes involved in DNA double-strand break (DSB) repair, P53 signaling, and BRCA1 in DNA damage response, including as its members *Atm*, *Atr*, *Prkdc*, *Rad50, Wrn*, *Msh3*, *Pms2*, and *Fancm* (Figure 4B; Tables S2 and S3). These results are consistent with evidence showing elevated DNA damage in the mouse brains with increased locomotor activities during the night (Suberbielle et al., 2013) and association of DNA repair function with sleep (Mourrain and Wang, 2019; Mure et al., 2018; Zada et al., 2021). Another feature of DP2 modules is their enrichment in genes involved in metabolism, such as spermine and spermidine degradation I and Leucine degradation in M9; fatty acid oxidation in M13; pyridoxal 5′-phosphate salvage pathway and AMP-activated protein kinase (AMPK) signaling for M30; and superpathway of cholesterol biosynthesis and AMPK signaling in M39 (Figure 4B; Table S3). Such findings are consistent with the circadian control of metabolism *in vivo* (Panda, 2016; Rijo-Ferreira and Takahashi, 2019). Finally, the top hub gene of the M9 module, *Cirbp* (Figure S12), is a cold-induced RNA binding protein that regulates expression of several clock genes as well as rapid-eye-movement (REM) sleep (Liu et al., 2013).

We next examined the enrichment terms of NP modules (Figure 4B; Tables S4 and S5). NP1 contains two modules (M25 and M54) that have a sharp peak at ZT12, a transition time from sleep to wake for the mice. Interestingly, these two modules contain genes that are regulated by the glucocorticoid hormone, as evidenced by the fact that their predicted upstream regulator is dexamethasone (Figure 4C). Consistent with this interpretation, the top hub gene of M25, *Sgk1* (Figure S12), is a glucocorticoid-responsive gene that is upregulated by stress and at the beginning of the rodent's active period at night (Hinds et al., 2017; Hor et al., 2019; Logan and McClung, 2019). Moreover, M54 is enriched in endothelial cell marker genes and genes involved in xenobiotic-metabolism signaling (Figures 2 and 4B).

NP2 contains three modules with a broad plateau between ZT15 and ZT0 that include genes known to be upregulated by neuronal activity and glutamate or GABA receptor agonists or antagonists (Figures 4B and 4C; Table S3; Kim et al., 2010; Zhang et al., 2007). M42 and M55 have predicted upstream regulators CREM and CREB1 (IPA, Figure 4B), which are critical for regulating activity- and circadian entrainment-dependent gene expression (Lonze and Ginty, 2002). Interestingly, the top hub gene of M55, *Sik1*, is essential for circadian clock entrainment (Figure S12; Table S5) (Jagannath et al., 2013). Consistent with the observations that genes upregulated by neuronal activities are enriched in the modules that peaked during the night, we also found that four DP2 modules that peaked during the daytime are significantly enriched in KCl-downregulated genes (Figures 4B and 4C; Table S5; e.g., Kim et al., 2010; Zhang et al., 2007).

(C) Overlaps of CoExMap modules (columns) with sets of genes upregulated around each of the 8 time points (rows). In the heatmap, the color represents hypergeometric test enrichment significance and numbers give the overlap sizes for cells for which the enrichment p value is less than 10^{-8}. Modules (columns) are ordered in the order of modules in hierarchical clustering in Figure 1B.

(D) UMAP representation of the network with genes significantly upregulated at one of the 8 time points indicated by a non-gray color. See also Figures S9–S11 and Table S4.



(legend on next page)

Finally, NP3 contains four modules that peak between ZT21 and ZT0. M14 is highly significantly enriched in protein-ubiquitination pathways, unfolded protein response (endoplasmic reticulum [ER] stress), dopamine-DARPP32 signaling, and protein processing in the ER (Figure 4C; Table S3). The upstream regulators of these pathways, e.g., *Creb1*, *Crem*, *Hsf1*, and *Xbp1*, are all M14 module genes. M14 module genes may have potential connections to sleep-related genes. First, M14 is significantly enriched in sleep-/wake-driven genes ($p = 2.5 \times 10^{-22}$) (Table S5; Figure 4D) (Hor et al., 2019). Second, its #3 hub gene, *Homer1a*, is a synaptic gene correlated with sleep loss (Figure S12) (Maret et al., 2007) and plays a critical role in sleep-dependent synaptic scaling (Diering et al., 2017). Third, M14 also contains other genes central to circadian clock regulation as well as sleep regulation (e.g., *Per2*, *Cry1*, *Abcc9*, *Csnk1a1* and *Sik3*) (Wang et al., 2018). Intriguingly, the #2 hub gene of M14, *Spred1*, a suppressor of Ras signaling (Wakioka et al., 2001), also appears to be a sleep-related gene in sparrows (Jones et al., 2008). Together, our results suggest an important role for M14 genes in nocturnal brain function and possibly in sleep/wake cycle regulation.

Gene coexpression network in the cortex is primarily driven by diurnal time

We next asked whether such diurnal time-driven gene coexpression is a general feature of transcriptomes from other brain regions. As a proof of concept, we performed WGCNA analysis of 100 cortical RNA-seq samples and identified 41 coexpression modules (Figure S13; Table S6). Importantly, 32 cortical modules showed significant overlap with the striatal modules (Figure S14A). Additionally, 9 cortical modules were strongly enriched in cell class or cell-type marker genes, including 4 modules in neuronal genes, one in oligodendrocyte, one in microglia, one in mixed astrocyte and microglia, and two in endothelial cells (Figure S14B).

We next generated and analyzed a cortical diurnal transcriptomic dataset to define genes that peaked at 8 different diurnal timepoints (Figure S15A). We found 23 cortical modules that are significantly enriched in cortical diurnally transcribed genes (Table S6). We created cortical CoExMap UMAPs and found that cortical diurnal genes that peaked at daytime or nighttime were located in the opposite domains on the UMAP, similar to our findings in the striatum (Figures 5A, S15B, and S15C). We defined the cortical diurnal modules based on their temporal kinetics and defined two day patterns and three night patterns (Figure 5B). By overlapping these modules with the diurnally varying modules from the striatum, we found strong overlap of modules that peaked in similar time windows during the day or night (Figure 5B), confirming the conservation of diurnal gene coexpression networks across two major brain regions.

Enrichment analysis of the cortical diurnal modules showed similar enrichment terms to those in the striatum (Table S6). For example, the cortical NP3 modules highly overlapped with the striatal M14 (FDR = 1.20E−210) and were enriched in the protein-ubiquitination pathway, AMPK signaling, and with upstream regulators of Crem, Creb1, bicuculline, and levodopa (Table S6). Moreover, the cortical DP1 and DP2 modules (5 in total) peaked at ZT6–ZT12 and significantly overlapped with striatal DP1 (M2) and DP2 modules (M9, M13, M39, and M40) (Figure 5B). These daytime-upregulated modules are enriched in genes involved in DNA repair, cyclins and cell-cycle regulation, insulin secretion-signaling pathways, superpathways of cholesterol biosynthesis, synaptogenesis signaling, and autophagy (Table S6). Thus, the core biological features of diurnal gene modules are also preserved across the two brain regions.

While we have thus far shown the similarities between the striatal and cortical CoExMap modules in terms of gene overlap and diurnal time-related genes, there are substantial tissue-specific features in gene coexpression modules across brain regions (Figures S14A and S16). Thus, our striatal and cortical CoExMaps are also valuable to study genes and modules that have tissue-specific cellular, molecular, and diurnal functions.

Because we had both striatal and cortical RNA-seq samples from 89 mice, we next explored whether the module eigengene expression values of diurnally varying modules (e.g., striatal M9, M13, or M14) in the two brain regions of the same mouse would correlate across our entire dataset. Our analyses confirmed that the eigengene expression levels for M9, M13, and M14 were highly positively correlated between the cortical and striatal datasets for the 89 mice (Figure 5C). These correlations support the notion that the transcriptome of each mouse, regardless of brain regions, carries shared "diurnal time-stamped" gene signatures.

Application of striatal CoExMap to inform gene signatures in Huntington's disease

We have annotated the majority of the striatal CoExMap modules (36/38) based on their enrichment with cell class or cell-type marker genes, molecular complexes, and diurnal time-dependent genes (Figure S17A). Moreover, we have mapped the modules with such annotations onto the striatal CoExMap UMAP clusters 1–4, providing a readily memorizable and visualizable association between the module annotations and their geographic locations on the UMAP (Figure S17B). We next tested the idea that the striatal CoExMap and its UMAP could be useful to interpret gene signatures based on the latter's enrichment patterns in our annotated modules and based on the locations of the gene signatures on the CoExMap UMAP.

Figure 4. Biological features of coexpression module-derived diurnal gene networks

(A) Module representatives of CoExMap modules in diurnal expression data, organized by day and night patterns (DP and NP, respectively). Each point represents the mean representative expression at the corresponding time.

(B) Enrichment of modules belonging to each DP and NP pattern in selected terms. The x axis represents the hypergeometric test p-value.

(C) Hypergeometric test enrichment significance of CoExMap modules (columns) in diurnally DE genes (blue color represents enrichment in gene sets up during the day, green color represents enrichment in gene sets up during the night) and neuronal activity-related sets (rows).

(D) Hypergeometric test enrichment significance of CoExMap modules in genes whose expression was best fit by sleep-dependent and circadian models (Hor et al., 2019). See also Figures S12 and Table S5.

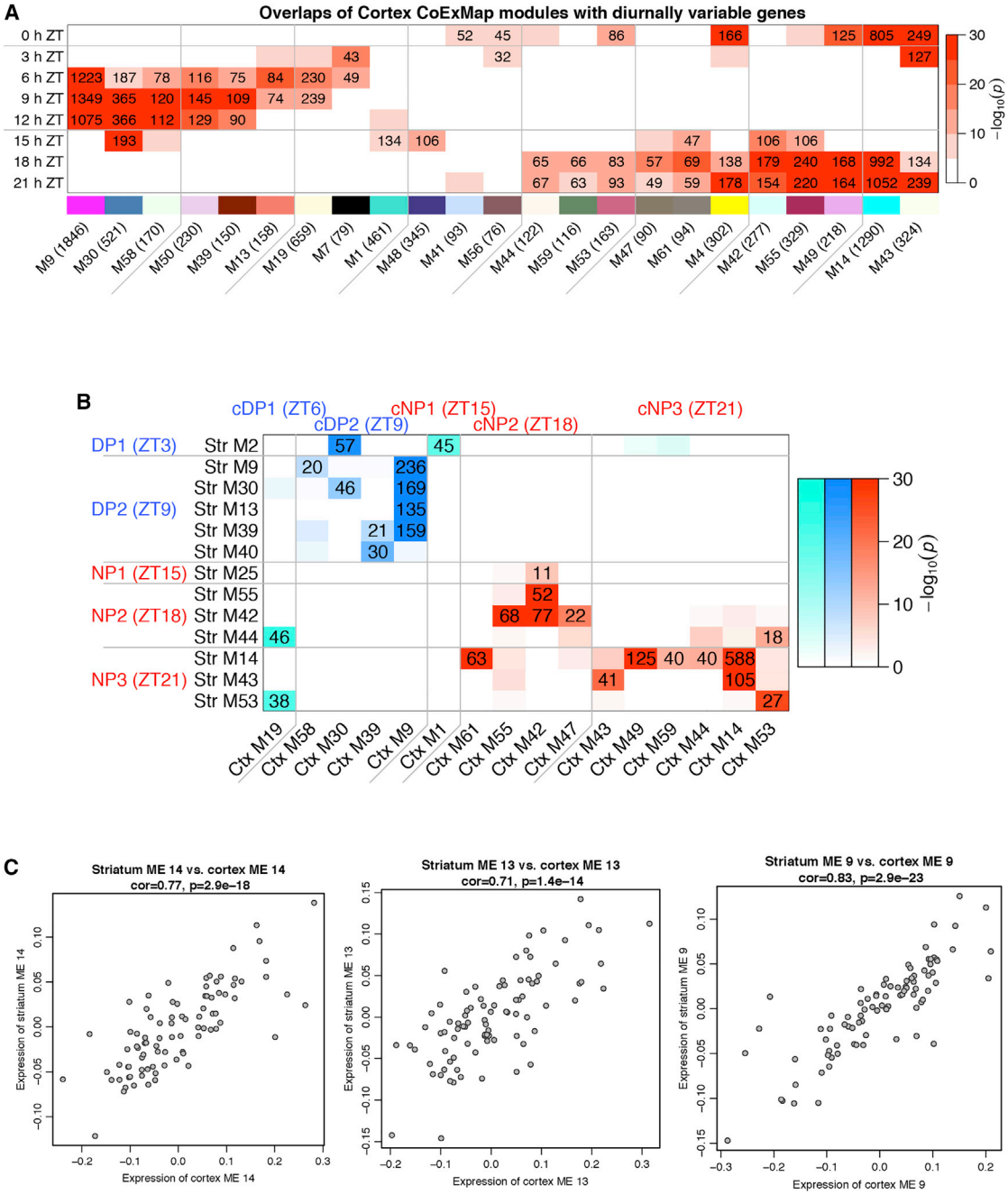


Figure 5. Gene coexpression network in the cortex is primarily driven by diurnal time and synchronized with the striatum

(A) Overlaps of diurnally enriched cortex modules (columns) with genes upregulated at specific times. Color represents hypergeometric test significance and numbers the overlap sizes.

(B) Overlaps of diurnally enriched cortex (columns) and striatum (rows) CoExMap modules. Blue (red) color indicates overlaps of modules peaking at night (day) in both regions; turquoise color indicates overlaps of modules that peak at night in one region and at day in the other.

(C) Scatterplots of module eigengenes of cortex and striatum modules M14, M13, and M9. Each point represents one of the 89 animals common to both the striatum and cortex datasets. See also Figures S13–S16 and Table S6. Title of each plot indicates the correlation and the corresponding student p value (n=89).

We first tested this concept using HD-related gene signatures. HD is a dominantly inherited neurodegenerative disorder caused by CAG repeat expansion in the *Huntingtin* (*HTT*) gene, and the repeat length is inversely correlated with the age at disease onset (Bates et al., 2015). A hallmark of HD neuropathology is the selective degeneration of striatal MSNs (Bates et al., 2015). Using an allelic series of murine *Huntingtin* (*Htt*) knockin mice, we previously defined *Htt* CAG repeat length-dependent

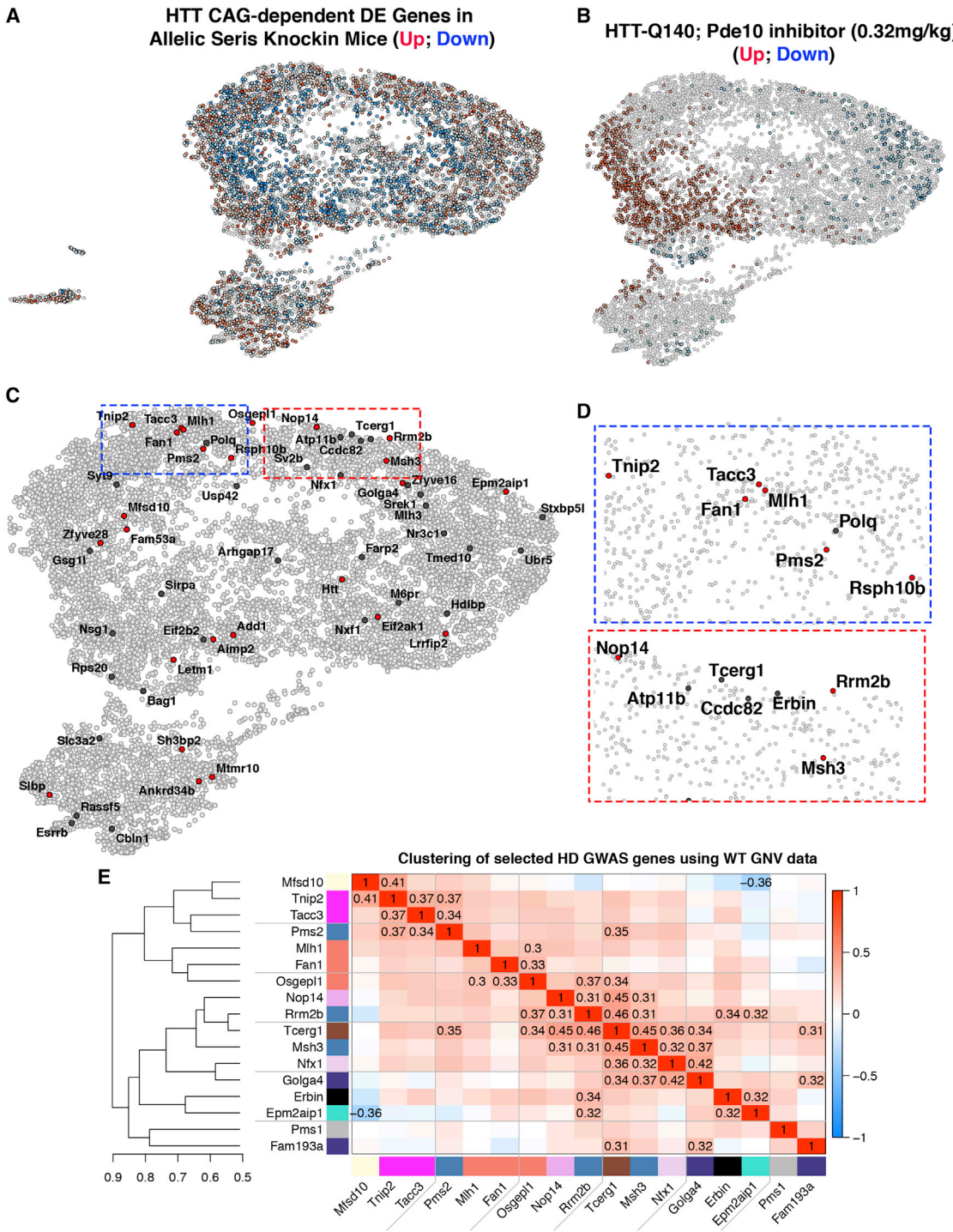


Figure 6. Application of CoExMap to inform genetic, transcriptomic, and pharmacological gene signatures in Huntington's disease

(A) CoExMap UMAP with genes significantly (FDR < 0.05) down- and upregulated with CAG length marked in blue and red, respectively.

(B) CoExMap UMAP with genes significantly (FDR < 0.05) down- and upregulated by treatment with 0.32 mg/kg of a Pde10a inhibitor in Q175 homozygous mice marked in blue and red, respectively.

(C) CoExMap UMAP with genes nearest to loci identified in Huntington's disease GWAS at $p < 10^{-5}$ ($p < 10^{-8}$) marked in black (red).

(D) Detailed views of the blue and red rectangle areas in (C).

(legend continued on next page)

dysregulated genes (Langfelder et al., 2016). We first examined the enrichment of m*Htt* CAG repeat length-dependent genes in our striatal CoExMap modules (Figure S18A). We found that downregulated genes are enriched in modules with MSN markers (M2, M45, M57), diurnal modules that peaked during the daytime (M40) or nighttime (M43, M14), and two other modules (M19 and M41). On the CoExMap UMAP, the downregulated genes appeared to be concentrated in the center of cluster 1, while upregulated genes were concentrated on the periphery of clusters 1, 2, and 3 (Figure 6A). The latter may be related to the upregulation of oligodendrocyte and astrocyte genes (Langfelder et al., 2016) as well as increased neurogenic genes in HD models (Kandasamy and Aigner, 2018).

We next asked whether the striatal CoExMap and its UMAP can be used to interpret gene signatures from pharmacological interventions in the HD mouse models. As a proof of concept, we mapped the striatal DE genes in the striatum of zQ175 knockin mice treated with three different doses of PDE10 inhibitor (PDEi), which partially rescues disease phenotypes in HD mice (Beaumont et al., 2016). Interestingly, we found that two doses of PDE10i (0.32 and 1 mg/kg), but not a third dose (3.2 mg/kg), reproducibly increased gene expression in a subset of modules downregulated by m*Htt* CAG expansion (Figure S18B, Table S7). Such partial rescue is readily visualizable on the CoExMap UMAP, as the DE genes in the PDE10i-treated HD mice only upregulated genes in the lower-left quadrant and downregulated genes on the upper-right quadrant of UMAP cluster 1, representing only a partial reversal of the dysregulated genes in HD knockin mice (Figures 6A and 6B; Figure S18).

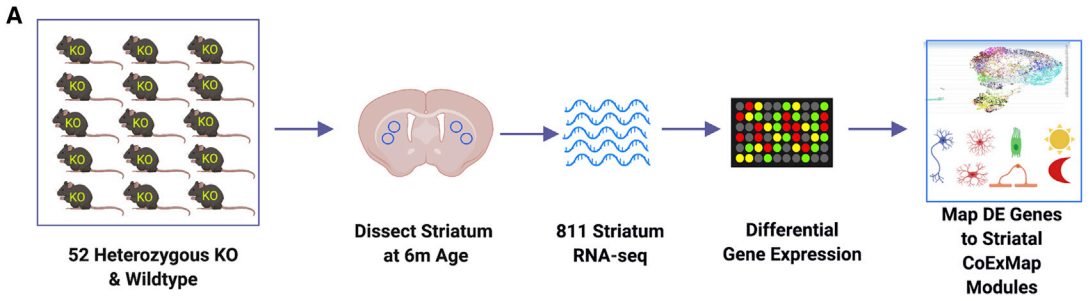
A third application of CoExMap UMAP in the context of HD is to examine the map location for genes linked to genomic loci identified in genome-wide association studies (GWAS) for modifiers of HD age of onset (Genetic Modifiers of Huntington's Disease Consortium, 2019). The prior studies revealed that HD GWAS loci genes are significantly enriched in DNA repair genes, and a subset of these genes (i.e., *MSH3*, *MSH2*, *MLH1*, *FAN1*) are known to regulate somatic instability of *mHTT* CAG repeat length (Wheeler and Dion, 2021; Goold et al., 2019; Manley et al., 1999). We found that M30, a module upregulated during daytime and sleep in mice (Figure 3C), is significantly enriched in murine homologs of HD GWAS genes (FDR = 0.0116), including as its member genes *Msh3*, *Pms2*, *Rrm2b*, and *Ccdc82*. We next mapped murine homologs of 75 HD GWAS modifier genes ($p < 1E{-}5$) (Genetic Modifiers of Huntington's Disease Consortium, 2019) onto the striatal CoExMap UMAP and found two regions of interest near the top of UMAP cluster 1, corresponding to M9, M13, and M30 territories (Figure 6C). Close examination of the first region showed proximity on the map for *Fan1*, *Mlh1*, *Tacc3*, *Pms2*, and *Polq*, and a second region showed proximity on the map for *Msh3*, *Rrm2b*, *Tcerg1*, *Ccdc82*, *Erbin*, and *Atp11b* (Figures 6C and 6D). Since the UMAP distance shows positive association with gene-gene correlation in our CoExMap modules (Figures 1C and S6), we next directly examined the correlation of these HD GWAS-related genes in our striatal transcriptomic dataset (Figure 6E). We confirmed significant positive correlation in gene expression in three groups of genes: *Pms2*, *Tacc3*, and *Tnip2* in the first group; *Mlh1*, *Fan1*, and *Osgepl1* in the second group; and *Msh3*, *Rrm2b*, and *Tcerg1* in the third group. Thus, our study demonstrates that a subset of HD GWAS-related DNA repair genes are significantly enriched in M30, a module upregulated in the daytime. Furthermore, our study reveals several HD GWAS genes (e.g., *Ccdc82*, *Tcerg1*, *Rrm2b*, *Tacc3*), with previously unclear roles in HD modification, to have significantly correlated expression with known DNA repair/HD GWAS genes, providing the impetus to investigate their potential roles in DNA handling in the context of HD.

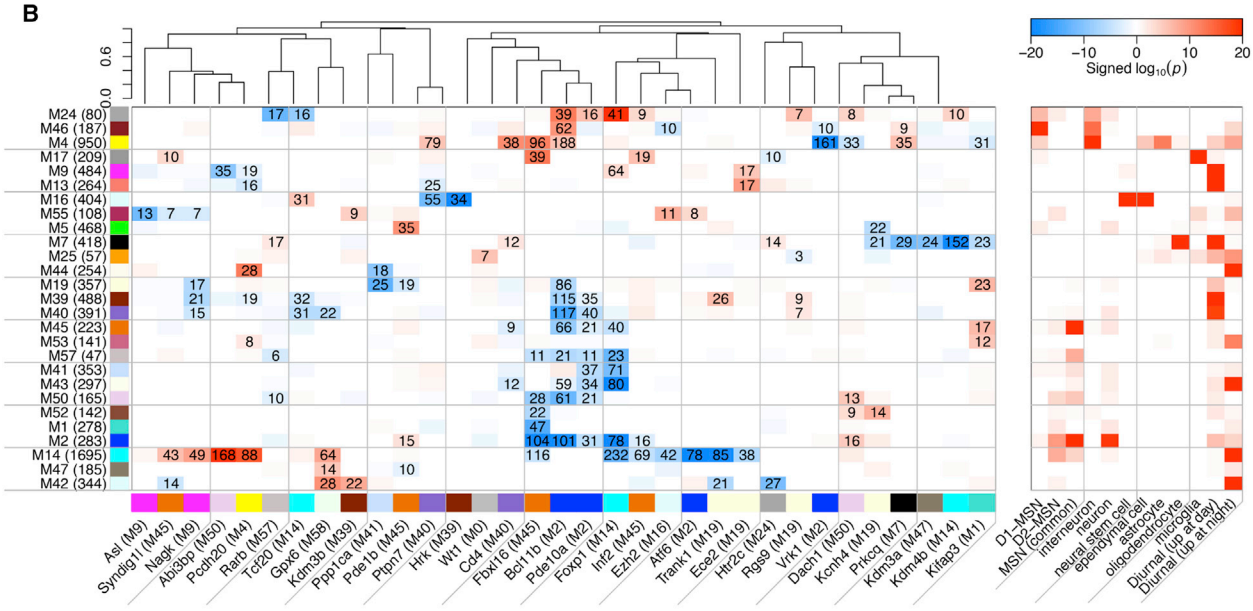
Application of striatal CoExMap to interpret differential gene expression signatures from 52 heterozygous knockout mice

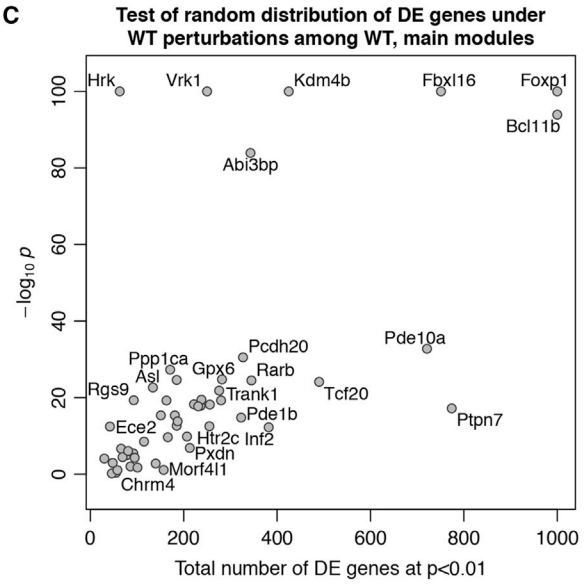
We next asked whether the striatal CoExMap could help interpret *in vivo* genetic perturbation gene signatures based on the annotated biology in the striatal CoExMap modules (Figures 1, 2, 3, 4, and S17). We tested this idea using transcriptomic datasets from striatal tissues dissected in the afternoons from KO-Het mice for 52 genes and their respective WT controls at 6 m of age (Figure 7A and Table S1; STAR Methods). Importantly, our study included 13 novel KO mouse lines (*Arpp19*, *Arpp21*, *Ddit4l*, *Ece2*, *Gpx6*, *Hrk*, *Kdm3b*, *Morf4l1*, *Pcdh20*, *Ppp1ca*, *Tcf20*, *Trank1*, *Zswim6*) created for this study (Table S1). We mapped the striatal DE genes from the 52 KO-Het transcriptomic study onto the CoExMap modules, using less stringent DE criterion of $p < 0.01$ to increase the statistical power to identify modules with significant DE gene overlap (Figure 7B). DE genes for 33 KO-Het mice had at least one significant enrichment ($p < 1E{-}5$) among the CoExMap modules (Figure 7B). Importantly, the majority of the KO-Het crosses (38 out of 52) had DE genes with a highly significant ($p < 1E{-}7$; chi-square test), non-random distribution of their DE genes among the CoExMap modules, with seven KO mouse lines having p values less than 1E−80 (Figure 7C and Table S8). This analysis suggests that the DE gene signatures from these KO-Het mice selectively affect certain striatal CoExMap modules.

We next assessed whether the distribution of KO-Het DE genes in the CoExMap modules could infer the biological impact of each genetic perturbation, using our annotated modules as a reference (Figures 7B and S17). Moreover, we hypothesized that such a CoExMap module-based DE gene mapping could help to cluster perturbations with similar gene signatures, hence connecting perturbations that may have similar molecular impacts in the striatum *in vivo*. The analysis resulted in several clusters of perturbations based on DE gene/module overlaps (Figure 8A and Table S9). The first group consisted of KO-Het of *Fbxl16*, *Bcl11b*, *Foxp1*, and *Pde10a*, all known MSN marker genes (Heiman et al., 2008), and they appeared to downregulate genes in modules enriched in MSN markers (e.g., M2, M45, and M57) (Figures 2 and 7B). The second cluster of KO-Het mice

(E) Heatmap representation of the correlation network of the largest cluster of HD GWAS genes. Specifically, we clustered the genes marked black or red in (A) using average linkage hierarchical clustering of correlation-based dissimilarity and determined clusters using constant-height tree cut at height 0.93 (corresponding to correlation 0.07). We retained the largest cluster for the plot shown here. Heatmap represents correlation; correlations with absolute values above 0.3 are shown explicitly. See also Figures S17 and S18 and Table S7.







(legend on next page)

(*Kdm4b*, *Prkcq*, *Kdm3a*, *Kifap3*, and *Kcnh4*) downregulated genes in M7, which is an oligodendrocyte gene-enriched module (Figures 2 and 7B). Additionally, two large KO-Het clusters either upregulated (*Syndig1l*, *Nagk*, *Abi3 bp*, *Pcdh20*, *Gpx6*) or downregulated (*Fbxl16*, *Foxp1*, *Inf2*, *Ezh2*, *Atf6*, *Trank1*, *Ece2*) genes in M14, the largest diurnal module (Figures 3C and 7B). Furthermore, *Vrk1* KO-Het selectively downregulated M4 module genes, which are enriched in interneuron and astrocyte marker genes, while those from *Hrk* and *Ptpn7* KO-Het mice significantly downregulated genes in M16, a module enriched in ependymal cell and neurogenesis genes (Figures 2 and 7B). Finally, we directly verified that the majority of KO-Het mice with DE genes mapped to modules annotated with a certain aspect of biology, e.g., cell types or diurnal time, do indeed dysregulate cell marker or diurnal time-dependent genes (Figures 8A and 8B).

Functional validation of Kdm4b and Prkcq in regulating oligodendrocyte differentiation *in vitro*

We next tested whether interpretation of KO-Het DE gene signatures based on CoExMap could help reveal their functional roles in brain cell type or diurnal biology. We first evaluated *Kdm4b* and *Prkcq* for their roles in oligodendrocyte differentiation *in vitro* (see STAR Methods) because their KO-Het DE genes are mapped to M7 and are enriched in oligodendrocyte genes. *Prkcq* is known to be selectively expressed in oligodendrocytes (Zhang et al., 2014b), and *Kdm4b* is a ubiquitously expressed gene encoding a histone H3K9 demethylase. We found that small interfering RNA (siRNA)-mediated knockdown of *Prkcq* or *Kdm4b*, but not scrambled siRNA controls, significantly reduced the differentiation of oligodendrocyte precursor cells (OPCs) into mature oligodendrocytes based on cellular morphology and expression of mature oligodendrocyte marker genes (He et al., 2017; Figures 8C–8E). This study confirms that two KO-Het genes with perturbation signatures that mapped to the striatal CoExMap M7 module indeed have functional roles in regulating oligodendrocyte differentiation.

A novel role for bipolar disorder GWAS gene Trank1 in regulating anxiety-like and nocturnal locomotor behaviors

Our novel KO-het for *Trank1* significantly affects genes in the M14 module, the largest striatal CoExMap module that peaks during the night. Human *TRANK1* is located at one of the most replicated GWAS loci for bipolar disorder (BP) (Ikeda et al., 2018; Li et al., 2020; Muhleisen et al., 2014; Stahl et al., 2019). However, the biological function of *Trank1* in the brain and its pathogenic role in BP remain unknown. Based on the enrichment of *Trank1* KO-Het DE gene signature in M14, we hypothesized that *Trank1* KO-Het and homozygous *Trank1* KO may affect nighttime behaviors and those associated with BP (e. g., anxiety). The *Trank1* KO mice showed normal performance in motor performance (rotarod test) (Figure S19A) (Gray et al., 2008) and depression-like behaviors (forced swimming test) (Figure S19B) Wang et al., 2014). Interestingly, in the light-dark box test (anxiety-like behaviors), the *Trank1* homozygous knockout (KO-Homo) showed a significant deficit in the time spent in the dark chamber, an indication of reduction in anxiety-like behaviors (Figure S19C). Importantly, in monitoring locomotor activity for 10 days (Whittaker et al., 2018), we found that *Trank1* KO-Homo showed significant hypolocomotion at nighttime (Figures 8F and 8G and S19D). Together, our study revealed an essential function of *Trank1* in regulating daily rhythmic locomotor behaviors and anxiety-like behavior and also provided new insights into a potential pathobiological role for *Trank1* in BP.

Online CoExMap UMAP as an annotated searchable resource for connecting gene signatures to gene coexpression neighborhoods, networks, and annotated biology

To facilitate the access and exploration of our striatal CoExMap UMAP, we created a user-friendly web tool (through https://www.hdinhd.org/) that allows Google Map-like zoomable searches for genes or gene sets and displays of their distribution on the map, enabling the interpretation of any gene or gene set based on the annotated biology of distinct map locations (i.e., modules) and known function of neighboring genes that are likely to be coexpressed. This online resource also contains an interactive table that allows various rankings of the striatal DE gene signatures for the 52 KO-Het perturbations based on their striatal CoExMap module enrichment. We envision that the CoExMap and its interactive online UMAP could be broadly useful to interpret brain gene signatures from perturbations or diseases and to connect them to annotated molecular, cellular, and diurnal biology defined by their significantly overlapped CoExMap modules.

DISCUSSION

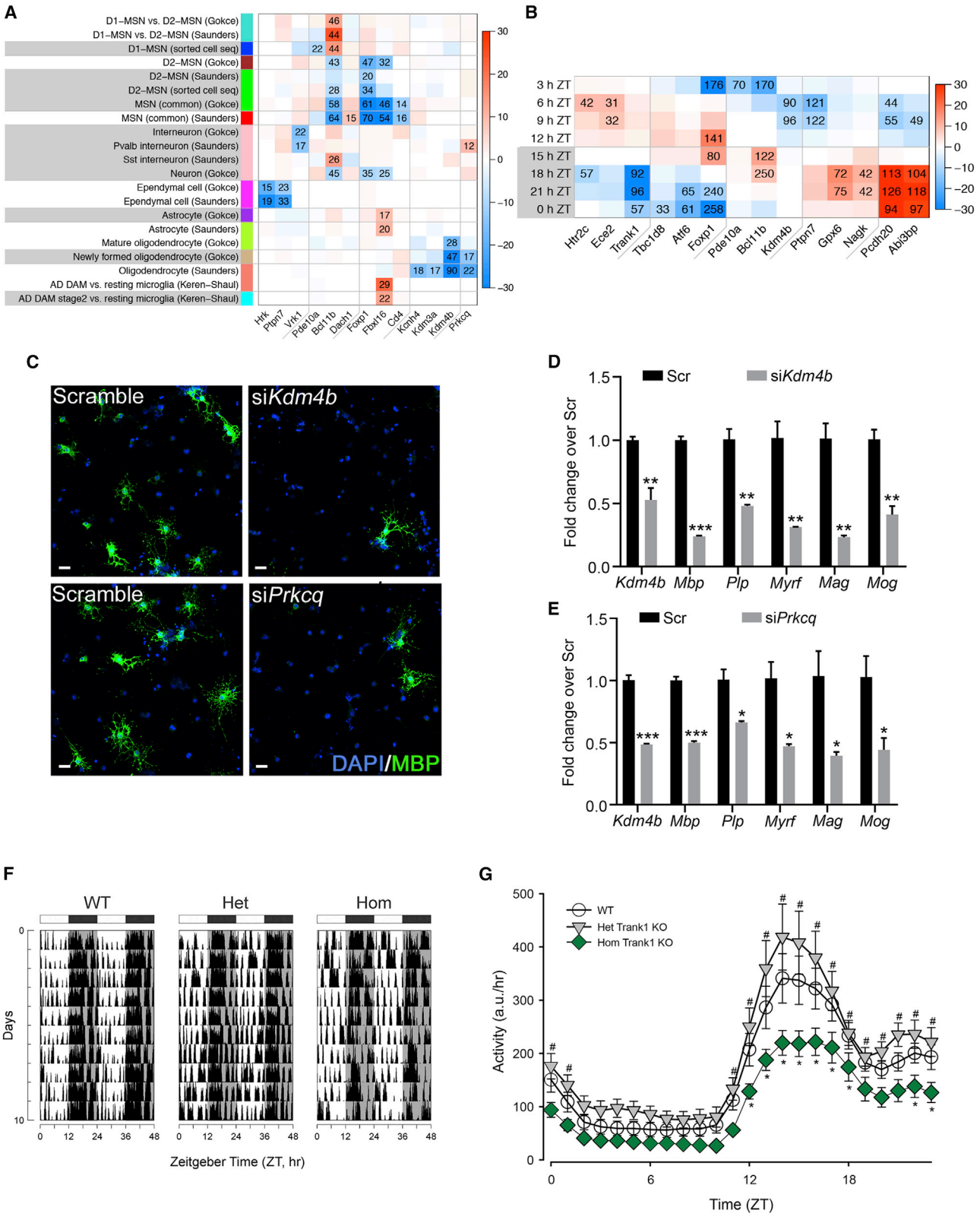
Our study concurs with prior findings that certain brain gene coexpression modules are enriched in markers of cell class or cell types (Oldham et al., 2008; Hawrylycz et al., 2012; Kelley et al.,

Figure 7. CoExMap-based interpretation of gene signatures from 52 heterozygous knockout mice

(A) Schematic diagram of perturbation study. WT controls and KO-Het mice for 52 genes were aged to 6 months and striata were sequenced. DE genes were identified and mapped onto CoExMap modules.

(B) Enrichment of modules in genes DE at $p < 0.01$ in KO-Het mice versus WT background (left heatmap) and in cell-type marker sets and diurnally variable genes (right heatmap). Each row corresponds to a striatum CoExMap module labeled by numeric label, number of genes, and module color. Each column corresponds to a heterozygous KO of the indicated gene, a cell type, or to diurnally variable genes. Color next to the gene symbol shows the striatum CoExMap module the gene is assigned to. In the left heatmap, color indicates "signed" enrichment significance ($-\log_{10}$ of the hypergeometric test enrichment p value, negative and positive for down- and upregulated genes, respectively). For each cell, we selected the direction (color) with the more significant enrichment p value. Only those modules and KO genes are shown that have at least one p value $< 10^{-5}$. Modules and perturbations are clustered using average linkage hierarchical clustering based on correlation of the signed significance. In the right heatmap, color represents $-\log_{10}$ of the hypergeometric test enrichment p value corrected for the number of marker sets for each cell type.

(C) Chi-square test of random distribution of DE genes for KO-Het perturbations versus WT controls among striatum CoExMap modules. The scatterplot shows the test significance as a function of the total number of genes DE at $p < 0.01$. See also Table S8.



(legend on next page)

2018). A novel finding here is that 12 out of 13 striatal modules are enriched in markers of more than one cell type. There are several possible explanations here. First, the cell types may be anatomically located in proximity and have fixed relative cell proportions (e.g., M16 module with neurogenic niche cells and ependymal cells). Second, the genes coexpressed in two different cell types may be co-regulated by the same upstream signals or represent a form of cell-cell communication. One module that may fit this mechanism is M4, which is highly enriched in both interneuron and astrocyte markers, which could be a result of interneuron-astrocyte signaling (Mederos et al., 2018; Shigetomi et al., 2011) or shared response to extracellular ATP signaling (Bowser and Khakh, 2004). Similarly, M11 contains both astrocyte and microglia marker genes and is preserved in the cortical network, and it may reflect crosstalk between these two cell types (Vainchtein and Molofsky, 2020). An intriguing finding is the mixed cell-type modules containing both the D1-MSN and D2-MSN marker genes (e.g., M2 and M45) (Figures 1B and 2A), which may be a result of M2, but not M45, being a diurnal module. Together, our study reveals that bulk tissue gene coexpression mapping may reveal co-regulated genes across cell types and provide molecular and cellular candidates to study cell-cell communications in the brain.

Our study demonstrates that certain molecular complexes are highly coexpressed in the mammalian brain. There are two possible sources for such molecular complex enrichment in coexpression modules. First, some of the protein complexes are selectively enriched in certain cell types that have specialized molecular complexes (e.g., synapse and potassium channel genes in M4 [interneurons], fatty acid metabolism and lysosome in M11 [astrocytes/microglia], innate immunity and interferon signaling in M12 [microglia], and cilia and microtubule proteins in M16 [ependymal cells]) (Figure S8C). However, there are several modules containing ubiquitously expressed molecular complexes, suggesting they are tightly coexpressed but at different levels across striatal samples. Examples of such modules are M17 (ribosome/mitochondria oxidative phosphorylation genes) and M5 (PD/mitophagy/autophagy/mitochondria genes). The molecular complexes in M5 are involved in two core functions that are highly relevant to PD (Pickrell and Youle, 2015), providing the impetus to study the M5 module hub genes for their roles in PD pathogenesis and therapeutics.

An important contribution of this study is the use of hundreds of daytime transcriptomic samples to define a full complement of diurnal gene coexpression modules for both the striatum and cortex in adult mice. An interesting question is how transcriptomes obtained from mice dissected during a restricted daytime window (1–5 pm) capture the full repertoire of 24-h diurnal gene networks embedded in the tissue. Our study showed a strong positive correlation of eigengenes of three diurnally variable modules between the striatal and cortical tissues, suggesting that the diurnal modules capture an expression "time stamp" that may vary with different kinetics for different diurnal modules within the narrow daytime dissection window. Additionally, our study provides a strong rationale for transcriptomic studies to be performed using tissues dissected from narrow time windows in the diurnal cycle in order to avoid the confounding effects of diurnally expressed genes. Together, our study provides the first experimental evidence to confirm the theory that diurnal time is a major driver of gene coexpression in mammalian tissues.

Our study provides a novel molecular framework to study the diurnal time-dependent gene expression and function in the mammalian brain. We first defined stable striatal and cortical coexpression modules and then used our experimental diurnal transcriptomic dataset to identify modules that peak or trough during the diurnal period. Using such an approach, our analysis is not limited by a pre-defined circadian or ultradian rhythmicity. Moreover, our approach can divide a large number of genes with similar daily periodicity into 2–6 coexpression modules enriched in biologically distinct terms. Perhaps the most important contribution of our work to the study of brain diurnal biology is the list of genes in each striatal and cortical diurnal module, ranked by their module hub gene status, and the biological pathways enriched for these modules. Together, they constitute unbiased entry points to study novel genes and gene networks related to diurnal time in the mammalian brain, an approach that can be readily expanded to study other mammalian tissues.

Some of the novel insights, based on the striatal and cortical diurnal modules, can be followed up to study the

Figure 8. Application of CoExMap to identify perturbations that disrupt cell-type signatures and diurnal rhythms

(A) Enrichment of cell-type marker sets in genes DE at $p < 0.01$ in heterozygous KO mice versus WT background. Each row corresponds to a cell-type marker set and each column to one dataset with heterozygous KO of the indicated gene. In the heatmap, color indicates "signed" enrichment significance ($-\log_{10}$ of the hypergeometric test p value, negative and positive in down- and upregulated genes, respectively). For each cell, we selected the direction (sign) with the more significant enrichment p value. Only those sets and KO genes are shown that have at least one p value $<10^{-8}$.

(B) Enrichment of diurnally DE gene sets in genes DE at $p < 0.01$ in heterozygous KO mice versus WT background. Each row corresponds to a set of genes upregulated (FDR < 0.05) around the indicated time. In the heatmap, color indicates "signed" enrichment significance.

(C) Immunofluorescence labeling for Mbp and DAPI in transfected rat oligodendrocytes with control siRNA versus siRNA targeting *Kdm4b* or *Prkcq*. Scale bar, 30 μm.

(D and E) qRT-PCR analyses of *Kdm4b*, *Mbp*, *Plp*, *Myrf*, *Mag*, and *Mog* in si*Kdm4b* (D) or si*Prkcq*-treated (E) rat oligodendrocytes. Data are presented as mean ± SEM.; n = 3 independent experiments; two-tailed unpaired Student's *t* test.

(F) Examples of cage activity rhythms recorded from WT, *Trank1* KO-Het, and KO-Homo mice. The activity levels in the actograms were normalized to the same scale (85% of the maximum of the most active individual). Each row represents two consecutive days, and the second day is repeated at the beginning of the next row. Gray shading in the actograms indicates lights off.

(G) Rhythms in cage activity in WT, *Trank1* KO-Het, and KO-Homo mice. Running averages (1 hr bin) of locomotor activity in different genotypes are plotted. ZT, Zeitgeber time. 24-h profiles of locomotor activity were analyzed using a two-way ANOVA with genotype and time as factors, followed by Holm-Sidak's multiple comparisons test, *p < 0.05. Data points are the mean ± SEM (WT: n = 10 animals; Het: n = 11; KO: n = 8). See also Figures S19 and Table S9.

role of individual genes or gene modules in the brain's separation of function based on time of day. For example, more studies should address why and how certain MSN identity genes in M2 peak at ZT3, which challenges the prior concept that neuronal identity genes are stably transcribed (Fishell and Heintz, 2013). The modules with a broad peak of daytime expression (DP2) provide independent support for emerging evidence suggesting elevated oligodendrocyte function (M7) (Artiushin and Sehgal, 2020; de Vivo and Bellesi, 2019), DNA repair (Mourrain and Wang, 2019; Mure et al., 2018; Bellesi et al., 2016), and metabolism (Panda, 2016) during the sleep and rest at daytime. However, unlike these prior studies that are often based on a small subset of genes, our striatal and cortical CoExMaps provide a rich list of diurnal module hub genes to pursue an in-depth, systematic study of genes and pathways essential to chronobiology in the mouse brain.

Another key finding is that genes up- and downregulated by KCl and synaptic activity in cultured neurons were enriched in night- and day-peaked diurnal modules, respectively. This finding is consistent with evidence that neuronal firing rate and synaptic homeostasis is promoted by wake and suppressed by sleep (Hengen et al., 2016; Torrado Pancheco et al., 2021; Cirelli, 2017; Diering et al., 2017). Thus, hub genes for these diurnal modules enriched with neuronal activity regulated genes (Figure 4C) are unbiased candidates to study synaptic homeostasis and neuronal plasticity throughout the daily cycle.

Finally, M14 has the most diurnally expressed genes and a strong link to brain function related to the awake and active phase. This module is enriched in multiple molecular pathways similar to those found active in nighttime in prior studies (Mure et al., 2018; Zhang et al., 2014a; Noya et al., 2019). Importantly, this module is significantly enriched in genes upregulated during sleep deprivation (with *Homer1* as one of the top hub genes) and has multiple sleep disorder-associated genes as members. Thus, M14 hub genes should be further investigated for their potential roles in brain function during the awake and active phase and possible relevance to sleep biology.

The connectivity map (CMAP) demonstrated that gene expression signatures from genetic or chemical perturbations of cultured cells can be used to identify previously unrecognized connections among these perturbagens based on high similarity of gene signatures (Lamb et al., 2006; Subramanian et al., 2017). Our study used extensive examples, from genetic as well as pharmacological gene signatures from HD mouse models to perturbation gene signatures from 52 heterozygous KO mice, to demonstrate the utility of CoExMap in interpreting perturbation gene signatures or other gene sets based on their enrichment in annotated CoExMap modules. Moreover, we developed an online searchable and zoomable striatal gene coexpression map (i.e., striatal CoExMap UMAP), described the fidelity of the map in preserving both local and global connectivity of the gene networks, and demonstrated its utility to interpret gene sets based on annotated biology associated with map location and local gene neighbors. Thus, analogous to the CMAP, mostly used for cellular perturbation signatures, CoExMap can readily interpret gene signatures from intact brain tissues based on their enrichment in different annotated modules to generate interpretable new knowledge and testable hypotheses.

Our study includes a rigorous demonstration of the utility of striatal CoExMap to help interpret *in vivo* genetic perturbation gene signatures (i.e., striatal DE genes) from 52 heterozygous KO mice. Crucially, this study demonstrated a highly nonrandom distribution of DE genes from the 52 KO-Het mice in the 38 CoExMap modules, suggesting that each genetic perturbation affects specific biological networks. Additionally, since the KO-Het gene signatures in bulk striatal tissue could be subtle, our CoExMap-based approach demonstrates a high sensitivity to detect biologically meaningful module enrichment, which is advantageous compared to conventional gene set enrichment analysis that uses gene sets generated mostly outside of the tissue or cell types of interest (e.g., cultured cancer cell lines).

An important application of CoExMap is to identify perturbations that disrupt diurnal gene expression and do so using transcriptomes derived from a narrow daytime window. To illustrate this point, we identified two major clusters of KO-Het gene signatures that either upregulated (5 genes) or downregulated (6 genes) the largest diurnal module, M14. We hypothesized, and our analysis confirmed, that these KO-Het gene signatures significantly overlapped with experimentally verified, diurnally expressed genes (Figure 8B). We provided evidence that one such KO gene, *Trank1*, was important for locomotor activity during nighttime and exhibits anxiety-reduction behaviors, providing novel insights into its potential role in bipolar disorders. Importantly, in all our perturbation RNA-seq studies, we carried out tissue dissection in a restricted time window during the daytime rather than periodically over a 24-h period, demonstrating that CoExMap is a simple and scalable tool that can use daytime transcriptomic data to connect genetic, pharmacological, and disease gene signatures to annotated diurnal molecular networks, even ones that peak in the nighttime.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Experimental animals for the striatal and cortical CoExMap and KO-het perturbation studies
  - Experimental animals for the diurnal transcriptomic study
- METHOD DETAILS
  - Collecting striatal and cortical tissues for RNA-sequencing for the striatal and cortical CoExMap and KO-het perturbation studies
  - Diurnal transcriptomic study
  - Monitoring of locomotor activity
  - Motor assessments using accelerated rotarod
  - Light-dark box test for anxiety-like behaviors

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.neuron.2022.09.028.

ACKNOWLEDGMENTS

The research was supported by CHDI Foundation, Inc. The Yang lab is also supported by NINDS (R01NS113612) and HDF. X.W.Y. is supported by the Terry Semel Chair in Alzheimer's Research and Treatment at UCLA. We thank UCLA Neuroscience Genomics Core (UNGC), Informatics Center for Neurogenetics and Neurogenomics (P30 NS062691), the Jackson Laboratories and UC Davies, and Rodent Genetics Core at Cedar-Sinai for their research support.

AUTHOR CONTRIBUTIONS

X.W.Y. conceptualized, designed, and supervised the study, and X.W.Y., N.W., and P.L. wrote the manuscript. X.W.Y., N.W., P.L., S.H., J.R., J.S.A., and T.F.V. participated in data interpretation. N.W. led a group of researchers, M.S., J.B.R., L.R., R.V., M.P., X.G., and S.Z., to perform the mouse model generation/import, colony management, tissue dissection, and RNA-seq studies. F.G. and P.L. performed RNA-seq data processing. P.L. performed all the bioinformatic analysis and generated relevant plots. X.L. constructed the UMAP illustration. L.V.I. created the searchable CoExMap website. A.J.M. and K.O. generated the striatal and cortical diurnal transcriptomic data. L.Z. and Q.R.L. performed oligodendrocyte culture experiments. T.K.T., L.R., and C.C. produced the 10-day behavioral study for *Trank1* KO mice.

DECLARATION OF INTERESTS

X.W.Y. is a Scientific Advisory Board (SAB) member for Ophidion Bio and a former SAB member for Triplet Therapeutics and Mitokinin, Inc. P.L. serves as an occasional bioinformatics consultant for The Bioinformatics CRO, Inc; Vynance Technologies, LLC; and FOXO Technologies, Inc.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research. One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in their field of research or within their geographical location.

Received: May 11, 2022
Revised: August 16, 2022
Accepted: September 22, 2022
Published: October 19, 2022

REFERENCES

Artiushin, G., and Sehgal, A. (2020). The glial perspective on sleep and circadian rhythms. Annu. Rev. Neurosci. *43*, 119–140. https://doi.org/10.1146/annurev-neuro-091819-094557.

Bates, G.P., Dorsey, R., Gusella, J.F., Hayden, M.R., Kay, C., Leavitt, B.R., Nance, M., Ross, C.A., Scahill, R.I., Wetzel, R., et al. (2015). Huntington disease. Nat. Rev. Dis. Primers *1*, 15005. https://doi.org/10.1038/nrdp.2015.5.

Beaumont, V., Zhong, S., Lin, H., Xu, W., Bradaia, A., Steidl, E., Gleyzes, M., Wadel, K., Buisson, B., Padovan-Neto, F.E., et al. (2016). Phosphodiesterase 10A inhibition improves cortico-basal ganglia function in huntington's disease models. Neuron *92*, 1220–1237. https://doi.org/10.1016/j.neuron.2016.10.064.

Bellesi, M., Pfister-Genskow, M., Maret, S., Keles, S., Tononi, G., and Cirelli, C. (2013). Effects of sleep and wake on oligodendrocytes and their precursors. J. Neurosci. *33*, 14288–14300. https://doi.org/10.1523/JNEUROSCI.5102-12.2013.

Bellesi, M., Bushey, D., Chini, M., Tononi, G., and Cirelli, C. (2016). Contribution of sleep to the repair of neuronal DNA double-strand breaks: evidence from flies and mice. Sci. Rep. *6*, 36804. https://doi.org/10.1038/srep36804.

Bowser, D.N., and Khakh, B.S. (2004). ATP excites interneurons and astrocytes to increase synaptic inhibition in neuronal networks. J. Neurosci. *24*, 8606–8620. https://doi.org/10.1523/JNEUROSCI.2660-04.2004.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinf. *14*, 128. https://doi.org/10.1186/1471-2105-14-128.

Cirelli, C. (2017). Sleep, synaptic homeostasis and neuronal firing rates. Curr. Opin. Neurobiol. *44*, 72–79. https://doi.org/10.1016/j.conb.2017.03.016.

de Vivo, L., and Bellesi, M. (2019). The role of sleep and wakefulness in myelin plasticity. Glia *67*, 2142–2152. https://doi.org/10.1002/glia.23667.

Diering, G.H., Nirujogi, R.S., Roth, R.H., Worley, P.F., Pandey, A., and Huganir, R.L. (2017). Homer1a drives homeostatic scaling-down of excitatory synapses during sleep. Science *355*, 511–515. https://doi.org/10.1126/science.aai8355.

Dong, J., and Horvath, S. (2007). Understanding network concepts in modules. BMC Syst. Biol. *1*, 24. https://doi.org/10.1186/1752-0509-1-24.

Doyle, J.P., Dougherty, J.D., Heiman, M., Schmidt, E.F., Stevens, T.R., Ma, G., Bupp, S., Shrestha, P., Shah, R.D., Doughty, M.L., et al. (2008). Application of a translational profiling approach for the comparative analysis of CNS cell types. Cell *135*, 749–762. https://doi.org/10.1016/j.cell.2008.10.029.

Fishell, G., and Heintz, N. (2013). The neuron identity problem: form meets function. Neuron *80*, 602–612. https://doi.org/10.1016/j.neuron.2013.10.035.

Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium (2019). CAG Repeat Not Polyglutamine Length Determines Timing of Huntington's Disease Onset. Cell *178*, 887–900. https://doi.org/10.1016/j.cell.2019.06.036.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol *5*, R80.

Geschwind, D.H., and Konopka, G. (2009). Neuroscience in the era of functional genomics and systems biology. Nature *461*, 908–915. https://doi.org/10.1038/nature08537.

Gokce, O., Stanley, G.M., Treutlein, B., Neff, N.F., Camp, J.G., Malenka, R.C., Rothwell, P.E., Fuccillo, M.V., Südhof, T.C., and Quake, S.R. (2016). Cellular Taxonomy of the Mouse Striatum as Revealed by Single-Cell RNA-Seq. Cell Rep. *16*, 1126–1137. https://doi.org/10.1016/j.celrep.2016.06.059.

Goold, R., Flower, M., Moss, D.H., Medway, C., Wood-Kaczmar, A., Andre, R., Farshim, P., Bates, G.P., Holmans, P., Jones, L., and Tabrizi, S.J. (2019). FAN1 modifies Huntington's disease progression by stabilizing the expanded HTT CAG repeat. Hum. Mol. Genet. *28*, 650–661. https://doi.org/10.1093/hmg/ddy375.

Gray, M., Shirasaki, D.I., Cepeda, C., André, V.M., Wilburn, B., Lu, X.H., Tao, J., Yamazaki, I., Li, S.H., Sun, Y.E., et al. (2008). Full-length human mutant huntingtin with a stable polyglutamine repeat can elicit progressive and selective neuropathogenesis in BACHD mice. J. Neurosci. *28*, 6182–6195. https://doi.org/10.1523/JNEUROSCI.0857-08.2008.

Graybiel, A.M., and Grafton, S.T. (2015). The striatum: where skills and habits meet. Cold Spring Harb. Perspect. Biol. *7*, a021691. https://doi.org/10.1101/cshperspect.a021691.

Hammond, T.R., Dufort, C., Dissing-Olesen, L., Giera, S., Young, A., Wysoker, A., Walker, A.J., Gergits, F., Segel, M., Nemesh, J., et al. (2019). Single-Cell RNA Sequencing of Microglia throughout the Mouse Lifespan and in the Injured Brain Reveals Complex Cell-State Changes. Immunity *50*, 253–271.e6. https://doi.org/10.1016/j.immuni.2018.11.004.

Hawrylycz, M.J., Lein, E.S., Guillozet-Bongaarts, A.L., Shen, E.H., Ng, L., Miller, J.A., van de Lagemaat, L.N., Smith, K.A., Ebbert, A., Riley, Z.L., et al. (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. Nature *489*, 391–399. https://doi.org/10.1038/nature11405.

He, D., Wang, J., Lu, Y., Deng, Y., Zhao, C., Xu, L., Chen, Y., Hu, Y.C., Zhou, W., and Lu, Q.R. (2017). lncRNA functional networks in oligodendrocytes reveal stage-specific myelination control by an lncOL1/Suz12 complex in the CNS. Neuron *93*, 362–378. https://doi.org/10.1016/j.neuron.2016.11.044.

Heiman, M., Schaefer, A., Gong, S., Peterson, J.D., Day, M., Ramsey, K.E., Suárez-Fariñas, M., Schwarz, C., Stephan, D.A., Surmeier, D.J., et al. (2008). A translational profiling approach for the molecular characterization of CNS cell types. Cell *135*, 738–748. https://doi.org/10.1016/j.cell.2008.10.028.

Hengen, K.B., Torrado Pacheco, A., McGregor, J.N., Van Hooser, S.D., and Turrigiano, G.G. (2016). Neuronal firing rate homeostasis is inhibited by sleep and promoted by wake. Cell *165*, 180–191. https://doi.org/10.1016/j.cell.2016.01.046.

Hinds, L.R., Chun, L.E., Woodruff, E.R., Christensen, J.A., Hartsock, M.J., and Spencer, R.L. (2017). Dynamic glucocorticoid-dependent regulation of Sgk1 expression in oligodendrocytes of adult male rat brain by acute stress and time of day. PLoS One *12*, e0175075. https://doi.org/10.1371/journal.pone.0175075.

Hor, C.N., Yeung, J., Jan, M., Emmenegger, Y., Hubbard, J., Xenarios, I., Naef, F., and Franken, P. (2019). Sleep-wake-driven and circadian contributions to daily rhythms in gene expression and chromatin accessibility in the murine cortex. Proc. Natl. Acad. Sci. USA *116*, 25773–25783. https://doi.org/10.1073/pnas.1910590116.

Hughes, M.E., Hogenesch, J.B., and Kornacker, K. (2010). JTK_CYCLE: an efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets. J. Biol. Rhythms *25*, 372–380. https://doi.org/10.1177/0748730410379711.

Ikeda, M., Takahashi, A., Kamatani, Y., Okahisa, Y., Kunugi, H., Mori, N., Sasaki, T., Ohmori, T., Okamoto, Y., Kawasaki, H., et al. (2018). A genome-wide association study identifies two novel susceptibility loci and trans population polygenicity associated with bipolar disorder. Mol. Psychiatry *23*, 639–647. https://doi.org/10.1038/mp.2016.259.

Jagannath, A., Butler, R., Godinho, S.I.H., Couch, Y., Brown, L.A., Vasudevan, S.R., Flanagan, K.C., Anthony, D., Churchill, G.C., Wood, M.J.A., et al. (2013). The CRTC1-SIK1 pathway regulates entrainment of the circadian clock. Cell *154*, 1100–1111. https://doi.org/10.1016/j.cell.2013.08.004.

Jones, S., Pfister-Genskow, M., Benca, R.M., and Cirelli, C. (2008). Molecular correlates of sleep and wakefulness in the brain of the white-crowned sparrow. J. Neurochem. *105*, 46–62. https://doi.org/10.1111/j.1471-4159.2007.05089.x.

Kandasamy, M., and Aigner, L. (2018). Neuroplasticity, limbic neuroblastosis and neuro-regenerative disorders. Neural Regen. Res. *13*, 1322–1326. https://doi.org/10.4103/1673-5374.235214.

Kelley, K.W., Nakao-Inoue, H., Molofsky, A.V., and Oldham, M.C. (2018). Variation among intact tissue samples reveals the core transcriptional features of human CNS cell classes. Nat. Neurosci. *21*, 1171–1184. https://doi.org/10.1038/s41593-018-0216-z.

Keren-Shaul, H., Spinrad, A., Weiner, A., Matcovitch-Natan, O., Dvir-Szternfeld, R., Ulland, T.K., David, E., Baruch, K., Lara-Astaiso, D., Toth, B., et al. (2017). A unique microglia type associated with restricting development of alzheimer's disease. Cell *169*, 1276–1290.e17. https://doi.org/10.1016/j.cell.2017.05.018.

Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182–187. https://doi.org/10.1038/nature09033.

Kreitzer, A.C., and Malenka, R.C. (2008). Striatal plasticity and basal ganglia circuit function. Neuron *60*, 543–554. https://doi.org/10.1016/j.neuron.2008.11.005.

Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.P., Subramanian, A., Ross, K.N., et al. (2006). The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. Science *313*, 1929–1935. https://doi.org/10.1126/science.1132939.

Langfelder, P., Cantle, J.P., Chatzopoulou, D., Wang, N., Gao, F., Al-Ramahi, I., Lu, X.H., Ramos, E.M., El-Zein, K., Zhao, Y., et al. (2016). Integrated genomics and proteomics define huntingtin CAG length-dependent networks in mice. Nat. Neurosci. *19*, 623–633. https://doi.org/10.1038/nn.4256.

Langfelder, P., and Horvath, S. (2007). Eigengene networks for studying the relationships between co-expression modules. BMC Syst. Biol. *1*, 54. https://doi.org/10.1186/1752-0509-1-54.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinf. *9*, 559. https://doi.org/10.1186/1471-2105-9-559.

Langfelder, P., and Horvath, S. (2012). Fast R Functions for Robust Correlations and Hierarchical Clustering. J Stat Softw *46*.

Langfelder, P., Luo, R., Oldham, M.C., and Horvath, S. (2011). Is my network module preserved and reproducible? PLoS Comput. Biol. *7*, e1001057. https://doi.org/10.1371/journal.pcbi.1001057.

Leek, J.T., and Storey, J.D. (2007). Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genet. *3*, 1724–1735. https://doi.org/10.1371/journal.pgen.0030161.

Li, W., Cai, X., Li, H.J., Song, M., Zhang, C.Y., Yang, Y., Zhang, L., Zhao, L., Liu, W., Wang, L., et al. (2020). Independent replications and integrative analyses confirm TRANK1 as a susceptibility gene for bipolar disorder. Neuropsychopharmacology *46*, 1103–1112. https://doi.org/10.1038/s41386-020-00788-4.

Liu, Y., Hu, W., Murakawa, Y., Yin, J., Wang, G., Landthaler, M., and Yan, J. (2013). Cold-induced RNA-binding proteins regulate circadian gene expression by controlling alternative polyadenylation. Sci. Rep. *3*, 2054. https://doi.org/10.1038/srep02054.

Lobo, M.K., Karsten, S.L., Gray, M., Geschwind, D.H., and Yang, X.W. (2006). FACS-array profiling of striatal projection neuron subtypes in juvenile and adult mouse brains. Nat. Neurosci. *9*, 443–452. https://doi.org/10.1038/nn1654.

Logan, R.W., and McClung, C.A. (2019). Rhythms of life: circadian disruption and brain disorders across the lifespan. Nat. Rev. Neurosci. *20*, 49–65. https://doi.org/10.1038/s41583-018-0088-y.

Lonze, B.E., and Ginty, D.D. (2002). Function and regulation of CREB family transcription factors in the nervous system. Neuron *35*, 605–623. https://doi.org/10.1016/s0896-6273(02)00828-0.

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

Maret, S., Dorsaz, S., Gurcel, L., Pradervand, S., Petit, B., Pfister, C., Hagenbuchle, O., O'Hara, B.F., Franken, P., and Tafti, M. (2007). Homer1a is

a core brain molecular correlate of sleep loss. Proc. Natl. Acad. Sci. USA *104*, 20090–20095. https://doi.org/10.1073/pnas.0710131104.

Manley, K., Pugh, J., and Messer, A. (1999). Instability of the CAG repeat in immortalized fibroblast cell cultures from Huntington's disease transgenic mice. Brain Res. *835*, 74–79. https://doi.org/10.1016/s0006-8993(99)01451-1.

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: uniform manifold approximation and projection for dimension reduction. Preprint at arXiv:1802.03426. https://doi.org/10.48550/arXiv.1802.03426.

McKenzie, A.T., Wang, M., Hauberg, M.E., Fullard, J.F., Kozlenkov, A., Keenan, A., Hurd, Y.L., Dracheva, S., Casaccia, P., Roussos, P., and Zhang, B. (2018). Brain cell type specific gene expression and co-expression network architectures. Sci. Rep. *8*, 8868. https://doi.org/10.1038/s41598-018-27293-5.

Mederos, S., González-Arias, C., and Perea, G. (2018). Astrocyte-neuron networks: a multilane highway of signaling for homeostatic brain function. Front. Synaptic Neurosci. *10*, 45. https://doi.org/10.3389/fnsyn.2018.00045.

Mourrain, P., and Wang, G.X. (2019). Sleep: DNA repair function for better neuronal aging? Curr. Biol. *29*, R585–R588. https://doi.org/10.1016/j.cub.2019.05.018.

Mühleisen, T.W., Leber, M., Schulze, T.G., Strohmaier, J., Degenhardt, F., Treutlein, J., Mattheisen, M., Forstner, A.J., Schumacher, J., Breuer, R., et al. (2014). Genome-wide association study reveals two new risk loci for bipolar disorder. Nat. Commun. *5*, 3339. https://doi.org/10.1038/ncomms4339.

Mure, L.S., Le, H.D., Benegiamo, G., Chang, M.W., Rios, L., Jillani, N., Ngotho, M., Kariuki, T., Dkhissi-Benyahya, O., Cooper, H.M., and Panda, S. (2018). Diurnal transcriptome atlas of a primate across major neural and peripheral tissues. Science *359*, eaao0318. https://doi.org/10.1126/science.aao0318.

Noya, S.B., Colameo, D., Brüning, F., Spinnler, A., Mircsof, D., Opitz, L., Mann, M., Tyagarajan, S.K., Robles, M.S., and Brown, S.A. (2019). The forebrain synaptic transcriptome is organized by clocks but its proteome is driven by sleep. Science *366*, eaav2642. https://doi.org/10.1126/science.aav2642.

Oldham, M.C., Konopka, G., Iwamoto, K., Langfelder, P., Kato, T., Horvath, S., and Geschwind, D.H. (2008). Functional organization of the transcriptome in human brain. Nat. Neurosci. *11*, 1271–1282. https://doi.org/10.1038/nn.2207.

Oldham, M.C., Langfelder, P., and Horvath, S. (2012). Network methods for describing sample relationships in genomic datasets: application to Huntington's disease. BMC Syst. Biol. *6*, 63. https://doi.org/10.1186/1752-0509-6-63.

Panda, S. (2016). Circadian physiology of metabolism. Science *354*, 1008–1015. https://doi.org/10.1126/science.aah4967.

Parikshak, N.N., Gandal, M.J., and Geschwind, D.H. (2015). Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. Nat. Rev. Genet. *16*, 441–458. https://doi.org/10.1038/nrg3934.

Pickrell, A.M., and Youle, R.J. (2015). The roles of PINK1, parkin, and mitochondrial fidelity in Parkinson's disease. Neuron *85*, 257–273. https://doi.org/10.1016/j.neuron.2014.12.007.

Rijo-Ferreira, F., and Takahashi, J.S. (2019). Genomics of circadian rhythms in health and disease. Genome Med. *11*, 82. https://doi.org/10.1186/s13073-019-0704-0.

Saunders, A., Macosko, E.Z., Wysoker, A., Goldman, M., Krienen, F.M., de Rivera, H., Bien, E., Baum, M., Bortolin, L., Wang, S., et al. (2018). Molecular diversity and specializations among the cells of the adult mouse brain. Cell *174*, 1015–1030.e16. https://doi.org/10.1016/j.cell.2018.07.028.

Shigetomi, E., Tong, X., Kwan, K.Y., Corey, D.P., and Khakh, B.S. (2011). TRPA1 channels regulate astrocyte resting calcium and inhibitory synapse efficacy through GAT-3. Nat. Neurosci. *15*, 70–80. https://doi.org/10.1038/nn.3000.

Stahl, E.A., Breen, G., Forstner, A.J., McQuillin, A., Ripke, S., Trubetskoy, V., Mattheisen, M., Wang, Y., Coleman, J.R.I., Gaspar, H.A., et al. (2019). Genome-wide association study identifies 30 loci associated with bipolar disorder. Nat. Genet. *51*, 793–803. https://doi.org/10.1038/s41588-019-0397-8.

Stuart, J.M., Segal, E., Koller, D., and Kim, S.K. (2003). A gene-coexpression network for global discovery of conserved genetic modules. Science *302*, 249–255. https://doi.org/10.1126/science.1087447.

Suberbielle, E., Sanchez, P.E., Kravitz, A.V., Wang, X., Ho, K., Eilertson, K., Devidze, N., Kreitzer, A.C., and Mucke, L. (2013). Physiologic brain activity causes DNA double-strand breaks in neurons, with exacerbation by amyloid-beta. Nat. Neurosci. *16*, 613–621. https://doi.org/10.1038/nn.3356.

Subramanian, A., Narayan, R., Corsello, S.M., Peck, D.D., Natoli, T.E., Lu, X., Gould, J., Davis, J.F., Tubelli, A.A., Asiedu, J.K., et al. (2017). A next generation connectivity map: L1000 platform and the first 1, 000, 000 profiles. Cell *171*, 1437–1452.e17. https://doi.org/10.1016/j.cell.2017.10.049.

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. *47*, D607–D613. https://doi.org/10.1093/nar/gky1131.

Takahashi, J.S. (2017). Transcriptional architecture of the mammalian circadian clock. Nat. Rev. Genet. *18*, 164–179. https://doi.org/10.1038/nrg.2016.150.

Torrado Pacheco, A., Bottorff, J., Gao, Y., and Turrigiano, G.G. (2021). Sleep Promotes downward firing rate homeostasis. Neuron *109*, 530–544.e6. https://doi.org/10.1016/j.neuron.2020.11.001.

Vainchtein, I.D., and Molofsky, A.V. (2020). Astrocytes and microglia: In sickness and in health. Trends Neurosci. *43*, 144–154. https://doi.org/10.1016/j.tins.2020.01.003.

Van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. Journal of machine learning research *9*, 2579–2605.

Wakioka, T., Sasaki, A., Kato, R., Shouda, T., Matsumoto, A., Miyoshi, K., Tsuneoka, M., Komiya, S., Baron, R., and Yoshimura, A. (2001). Spred is a sprouty-related suppressor of Ras signalling. Nature *412*, 647–651. https://doi.org/10.1038/35088082.

Wang, N., Gray, M., Lu, X.H., Cantle, J.P., Holley, S.M., Greiner, E., Gu, X., Shirasaki, D., Cepeda, C., Li, Y., et al. (2014). Neuronal targets for reducing mutant huntingtin expression to ameliorate disease in a mouse model of Huntington's disease. Nat. Med. *20*, 536–541. https://doi.org/10.1038/nm.3514.

Wang, J.K.T., Langfelder, P., Horvath, S., and Palazzolo, M.J. (2017). Exosomes and Homeostatic Synaptic Plasticity Are Linked to Each other and to Huntington's, Parkinson's, and Other Neurodegenerative Diseases by Database-Enabled Analyses of Comprehensively Curated Datasets. Front Neurosci *11*, 149.

Wang, Y., Lv, D., Liu, W., Li, S., Chen, J., Shen, Y., Wang, F., Hu, L.F., and Liu, C.F. (2018). Disruption of the circadian clock alters antioxidative defense via the SIRT1-BMAL1 pathway in 6-OHDA-induced models of Parkinson's disease. Oxid. Med. Cell. Longev. *2018*, 4854732. https://doi.org/10.1155/2018/4854732.

Wheeler, V.C., and Dion, V. (2021). Modifiers of CAG/CTG Repeat Instability: Insights from Mammalian Models. J. Huntingtons Dis. *10*, 123–148. https://doi.org/10.3233/JHD-200426.

Whittaker, D.S., Loh, D.H., Wang, H.B., Tahara, Y., Kuljis, D., Cutler, T., Ghiani, C.A., Shibata, S., Block, G.D., and Colwell, C.S. (2018). Circadian-based treatment strategy effective in the BACHD mouse model of huntington's disease. J. Biol. Rhythms *33*, 535–554. https://doi.org/10.1177/0748730418790401.

Winden, K.D., Oldham, M.C., Mirnics, K., Ebert, P.J., Swan, C.H., Levitt, P., Rubenstein, J.L., Horvath, S., and Geschwind, D.H. (2009). The organization of the transcriptional network in specific neuronal classes. Mol. Syst. Biol. *5*, 291. https://doi.org/10.1038/msb.2009.46.

Yu, W., Clyne, M., Khoury, M.J., and Gwinn, M. (2010). Phenopedia and Genopedia: disease-centered and gene-centered views of the evolving knowledge of human genetic associations. Bioinformatics *26*, 145–146. https://doi.org/10.1093/bioinformatics/btp618.

Zada, D., Sela, Y., Matosevich, N., Monsonego, A., Lerer-Goldshtein, T., Nir, Y., and Appelbaum, L. (2021). Parp1 promotes sleep, which enhances DNA

repair in neurons. Mol. Cell *81*, 4979–4993.e7. https://doi.org/10.1016/j.molcel.2021.10.026.

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Stat. Appl. Genet. Mol. Biol. *4*, Article17. https://doi.org/10.2202/1544-6115.1128.

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

Zhang, S.J., Steijaert, M.N., Lau, D., Schütz, G., Delucinge-Vivier, C., Descombes, P., and Bading, H. (2007). Decoding NMDA receptor signaling: identification of genomic programs specifying neuronal survival and death. Neuron *53*, 549–562. https://doi.org/10.1016/j.neuron.2007.01.025.

Zhang, Y., Chen, K., Sloan, S.A., Bennett, M.L., Scholze, A.R., O'Keeffe, S., Phatnani, H.P., Guarnieri, P., Caneda, C., Ruderisch, N., et al. (2014b). An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. J. Neurosci. *34*, 11929–11947. https://doi.org/10.1523/JNEUROSCI.1860-14.2014.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Organisms/strains		
Mouse: Kcnh4 KO	EMMA	RRID:IMSR_EM:08235
Mouse: Kdm4b KO	EMMA	RRID:IMSR_EM:05854
Mouse: Kifap3 KO	EMMA	RRID:IMSR_EM:11555
Mouse: Nagk KO	EMMA	RRID:IMSR_EM:05309
Mouse: P4ha1 KO	EMMA	RRID:IMSR_EM:08331
Mouse: Pxdn KO	EMMA	RRID:IMSR_EM:01274
Mouse: Sh2d5 KO	EMMA	RRID:IMSR_EM:07282
Mouse: Asl KO	JAX	RRID:IMSR_JAX:018830
Mouse: Atf6 KO	JAX	RRID:IMSR_JAX:028253
Mouse: Bcr KO	JAX	RRID:IMSR_JAX:026396
Mouse: Cd4 KO	JAX	RRID:IMSR_JAX:002663
Mouse: Chn1 KO	JAX	RRID:IMSR_JAX:026751
Mouse: Ezh2 KO	JAX	RRID:IMSR_JAX:022616
Mouse: Htr2c KO	JAX	RRID:IMSR_JAX:002627
Mouse: Pde10a KO	JAX	RRID:IMSR_JAX:008210
Mouse: Rarb KO	JAX	RRID:IMSR_JAX:022999
Mouse: Rest KO	JAX	RRID:IMSR_JAX:024549
Mouse: Ryr1 KO	JAX	RRID:IMSR_JAX:025199
Mouse: Scn4b KO	JAX	RRID:IMSR_JAX:027530
Mouse: Wt1 KO	JAX	RRID:IMSR_JAX:010911
Mouse: Abi3 bp KO	MMRRC	RRID:MMRRC_046427-UCD
Mouse: Bcl11b KO	MMRRC	RRID:MMRRC_046780-UCD
Mouse: Chrm4 KO	MMRRC	RRID:MMRRC_047100-UCD
Mouse: Fbxl16 KO	MMRRC	RRID:MMRRC_047668-UCD
Mouse: Foxp1 KO	MMRRC	RRID:MMRRC_047754-UCD
Mouse: Gsg1l KO	MMRRC	RRID:MMRRC_042126-JAX
Mouse: Hipk4 KO	MMRRC	RRID:MMRRC_048067-UCD
Mouse: Inf2 KO	MMRRC	RRID:MMRRC_048212-UCD
Mouse: Itpr1 KO	MMRRC	RRID:MMRRC_030220-MU
Mouse: Kdm3a KO	MMRRC	RRID:MMRRC_036474-UNC
Mouse: Pde1b KO	MMRRC	RRID:MMRRC_011635-UNC
Mouse: Prkcq KO	MMRRC	RRID:MMRRC_049206-UCD
Mouse: Rasgef1a KO	MMRRC	RRID:MMRRC_042198-JAX
Mouse: Rgs9 KO	MMRRC	RRID:MMRRC_031829-UNC
Mouse: Syndig1l KO	MMRRC	RRID:MMRRC_049952-UCD
Mouse: Tbc1d8 KO	MMRRC	RRID:MMRRC_042140-JAX
Mouse: Vrk1 KO	MMRRC	RRID:MMRRC_050329-UCD
Mouse: Dach1 KO	RIKEN	RRID:IMSR_RBRC06804
Mouse: Ptpn7 KO	RIKEN	RRID:IMSR_RBRC00757
Mouse: Arpp19 KO	This paper	N/A
Mouse: Arpp21 KO	This paper	N/A
Mouse: Ddit4l KO	This paper	N/A
Mouse: Ece2 KO	This paper	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Mouse: Gpx6 KO	This paper	N/A
Mouse: Hrk KO	This paper	N/A
Mouse: Kdm3b KO	This paper	N/A
Mouse: Morf4l1 KO	This paper	N/A
Mouse: Pcdh20 KO	This paper	N/A
Mouse: Ppp1ca KO	This paper	N/A
Mouse: Tcf20 KO	This paper	N/A
Mouse: Trank1 KO	This paper	N/A
Mouse: Zswim6 KO	This paper	N/A
Mouse: Q140 KI	JAX	RRID:IMSR_JAX:027409
Mouse: EIIa-cre	JAX	RRID:IMSR_JAX:003724
Mouse: C57BL/6J	JAX	RRID:IMSR_JAX:000664
Chemicals, antibodies, kits		
Paraformaldehyde	003724	Cat# P6148
Proteinase K	Sigma-Aldrich	Cat# P6556
Anti-Myelin Basic Protein Antibody	Sigma Aldrich	Cat# PA1-10008; RRID:AB_1077024
siRNA for mouse Kdm4b	Sigma Aldrich	Cat# NM_172132
siRNA for mouse Prkcq	Sigma Aldrich	Cat# NM_008859
ProLong Gold Antifade Mountant with DAPI	Invitrogen	Cat# P36935
QIAshredder	QIAGEN	Cat# 79656
RNAeasy Plus Micro Kit	QIAGEN	Cat# 74034
TruSeq Stranded mRNA LT Prep Kit	Illumina	Cat# RS-122-2101
Lipofectamine RNAiMAX	Life Technologies	Cat# 1377850
Deposited data		
Raw data files for CoExMap RNA-seq	This paper	GSE149900
Raw data files for circadian RNA-seq	This paper	GSE151565
Software and algorithms		
WGCNA: an R package for weighted correlation network analysis	UCLA	https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/; RRID:SCR_003302
ImageJ	NIH	https://imagej.nih.gov/ij/index.html; RRID:SCR_003070
STAR	N/A	https://github.com/alexdobin/STAR; RRID:SCR_015899
HTSeq	N/A	https://htseq.readthedocs.io/en/release_0.10.0/; RRID:SCR_005514
GraphPad Prism 7	GraphPad Software	https://www.graphpad.com; RRID:SCR_002798
Adobe Photoshop CC19	Adobe	https://www.adobe.com; RRID:SCR_014199
Additional resources		
Online CoExMap database	UCLA & CHDI	https://www.hdinhd.org/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact Dr. X. William Yang (xwyang@mednet.ucla.edu).

Materials availability

Mouse lines generated in this study are in the process to be deposited to MMRRC for distribution to the scientific community. Before the lines become available to public, limited stock will be available upon request following a completed material transfer agreement.

Data and code availability

CoEx RNAseq data has been deposited within the Gene Expression Omnibus (GEO) repository (www.ncbi.nlm.nih.gov/geo), accession number GSE149900. The circadian mRNA sequencing data is also available as GSE151565. We also created an online searchable CoExMap (through https://www.hdinhd.org/). Most codes are listed under "Software and Algorithms" in the "Key Resource Table", other algorithms are under "Quantification and statistical analysis" section of STAR Method. Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Experimental animals for the striatal and cortical CoExMap and KO-het perturbation studies

The mice use for the striatal and cortical CoExMap and KO-het perturbation studies were at 6-month age and sex balanced (832 males and 832 females). There is no influence or association of sex on our findings (see Figure S5). 39 KO mouse lines were imported from public repositories (i.e., JAX, EMMA, MMRRC and RIKEN). Thirteen KO mouse lines (*Arpp19*, *Arpp21*, *Ddit4l*, *Ece2*, *Gpx6*, *Hrk*, *Kdm3b*, *Morf4l1*, *Pcdh20*, *Ppp1ca*, *Tcf20*, *Trank1*, *Zswim6*) were generated by Yang Lab using the CRISPR/Cas9 method at the Jackson Laboratory's Customer Model Generation Core (9 lines) and at Mouse Biology Program at University of California, Davies (4 lines; Table S1). 6 lines are conditional knockout mice (*Asl*, *Atf6*, *Chn1*, *Ezh2*, *Rest*, *Scn4b*). Male mice in these 6 lines were first bred to E2a-Cre female (JAX 003724), heterozygous offspring carrying post-Cre (null) allele were used for further crossing. We bred heterozygous KO mice for 52 genes (Table S1) to Q140 KI mice to generate offspring with 4 different genotypes. For each crossing, pups were weaned and ear tagged at around 3 weeks of age, and genomic DNAs extracted from ear biopsies were used for genotyping PCR. All mice were maintained and bred under standard conditions consistent with National Institutes of Health guidelines and approved by the University of California, Los Angeles Institutional Animal Care and Use Committees. The cages were maintained on a 12:12 light/dark cycle, with food and water ad lib.

Experimental animals for the diurnal transcriptomic study

Seventy-two male C57Bl/6J mice (JAX:000664) at 26 weeks age were used for the diurnal transcriptomic study. We selected to use male mice to avoid hormonal effects on diurnal activity in female mice. The mouse housing and tissue collection for diurnal transcriptomic study were performed at Cambridge University, United Kingdom under animal protocols approved by the Institutional Animal Care and Use Committees. The cages were maintained on a 12:12 light/dark cycle, with food and water ad lib.

METHOD DETAILS

Collecting striatal and cortical tissues for RNA-sequencing for the striatal and cortical CoExMap and KO-het perturbation studies

Four female and four male mice from wildtype and heterozygous KO (KO-Het) mice between 26 and 27 weeks of age were used for transcriptomic study, thus a total of 8 WT and 8 KO-Het were profiled from each cross. Striatal tissues were harvested from mice strictly in the afternoon time window between ZT7 and ZT11 and fresh frozen on dry ice. We also dissected cortex from 100 mice (89 of which also have striatum expression data) and performed RNA-seq study. Total RNA was extracted using RNeasy kit (Qiagen). All RNA samples had RIN>8.0, with average RIN $\approx$ 8.4. Library preparation and RNA sequencing were performed by the UCLA Neuroscience Genomics Core (UNGC). Libraries were prepared using the Illumina TruSeq RNA Library Prep Kit v2 and sequenced on an Illumina HiSeq4000 sequencer using strand-specific, paired-end, 69-mer sequencing protocol to a minimum read depth of 30 million reads per sample. Reads were aligned to mouse genome mm10 using the STAR aligner with default settings. Read counts for individual genes were obtained using HTSeq. RNAseq data has been deposited within the Gene Expression Omnibus (GEO) repository (www.ncbi.nlm.nih.gov/geo), accession number GSE149900.

Diurnal transcriptomic study

Male C57Bl/6J mice were purchased from the Jackson Lab (JAX:000,664). Mice aged 22 weeks were singly housed within circadian cabinets under 12:12LD cycle with different light cycle settings on each of 4 cabinets. This was to reduce disturbance to mice in the cabinets and to allow mice to be harvested from different parts of the light cycle at the same time. At 26 weeks the mice were sacrificed, and tissue collected for RNA sequencing. At least 6 mice were harvested at 3 hourly intervals across a 36-h window. At each tissue collection point, 2 mice were chosen pseudorandomly from different cabinets. Before the cabinet was opened and mice were removed, the actograms were checked so their rest-activity rhythms of all mice were consistent with the light period: only mice that were active in the dark period and non-active (that had been resting for >15 min) in the light period were used. Dissection was rapid (less than 10 min). Tissue specimens were removed, flash frozen on dry ice, and frozen samples were shipped to Expression Analysis, Inc. (Durham, NC) for RNA-seq study. Briefly, total RNA was harvested using the Qiagen miRNeasy kit. RNA quality and integrity were

monitored via Agilent Bioanalyzer. A minimal RIN score of greater than 7.0 was used as a criterion for subsequent RNA-seq study (156 samples met the criterion and with an average RIN score of 8.70; 29 samples did not meet the criterion). A minimum of 3.75 ug of total RNA were used for library construction. RNA libraries preparation and sequencing were performed following standard Illumina protocols. The following numbers of mouse brain tissue samples per ZT time were used for our diurnal time transcriptomic analysis: *(i). Striatum*: ZT0 (9 samples); ZT3 (8 samples); ZT6 (11 samples); ZT9 (9 samples); ZT12 (11 samples); ZT15 (2 samples); ZT18 (4 samples); ZT21 (5 samples). *(ii). Cortex:* ZT0 (11 samples); ZT3 (12 samples); ZT6 (11 samples); ZT9 (11 samples); ZT12 (12 samples); ZT15 (6 samples); ZT18 (6 samples); ZT21 (6 samples). GEO identifier for the diurnal mRNA data: GSE151565.

Monitoring of locomotor activity

Adult mice were singly housed in cages containing IR motion sensors (Mini Mitter, Bend, OR), and locomotor activity was recorded as previously described (Whittaker et al., 2018). Mice were entrained to a 12:12 h light:dark (LD) cycle for a minimum of 2 weeks prior to collection of 10–14 days of data under LD conditions. Locomotor activity data were recorded using Mini Mitter (Bend, OR) data loggers in 3 min bins, and 10 days of data under each condition were averaged for analysis. Free-running period (tau, τ) was determined using the $\chi 2$ periodogram and the power of the rhythm was determined by multiplying the amplitude, Qp, by 100/n, where n = number of datapoints examined using the El Temps program (A. Diez-Noguera, Barcelona, Spain). Activity amount was determined by averaging 10 days of wheel revolutions (rev/hr). Activity duration (alpha, α) was determined by the duration of activity over the threshold of the mean using an average waveform of 10 days of activity. Nocturnality was determined from the average percentage of activity conducted during the dark. Precision was determined by calculating the daily variation in onset from best-fit regression line drawn through 10 days of activity in both LD and DD conditions using the Clocklab (Actimetrics, Wilmette, IL) program. Fragmentation was defined by bouts/day with a maxgap of 21 min.

Motor assessments using accelerated rotarod

Accelerating rotarod testing was performed in Trank1 knockout mice at 6 months of age based on our published protocol (Gray et al., 2008). Briefly, mice were first trained on an Ugo Basile Accelerating Rotarod for three trials per day for 2 consecutive days. On each training trial, mice were placed on the slowly rotating axle, at a constant speed of 4 rpm for 60 s (day 1) to 120 s (day 2). The mice were placed back on the axle if they fell off during the training trials. The mice were then tested for three trials per day for 3 consecutive days on accelerating mode (4 rpm–40 rpm in 5 min). For each mouse, at least 10-min rest was guaranteed between two trials. 5 min was the maximal time for each testing trial. If a mouse dropped off in less than 20 s, it was re-tested once. Any latency to fall equal or greater than 20 s was recorded as it was.

Light-dark box test for anxiety-like behaviors

Anxiety-like behavior was assessed in the light/dark box exploration test in Trank1 knockout mice at 6 months of age based on our published protocol (Wang et al., 2014). The light/dark exploration box is made of plexiglass with a clear (light) side (27 × 27 × 30 cm) and a smaller, fully opaque (dark) side (18 × 27 × 30 cm) separated by a partition with a small opening. The dark side is completely closed to light. The lighted side is illuminated with a 60-watt light bulb or light with an illuminance of about 400 lux. Light/dark exploration was performed during the light phase of the light/dark cycle. Mice were placed in testing room at least 1 h before testing. The mouse was put at the center of the clear side at beginning, tested for 10 min and assessed for movements from one side to another. Initial latency to dark, time in dark, transitions between light and dark were recorded.

Forced swimming test for depression-like behaviors

Forced swimming test was performed in Trank1 knockout mice at 6 months of age based on our published protocol (Wang et al., 2014). Mice were allowed to acclimate to experimental room at least 1 h before test. One mouse was put into a plexiglass tank or a transparent plastic cylinder (15 cm diameter) filled with water (23–25°C) to a height of 25 cm and videotaped for 6 min. The last 4 min of video was used to score for immobility, which was defined as absence of all movement except motions required keeping the head above the water. Time spent immobile was recorded and analyzed.

Oligodendrocyte culture, transfection, and quantification

Isolation and culture of oligodendrocyte precursor cells (OPCs) from prenatal rat brains was performed as previously described (He et al., 2017). siRNAs for *Kdm4b* and *Prkcq* were purchased from Sigma, St Louis, MO. Rat OPCs were transfected with siRNAs using Lipofectamine RNAiMAX (Life Technologies, cat# 1377850) according to manufacturer's protocol. The cells were harvested 72 h after transfection and processed for qRT-PCR or immunostaining analysis using anti-Mbp antibodies (Sigma, Cat #PA1-10008, 1:500).

QUANTIFICATION AND STATISTICAL ANALYSIS

Overview of data analysis

We first provide a brief overview of the major steps of expression data preprocessing and network analysis. Raw reads were aligned to mouse genome mm10 and then summarized to genes. The count data were filtered to remove non- and low-expressed genes, counts were transformed using variance stabilization and outlier samples were removed. Data were adjusted for batch effects

and other technical covariates. Next, we constructed a weighted co-expression network and identified modules. We carried out the enrichment analysis of module genes in in-house and literature gene sets representing biologically informative categories. We next calculated module eigengenes and analyzed module eigengene correlations and created module eigengene networks. We then defined day and night pattern (DP and NP) module groups using hierarchical clustering of module representatives in diurnal variation data and studied overlaps of cortex and striatum DP and NP groups. Although none of these steps depend on our use of UMAP for visualization, we found that the UMAP helpful in illustrating some of the salient findings.

Preprocessing of striatum and cortex RNA-seq data for the striatal and cortical CoExMap and KO-het perturbation studies

We preprocessed striatum data in two steps. In the first step, samples from each KO-Het cross were treated as a separate dataset. We retained mRNA profiles whose observed counts per million reads are at least 1 in at least one-quarter of the samples and transformed the raw counts using variance stabilization (function varianceStabilizingTransformation in R package DESeq2). To check for potential outlier samples, we used a modified version of the sample network methodology originally described in (Oldham et al., 2012). Specifically, to quantify inter-sample connectivity, we used Euclidean inter-sample distance based on the scaled profiles of the 8000 genes with highest mean expression. The inter-sample connectivities k were transformed to Z scores using robust standardization

$$Z_a = \frac{k_a - median(k)}{1.4826 MAD(k)},$$

where index a labels samples, MAD is the median absolute deviation, a robust analog of standard deviation, and the constant 1.4826 ensures asymptotic consistency (approximate equality of MAD and standard deviation for large, normally distributed samples). Within each KO-Het data set, we removed outliers generally using thresholds $Z = -6$ (the threshold was varied somewhat to account for the presence of technical effects in some of the batches). In certain sets we observed within-batch technical effects that were accounted for using the leading factor determined by Surrogate Variable Analysis (SVA) (Leek and Storey, 2007).

For the combined striatum data set, we repeated the preprocessing starting from the raw counts from outlier-removed samples, filtered again using the theshold of 1 CPM in at least ¼ of the samples and used a relatively high outlier removal threshold of $Z = -10$ to avoid removing entire batches. After outlier removal, the data were corrected using Emprical Bayes-moderated linear models for sex, batch and for the leading surrogate variable from selected batches. This procedure resulted in 407 samples and 15974 genes retained for network analysis.

The cortex expression data were processed in a single step since they were all sequenced in a single dataset. After filtering and variance stabilization, outliers were removed using the threshold $Z = -6$ and data were adjusted for sex and the original KO-Het dataset from which the mice were taken. These steps resulted in retaining 15793 genes in 100 samples for further analysis.

Calculation of outlier suppression weights

To further minimize the effect of outlier measurements that may remain even after removing outlier samples, we used individual observation weights constructed as follows. Tukey bi-square-like weights λ are calculated for each (variance-stabilized) observation x, as

$$\lambda = \left(1 - u^2\right)^2$$

where $u = \min(1, |x\text{-}m|/(9MMAD))$. For the individual KO-Het data sets, m and $MMAD$ are median and modified median absolute deviation of the expression values x. $MMAD$ is initially set to equal the median absolute deviation of x and the following conditions are checked for each gene separately: (1) 10^{th} percentile of the weights λ is at least 0.1 (that is, the proportion of observations with weights <0.1 is less than 10%) (Langfelder and Horvath, 2012) and (2) in each genotype (i.e., generally 8 KO-Het samples and 8 WT controls), the proportion of observation with weight < 0.9 is at most 0.4. If both conditions are checked, $MMAD$ equals MAD. If either condition is not met, $MMAD$ equals the lowest value for which both conditions are met. This ensures that in each genotype at least 60% of the observations have high weights and prevents the down-weighing of entire genotypes whose expression levels are very different from the rest.

For the combined striatum and cortex data, $MMAD$ is initially set to the median absolute deviation of the expression values x adjusted for batch and technical covariates. For each gene, we check whether 10^{th} percentile of the weights λ is at least 0.1 (that is, the proportion of observations with weights <0.1 is less than 10%). If this condition is met, $MMAD$ equals MAD; otherwise, $MMAD$ equals the smallest value for which the condition is met.

DE analysis of KO-Het striatum samples

DE testing was carried out using R package DESeq2 (Love et al., 2014) version 1.20 using default settings except when we disabled outlier replacement (because outliers were suppressed using weights) and disabled independent filtering (since we have removed non- and low-expressed genes in preprocessing). In data sets with obvious technical effects we used the leading SVA factor (Surrogate Variable) as a covariate. For enrichment and overlap calculations, we retained genes DE at $p < 0.01$. the rationale for this

permissive threshold is to retain a reasonable number of genes for overlaps (at the expense of more false positives; however, false positives are not expected to be concentrated in a specific CoExMap module).

Preprocessing of diurnal transcriptomic data

We applied essentially the same preprocessing steps to the diurnal expression data as well, with the following modifications. The threshold for removing outlier samples was 5. In the calculation of the weights λ, we adjusted *MAD* such that (1) 10^{th} percentile of the weights λ is at least 0.1 (that is, the proportion of observations with weights <0.1 is less than 10%) and (2) at each time point, the proportion of observation with weight <0.9 is at most 0.4. This ensures that at each time point at least 60% of the observations have high weights and prevents the down-weighing of entire time points whose expression levels are very different from the rest. Furthermore, we used Surrogate Variable Analysis (SVA) (Leek and Storey, 2007) on the variance-stabilized data to identify unwanted latent variation. We used the leading 6 surrogate variables (SVs) in striatum and 2 in cortex as covariates in association analysis of individual genes, and we adjusted the variance-stabilized data for these surrogate variables before WGCNA.

Association analysis for diurnal expression data

To improve statistical power to detect diurnal changes, we tested association of each gene with 4 variables defined as follows: for each of ZT times τ=0, 3, 6 and 9 h, the variable I_τ is defined as $I_\tau(t) = \cos[2\pi(t-\tau)/24]$ except it is undefined (value NA in R) when $|t-\tau| = 6$ or 18. In other words, the variable I_τ equals 1 for $t = \tau$, $\sqrt{2}/2 \approx 0.7$ when t and τ differ by 3 h, $-\sqrt{2}/2 \approx -0.7$ when t and τ differ by 9 h and -1 when t and τ differ by 12 h. A positive (negative) association with I_τ indicates that a gene is upregulated (downregulated) around time τ, or, equivalently, downregulated (upregulated) around time $\tau + 12h$. The advantage of testing association with I_τ rather than DE at time τ vs. all other times is better power for detecting periodic oscillations. Additionally, it is less susceptible to loss of power at time points with a small number of samples.

Association testing was carried out using R package DESeq2 (Love et al., 2014) version 1.20.0 using default settings except we disabled outlier replacement (because outliers were suppressed using weights) and disabled independent filtering (since we have removed non- and low-expressed genes in preprocessing). The significance threshold of FDR = 0.05 was used to call significantly associated genes.

We used the JTK_CYCLE (Hughes et al., 2010) as implemented in function meta2d in R package MetaCycle. This method uses the Jonckheere-Terpstra-Kendall (JTK) algorithm, which in turn is an application of the Jonckheere-Terpstra (JT) test (a test for detecting monotonic orderings of data across ordered independent groups). The JTK algorithm is applied to a range of hypothesized orderings corresponding to a range of user-defined period lengths and phases, in effect finding the best combination of phase and period that maximizes Kendall correlation between the expression of each tested cyclical ordering (Hughes et al., 2010).

Weighted gene Co-expression network analysis

We used weighted correlation with individual sample weights determined as described above and the "signed hybrid" network in which negatively correlated genes are considered unconnected. We used a soft thresholding power of 4 for the striatum data and 5 for cortex. These choices reflect the higher number of samples in striatum.

To improve robustness of the identified modules, we used a novel subsampling-based module merging criterion in conjunction with Dynamic Tree Cut (Langfelder and Horvath, 2007). Dynamic Tree Cut operates by traversing the hierarchical clustering tree (dendrogram) and merging pairs of branches that do not satisfy all criteria for being called separate modules. The standard criteria include number of objects on each branch (i.e., minimum module size), mean dissimilarity among a small number of "core" objects in each branch and the difference ("gap") between the mean dissimilarity of core objects and the merging height of the two clusters. In addition to these standard criteria (always used in Dynamic Tree Cut), our module identification uses two additional criteria. The first is that the correlation of the first singular vectors ("eigengenes") of the two branches must be below a threshold, in this work set at 0.9. The second criterion evaluates how separable the two branches are with respect to a separate set of cluster labels. Specifically, consider two branches A and B and a separate clustering. The central idea is to count the fractions of A and B that are covered by clusters in the separate clustering that have substantial overlap with both A and B. If this fraction is low in both branches, they can be considered separate modules. If the fraction is high in either of the two branches, it means the branches are not really separable with respect to the separate clustering. When multiple separate clusterings are available, the separability measures can be averaged over the clusterings.

We now describe the dissimilarity measure in detail. Denote the number of objects (genes) on branches A and B by n_A and n_B, respectively. Assume that we also have a separate clustering with cluster labels c_i (index i labels the objects). Denote by C the clusters that overlap with both A and B. For each cluster we calculate the size of its overlaps with A and B, denoted by o_{Ac} and o_{Bc}. We then define the ratios $r_{Ac} = \max[0.5, o_{Ac}/(o_{Ac} + o_{Bc})]$ and $r_{Bc} = \max[0.5, o_{Bc}/(o_{Ac} + o_{Bc})]$. When most of the objects in cluster c are in one of the two clusters, one of the ratios will be close to 1; when the objects in cluster c are equally distributed among the two clusters, both ratios will be close to 0.5. We next form sums $S_A = \sum_c o_{Ac} r_{Ac} + \left(n_A - \sum_c o_{Ac}\right)$ and $S_B = \sum_c o_{Bc} r_{Bc} + \left(n - \sum_c o_{Bc}\right)$. The first term in these sums essentially weighs the ratios by the overlap sizes. The second term adds the numbers of objects in clusters that overlap with either A or B but not both.

Consider a case where both branches A and B are covered by one or more clusters such that about the same number of each cluster objects are in A and B. Thus, A and B are not distinguishable; the sums are $S_A \approx n_A/2$ and $S_B \approx n_B/2$, respectively. Conversely, consider a case where most of branch A is covered by clusters that have no or few objects on branch B and vice-versa. In this case, the branches are distinguishable; the sums approximately equal the numbers of objects on the respective branch. Thus, we define a separability statistic for each branch as $D_A = 2S_A/n_A - 1$ and analogously for the B branch. When multiple clusterings are available, the separability D_A and D_B are averaged over all clusterings. The mutual separability of branches A and B is then defined as the minimum of D_A and D_B. This statistic is used in Dynamic Tree Cut: when two branches merge on the clustering tree, they are merged if the mutual separability is below a user-specified threshold, otherwise they are kept as separate clusters.

To generate the extra clustering assignments, we carried out independent WGCNA analyses on 50 sets of 256 samples randomly selected (without replacement) from the full set of 407 samples (the fraction of retained samples is approximately 0.63). We then used the 50 resulting module assignments as inputs to Dynamic Tree Cut in the final module identification on the full 407 samples, using the merging threshold of 0.6 for the branch separability D in addition to previously available module identification arguments (deepSplit = 2, minimum module size of 30 and module merging cut height of 0.1). To obtain a similar number of modules in the cortex data, we used a threshold of 0.57 for D, deepSplit of 2.5 and module merging cut height of 0.19.

Module eigengenes and fuzzy module membership

Since each module groups together correlated genes, it makes sense to represent each module by a single representative expression profile called module eigengene. The module eigengene is defined as the first singular vector of the standardized module expression matrix; this vector explains most of the variance of the module. Module eigengenes lead to a natural measure of similarity (membership) of all individual genes to all modules (Dong and Horvath, 2007). A continuous ("fuzzy") measure of module membership of gene i in module I is defined as

$$kME_i^I = \mathrm{cor}(x_i, E^I),$$

where x_i is the expression profile of gene i and E^I is the eigengene of module I. The value of module membership lies between -1 and 1. Higher kME indicate that the expression profile of gene i is similar to the summary profile of module I. Since we use signed networks here, we consider module membership near -1 low. The advantage of using a correlation to quantify module membership is that the corresponding statistical significance (p -values) can be easily computed. Genes with highest module membership are called hub genes. Hub genes are centrally located inside the module and represent the expression profiles of the entire module. Some genes may have high continuous module membership in two or more modules and may, in this sense, be considered members of (or intermediate between) several modules.

Module representatives of CoExMap modules in diurnal expression study

To study the diurnal variation of CoExMap modules, we defined, for each of the 18 diurnally-enriched CoExMap modules, a module representative expression profile by averaging scaled (i.e., centered to mean 0 and scaled to variance 1) expression profiles of genes in each module with weights proportional to the fuzzy module membership kME raised to the power 4. By construction, the module representatives are centered (i.e., have mean 0) and hence can be considered to represent expression relative to the mean. For each module representative we calculated the mean expression at each of the 8 time points as well as within-time point standard deviation, defined as the standard deviation of the residuals of expression when the means at each time point are subtracted (in other words, residuals of the expression regressed on time point viewed as a categorical variable). We then retained the 16 modules for which the difference between the maximum and minimum mean expression (across the 8 time points) is at least twice the within-time point standard deviation.

Module replicability from module assignments of top hub genes

While module preservation calculations provide evidence that genes of a module are correlated more strongly than expected in an independent data set, here we want to answer a related but distinct question of how to measure, at the level of individual modules, whether a module than can be said to correspond to a given reference module can be found in a *de novo* WGCNA on independent data. Because interpretation of overlap p values and sizes depends strongly on module size (i.e., for the same fraction of overlapping genes, p values tend to become much more significant for larger modules), we decided to focus on a constant number of the top hub genes in each original module. We chose the number to be 50 because the smallest module contains close to 50 genes; for the two modules that have less than 50 genes, we retain the entire module. Given WGCNA modules from a replication analysis, we calculate overlaps of the top hubs in each original module with the replication modules and choose the module with the highest number of overlapping genes (which is usually but not necessarily also the module with the statistically strongest overlap). We then require that the fraction of the top hubs overlapping with the selected module is above a certain threshold (e.g., 0.5) and the overlap p value is $p < 10^{-20}$.

Using our set of 103 replication samples, we carried out 5 WGCNA analyses on randomly chosen subsets of 15, 20, 30, 40, 60, 80 and 103 samples using the same approach as for the discovery set of 407 samples, including the module stability analysis, for a total of 35 replication WGCNA analyses. Since the module stability involves randomly chosen subsets, the 5 analyses on the full set of 103

samples each return a slightly different module assignment. We varied network construction parameters so that the number of identified modules remained between 35 and 50, comparable to the 38 modules found in the full analysis of the 407 samples. Each of the 35 replication analyses results in an overlap fraction and log p value or each of the 38 original modules. We then averaged the overlap fractions and log p values for the 5 analyses at each sample size, resulting in one mean fraction and mean significance (log p value) for each of the original modules at each replication sample size.

Gene set enrichment calculations

We used the freely available, open source R package anRichment (https://labs.genetics.ucla.edu/horvath/htdocs/CoexpressionNetwork/GeneAnnotation/) to calculate the enrichment the WGCNA modules in a large compendium of reference gene sets that includes the following collections: (1) Cell type and brain region markers collected from various bulk and single cell type studies; (2) Sets of DE genes and WGCNA modules from analysis of diurnal expression variation; (3) Sets of DE genes in various patient and model organism studies of Huntington's Disease curated from the literature by CHDI and available on HDinHD.org under Downloads as HD Gene Set Enrichment Library; (4) Sets of DE genes and WGCNA modules from an aging study in the striatum (unpublished); (5) WGCNA modules from our analysis of a series of mouse HD models with expanded CAG tract in the Huntingtin (*Htt*) gene (Langfelder et al., 2016); (6) Gene Ontology (GO) sets available in annotation packages from the Bioconductor project (Gentleman et al., 2004); (7) Pathway gene sets (KEGG, Reactome, BioCyc, Lipid Maps) provided by NCBI BioSystems (https://www.ncbi.nlm.nih.gov/Structure/biosystems/docs/biosystems_about.html, downloaded 2017/10/20); (8) Molecular Signatures Database (MSigDB) version 6.2 (https://www.gsea-msigdb.org/gsea/msigdb/) excluding positional sets (C1), curated gene sets from online pathway databases, publications in PubMed, and knowledge of domain experts" (C2) and GO (C5); (9) genomic position collection in which each set contains genes in 5 Mb windows; (10) CDC Phenopedia gene sets (Yu et al., 2010); (11) DE sets, WGCNA modules and other gene sets collected from the literature that we have found useful in our research; (12) protein-protein interactor (PPI) sets curated from the literature primarily for Huntington's disease research (Wang et al., 2017); (13) ChEA 2016, ENCODE histone modifications 2015, ENCODE transcription factor ChIP-seq 2015 and mirTarBase libraries from Enrichr (Chen et al., 2013).

Fisher exact test (equivalently, the hypergeometric test) was used to evaluate overlap significance. The set of genes in each WGCNA was used as background and Bonferroni and FDR multiple testing correction was carried out using all reference gene sets and all modules (separately for each WGCNA).

Construction of cell type and cell cluster marker sets from single cell RNA sequencing data

Gokce et al. (2016): Table S4 of this study contains mean expression of all genes across all clusters (neurons are differentiated here into D1-MSN, D2-MSN and interneurons). We added a column for ''common MSN'' by taking the minimum of the D1-MSN and D2-MSN entry for each gene. The rationale is to define (common D1-and D2-) MSN-specific genes. Next, for each gene and each cluster, we calculated the fold changes between this cluster and all other clusters. Fold change between mean log-expression x and y is defined as $FC = (2^x - 1 + \varepsilon)/(2^y - 1 + \varepsilon)$ where $\varepsilon = 0.01$ is a regularization term. For each cluster, we calculate the minimum of fold changes between this cluster and all other clusters. For each cluster, we then retained genes that satisfy the following criteria: mean ($\log_2$) expression in the cluster $x > log_2(2 + 1)$ and minimum of fold changes against all other clusters at least 1.8. When more than 200 genes satisfy these criteria, we retain the top 200 genes ordered by decreasing minimum fold change.

In addition to cluster-specific gene sets, we also created marker gene sets for groups of related clusters: common MSNs (D1-MSN and D2-MSN clusters), common neurons (D1-MSN, D2-MSN, interneuron clusters) and immune cells (microglia and macrophage clusters). Group markers were selected using the minimum mean expression $x > log_2(2 + 1)$ in all clusters within the group, fold change of at least 1.8 between any in-group cluster vs. any out-group cluster and mean absolute fold change among all in-group clusters less than 2 (this criterion excludes genes strongly DE among clusters within a group).

We further created one set of upregulated and one of down-regulated genes in D1-MSNs vs. D2-MSNs. For these two sets, we selected at most 200 genes with mean expression and fold change for D1-vs. D2-MSN or vice versa of at least 1.8.

Saunders et al. (2018): We downloaded the raw counts and sample information from www.dropviz.org in January 2020. Data from each brain region was processed and gene sets were created separately (i.e., references to all cells in the following mean all cells in a particular brain region).

We first excluded samples not assigned to annotated clusters. We then excluded genes whose sum of raw count across all cells was 50 or less as well as mitochondrial genes (genes whose symbol starts with ''mt-''). We next excluded cells with extremely high total counts (more than 95th percentile of the distribution of total counts in all cells) as well as cells with low total counts, defined as the mode of the distribution of raw reads for each cell. For the latter, we created a histogram of total cell read counts with 1,000 breaks and took the mode as the location of the histogram bin with the highest count. We then normalized all cells to the same total count (defined as the geometric mean of the total counts across all retained cells and genes) and finally transformed the data using $log_2(x + 1)$.

We calculated means and standard errors for all genes in all clusters and carried out unequal variance t-tests of differential expression among all pairs of clusters. Clearly, t-test is not the ideal approach to count data and perhaps even more so for single cell RNA-seq, but we expect it to perform well enough for generating sets of marker genes where the differential expression (fold change) threshold is relatively high and the number of samples (cells) in each cluster is large. We also calculated the corresponding fold changes between pairs of clusters. Fold change between mean expression x and y is defined as $FC = (2^x - 1 + \varepsilon)/(2^y - 1 + \varepsilon)$ where

$\varepsilon = 0.01$ is a regularization term. For each cluster, we calculate the minimum of fold changes between this cluster and all other clusters. For each cluster, we then retained genes that satisfy the following criteria: mean ($\log_2$) expression in the cluster $x > log_2(1.01)$, t-test p value $p < 10^{-8}$ and minimum of fold changes against all other clusters at least 1.5. When more than 200 genes satisfy these criteria, we retain the top 200 genes with highest fold changes. Conversely, if fewer than 20 genes satisfied the criteria, we increased the p value threshold by multiplying it by 10 until at least 20 genes satisfied the criteria.

In addition to cluster-specific gene sets, we also created marker gene sets for groups of related clusters: common MSNs (D1-MSN and D2-MSN clusters), common interneurons (Sst, Chat, Pvalb interneurons), common neurons (D1-MSN, D2-MSN, all 3 interneuron clusters) as well as certain other groups (e.g., endothelial stalk and endothelial tip where both clusters are present). Group markers were selected using the same criteria as individual cluster markers, applied to all clusters within group vs. all clusters not in the group and an additional criterion of mean absolute fold change among all in-group clusters less than 1.5 (this criterion excludes genes strongly DE between clusters within a group).

We further created one set of upregulated and one of down-regulated genes in D1-MSNs vs. D2-MSNs. For these two sets, we selected genes with mean expression $x > log_2(1.02)$ and the same fold change, p value and marker set size (gene count) criteria as individual cluster markers. For these two marker sets, only the D1-MSN and D2-MSN clusters were considered, that is, genes were selected without regard for DE vs. other clusters.

Keren-Shaul et al. (2017): Keren-Shaul et al. provide two types of supplementary tables, one that contains mean expression (UMI counts) across clusters, and one with differential expression (p values are always provided but mean expression and fold changes are not always present) for contrasts that are separate from the main cluster comparisons.

For tables that contain mean expression, we used correlation-based average linkage hierarchical clustering to cluster the mean expression profiles and define meta-clusters as groups of clusters with average correlation >0.89. From each meta-cluster, we selected the first cluster and removed the rest from the analysis. In the AD microglia data, this removed clusters Granulocytes 2 (cluster X), Immature B-cells (cluster VII) and Microglia 2 (cluster II). In ALS microglia, this procedure removed clusters Granulocytes 2 (cluster XIII), Monocytes 3 (cluster XI) and B-cells 2 (cluster V).

We then selected those marker genes for each cluster that has mean UMI count in the cluster was at least 0.5 and at least 1.5 times larger than the largest mean UMI count in all other (retained) clusters.

We define fold change between mean expression x and y as $FC = (x + \varepsilon)/(y + \varepsilon)$ where $\varepsilon = 0.2$ is a regularization term that shrinks fold changes for genes with low expression. If less than 20 genes satisfied the fold change criterion, we took the top 20 genes with largest fold changes as the marker set. We also set an upper limit of the marker set sizes at 200 but none of the clusters had more than 200 genes with mean UMI >0.5 and fold change >1.5.

We selected top genes for DE tests based on similar criteria (mean UMI >0.5, fold change >1.5, where mean UMI and fold change are available) plus we required that the nominal DE p value is $p < 10^{-8}$.

UMAP representation of the coexpression network

We first retained only genes assigned to a module (these genes have a fuzzy module membership $kME \geq 0.3$. We then used the *kME* values of the retained genes in all modules as input to the R package umap implementation of UMAP (Uniform Manifold Approximation and Projection), a method for generating a lower-dimensional representation of high-dimensional data (McInnes et al., 2018). This essentially amounts to a two-stage dimensional reduction, first reducing the expression data from 407 samples to fuzzy module membership in 38 modules (i.e., 38 values or "samples" for each gene), then further reducing the 38-dimensional space to 2 dimensions using UMAP. The rationale for using *kME* values instead of expression data is to create a two-dimensional representation of the network with an emphasis on intra-modular connectivity.