Circadian rhythms in multiple behaviors depend on sex, neuropeptide signaling, and ambient light

Larissa Rays Wahba[1,*], Blanca Perez[2,*], KL Nikhil[1], Erik D. Herzog[1], and Jeff R. Jones[2]

Abstract

Organisms have evolved circadian rhythms in behavior to anticipate daily opportunities and challenges such as mating and predation. However, the ethological investigation of behavioral rhythms has been traditionally limited to studying easy-to-measure behaviors (such as locomotor activity) on a circadian timescale or difficult-to-measure behaviors with limited temporal resolution. Here we sought to examine eight overt behaviors never before studied as a function of time of day, sex, light cycle, and neuropeptide signaling. We hypothesized that sex and neuropeptide signaling-dependent differences in daily behaviors have been largely missed because of the focus on running wheel activity in rodents. To address this hypothesis, we used high-throughput machine learning to automatically score complex behaviors from millions of video frames of singly housed, freely behaving male and female mice. Automated predictions for each of the eight behaviors correlated highly with consensus labels by trained human classifiers. We discovered reliable daily rhythms in eating, drinking, grooming, rearing, nesting, digging, exploring, and resting behaviors that persisted in constant darkness. We found that the overall frequency of most behaviors was predominantly affected by light cycle, but the amplitude and peak time of circadian rhythms in multiple behaviors were each dramatically influenced by neuropeptide signaling and sex. We conclude that machine learning can be used to reveal novel daily rhythms in behaviors that depend on sex, neuropeptide signaling, and ambient light and will allow for the rapid circadian phenotyping of mice with different genotypes or disorders.

[1]Department of Biology, Washington University in St. Louis, St. Louis, MO, USA. [2]Department of Biology, Texas A&M University, College Station, TX, USA. *These authors contributed equally to this work.
Email: jjones@bio.tamu.edu

Introduction

Understanding the genetic, neural, and ethological mechanisms that temporally organize behavior is a fundamental goal of fields including neuroscience, motion science and circadian biology. One such behavior, locomotor activity, has been highly studied at a wide range of timescales using running wheels or infrared beam breaks in animals or wrist actigraphy in humans (Ancoli-Israel et al., 2003; Jud et al., 2005). Other behaviors such as eating and drinking can be measured over time with optical or electrical sensors (Schwartz and Zimmerman, 1990; Pendergast et al., 2013). However, determining the temporal structure of more complex behaviors typically requires manual scoring by trained observers of videos recorded over long timescales and, consequently, has not been widely adopted. To overcome the challenges of labor intensive and subjective manual behavior scoring, we sought to adapt a machine learning approach to simultaneously track several behaviors in individual mice over multiple days and conditions with high temporal resolution.

We chose to investigate eight behaviors that encompass an individual mouse's complete behavioral repertoire: eating, drinking, grooming, rearing, nesting, digging, exploring, and resting. We selected these comprehensive behaviors even though some can be further subdivided because together they describe the full extent of a singly housed mouse's "ethogram," including maintenance, nesting, exploratory, and inactive behaviors (Garner, 2017). Many of these behaviors have been quantified using various machine learning approaches that use markerless pose estimation including DeepLabCut, SimBA, and MARS (Pereira et al., 2020). However, these methods have not yet been applied to understanding the temporal organization of behavior.

To address this critical problem, we used DeepEthogram, a recently developed general purpose machine learning classifier that uses raw pixel values of videos instead of pose estimation to predict behavior with high accuracy (Bohnslav et al., 2021). Importantly, when given a video input, a trained DeepEthogram model will output a temporally sequenced binary matrix - an ethogram - that indicates if a given behavior is present or absent in a given frame. We could then use this ethogram to determine if and how behavioral patterns change over time. Specifically, we focused our analysis on how these eight behaviors changed on a circadian, or near-24 h, timescale. Circadian rhythms provide an excellent opportunity to test our model on data collected over multiple days and are also ethologically relevant, as they have evolved to provide an adaptive advantage to animals by allowing them to anticipate regular changes to their environment such as food availability, predator avoidance, and mate selection (Yerushalmi and Green, 2009).

In mammals, circadian rhythms in behavior and physiology are coordinated by the central pacemaker, the suprachiasmatic nucleus (SCN) (Hastings et al., 2018). To accomplish this, the SCN must be synchronized by a population of neuropeptidergic vasoactive intestinal peptide (VIP)-producing neurons that are critical for the circadian organization of physiology (Jones et al., 2018; Mazuski et al., 2018; Todd et al., 2020). Mice genetically deficient for *Vip* or its receptor exhibit disrupted circadian rhythms in numerous physiological processes including glucocorticoid production, metabolism, cardiovascular function, and body temperature (Bechtold et al., 2008; Loh et al., 2008, 2014; Schroeder et al., 2011). VIP is also required for normal circadian rhythms in locomotor behavior, as *Vip*-deficient mice fail to fully entrain their wheel-running activity to a 12 h:12 h light:dark cycle (LD) and exhibit attenuated, or, frequently, arrhythmic wheel-running activity rhythms in constant darkness (DD) (Colwell et al., 2003; Aton et al., 2005). However, whether this neuropeptide is essential for daily rhythms in other complex behaviors is unknown.

Males and females also differ in the temporal patterning of numerous physiological processes (Krizo and Mintz, 2014; Joye and Evans, 2022). Clear sex differences have been identified in, for example, daily rhythms in glucocorticoid production, cardiovascular function, body temperature, and immune function (Walton et al., 2022). Sex differences have also been observed in locomotor activity rhythms, although these behavioral differences are much more subtle. For instance, male mice show a greater precision of locomotor activity onsets in LD and female mice show a longer locomotor activity duration in DD (Kuljis et al., 2013). These differences are likely due to differences in levels of circulating sex hormones and sex hormone expression (Bailey and Silver, 2014). However, it is unclear if there are more overt sex differences in other behavioral rhythms that have not yet been investigated.

Here, we tested the hypotheses that neuropeptide signaling, sex, and ambient light affect daily patterns of behavior. We used machine learning to measure behaviors never before studied simultaneously as a function of time of day in male and female wild-type and *Vip*-deficient mice housed in LD and in DD. We found that the overall frequency of most behaviors was predominantly affected by light cycle rather than VIP or sex. However, we discovered that the amplitude and peak time of circadian rhythms in multiple behaviors were each dramatically influenced by neuropeptide signaling and sex. We conclude that several novel behavioral rhythms depend on neuropeptide signaling, sex, and light, and that machine learning can be used to discover circadian phenotypes that were previously difficult or impossible to observe.

Results

Machine learning can reliably classify behaviors.

To validate the use of machine learning to accurately identify behaviors, we recorded videos from mice singly housed in custom-built cages (**Supplementary Fig. 1a**, for full details see Methods). We manually labeled a set of frames as one of eight behaviors (eating, drinking, grooming, rearing, nesting, digging, exploring, or resting). We used these labeled frames to train our model (**Supplementary Fig. 1b**) to automatically predict behaviors in our recorded videos (**Supplementary Fig. 1c**).

We then used our model to infer behaviors on a manually labeled 24 h video of a freely behaving mouse. We compared the inferred labels to our manual labels and calculated an F1 score that accounted for both true and false positives and negatives. Our model had an F1 score of ≥ 75% for 7 out of 8 behaviors, and an F1 score of 52% for the "rearing" behavior. To determine how these F1 values compared to chance, we simulated ten randomly generated arrays of behaviors and calculated their F1 scores versus our ground truth manually labeled frames. Chance F1 values for our simulated random arrays were much smaller than the calculated F1 scores from our model: 14.1 ± 1.8 % (eating, mean ± SEM), 0.4 ± 0.9 % (drinking), 20.3 ± 3.9 % (grooming), 0.4 ± 1.3% (rearing), 4.6 ± 0.7 ± (nesting), 9.9 ± 1.8 % (digging), 7.5 ± 1.3 % (exploring), and 43.8 ± 12.8 % (resting). We also inferred the same video twice through our model and found that reproducibility was high (>90%) between inferences.

Next, we wanted to determine how well our model predicted behaviors compared to naive human classifiers. Three trained human classifiers independently labeled behaviors in four 5 min videos of a wild-type mouse. The labeled videos were each empirically compared by all four human classifiers to generate a ground-truth reference standard for each of the four videos that

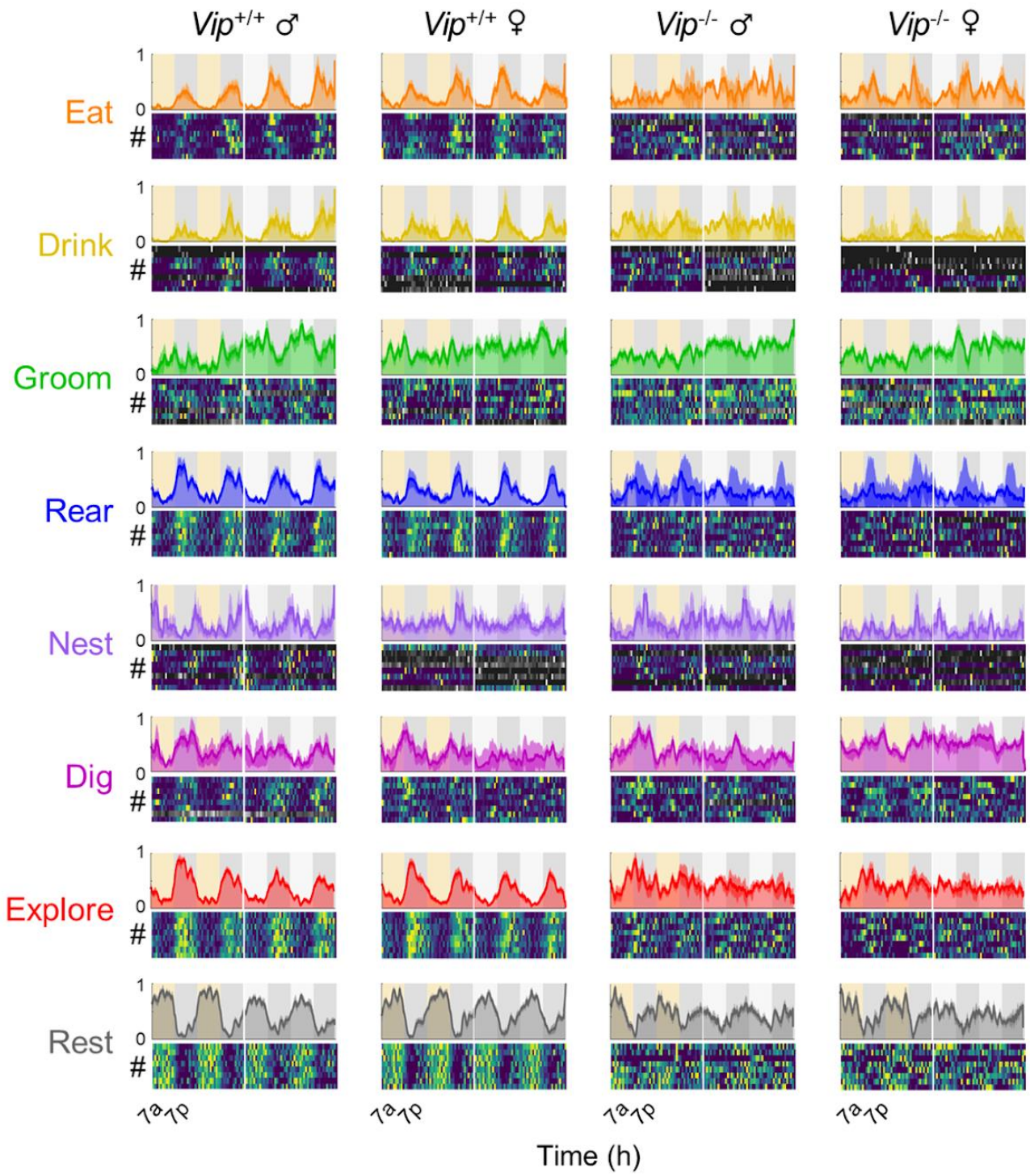

Figure 1. Circadian rhythms in multiple behaviors in male and female wild-type and *Vip*-deficient mice. *Top*, ethograms of average behaviors (colored lines and dark colored shading, mean ± SEM) from mice (n = 32; 8 wild-type male, 8 wild-type female, 8 *Vip*$^{-/-}$, 8 *Vip*$^{-/-}$ female) recorded over 96 h in a 12:12 light:dark cycle (LD, yellow and dark gray background) and in constant darkness (DD, light and dark gray background). For visualization, data were smoothed with a 1 h running average, Y axes were individually scaled by behavior, and areas under the curves were shaded. *Bottom*, raster plots of behaviors from each individual mouse in LD and DD; warmer colors indicate greater activity. For visualization, behaviors were plotted in 1 h bins. Activity profiles of individual mice were scored as either rhythmic (color) or arrhythmic (grayscale).

minimized intra-classifier variability (<3% of all frames). We then had four naive human classifiers and our model label each of the four videos and compared the human labeled behaviors and model labeled behaviors to those labeled in the reference standard (**Supplementary Fig. 1d**). We found that accuracy (defined as percentage of labeled frames that were identical to the

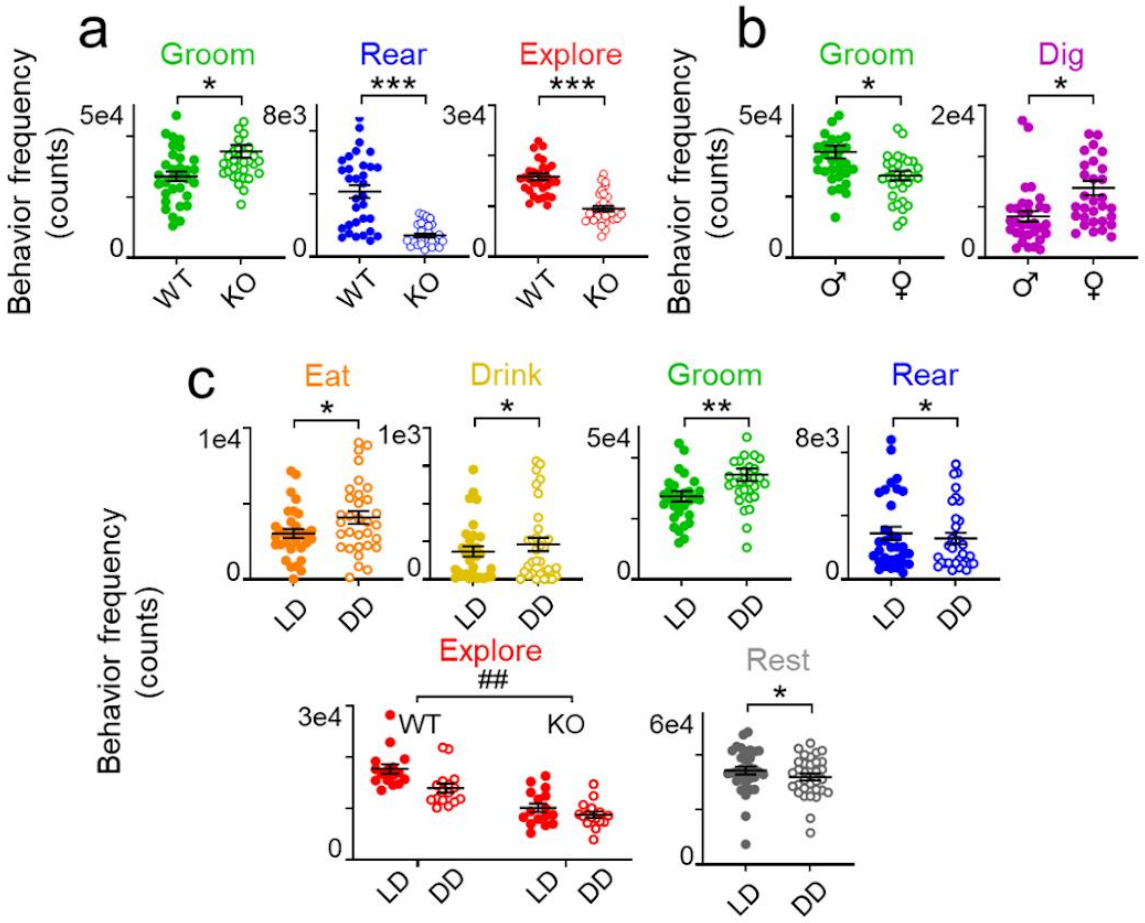

Figure 2. Behavior frequency differs by genotype, sex, and light cycle. Behavior frequencies (total activity counts) that significantly differ by genotype, sex, or light cycle in individual mice. **a)** Behavior frequencies for individual wild-type (WT; male and female, housed in a 12:12 light:dark cycle (LD) and in constant darkness (DD); closed circles) and *Vip*-deficient (KO; male and female, housed in LD and in DD; open circles) mice. **b)** Behavior frequencies for individual male (WT and KO, housed in LD and in DD; closed circles) and female (WT and KO, housed in LD and in DD; open circles) mice. **c)** Behavior frequencies for individual mice housed in LD (WT and KO, male and female; closed circles) and in DD (WT and KO, male and female; open circles). *, significant difference in behavior frequencies between groups within a genotype, sex, or light cycle, Three-Way Repeated Measures ANOVA, $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. ##, significant interaction between behavior frequencies across genotype, sex, or light cycle, $p < 0.01$. Lines depict mean ± SEM.

reference standard) for each behavior did not differ significantly between videos labeled by human classifiers or our model ($p \geq 0.500$ for each behavior, Two-Way ANOVA).

Finally, we wanted to compare the results of our model to results from an independent automated behavior analysis program. We used MouseActivity (Zhang et al., 2020) to analyze the position of a mouse and automatically calculate several variables including distance traveled and velocity over time in four video clips previously labeled by our model. To directly compare MouseActivity-predicted behaviors with the predictions of our model, we thresholded the MouseActivity output by defining "exploring" behavior as a minimum path length of 9 mm and "resting" behavior as a frame with no observed movement. We found remarkably high (>80%) agreement between our model's and MouseActivity's predictions of whether a given frame contained exploring or resting behavior (**Supplementary Fig. 1e**). Together, these methods of validation demonstrate that our model can reliably and accurately predict behaviors in our videos.

Behavior frequency differs by genotype, sex, and light cycle.

We next wanted to use our validated model to determine the temporal distribution of a mouse's behavioral repertoire over multiple days. To do this, we recorded videos of freely-behaving male and female wild-type and *Vip*$^{-/-}$ mice (n = 32; 8 wild-type male, 8 wild-type female, 8 *Vip*$^{-/-}$, 8 *Vip*$^{-/-}$ female) over 96 h in a 12:12 light:dark cycle (LD) and in constant darkness (DD). We then used machine learning to automatically classify which of eight behaviors were present in each frame of the recorded videos, as described above. We transformed this binary classification matrix into an ethogram depicting the occurrence of each behavior over time both for each individual mouse and for mice averaged across genotype and sex (**Fig. 1**).

Because we noticed that certain behaviors (e.g., resting) occurred much more frequently than other behaviors (e.g., nesting) across all animals of either genotype or sex, we quantified behavior frequency, which we defined as the summed total number of frames containing a given behavior for each mouse in 2 days in LD and in 2 days of DD (**Fig. 2, Supplementary Fig. 2, Supplementary Table 1**). We observed that behavior frequency differed by genotype, sex, and

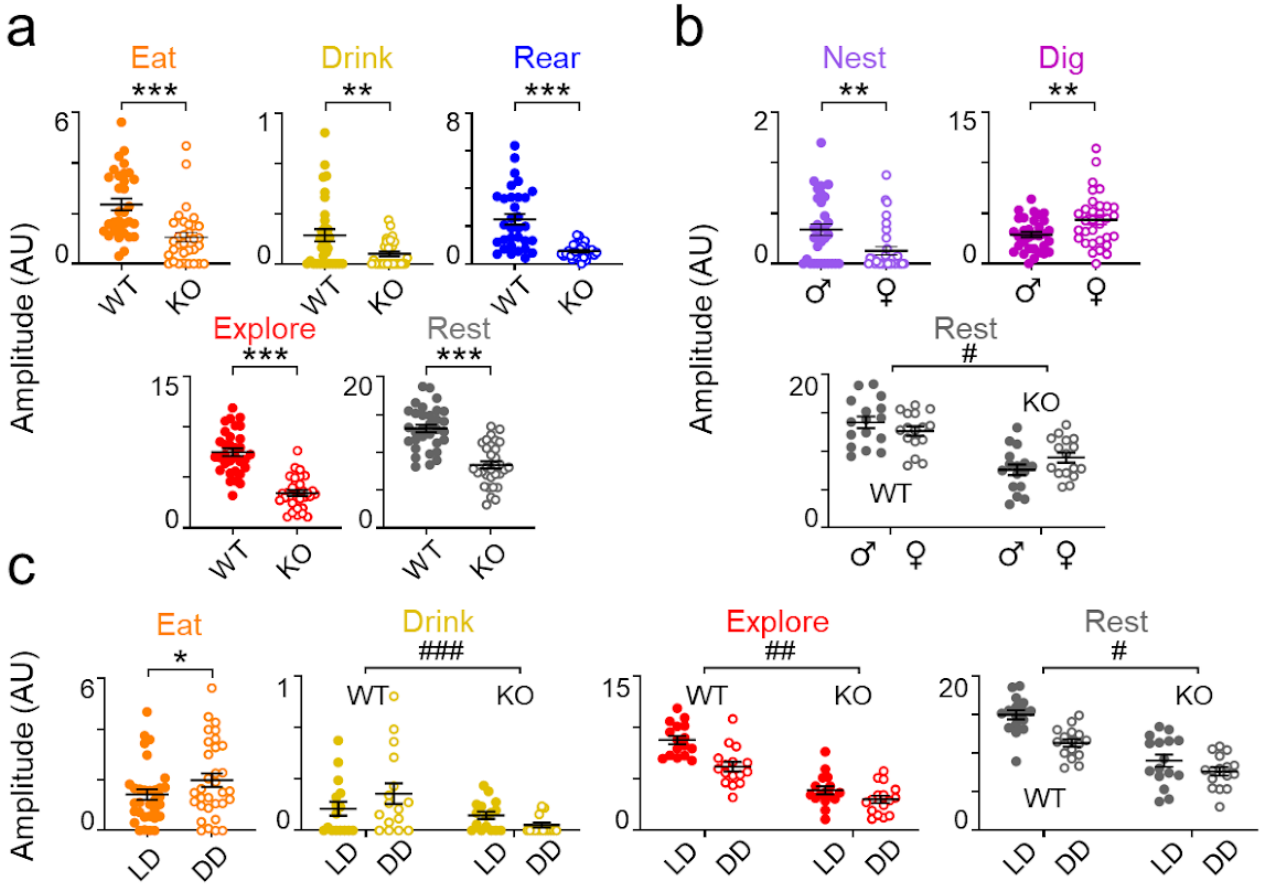

Figure 3. Behavioral rhythm amplitudes differ by genotype, sex, and light cycle. Behavior rhythm amplitudes (in arbitrary units) that significantly differ by genotype, sex, or light cycle in individual mice. **a)** Behavioral rhythm amplitudes for individual wild-type (WT; male and female, housed in a 12:12 light:dark cycle (LD) and in constant darkness (DD); closed circles) and *Vip*-deficient (KO; male and female, housed in LD and in DD; open circles) mice. **b)** Behavior rhythm amplitudes for individual male (WT and KO, housed in LD and in DD; closed circles) and female (WT and KO, housed in LD and in DD; open circles) mice. **c)** Behavior rhythm amplitudes for individual mice housed in LD (WT and KO, male and female; closed circles) and in DD (WT and KO, male and female; open circles). *, significant difference in behavior rhythm amplitudes between groups within a genotype, sex, or light cycle, Three-Way Repeated Measures ANOVA, $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. #, significant interaction between behavior rhythm amplitudes across genotype, sex, or light cycle, $p < 0.05$, ## $p < 0.01$, ### $p < 0.001$. Lines depict mean ± SEM.

light cycle. Specifically, we found that wild-type mice of either genotype in either light cycle rear and explore more frequently, but groom less frequently, than *Vip*$^{-/-}$ mice (**Fig. 2a**). We also found that male mice of either sex in either light cycle groom more frequently, but dig less frequently, than female mice (**Fig. 2b**). Importantly, these results indicate that there is no difference in the frequency of occurrence of most behaviors between wild-type and *Vip*$^{-/-}$ mice (5/8 behaviors similar) or male and female mice (6/8 behaviors similar).

Conversely, we found that the frequency of occurrence of most behaviors differed between mice housed in LD and in DD (2/8 behaviors similar, **Fig. 2c**). We found that mice of either genotype or sex in LD rear and rest more frequently, but eat, drink, and groom less frequently, than mice in DD. Furthermore, wild-type mice of either sex explored more frequently in LD than in DD, but *Vip*$^{-/-}$ mice of either sex did not differ in exploring behavior in either light cycle. Together, these results demonstrate that light, but, surprisingly, not sex or genotype, affects the frequency of most behaviors.

Behavior rhythm amplitudes differ by genotype, sex, and light cycle.

We noticed that even though overall behavior frequency was mostly similar between male and female wild-type and *Vip*$^{-/-}$ mice, the distribution of each behavior seemed to vary with time of day across each genotype, sex, and light cycle. To quantify this, we used circadian analysis to detect daily rhythms in each behavior across individual mice (**Fig. 1**). We found that most individual mice of each genotype and sex exhibited significant diurnal (in LD) and circadian (in DD) rhythms for each behavior regardless of which method we used to determine rhythmicity (**Supplementary Fig. 3**).

We next determined the amplitude of each behavioral rhythm for each mouse, which we defined as half the distance from the peak to the trough of a fitted cosine wave (Yang and Su, 2010). We defined the amplitude of arrhythmic behaviors for each mouse as zero. Amplitude is analogous to the "strength" of a particular behavioral rhythm; that is, a rhythm with a higher amplitude has more defined bouts of activity and inactivity over the 24 h day. We observed that unlike behavior frequency, the amplitudes of each behavioral rhythm varied widely by genotype, sex, and light cycle (**Fig. 3, Supplementary Fig. 4, Supplementary Table 2**). Specifically, we found that wild-type mice of either sex in either light cycle had much higher amplitude

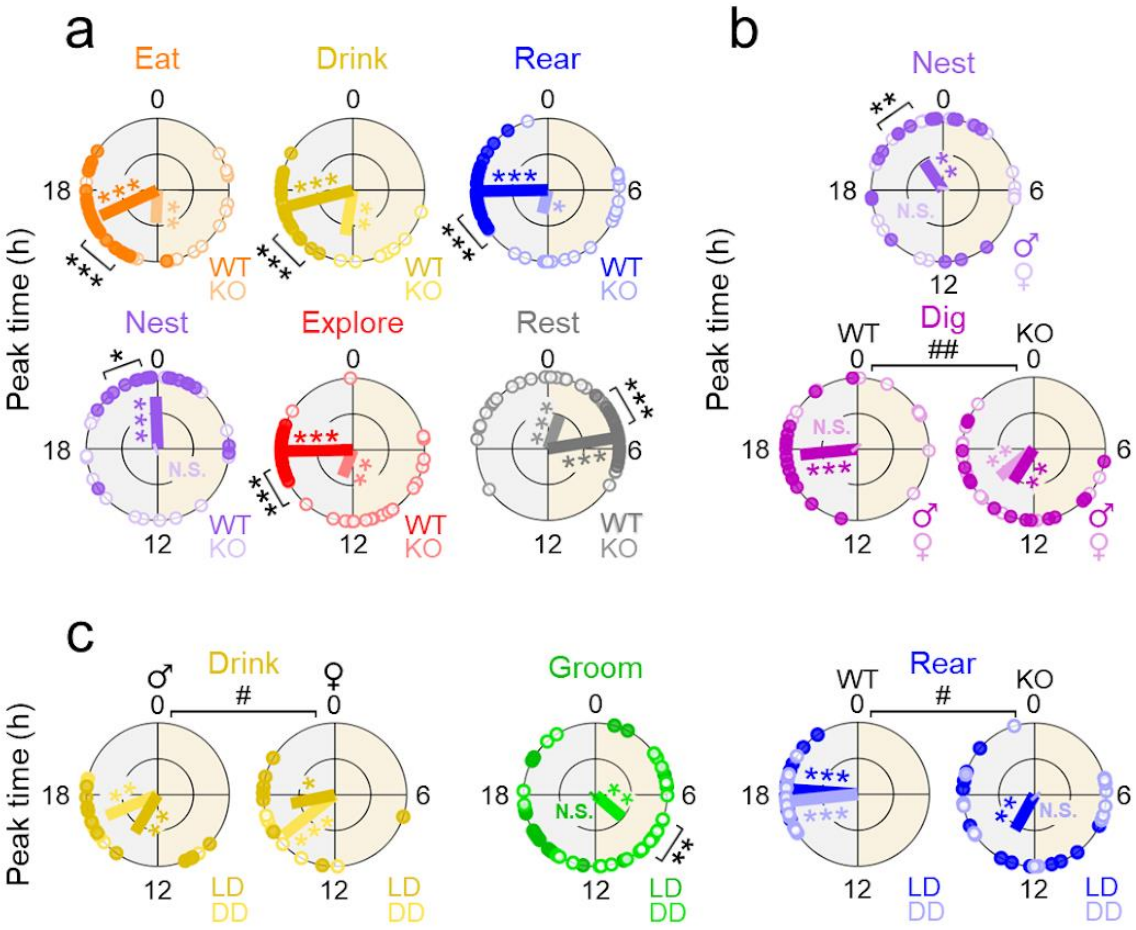

Figure 4. Peak times of behavioral rhythms differ by genotype, sex, and light cycle. Rayleigh plots depicting peak times of behavioral rhythms (in hours) that significantly differ by genotype, sex, or light cycle in individual mice. Direction of the line depicts the mean phase of all rhythmic mice. Length of the line ranges from 0 to 1, where 0 = mice peak randomly around the day and 1 = all mice peak at the same time. **a)** Peak times of behavioral rhythms for individual wild-type (WT; male and female, housed in a 12:12 light:dark cycle (LD) and in constant darkness (DD); closed circles) and *Vip*-deficient (KO; male and female, housed in LD and in DD; open circles) mice. **b)** Peak times of behavioral rhythms for individual male (WT and KO, housed in LD and in DD; closed circles) and female (WT and KO, housed in LD and in DD; open circles) mice. **c)** Peak times of behavioral rhythms for individual mice housed in LD (WT and KO, male and female; closed circles) and in DD (WT and KO, male and female; open circles). Colored *, significant clustering of peak times across mice, Rayleigh test, $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Black *, significant difference in peak times of behavioral rhythms between groups within a genotype, sex, or light cycle, Multi-Way Circular ANOVA, $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. #, significant interaction between peak times of behavioral rhythms across genotype, sex, or light cycle, $p < 0.05$; ## $p < 0.01$. Lines depict mean ± SEM.

eating, drinking, rearing, exploring, and resting rhythms than *Vip*$^{-/-}$ mice (**Fig. 3a**). The decrease in circadian amplitude we observed for most behaviors in individual *Vip*$^{-/-}$ mice is similar to the reduction in wheel-running activity rhythm amplitude previously observed in these animals (Aton et al., 2005). This suggests that SCN desynchronization due to disrupted VIP signaling dampens the amplitude of, but does not eliminate, most behavioral rhythms.

We also found that male mice of either genotype in either light cycle had higher amplitude nesting rhythms, but lower amplitude digging rhythms, than female mice (**Fig. 3b**). Additionally, wild-type male and female mice in either light cycle had similar amplitude resting rhythms, but *Vip*$^{-/-}$ male mice had lower amplitude resting rhythms than female mice. These results suggest that even though most behavioral rhythms are similar between individual male and female mice regardless of genotype or light cycle, there are pronounced sex differences in specific behavioral rhythms (nesting and digging) that had not previously been identified.

Finally, we found that mice of either genotype and sex had lower amplitude eating rhythms in LD than in DD (**Fig. 3c**). We also found that several behavioral rhythms showed a distinct interaction between genotype and light cycle. Specifically, there were no differences in drinking, exploring, or resting rhythm amplitudes between *Vip*$^{-/-}$ mice of either sex in LD and in DD. However, wild-type mice of either sex in LD had higher amplitude exploring and resting rhythms, but lower amplitude drinking rhythms, than wild-type mice of either sex in DD. This genotype difference in light-induced masking (or amplification/ suppression) of behavioral rhythm amplitudes suggests that, as has been previously reported for wheel-running activity (Colwell et al., 2003), *Vip*-deficient mice exhibit deficits in photoentrainment. Together, these results demonstrate that many genotype, sex, and light cycle differences in complex

behaviors are only evident when observed on a circadian timescale.

Peak times of behavioral rhythms differ by genotype, sex, and light cycle.

Next, we wanted to determine if the peak time, or phase, of each behavioral rhythm differed by genotype, sex, or light cycle. Phase describes how reliably synchronized a behavioral rhythm is across mice; accordingly, the "true" endogenous phase of a behavioral rhythm likely occurs when multiple mice exhibit similar peak times for a given behavior. We again used circadian analysis to quantify the phase of each behavioral rhythm for each mouse, which we defined as the time series value of the peak of a fitted cosine wave (Yang and Su, 2010). Phase was undefined for arrhythmic behaviors from individual mice and, as such, we excluded these behaviors from analysis.

We found that wild-type male mice exhibited synchronized rhythms (peaking at the same time across all mice) in 7 out of 8 behaviors in LD and in DD (**Supplementary Fig. 5**). In these mice, resting rhythms peaked around noon while most other behavioral rhythms peaked around midnight. Surprisingly, wild-type male nesting rhythms peaked at dawn in each light cycle. Wild-type female mice also exhibited synchronized rhythms in most (5 out of 8) behaviors in LD and in DD. However, unlike with males, female digging and nesting rhythms were each desynchronized and peaked randomly throughout the day. These results from wild-type mice sharply contrast with what we observed for behavioral rhythm synchrony across *Vip*-deficient mice. In LD, *Vip*$^{-/-}$ male and female mice exhibited synchronized rhythms in most (5 out of 8) or some (4 out of 8) behaviors, respectively. However, strikingly, in DD, almost no behaviors were synchronized across *Vip*$^{-/-}$ male (1 out of 8) or female (0 out of 8) mice.This suggests that *Vip*$^{-/-}$ mice, though individually rhythmic in most behaviors, are desynchronized from one another and not entrained to the external LD cycle.

We next compared peak times of behavioral rhythms of individual mice and found that they varied widely by genotype, sex, and light cycle, similar to what we observed for behavioral rhythm amplitudes in individual mice (**Fig. 4, Supplementary Fig. 6, Supplementary Table 3**). Specifically, we found that rhythms in eating, drinking, rearing, nesting, exploring, and resting behaviors in wild-type mice of either sex in either light cycle peaked later in the day than those in *Vip*$^{-/-}$ mice, although *Vip*$^{-/-}$ nesting rhythms were not significantly synchronized across mice (Rayleigh test, $p > 0.05$; **Fig. 4a**). The earlier peak time of these behavioral rhythms in individual *Vip*$^{-/-}$ mice is similar to the phase advance in wheel-running activity previously observed in these animals (Colwell et al., 2003). This suggests that the reduced ability of these animals to entrain to the external light:dark cycle advances the phase of most, but not all, behavioral rhythms.

We also found that nesting behavior rhythms in male mice of either genotype in either light cycle peaked later in the day than those in female mice (**Fig. 4b**), although these rhythms in females were not significantly synchronized across mice (Rayleigh test, $p > 0.05$). Furthermore, digging behavior rhythms in wild-type male mice in either light cycle peaked earlier in the day than those in wild-type female mice, although these rhythms in females were also not significantly synchronized across mice (Rayleigh test, $p > 0.05$). The peak time of digging behavior rhythms did not differ between male and female *Vip*$^{-/-}$ mice in either light cycle. As with behavioral rhythm amplitude, these results again highlight that there are profound sex differences in specific behavioral rhythms (nesting and digging).

Finally, we found that grooming behavior rhythms in mice of either genotype or sex peaked later in the day in LD than in DD (**Fig. 4c**), although these rhythms were not significantly synchronized across mice in LD (Rayleigh test, $p > 0.05$). We also found that drinking rhythms in male mice of either genotype peaked earlier in the day in LD than in DD. Conversely, drinking rhythms in female mice of either genotype peaked later in the day in LD than in DD. Furthermore, rearing behavior rhythms in *Vip*$^{-/-}$ mice of either sex peaked later in the day in LD than in DD, although these rhythms in DD were not significantly synchronized across mice (Rayleigh test, $p > 0.05$). The peak time of rearing behavior rhythms did not differ between wild-type mice of either sex in LD or in DD. This suggests that ambient light has only a modest effect on the timing of endogenously generated behavioral rhythms. Together, these results demonstrate that genotype, sex, and light cycle each influence the time of day at which a given behavior is most likely to occur.

Discussion

To test the hypotheses that neuropeptide signaling, sex, and ambient light affect daily patterns of behavior, we used machine learning to automatically measure circadian rhythms in behavior from male and female wild-type and *Vip*$^{-/-}$ mice over several days in LD and DD. The frequency of occurrence of most behaviors was not largely affected by circadian genotype or sex but was strongly influenced by light cycle. Conversely, the amplitudes and peak times of circadian rhythms in several behaviors varied by circadian genotype, sex, and light cycle. We conclude that several previously unstudied behavioral rhythms depend on sex, neuropeptide signaling, and ambient light. Our identification of these novel behavioral rhythms that differ by circadian genotype, sex, and light cycle and our implementation of a method to automatically classify circadian rhythms in behavior will provide a foundation for future studies investigating the temporal organization of other complex behaviors.

We found that each measured behavior was typically rhythmic in individual animals of each genotype and sex in both LD and DD regardless of which rhythmicity analysis method we used (JTK Cycle, Cosinor, Lomb-Scargle periodogram, empirical JTK Cycle, or ARSER) (Refinetti et al., 2007; Zielinski et al., 2014; Wu et al., 2016). These algorithms may have a propensity for false positives with our data because, for the large number of time points (172,800 frames) we recorded within a limited sampling window (48 h), inactivity/activity (sleep/wake) could cause a behavior to appear rhythmic. However, the rhythmicity we observed in a majority of behaviors in a majority of individual mice is consistent with several previous studies. For example, wheel-running activity is rhythmic in male and female wild-type mice in LD and DD (Kuljis et al., 2013), and wheel-running activity is rhythmic in most *Vip*$^{-/-}$ mice in LD and for the first several days in DD (Aton et al., 2005). It is possible that the rhythms we observed in some or all behaviors in our *Vip*$^{-/-}$ mice would disappear after a longer time in DD.

These behavioral rhythms we observed in individual mice occurred at "logical" times and were not randomly scattered throughout the day. For instance, exploring and resting rhythms were antiphase, occurring roughly 12 h apart in individual mice regardless of genotype, sex, or light cycle. We also found that within individual animals most behavioral rhythms other than nesting and resting peaked at approximately the same time of day. This suggests that unlike physiological rhythms that peak around the clock (Perreau-Lenz et al., 2004), behavioral rhythms are largely constrained to the middle of the night in nocturnal mice. This may be simply due to sleep restricting behavior to essentially only half of the circadian day while physiological processes can typically persist during the inactive phase. However, we found that nesting rhythms in individual wild-type male mice reliably peak at dawn. As nesting is a sleep-preparatory

behavior (Eban-Rothschild et al., 2016) and mice sleep during the day, future studies should distinguish between the role of sleep need versus circadian output in producing this ethologically-relevant rhythm.

We found that several behavioral rhythms differed by circadian genotype. For instance, rhythms in eating, drinking, rearing, exploring, and resting behaviors were each dramatically lower in amplitude in *Vip*-deficient mice than in wild-type mice. This is consistent with previous studies that found that wheel-running activity, sleep, and feeding rhythms in *Vip*$^{-/-}$ mice each have lower amplitudes than rhythms in wild-type mice (Colwell et al., 2003; Bechtold et al., 2008; Hu et al., 2011), but alterations in daily rhythms in drinking and rearing behaviors in these animals have not been previously reported. Similarly, rhythms in eating, drinking, rearing, nesting, exploring, and resting behaviors peaked earlier in the day in *Vip*-deficient mice than in wild-type mice. This is again consistent with previous studies that found that wheel-running activity and sleep rhythms in *Vip*$^{-/-}$ mice peak earlier than rhythms in wild-type mice (Colwell et al., 2003; Hu et al., 2011), but phase differences in eating, drinking, rearing, and nesting have not been previously reported. Importantly, this study is the first to simultaneously measure each of these behaviors in *Vip*$^{-/-}$ mice. This parallel analysis is critical to understand how disrupted neuropeptide signaling in the SCN affects the temporal sequencing of multiple behaviors.

We also found that rhythms in digging and nesting behaviors differed between individual male and female mice. The sex differences we observed in the amplitude or phase of these behavioral rhythms are entirely novel and have not been previously reported. However, other studies have found that individual female mice show a longer duration of wheel-running activity (comparable to our "exploring" behavior) than male mice in DD, but individual male mice show a greater precision of wheel-running activity than female mice in LD (Kuljis et al., 2013). We were unable to directly compare our results to these findings because we only measured behaviors over two days in LD and two days in DD. However, we did observe that the duration of exploring rhythms in female mice was indeed slightly longer than in male mice (14.6 ± 0.5 h versus 13.7 ± 0.6 h; unpaired Welch's *t* test, $p < 0.01$). These results emphasize that measuring circadian rhythms in behaviors other than locomotor activity can reveal critical unseen differences between males and female mice and, most likely, between other groups of experimental animals.

Intriguingly, the circadian genotype and sex differences we observed were only present in a subset of an individual mouse's entire behavioral repertoire. The amplitudes of grooming, nesting, and digging rhythms and the phases of grooming and digging rhythms do not differ by circadian genotype. Similarly, the amplitudes and phases of all behavioral rhythms but digging and nesting do not differ by sex. This suggests that the brain regions and neural circuits that regulate each of these behaviors may be differentially influenced by the daily timing signal that originates from the SCN. For example, the circuit that regulates digging rhythms may be relatively robust to a desynchronized, low-amplitude input arising from a *Vip*-deficient SCN, but the circuit that regulates exploring rhythms may be more directly affected by *Vip* deficiency. Similarly, circuits regulating digging and nesting behaviors may respond differently to SCN input in male and female mice, perhaps due to a differential expression of estrogen or androgen receptors. Future experiments should investigate the role of local clocks and sex hormone receptor expression in brain circuits that are known to be associated with each of these behaviors.

We found that circadian genotype, sex, and light cycle also affected the synchronization of behaviors across mice. For instance, *Vip*$^{-/-}$ mice of either sex only exhibited synchronized behavioral rhythms in LD but not DD. This suggests that *Vip*-deficient mice are not entrained to the LD cycle, causing their behavioral rhythms to peak at random times in DD, which we observe as desynchrony across mice. Furthermore, nesting and digging rhythms were desynchronized across female mice but were synchronized across male mice. Indeed, nesting rhythms in male mice had their own unique phase compared to all other behaviors, peaking synchronously around dawn. The mechanism underlying this desynchrony in some, but not all behavioral rhythms across female, but not male, mice is unclear, but may be due to changes in specific circuits that respond differently to circulating sex hormones.

Our observation that resting rhythms are synchronized across *Vip*$^{-/-}$ male mice but nesting rhythms are desynchronized helps discern whether the sleep-preparatory nesting rhythm is regulated by sleep or the circadian system. In wild-type male mice, rhythms in nesting behavior peak about 6 h before resting rhythms. If nesting was entirely regulated by sleep need, we would expect to see synchronized rhythms in nesting behavior in *Vip*$^{-/-}$ male mice peak several hours before resting rhythms, which are still synchronized across animals even though the daily timing signal from the *Vip*-deficient SCN is disrupted. Instead, nesting rhythms are desynchronized in the absence of *Vip* signaling. This suggests that the daily occurrence of nesting behavior at dawn is regulated by the circadian clock.

Finally, we identified sex differences in the frequencies of grooming and digging behaviors that are consistent with previous studies that also identified sex differences in grooming and digging, albeit within a much shorter temporal window (Geuther et al., 2021; Pond et al., 2021). Surprisingly, we also observed differences in the frequency of grooming, rearing, and exploring behaviors between *Vip*-deficient and wild-type mice. Behavior frequency is independent of circadian time and, consequently, SCN desynchrony caused by disrupted neuropeptide signaling in *Vip*$^{-/-}$ mice should theoretically have no effect on this measurement. There is limited evidence that the SCN can influence certain behaviors independently of its role in rhythm generation (Yu et al., 2017). Alternatively, VIP is also expressed in other neurons, such as those in the olfactory bulb and cortex (Lein et al., 2007). Disrupted neuropeptide signaling in these circuits could result in our observed circadian genotype-dependent differences in behavior frequency.

In this study, we used machine learning to automatically identify differences in the temporal organization of behavior due to neuropeptide signaling, sex, and ambient light. This approach can readily be expanded to address other critical questions in neuroscience and circadian biology, including the ethological investigation of other behavioral rhythms in videos of mice recorded in the laboratory and, potentially, in the wild. Notably, machine learning can also be used for the rapid circadian phenotyping of mice with different genotypes or disorders. Current approaches almost universally measure changes to wheel-running activity rhythms as evidence that a mutation, disease, or drug influences circadian behavior. Here, we found that some, but not all, behavioral rhythms differ by sex and with neuropeptide signaling. It is therefore likely that a given experimental treatment may cause circadian alterations in behaviors other than, or in addition to, wheel-running activity. Machine learning can be used to study these circadian behaviors that were previously difficult or impossible to observe.

Methods

Animals. Prior to recording, we group-housed male and female wild-type (*Vip*$^{+/+}$, n = 8, 8) and *Vip*$^{-/-}$ ((Colwell et al., 2003), n = 8, 8) mice in their home cages in a 12h:12h light:dark cycle (LD, where lights on is defined as zeitgeber time (ZT) 0; light

intensity ~2 x 10^{14} photons/cm^2/s) at constant temperature (~23°C) and humidity (~40%) with food and water provided ad libitum. All mice were between 6 and 12 weeks old at the time of recording. All experiments were approved by and performed in accordance with the guidelines of Texas A&M University and Washington University's Institutional Animal Care and Use Committees.

Experimental housing. We transferred individual mice from their home cages to custom-built recording cages (**Supplementary FIg. 1a**) inside light-tight, temperature-and-humidity controlled circadian cabinets for the duration of our experiments. The cages are built out of transparent 6 mm-wide acrylic and are approximately 21 cm in length, width, and height. In one of the cage walls, we installed an acrylic food hopper with a slotted opening of around 15 x 10 cm that we filled with food pellets. On the outside of the wall immediately adjacent to the food hopper, we installed a water bottle with a metal spout that protruded about 1.5 cm into the cage. We covered the cage floor (around 440 cm^2 of explorable space) with about 1 cm of standard corn cob bedding and placed half of a cotton nestlet in the center of the cage. In the light phase in LD, we illuminated the cabinets with broad-spectrum white light (~6 x 10^{13} photons/cm^2/s measured at the cage floor). In the dark phase in LD and in DD, we illuminated the cabinets with dim red light (660 nm, ~3 x 10^{13} photons/cm^2/s). We acclimated mice in the recording cages for 20-24 hours before the start of each experiment.

Video recording. We positioned networked video cameras (DCS-932L and DCS-933L, D-Link) equipped with fish-eye lenses directly above the recording cages such that all four corners of the cage, the food hopper, and water spout were all visible in the recorded video and the mouse and nesting material were in focus (**Supplementary Fig. 1a**). We continually recorded grayscale videos (640 x 480-pixel resolution, 30 frames per second) of singly-housed mice as they freely behaved continually over four days, two days in 12h:12h LD and two days in DD. We then exported the recorded videos as .asf files using D-ViewCam software, downsampled the videos to 1 fps using the open-source video editing program FFMpeg, and converted the videos to .mp4 files using FFMpeg. Finally, we cropped the converted videos to the extent of the cage floor using the open-source video editing program Handbrake.

Model training and inference. We installed DeepEthogram (Bohnslav et al., 2021) from source on a custom-built machine learning computer (12-core AMD Ryzen 9 5900X CPU, 32 GB RAM, NVIDIA GeForce RTX 3090 with 24 GB VRAM) according to the DeepEthogram Github installation guide (http://github.com/jbohnslav/deepethogram/). After installation, we loaded a set of training videos (three 24 h, 86,400 frame-long videos of an individual wild-type male mouse and five 10 min, 600 frame-long videos of female wild-type and male and female *Vip*$^{-/-}$ mice) into the DeepEthogram GUI.

To train the DeepEthogram models, we first chose eight complex motor behaviors that encompass the majority of a mouse's daily activity: eating, drinking, grooming, rearing, nesting, digging, exploring and resting. We manually labeled each frame of the training videos using the GUI as one of these behaviors or background (**Supplementary Fig. 1b**) based on consensus between three trained classifiers and the behavioral descriptions listed on the Stanford University Mouse Ethogram index (Garner, 2017). While we labeled our training videos, we began training the "flow generator" convolutional neural network (CNN), which is pretrained on the Kinetics700 video dataset and requires no user input. The flow generator uses the MotionNet architecture to estimate motion ("optic flow") from our training video frames. Next, we trained a two-stream CNN, a "feature extractor," on the output from the flow generator and our manually labeled frames. The feature extractor, also pre trained on the Kinetics700 video dataset, uses the ResNet50 architecture to determine the probability of a behavior being present on a particular frame based on a low-dimensional set of temporal (optic flow) and spatial (labeled frames) features. Finally, we used our feature extractor outputs to train a "sequence model" CNN that uses the Temporal Gaussian Mixture CNN which has a large temporal receptive field and can thus further refine the model predictions using a longer timescale to provide "context" for a given behavior frame.

After training our models, we uploaded our experimental videos using the DeepEthogram GUI. In total, we uploaded 32 videos (male and female, wild-type and *Vip*$^{-/-}$) each containing 345,600 frames (96 h). For each video, we inferred behaviors on each frame using our trained feature extractor and sequence models in the DeepEthogram GUI (**Supplementary Fig. 1c**). Because the original DeepEthogram model was trained on datasets in which behaviors were not mutually exclusive (that is, multiple behaviors could occur on a given frame), we changed the final activation value for our feature extractor and sequence model CNNs from "sigmoid" to "softmax" during training. A softmax, or normalized exponential activation, function requires that the probability of predicted behaviors on a given frame sum to one and thus, for our inferred videos, each frame was uniquely labeled as one behavior.

Analysis. We determined circadian rhythmicity using three methods (**Supplementary Fig. 3**): Cosinor analysis (Refinetti et al., 2007) in Matlab and Lomb-Scargle Periodogram and ARSER in Metacycle (Wu et al., 2016). We also performed empirical JTK Cycle in BioDare2 (Zielinski et al., 2014) on data in 5 min bins instead of on the entire dataset because the current release of JTK Cycle struggles to handle data of this length. We ultimately used rhythmicity, amplitude, and phase predictions by ARSER to perform comparisons of circadian rhythms. We calculated rhythm amplitudes and peak times for individual mice using the amplitude and phase values determined by ARSER. We calculated total activity by summing the total frames exhibiting a given behavior for individual mice in LD and in DD. We calculated transition scores, which we defined as the probability that a given behavior will follow another behavior within 10 or fewer seconds, using a custom MATLAB script.

We performed the following statistical tests in Prism 9.0 (GraphPad, San Diego, CA): Two-Way ANOVA, Three-Way Repeated Measures ANOVA. We performed the following statistical analyses using the Circular Statistics Toolbox (Berens, 2009) in Matlab (Mathworks, Natick, MA): Rayleigh test, multi-way circular ANOVA, and Watson-Williams test. We performed the following statistical analysis using the Statistics Toolbox in Matlab: multi-way ANOVA with post hoc Tukey's HSD test. We used Shapiro–Wilk and Brown–Forsythe tests to test for normality and equal variances, defined α as 0.05, and presented all data as mean ± SEM.

References

Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W, Pollak CP (2003) The role of actigraphy in the study of sleep and circadian rhythms. Sleep 26:342–392.

Aton SJ, Colwell CS, Harmar AJ, Waschek J, Herzog ED (2005) Vasoactive intestinal polypeptide mediates circadian rhythmicity and synchrony in mammalian clock neurons. Nat Neurosci 8:476–483.

Bailey M, Silver R (2014) Sex differences in circadian timing systems: implications for disease. Front Neuroendocrinol

35:111–139.

Bechtold DA, Brown TM, Luckman SM, Piggins HD (2008) Metabolic rhythm abnormalities in mice lacking VIP-VPAC2 signaling. Am J Physiol Regul Integr Comp Physiol 294:R344–R351.

Berens P (2009) CircStat: a MATLAB toolbox for circular statistics. J Stat Softw 31:1–21.

Bohnslav JP, et al. (2021) DeepEthogram, a machine learning pipeline for supervised behavior classification from raw pixels. Elife 10:e63377.

Colwell CS, et al. (2003) Disrupted circadian rhythms in VIP- and PHI-deficient mice. Am J Physiol Regul Integr Comp Physiol 285:R939–R949.

Eban-Rothschild A, Rothschild G, Giardino WJ, Jones JR, de Lecea L (2016) VTA dopaminergic neurons regulate ethologically relevant sleep-wake behaviors. Nat Neurosci 19:1356–1366.

Garner J (2017) Mouse Ethogram. Available at: https://mousebehavior.org/ethogram/ [Accessed August 17, 2022].

Geuther BQ, Peer A, He H, Sabnis G, Philip VM, Kumar V (2021) Action detection using a neural network elucidates the genetics of mouse grooming behavior. Elife 10:e63207.

Hastings MH, Maywood ES, Brancaccio M (2018) Generation of circadian rhythms in the suprachiasmatic nucleus. Nat Rev Neurosci 19:453–469.

Hu W-P, Li J-D, Colwell CS, Zhou Q-Y (2011) Decreased REM sleep and altered circadian sleep regulation in mice lacking vasoactive intestinal polypeptide. Sleep 34:49–56.

Jones JR, Simon T, Lones L, Herzog ED (2018) SCN VIP Neurons Are Essential for Normal Light-Mediated Resetting of the Circadian System. J Neurosci 38:7986–7995.

Joye DAM, Evans JA (2022) Sex differences in daily timekeeping and circadian clock circuits. Semin Cell Dev Biol 126:45–55.

Jud C, Schmutz I, Hampp G, Oster H, Albrecht U (2005) A guideline for analyzing circadian wheel-running behavior in rodents under different lighting conditions. Biol Proced Online 7:101–116.

Krizo JA, Mintz EM (2014) Sex differences in behavioral circadian rhythms in laboratory rodents. Front Endocrinol 5:234.

Kuljis DA, et al. (2013) Gonadal- and sex-chromosome-dependent sex differences in the circadian system. Endocrinology 154:1501–1512.

Lein ES, et al. (2007) Genome-wide atlas of gene expression in the adult mouse brain. Nature 445:168–176.

Loh DH, Abad C, Colwell CS, Waschek JA (2008) Vasoactive intestinal peptide is critical for circadian regulation of glucocorticoids. Neuroendocrinology 88:246–255.

Loh DH, et al. (2014) Disrupted reproduction, estrous cycle, and circadian rhythms in female mice deficient in vasoactive intestinal peptide. J Biol Rhythms 29:355–369.

Mazuski C, et al. (2018) Entrainment of Circadian Rhythms Depends on Firing Rates and Neuropeptide Release of VIP SCN Neurons. Neuron 99:555–563.e5.

Pendergast JS, Branecky KL, Yang W, Ellacott KLJ, Niswender KD, Yamazaki S (2013) High-fat diet acutely affects circadian organisation and eating behavior. Eur J Neurosci 37:1350–1356.

Pereira TD, Shaevitz JW, Murthy M (2020) Quantifying behavior to understand the brain. Nat Neurosci 23:1537–1549.

Perreau-Lenz S, Pévet P, Buijs RM, Kalsbeek A (2004) The biological clock: the bodyguard of temporal homeostasis. Chronobiol Int 21:1–25.

Pond HL, Heller AT, Gural BM, McKissick OP, Wilkinson MK, Manzini MC (2021) Digging behavior discrimination test to probe burrowing and exploratory digging in male and female mice. J Neurosci Res 99:2046–2058.

Refinetti R, Lissen GC, Halberg F (2007) Procedures for numerical analysis of circadian rhythms. Biol Rhythm Res 38:275–325.

Schroeder A, Loh DH, Jordan MC, Roos KP, Colwell CS (2011) Circadian regulation of cardiovascular function: a role for vasoactive intestinal peptide. Am J Physiol Heart Circ Physiol 300:H241–H250.

Schwartz WJ, Zimmerman P (1990) Circadian timekeeping in BALB/c and C57BL/6 inbred mouse strains. J Neurosci 10:3685–3694.

Todd WD, et al. (2020) Suprachiasmatic VIP neurons are required for normal circadian rhythmicity and comprised of molecularly distinct subpopulations. Nat Commun 11:4410.

Walton JC, Bumgarner JR, Nelson RJ (2022) Sex differences in circadian rhythms. Cold Spring Harb Perspect Biol 14: 039107.

Wu G, Anafi RC, Hughes ME, Kornacker K, Hogenesch JB (2016) MetaCycle: an integrated R package to evaluate periodicity in large scale data. Bioinformatics 32:3351–3353.

Yang R, Su Z (2010) Analyzing circadian expression data by harmonic regression based on autoregressive spectral estimation. Bioinformatics 26:i168–i174.

Yerushalmi S, Green RM (2009) Evidence for the adaptive significance of circadian rhythms. Ecol Lett 12:970–981.

Yu Y-Q, Barry DM, Hao Y, Liu X-T, Chen Z-F (2017) Molecular and neural basis of contagious itch behavior in mice. Science 355:1072–1076.

Zhang C, Li H, Han R (2020) An open-source video tracking system for mouse locomotor activity analysis. BMC Res Notes 13:48.

Zielinski T, Moore AM, Troup E, Halliday KJ, Millar AJ (2014) Strengths and limitations of period estimation methods for circadian data. PLoS One 9:e96462.