

Cognition and Behavior

Hierarchical Compression Reveals Sub-Second to Day-Long Structure in Larval Zebrafish Behavior

Marcus Ghosh and Jason Rihel

<https://doi.org/10.1523/ENEURO.0408-19.2020>

Department of Cell and Developmental Biology, University College London, London WC1E 6BT, United Kingdom

Abstract

Animal behavior is dynamic, evolving over multiple timescales from milliseconds to days and even across a lifetime. To understand the mechanisms governing these dynamics, it is necessary to capture multi-timescale structure from behavioral data. Here, we develop computational tools and study the behavior of hundreds of larval zebrafish tracked continuously across multiple 24-h day/night cycles. We extracted millions of movements and pauses, termed bouts, and used unsupervised learning to reduce each larva's behavior to an alternating sequence of active and inactive bout types, termed modules. Through hierarchical compression, we identified recurrent behavioral patterns, termed motifs. Module and motif usage varied across the day/night cycle, revealing structure at sub-second to day-long timescales. We further demonstrate that module and motif analysis can uncover novel pharmacological and genetic mutant phenotypes. Overall, our work reveals the organization of larval zebrafish behavior at multiple timescales and provides tools to identify structure from large-scale behavioral datasets.

Key words: behavioral dynamics; sleep; zebrafish

Significance Statement

Behavior is dynamic and not only can change from one second to the next but also can unfold over many hours or even days. Understanding how behavior is organized on these different timescales is a critical task in neuroscience, because the constraints on and patterns of behavior offer important clues about the underlying computations being performed in the brain. The analysis tools we develop in this manuscript and apply from sub-second to day-long larval zebrafish behavior expands our understanding of how behavioral patterns change at multiple timescales. The computational metrics we describe can now be used to understand the behavioral consequences of psychotropic drugs or genetic lesions associated with neurodevelopmental and neuropsychiatric disorders.

Introduction

To survive, animals must coordinate patterns of action and inaction in response to their environment. These actions and inactions, which together we will define as behavior, result from some function incorporating internal (e.g., transcriptional, hormonal, or neuronal activity) and external (e.g., time of day or temperature) state. Thus, behavioral descriptions provide insight into the underlying mechanisms that control behavior and are a necessary step in understanding these systems (Krakauer et al., 2017).

Animal behavior, however, typically has many degrees of freedom and evolves over multiple timescales from milliseconds (Wiltschko et al., 2015) to days (Proekt et al., 2012; Fulcher and Jones, 2017) and even across an animal's entire lifespan (Jordan et al., 2013; Stern et al., 2017). As such, quantitatively describing behavior remains both conceptually and technically challenging (Berman, 2018; Brown and de Bivort, 2018). Inspired by early ideas from ethology (Lashley, 1951; Tinbergen, 2010), one approach is to describe behavior in terms of

Received October 4, 2019; accepted February 26, 2020; First published April 2, 2020.

The authors declare no competing financial interests.

Author contributions: M.G. and J.R. designed research; M.G. performed research; J.R. contributed unpublished reagents/analytic tools; M.G. and J.R. analyzed data; M.G. and J.R. wrote the paper.

simple modules that are arranged into more complex motifs. Behavioral modules are often defined from postural data as stereotyped movements, such as walking in *Drosophila* (Berman et al., 2014; Vogelstein et al., 2014; Robie et al., 2017) and mice (Wiltschko et al., 2015), while behavioral motifs are defined as sequences of modules, which capture the patterns inherent to animal behavior, such as grooming in *Drosophila* (Berman et al., 2014, 2016).

Zebrafish larvae have emerged as a powerful model organism in neuroscience, owing to their genetic tractability (Howe et al., 2013), translucency (Vanwalleghem et al., 2018), and amenability to pharmacological screening (Rihel and Ghosh, 2015). In terms of behavior, larvae exhibit an alternating sequence of movements and pauses, termed bouts. This structure is particularly suited to modular description as individual bouts can be easily segmented and it is relatively easy to acquire many examples from even a single animal due to the high frequency of their movement (Kim et al., 2017). Leveraging these advantages, recent work used unsupervised learning to uncover a locomotor repertoire of 13 swim types in larval zebrafish, including slow forward swims and faster escape swims (Marques et al., 2018). However, the inactive periods between swim bouts were not considered, despite reflecting behavioral states such as passivity in the face of adversity (Mu et al., 2019) or even sleep (Prober et al., 2006).

To explore an animal's full behavioral repertoire, from fast movements to sleep, it is necessary to study behavior over long timescales. To date, however, module and motif descriptions of behavior have been developed from videos 15 min (Vogelstein et al., 2014; Wiltschko et al., 2015; Robie et al., 2017) to 2 h (Marques et al., 2018) in length. Consequently, most identified behavioral structure has been on the order of milliseconds, and the existence of longer-timescale structure, on the order of minutes to hours has remained largely unexplored. The development of methods to extract multi-timescale structure from long-timescale recordings would open avenues to explore questions about how behavior varies across the day/night cycle and develops across an animal's lifespan. Furthermore, as pharmacologically or genetically induced behavioral phenotypes can differ at different times of the day/night cycle in zebrafish larvae (Rihel et al., 2010a; Hoffman et al., 2016), a long-timescale approach would provide valuable phenotyping information.

Currently, the limiting factor in scaling these methods is the volume of data, owing to the high framerates and

dimensionality required to estimate animal posture. Here, we present an alternative approach in which we trade dimensionality for scale by building a module and motif description of larval zebrafish behavior from a one-dimensional behavioral parameter recorded over time. Specifically, we used a high-throughput behavioral set-up (Rihel et al., 2010b) to continuously monitor the activity of hundreds of zebrafish larvae across multiple days and nights. To identify multi-timescale behavioral structure, we developed a three-step computational approach. First, we used unsupervised learning to identify a set of 10 behavioral modules that describe both active and inactive bout structure. Second, we applied a compression algorithm (Nevill-Manning and Witten, 2000) to our module data to compile a library of almost 50,000 motifs, revealing behavioral patterns organized across sub-second to minute timescales. Finally, we used a supervised learning algorithm (Peng et al., 2005) to identify motifs from the library used at particular times of the day/night cycle. To test the ability of our approach to detect biologically relevant phenotypes, we also studied the behavior of larvae exposed to the seizure-inducing drug, pentylenetetrazol (PTZ; Baraban et al., 2005), the sedating drug, melatonin (Zhdanova et al., 2001), and mutants of the hypocretin receptor (*hcrtr*; Yokogawa et al., 2007), the loss of which is associated with narcolepsy in humans (Lin et al., 1999) and altered bout structure in zebrafish (Yokogawa et al., 2007; Elbaz et al., 2012). We found that our computational approach could readily detect both compound dose and mutant specific differences in module and motif usage, demonstrating the biological relevance of our behavioral description.

Ultimately, our work reveals the organization of larval zebrafish behavior at sub-second to day-long timescales and provides new computational tools to identify structure from large-scale behavioral datasets.

Materials and Methods

Animal husbandry

Adult zebrafish were reared by University College London (UCL) Fish Facility on a 14/10 h light/dark cycle (lights on: 9 A.M. to 11 P.M.). To obtain embryos, pairs of adult males and females were isolated overnight with a divider that was removed at 9 A.M. the following morning. After a few hours, fertile embryos were collected and sorted under a bright-field microscope into groups of 50 embryos per 10-cm Petri dish filled with fresh fish water (0.3 g/l Instant Ocean). Plates were kept in an incubator at 28.5°C on a 14/10 h light/dark cycle. Using a Pasteur pipet under a bright-field microscope, debris was removed from the plates and the fish water replaced each day. All work was in accordance with the United Kingdom Animal Experimental Procedures Act (1986) under Home Office Project License 70/7612 awarded to J.R.

Behavioral setup

For all behavioral experiments a Pasteur pipet was used to transfer single zebrafish larvae [aged 4–5 d post fertilization (dpf)] into the individual wells of a clear 96-square well plate (7701-1651; Whatman); then each well was

This work was supported by a Medical Research Council Doctoral Training grant (M.G.), a UCL Excellence fellowship (J.R.), a Wellcome Trust Investigator Award (J.R.), and a European Research Council starting grant (J.R.).

Acknowledgements: We thank members of the University College London (UCL) fish floor for insightful discussions, Ida Barlow and François Kroll for comments on this manuscript, and the UCL Fish Facility and staff for support.

Correspondence should be addressed to Jason Rihel at j.rihel@ucl.ac.uk.

<https://doi.org/10.1523/ENEURO.0408-19.2020>

Copyright © 2020 Ghosh and Rihel

This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

filled with 650 μ l of fish water. For experiments longer than 24 h, larvae were plated at 4 dpf and tracking was started the same day. For the duration of these experiments, evaporated fish water was replaced each morning between 9 and 9:30 A.M. For the wild-type experiments, each plate was covered with a plastic lid (4311971; Applied Biosystems) to prevent evaporation and to negate the need to replenish the fish water. For the 24-h small molecule experiments (melatonin and PTZ), larvae were plated at 5 dpf and the plates were left overnight in a 28.5°C 14 h/10 h light/dark incubator. The following morning each plate was transferred to a behavior setup where larvae were dosed, between 9 and 10 A.M., immediately after which behavioral recordings were started and run for 24 h. Following two control rows, each row was dosed with a successively higher concentration of test compound. Larvae were not fed either before or during experiments.

To record each animal's behavior, each plate was placed into a Zebrabox (ViewPoint Life Sciences) running quantization mode with the following settings: detection sensitivity, 15; burst, 50; and freezing, 4. All experiments were conducted on a 14/10 h light/dark cycle (lights on at 9 A.M. to 11 P.M.) with constant infrared illumination. All experiments were recorded at 25 Hz. Larvae were tracked continuously for 24–73 h, after which all larvae unresponsive to touch with a 10- μ l pipette tip were presumed sick or dead and excluded from subsequent analysis. In the wild-type experiments we additionally excluded wells in which a bubble had formed under the plastic lid. The number of larvae excluded/used for each set of experiments is as follows: wild type, 164/288 (all exclusions were due to bubble artefacts); hypocretin, 0/180; melatonin, 0/96; PTZ, 44/96 (doses of 10, 15, and 30 mM were lethal to all tested larvae over 24 h). Following each experiment, larvae were euthanized with an overdose of 2-phenoxyethanol (Acros Organics).

Fish lines

The term wild type refers to the AB x TUP LF zebrafish strain. This line was used for the wild-type experiments, as well as the melatonin and PTZ dose-response curves. *hcrtr* (ZFIN ID: hu2098; Yokogawa et al., 2007) experiments were conducted on embryos collected from heterozygous in-crosses, with larvae genotyped using KASP primers (LGC Genomics) post-tracking. KASP results were validated by comparison to PCR-based genotyping of samples from each KASP classified genotype.

hcrtr genotyping

DNA extraction

Following each *hcrtr* experiment, each larva was euthanized in its well (as above) and DNA was extracted using HotSHOT DNA preparation (Truett et al., 2000). Larval samples were transferred to the individual wells of a 96-well PCR plate. Excess liquid was pipetted from each well before applying 50 μ l of 1 \times base solution (1.25 M KOH and 10 mM EDTA in water). Plates were heat sealed and incubated at 95°C for 30 min, then cooled to room temperature before the addition of 50 μ l of 1 \times neutralization solution (2 M Tris-HCl in water).

PCR

The following reaction mixture per sample was prepared on ice in a 96-well PCR plate: 18.3 μ l PCR mix (2 mM MgCl₂, 14 mM, pH 8.4, Tris-HCl, 68 mM KCl, 0.14% gelatin in water, autoclaved for 20 min, cooled to room temperature, chilled on ice, then we added: 1.8% 100 mg/ml BSA and 0.14% 100 mM d [A, C, G, T] TP), 0.5 μ l of forward and reverse primers (20 μ M), 5.5 μ l water, 0.2 μ l of Taq polymerase, and 3.0 μ l of DNA. Next, each plate was heat sealed and placed into a thermocycler, set with the following program: 95°C, 5 min; 44 cycles: 95°C, 30 s; 57°C, 30 s; and 72°C, 45 s; then 72°C, 10 min and 10°C until collection. Finally, samples were mixed with 6 \times loading buffer [colorless buffer: Ficoll-400, 12.5 g, Tris-HCl (1 M, pH 7.4) – 5 ml, EDTA (0.5 M) – 10 ml, to 50 ml in pure water; heated to 65°C to dissolve, per 10 ml of colorless buffer 25 mg of both xylene cyanol and orange G were added, then diluted to 6 \times] and run on agarose gels (1% – 2%) with 4% GelRed (Biotium).

hcrtr forward primer: 5'-CCACCCGCTAAAATTCAAAAGCACTGCTAAC-3'; *hcrtr* reverse primer: 5'-CATCACAGCGGTGAACAGG-3'.

PCR information

PCR products were digested with DdeI at 37°C to produce a 170-bp band in the wild-type animals and 140- and 30-bp bands in *hcrtr* mutants.

KASP

KASP genotyping was conducted in white, low profile PCR plates on ice with six wells allocated 50:50 for positive and negative controls. The following reaction mixture was prepared per sample: 3.89 μ l of 2 \times KASP reaction mix, 0.11 μ l KASP primers, 1.0 μ l water, and 3.0 μ l DNA. Plates were then heat sealed and placed into a thermocycler with the following thermal cycling program: 94°C, 15 min; 10 cycles: 94°C, 20 s; 61–53°C (dropping 0.8°C per cycle), 60 s; 26 cycles: 94°C, 20 s; 53°C, 60 s, then 10°C until collection.

Following thermal cycling we used a fluorescence reader (Bio-Rad CFX96 Real-Time System) and Bio-Rad CFX Manager software (version 3.1) to automatically determine each sample's genotype from a 2D scatter plot of fluorescence in each channel. From this scatter plot, outlying samples of unclear genotype were manually excluded from subsequent analysis.

KASP assay ID: 554-0090.1

KASP Flanking Sequence (alternative allele shown in square brackets, with a forward slash indicating a deletion in the alternative allele): 5'-ACCGCTGGTATGCGATCTGCCACCCGCTAAAATTCAAAAGCACTGCTAAA[A/T]GAGCCGCAAGAGCATC GTGCTGATCTGGCTGGTGTCCCTGCATCATGATG-3'.

Pharmacology

0.15 M melatonin and 1 M PTZ (M5250 and P6500; Sigma) stock solutions were made in DMSO and sterile water, respectively. Behavioral testing concentrations for each compound were selected based on Rihel et al. (2010a). For behavior experiments, each animal in a well

with 650 μ l of fish water was dosed with 1.3 μ l of either vehicle control or compound at 500 \times concentration, resulting in a 1 in 500 dilution and thus the desired testing concentration.

Computing

Hardware

A desktop computer with 16 GB of RAM was used for most data analysis, figure production, and writing. For two-time intensive steps, hierarchical compression of full module sequences (Batch_Compress.m) and normalizing the behavioral motif counts (Batch_Grammar_Freq.m), data were run in parallel, with a worker for every animal, on the UCL Legion Cluster (Research Computing Services, UCL).

Software

All software used for data handling, analysis, and the production of figures is available at https://github.com/ghoshm/Structure_Paper. The codes are also available as Extended Data 1 and Extended Data 2.

Processing behavioral data

See Extended Data Figure 1-3 for a flow diagram describing behavioral data acquisition and analysis. All custom behavioral analysis software was written and run in MATLAB 2016b-2018a (MathWorks). The suffixes .m and .mat denote MATLAB code and MATLAB data files, respectively.

Behavioral data were recorded by subtracting subsequent pairs of frames from each other and determining the number of pixels that changed intensity within each well between each pair of frames, termed Δ pixels. To acquire behavior data, each Zebrafish was setup using ViewPoint's ZEBRALAB software (version 3.22), which outputs a .xls and a .raw file (ViewPoint specific format) per experiment. Each behavior .xls file was reorganized into a .txt file using the function perl_batch_192.m (Jason Rihel). For each experiment a .txt metadata file assigning each animal to an experimental group, for example genotype, was manually produced. To replicate the previous analysis methodology, as in Extended Data Figure 1-1D, behavior and metadata .txt files were input to the function sleep_analysis2.m (Jason Rihel).

To assess data on a frame by frame basis, each experiment's .raw file, which was output from Viewpoint's Zebrafish, was exported within the ZEBRALAB software to thousands of .xls files. Each .xls file contained 50,000 rows and 21 columns, with data from any given well listed approximately every 192 rows, as the setup always assumes recordings are from two 96-well plates. This formatting is, however, only approximate as infrequently the well order is erroneously non-sequential; these rows were termed ordering errors. Each .xls file is formatted with 21 columns, of which three contain useful data: type, notes when ViewPoint defined data acquisition errors occurred; location, denotes which well the data came from; and data1, records the Δ pixel value from that well for that time point.

The function Vp_Extract.m was used to reformat the .xls files from each experiment to single frame by fish matrices, from which each animal's behavior was quantified. Vp_Extract.m requires three inputs to be selected: a folder

containing the .xls files; a .txt behavior file output from perl_batch_192.m; and a .txt metadata file. To ensure that each animal has the same number of frames, frames with ViewPoint defined errors or ordering errors (which are automatically detected by Vp_Extract.m) are discarded. A maximum Δ pixels value can be set and active bouts containing even a single frame with a higher Δ pixels value than this are set to zero for the entire duration of the bout. Here a maximum Δ pixels threshold of 200 was set. This value was determined from manual inspection of the dataset as well as by comparisons of these data to data recorded from plates with no animals in. Following this step, the maximum Δ pixels value observed was 165. Most Δ pixels values were between 3 (minimum observed value) and 30 (Extended Data Fig. 1-2A). Time periods during which water is being replenished are automatically detected and set to a Δ pixels value of zero. These time periods are noted and excluded from later analysis. The function outputs .mat files for subsequent analysis. Either single or multiple .mat files output from Vp_Extract.m were input to Vp_Analyse.m and Bout_Clustering.m.

Vp_Analyse.m was used to compare general activity levels and bout features across time and between groups. The function has two options. The first allows for specific days and nights of interest to be cropped from the data. The second determines how experimental repeats are handled, treating the data as either a single merged dataset or as separate datasets. In the latter case, each experimental repeat is plotted with the same color scheme as the first experiment, with progressive shading for each repeat. Additionally, the N -way ANOVA comparisons include a repeat factor, which can be used to determine whether results are consistent across experimental repeats. Vp_Analyse.m outputs two statistics results structures: twa, N -way ANOVA comparison results; and kw, two-sample Kolmogorov-Smirnov test results. Vp_Analyse.m outputs figures showing each group's activity (Fig. 1D,E) and bout features (Extended Data Fig. 1-2) over time.

The script Bout_Clustering.m was used to cluster all active and inactive bouts into behavioral modules, as well as to compare the resultant modules. To cluster the data an evidence accumulation approach is used (Fred and Jain, 2002, 2005) implemented by the custom MATLAB function gmm_sample_ea.m. Bout_Clustering.m produces figures (Extended Data Fig. 2-1) and statistically compares the modules. The MATLAB workspace output from Bout_Clustering.m can be input to either Bout_Transitions.m or Bout_Transitions_Hours.m.

The function gmm_sample_ea.m clusters data using an evidence accumulation approach (Fred and Jain, 2002, 2005) through which the results of multiple Gaussian mixture models are combined to generate an aggregate solution. This process is executed through the following six steps. First, a sample of "probe points" are randomly sampled from the data. The number of probe points to sample is user defined. Second, values of K and sample sizes are uniformly sampled from user set ranges. The values of K are used to set the number of mixture components for each mixture model. The sample sizes determine the number of points, randomly sampled from

the data that each mixture model is fit to. Third, a Gaussian mixture model is iteratively fit to the sampled data with K components. Each probe point is assigned to the component with the highest corresponding posterior probability and evidence is accumulated on the probe points; evidence is defined as pairwise co-occurrences in the same component. Fourthly, the evidence accumulation matrix is hierarchically clustered, and the final number of clusters is determined by using the maximum differentiated linkage distance to cut the resultant dendrogram. The linkage metric used is a user-defined option. Fifthly, the clusters are normalized for size by randomly sampling the number of points in the smallest cluster, from each cluster. Finally, all data points are assigned to these final size normalized clusters using the mode cluster assignment of the k -nearest neighbors, with k being user defined.

The script `Bout_Transitions.m` takes the MATLAB workspace output from `Bout_Clustering.m` as an input and compresses each animal's full module sequence to generate a library of behavioral motifs. The number of occurrences of each motif are counted and normalized by comparison to paired shuffled data. Finally, a supervised learning algorithm is applied to identify context specific behavioral motifs. For two-time intensive steps, hierarchical compression of full module sequences (`Batch_Compress.m`) and normalizing the behavioral motif counts (`Batch_Grammar_Freq.m`), data were manually copied (via MobaXterm, Personal Edition v10.5) to UCL Legion Cluster (Research Computing Services, UCL) and processed in parallel with a worker for every fish. MATLAB code for hierarchical compression is described in Gomez-Marin et al. (2016). MATLAB code for submitting these jobs to Legion, analyzing data and retrieving results is available at https://github.com/ghoshm/Legion_Code. Ultimately, `Bout_Transitions.m` outputs a library of behavioral motifs and motif-related figures (Fig. 3).

The script `Bout_Transitions_Hours.m` compresses blocks of 500 modules for statistical comparisons, uses the motif library from `Bout_Transitions.m` to count the occurrence of each motif every hour, normalizes these counts to paired shuffled data and finally uses supervised learning to identify hour specific behavioral motifs. As with `Bout_Transitions.m` behavioral motifs are normalized, via `Batch_Grammar_Freq.m`, using UCL Legion Cluster. `Bout_Transitions_Hours.m` outputs figures (Fig. 4D) and statistics.

Processed behavioral data are available at: <https://zenodo.org/record/3344770#.XYyWYyhKiUk>. Raw data are available upon request.

Behavioral data analysis

Δ Pixels

At the acquisition stage, Δ pixels data were filtered by the software (ViewPoint) such that each frame for a given well was scored as either zero or higher. In the absence of movement within a well, and hence no pixels changing intensity, Δ pixels values of zero were recorded. These periods were termed inactive bouts and were defined as any single or consecutive frames with Δ pixels values equal to zero. The length of each inactive bout was used as a descriptive feature. When there were movements

within a well, Δ pixels values greater than zero were recorded. These periods were termed active bouts and were defined as any single or consecutive frames with Δ pixels values greater than zero. Six features were used to describe each active bout: length, mean, SD, total, minimum, and maximum. These features, as well as the number of active bouts, percentage of time spent active and total Δ pixels activity, were compared between conditions, e.g., day and night and dose of drug, in two ways using the function `Vp_Analyse.m`.

To compare the distribution of values for each feature between conditions, a probability density function (pdf) was fit to each animal's data and the mean shape of each condition's pdf was compared using a two-sample Kolmogorov–Smirnov test (Extended Data Fig. 1-2A). To compare each feature's average values between conditions, mean values were taken from each animal, and N -way ANOVA was computed. The following factors, when relevant, were included and full interaction terms were calculated: condition, e.g., mutant and wild type; time, e.g., day and night; development, defined as a consecutive day and night; and experimental repeat, i.e., which experimental repeat a datapoint came from. For experiments with multiple repeats, the lack of an interaction effect between the comparison of interest and experimental repeat factor was considered as evidence of a consistent result.

Note that our analysis relies on periods of inactivity registering Δ pixels values of zero. Consequently, to be compatible with our analysis code, other data may require post-recording filtering.

Clustering

To cluster the bouts, the script `Bout_Clustering.m` was used. First, matrices of bouts by features were constructed (active matrix, $30,900,018 \times 6$; inactive matrix, $30,900,418 \times 1$). To prepare the active data for clustering, each animal's data were individually normalized by calculating z scores using Equation 1, which illustrates how every bout (i) from each animal (f) was normalized by first subtracting the mean of this animal's bout features (\bar{x}_f) from the bout and then dividing by the SD of each bout feature for this animal σ_f :

$$Z_i = \frac{x_i - \bar{x}_f}{\sigma_f}. \quad (1)$$

Active bout features across all animals were then centered by subtracting each feature's mean value from every bout, and principal component analysis (PCA) was used to reduce the data to three dimensions, the knee point of the scree plot, which together explain 97.5% of the variance (Extended Data Fig. 2-1A).

Next, the active and inactive bouts were separately clustered using an evidence accumulation-based approach (Fred and Jain, 2002, 2005) implemented by the function `gmm_sample_ea.m`. First, 40,000 probe points were randomly sampled from the data. Next, for 200 iterations, another group of points were randomly sampled and fit with a Gaussian mixture model with a random number of clusters. For each iteration, these two parameters

varied uniformly in the following ranges: the number of points sampled, 40,000–100,000; the number of clusters fit, 2–20. Each mixture model was fit using MATLAB's fitgmdist function (MATLAB, Statistics and Machine Learning Toolbox) with full, regularized, independent covariance matrices and initialized using the k-means++ algorithm (Arthur and Vassilvitskii, 2007). Each mixture model was fit five times and the one with the largest log-likelihood was retained. Once each model had been fit, each probe point was assigned to the component with the largest posterior probability, and evidence in the form of pairwise occurrence in the same cluster was accumulated on the probe points. Once the 200 mixture models had been fit, average link clustering was applied to the evidence accumulation matrix and the final number of clusters determined based on maximum cluster lifetime. Next, the resultant clusters were normalized for size by randomly selecting the number of points in the smallest cluster from each cluster (5983 active, 614 inactive bouts). Finally, all points were assigned to the size normalized clusters using the mode cluster assignment of the 50 nearest neighbors for every point.

Hierarchical compression

Clustering reduced each animal's behavior to a non-repetitive sequence of active and inactive bouts, termed modules. On average this reduced each wild-type sequence length by 96%, from 6,308,514 frames to 236,636 modules, easing the computational demands of compressing these sequences.

To compress modular sequences, an offline compressive heuristic (Nevill-Manning and Witten, 2000) was used (Eq. 2). At each iteration (i) of the algorithm, the most compressive motif was defined as the motif which made the most savings, a balance between the length of the motif (W) and the number of times it occurred in the sequence (N), which also considered the combined cost of adding a new motif to the dictionary ($W + 1$) and of introducing a new symbol into the sequence ($+N$) at every occurrence of this motif in the sequence:

$$\text{Savings}_i = WN - (W + 1 + N). \quad (2)$$

The overall compressibility of a given input sequence was calculated by summing these savings across all iterations and dividing this total by the length of the original input sequence (in modules). This process resulted in a compressibility metric that ranged from 0 to 1 (low-high compressibility). To reduce computational time, motifs of a maximum of 10 modules long were sought, although the hierarchical nature of the algorithm enabled the identification of longer motifs through nesting. To generate the common motif library, the motifs obtained from compression of every animal's full module sequence (Batch_Compress.m) were merged, and then all unique motifs were kept (Bout_Transitions.m). To generate sets of paired control sequences for every animal, each animal's module sequence was divided into sequential day and night or hourly segments and the modules within each of these windows was shuffled 10 times, maintaining the active/inactive transition structure (Bout_Transitions.m). As compressibility varies non-linearly with uncompressed

sequence length (Extended Data Fig. 3-1B), to enable comparisons between samples with different numbers of modules, non-overlapping blocks 500 modules long were compressed (Bout_Transitions_Hours.m).

Supervised motif selection

To identify both which and how many motifs were required to distinguish between behavioral contexts (e.g., day and night), the following approach was executed by the function Batch_Grammar_Freq.m. First, the number of occurrences of every motif from the common motif library was counted in every real and shuffled modular sequence. Next, to calculate enrichment/constraint scores for every motif, the deviation of the real from shuffled counts, as well as the deviation of each shuffle from the other shuffles, was calculated (Eq. 3). For a given animal and time window, e.g., day or night, the mean number of times motif (i) was counted in the shuffled data (\bar{s}_i) was subtracted from the real number of counts (x_i) and divided by the SD of the shuffled counts (σ_{si}):

$$Z_i = \frac{x_i - \bar{s}_i}{\sigma_{si}}. \quad (3)$$

For example, for a given animal and time window, if a motif was used 10 times in the real data and a mean of 4.5 times in the 10 sets of paired shuffled data (SD 1.58), enrichment/constraint would equal 3.48:

$$3.48 = \frac{10 - 4.5}{1.58}.$$

When comparing the shuffled data to itself, each shuffle (now x_i) was excluded from \bar{s}_i and σ_{si} . Infinite values occurred when there was no SD in the σ_{si} counts and thus σ_{si} equaled zero. To facilitate subsequent working, infinite values were replaced with a constant value of $\sqrt{1 + \text{number of shuffled counts}} = \sqrt{11} = \pm 3.32$. Note that in the real data, infinite values constituted only 2.2% of all enrichment/constraint scores.

For any given comparison, motif library enrichment/constraint scores for the relevant animals were formatted into a matrix of samples by motifs (Fig. 4B). Scores for each motif (column) were normalized by subtracting each column's mean score and dividing by each column's SD. A supervised feature selection algorithm (Peng et al., 2005) was applied to these matrices to select the top 250 maximally relevant and minimally redundant (mRMR) motifs. For each comparison we defined the first motif chosen by mRMR as the best motif. To determine the best motifs among test conditions, we excluded control data. For example, to determine the best 0.01 μM melatonin motif, we compared this group's data to all larvae dosed with melatonin at other doses (Fig. 5A). For completeness, we still plot the control enrichment/constraint scores for these motifs.

To determine how many of the 250 motifs chosen by mRMR were necessary for accurate classification, linear discriminant analysis classifiers were trained on data using 10-fold cross validation as sequential mRMR motifs were added, and classification error mean and SD were calculated. The MATLAB function fitcdiscr (Statistics and

Machine Learning Toolbox) was used to implement these steps. Finally, to determine how many motifs were necessary for a given comparison, classification error curves were smoothed with a running average three motifs wide and the number of motifs at which the minimum classification error occurred was identified (Extended Data Fig. 4-1A). To evaluate classifier performance, the results of each classifier were compared with a majority class classifier whose performance depended on the ratio of samples of each class. For example, in a dataset with two labels at a ratio of 0.1:0.9, the majority class classifier would consistently assign the latter label and achieve a classification error of 10% (\pm SE of proportion). Additionally, we compared each classifier's performance to a set of 10 classifiers built using the same number of motifs, although randomly selected. For example, for a classifier which achieved its minimal classification error using 15 motifs, we randomly selected 15 motifs and trained classifiers as above. We repeated this process 10 times per classifier and report the error and SD across these 10 repeats.

Results

Behavior at scale

Larval zebrafish behavior consists of an alternating sequence of movements and pauses, termed bouts, that are organized at sub-second timescales. To capture this structure from high-throughput, long-timescale experiments, we used a 96-well plate set-up with a single larva housed in each well (Extended Data Fig. 1-1A) and as a proxy for movement recorded the number of pixels that changed intensity within each well between successive pairs of frames, a metric we term Δ pixels. We built on previous work using this set-up (for review, see Barlow and Rihel, 2017; Oikonomou and Prober, 2017) by analyzing Δ pixels data at 25 Hz, rather than in 1-min bins. When recorded in this way, Δ pixels data are an alternating sequence of positive values representing movement magnitude and zeros representing periods of inactivity (Fig. 1A; Movie 1). We defined active bouts as any single or consecutive frames with non-zero Δ pixels values and described each bout using several features including the mean and SD of Δ pixels values across the bout (Fig. 1A). We defined inactive bouts as any single or consecutive frames with zero Δ pixels values and described each inactive bout using its length (Fig. 1A). We found no correlation of active bout initiation between larvae in adjacent wells of the 96-well plate (Extended Data Fig. 1-1B), suggesting no interactions among larvae.

Using this approach, we first assessed the behavior of wild-type larvae across a 14/10 h day/night cycle (Extended Data Fig. 1-2A). During the day, wild-type larvae initiated bouts at a mean frequency of 0.89 Hz (Fig. 1B), a rate consistent with other assays (Kim et al., 2017), and tended to use short, sub-second long inactive bouts (Fig. 1C). During the night, larvae displayed a mean bout frequency of only 0.07 Hz (Fig. 1B) and used longer inactive bouts, on the order of seconds to minutes (Fig. 1C). Together, these differences resulted in a diurnal pattern of activity (Fig. 1D). These results are broadly consistent

with those from analysis of binned Δ pixels data (Barlow and Rihel, 2017; Oikonomou and Prober, 2017), with the addition of sub-second resolution and an increase in accuracy, as determined by intra-fish comparisons between the methods (Extended Data Fig. 1-1C,D).

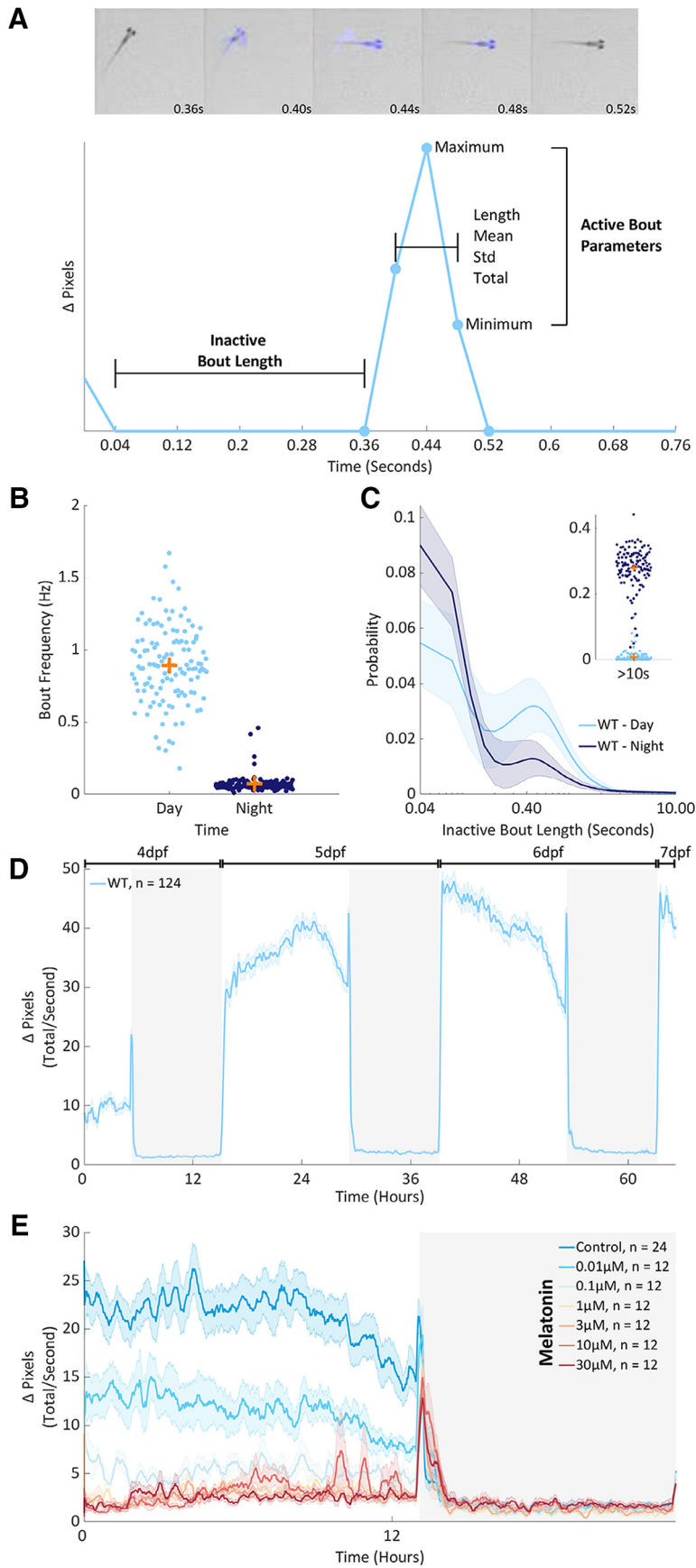
Next, we extended our approach to examine the behavioral effects of pharmacological and genetic manipulations. Melatonin, which is strongly hypnotic in zebrafish (Rihel et al., 2010a), dose dependently decreased larval activity (Fig. 1E) by decreasing the number, magnitude, and length of active bouts and by inducing longer inactive bouts (Extended Data Fig. 1-2B). The epileptogenic drug PTZ (Extended Data Fig. 1-1E) altered both active and inactive bout parameters (Extended Data Fig. 1-2C), eliciting on average longer, lower amplitude active bouts and longer inactive bouts during the day. Finally, homozygous *hcrtr*^{-/-} mutants had only subtle differences in active bout structure, with shorter mean active bout length and lower active bout total and SD, compared with both wild-type *hcrtr*^{+/+} and heterozygous *hcrtr*^{+/-} siblings, which did not differ from one another by any metrics (Extended Data Fig. 1-2D).

Collectively, these results quantitatively demonstrate the advantages of assessing Δ pixels data on a frame by frame basis and provide insight into the behavior of wild-type zebrafish larvae across the day/night cycle as well as those subject to pharmacological or genetic manipulations.

Module usage varies with behavioral context

Recent work has demonstrated that larval activity can be classified using unsupervised learning into 13 distinct bout types that represent different swimming movements (Marques et al., 2018). A full description of larval behavior, however, requires quantification of both the movements and pauses that they execute. Thus, we sought to determine whether distinct active or inactive bout types, which we termed modules, were identifiable from our data, and if module usage depended on behavioral context.

To address these questions, we separately clustered the active and inactive bouts (combined across experiments a total of 30,900,018 active and 30,900,418 inactive bouts) using an evidence accumulation-based clustering algorithm (see Materials and Methods). In brief, 200 Gaussian mixture models were built from each data set, then the results of these models were combined to generate aggregate solutions. This clustering method identified five active and five inactive modules (Fig. 2A,B; Extended Data Fig. 2-1), which we separately labeled from 1 to 5 from the shortest to longest mean bout length. The active modules, which formed discrete peaks in bout feature space (Extended Data Fig. 2-1B), corresponded to different shapes of Δ pixel changes in terms of amplitude and length (Fig. 2A; Extended Data Fig. 2-2A), while the inactive modules consisted of different lengths of inactivity (Fig. 2B; Extended Data Fig. 2-2A). The shortest inactive module (module 1) had a mean length of 0.06 s and ranged from a minimum of 0.04 s (our sampling limit) to a maximum of 0.12 s. In contrast, the longest inactive module (module 5) had a mean length of 96 s and covered a



continued

Figure 1. Behavior at scale. **A**, top panel, Five consecutive frames from an individual well of a 96-well plate as a 6 dpf zebrafish larva performs a swim bout. Blue highlights pixels that change intensity between frames (Δ pixels). Lower panel. A Δ pixels time series from the larva above. Highlighted are the features that describe each active and inactive bout. **B**, Mean bout frequency (Hz) recorded from individual larvae at 5 and 6 dpf during the day (light blue) and the night (dark blue). Each dot is 1 of 124 wild-type larvae. The orange crosses mark the population means. **C**, The probability of observing different lengths of inactivity during the day (light blue) or the night (dark blue) at 5 and 6 dpf. Each larva's data were fit by a pdf. Shown is a mean pdf (bold line) and SD (shaded surround) with a log scale on the x-axis cropped to 10 s. Inset, The total probability of inactive bout lengths longer than 10 s, per animal. **D**, The mean activity of 124 wild-type larvae from 4 to 7 dpf, on a 14/10 h light/dark cycle. Data for each larva was summed into seconds and then smoothed with a 15-min running average. Shown is a summed and smoothed mean Δ pixels trace (bold line) and SEM (shaded surround). **E**, Average activity across one day (white background) and night (dark background) for larvae dosed with either DMSO (control) or a range of melatonin doses immediately before tracking at 6 dpf. Data were summed and smoothed as in **D**. The number of animals per condition is denoted as n . Extended Data Figures 1-1, 1-2, 1-3 support Figure 1.

huge range of values from a minimum of 20 s to a maximum of 8.8 h.

To examine how module usage varied across time, we represented each larva's behavior as an alternating sequence of active and inactive modules (Fig. 2C; Movie 2). In the wild-type data, module usage varied with time of day (Fig. 2D). For example, the probability of observing inactive module 2, which consists of typical day pause lengths (0.16–1.16 s), was on average 0.6 during the day and only 0.24 during the night, when inactive modules 1, 4, and 5 became more likely (Fig. 2D). To reveal finer-grain temporal dynamics, we also examined each module's mean frequency over time (Fig. 2E). In general, both the active and the short inactive modules had high frequencies during the day, peaking at the light/dark transition as the larvae responded to the sudden change in illumination. In contrast, the only module with a peak in frequency at the dark-to-light transition was inactive module 4 (3.72–20 s), which also had an increased frequency approaching the light-to-dark transition. Together, these results reveal that zebrafish employ different bout types in a time of day/night dependent manner.

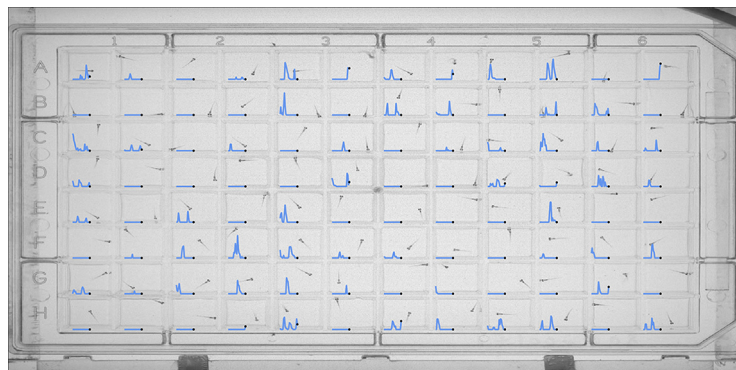
Next, we examined the impact of pharmacological and genetic manipulations on bout type usage. Larvae dosed with melatonin showed a shift toward using shorter active modules and longer inactive modules (Extended Data Fig. 2-2B). In PTZ-dosed larvae, there were also shifts in active module probability. Particularly notable was the complete exclusion of active module 1 in 27 of the 28 (96.4%) PTZ-dosed larvae, while control larvae used this module

with 0.12 probability during the day and 0.22 during the night (Extended Data Fig. 2-2C). These shifts likely reflect the chaotic, seizure-like swimming observed in PTZ-treated larvae (Baraban et al., 2005), although no single active module clearly captured these behavioral seizures. PTZ also increased the probability of the shortest inactive (module 1) as well as the two longest inactive modules (modules 4 and 5), the latter of which are likely to correspond to the interictal bouts of inactivity associated with seizures (Extended Data Fig. 2-2C). Conversely, *hcrtr* mutants exhibited no differences in either active or inactive module probabilities compared with their wild-type siblings (Extended Data Fig. 2-2D), demonstrating that bout type usage is similar between these mutants and wild-type animals across the day/night cycle.

Collectively, these results reveal that zebrafish behavior in this assay can be described by five types of active and five types of inactive modules, the usage of which varies with behavioral context. Interestingly, in many contexts, both active and inactive module probabilities were shifted, suggesting that these module types may co-vary, perhaps by being arranged into recurrent sequences.

Hierarchical compression reveals structure in zebrafish behavior

From a set of behavioral modules, an animal could structure their behavior in a range of ways. At one end of this spectrum, successive modules could be organized completely randomly, such that prior modules exert no influence on future module selection. At the other end,



Movie 1. High-throughput behavioral tracking. A video of 96, 6 dpf zebrafish larvae swimming in our rig. The last 1 s of each larva's Δ pixels data is plotted over each well. This video was filmed at 25 Hz and is played back in real time. [View online]

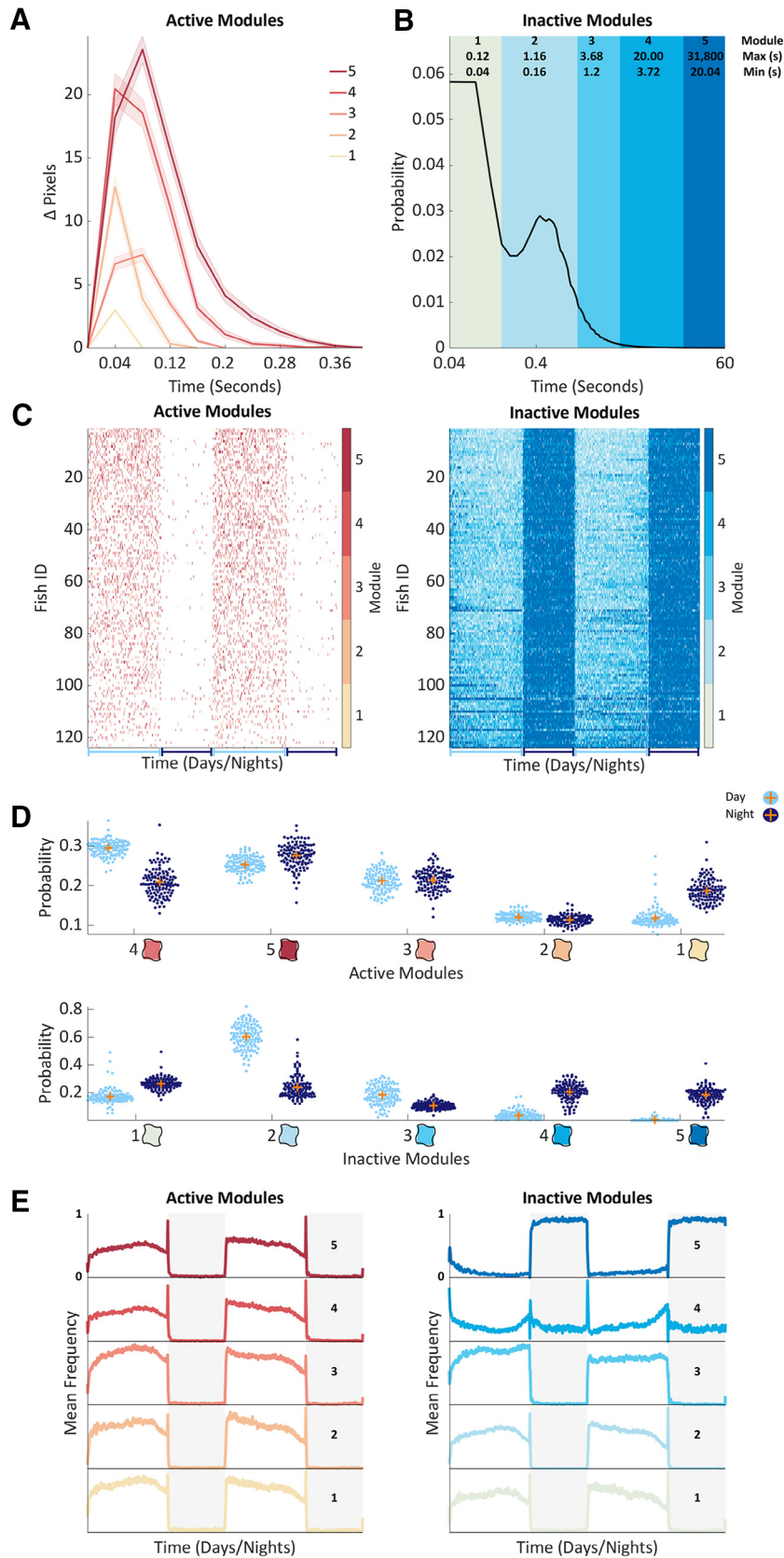


Figure 2. Unsupervised learning identifies contextual behavioral modules. **A**, Average Δ pixels changes for each active module. Shown is the mean (bold line) and SEM (shaded surround) of 100 bouts randomly sampled from each module from one representative larva. Modules are numbered and colored by average module length across all animals, from shortest (1) to longest (5). **B**, A

continued

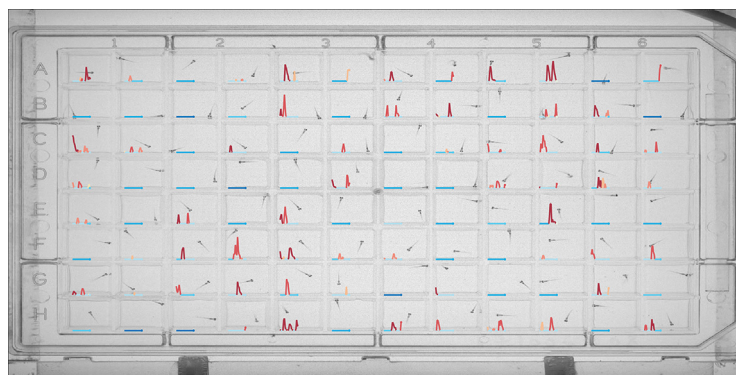
probability density curve showing the distribution of inactive bout lengths in seconds, on a log x-axis cropped to 60 s. Modules are numbered and colored from shortest (1) to longest (5) mean length (see legend for each module's minimum and maximum bout length). **C**, Matrices showing the active (left) or inactive (right) module assignment of every frame (x-axis) for each of 124 wild-type larvae (y-axis) across the 14-h days (light blue underlines) and 10-h nights (dark blue underlines) from 5 to 6 dpf. Larvae were sorted by total number of active modules from highest (top) to lowest (bottom). Modules are colored according to the adjacent colormaps. **D**, Average active (upper) and inactive (lower) module probability during day (light blue) and night (dark blue) 5 and 6 of development. Each of 124 wild-type animals is shown as a dot and orange crosses mark the population means. Active modules are sorted by mean day probability from highest to lowest (left to right). Inactive modules are sorted by mean length from shortest to longest (left to right). The blobs correspond to the color used for each module in other figures. **E**, The mean frequency of each active (left) and inactive (right) module across days 5 and 6 of development. Shown is a mean smoothed with a 15-min running average, re-scaled to 0–1. Days are shown with a white background, nights with a dark background. Modules are sorted from shortest to longest (lower to upper panels). Extended Data Figures 2-1, 2-2 support Figure 2.

module selection could be fully deterministic with a particular module always following another. Rather than being fixed, however, it is likely that animals adapt their behavioral structure in response to changing internal or external states. We sought to map the structure of zebrafish behavior in different contexts by examining the presence and organization of module sequences, which could provide insight into the mechanisms governing behavior. To do this, we used a compression algorithm (Nevill-Manning and Witten, 2000) as Gomez-Marin et al. (2016) used to discover structure in *Caenorhabditis elegans* postural data. When applied to our dataset (Fig. 3A), this algorithm iteratively identified motifs from each larva's modular sequence and returned two outputs, compressibility, a measure of each larva's behavioral structure, and a library of identified recurrent module sequences, termed motifs.

To quantify the structure of zebrafish behavior, we first compressed every animal's full modular sequence, which in wild-type animals were on average 236,636 modules long across 70 h. To determine whether the resultant compression values indicated more structure than would be expected based on either the distribution or the transition structure of the active-to-inactive modules, we compared each larva's compressibility to that of 10 sets of paired shuffled data. All wild-type larvae were more compressive than their paired shuffled data, demonstrating that their behavior is more structured than expected from modular probabilities alone (Extended Data Fig. 3-1A). Compressibility, however, varies non-linearly with input

sequence length, as longer sequences will be more likely to contain motifs (Extended Data Fig. 3-1B). Thus, to enable comparisons between samples with different numbers of modules, we compressed non-overlapping 500 module blocks of sequence per larva. This approach revealed that compressibility was higher during the day than the night (Fig. 3B). To determine whether these differences were primarily due to the presence of behavioral motifs or instead were a consequence of differences in module distribution, we also compared the difference in compressibility (Δ compressibility) between each animal's real and shuffled data. This approach revealed that the compressibility difference between the day and the night is predominantly due to differences in module distribution (Extended Data Fig. 3-1E). To reveal finer-grain temporal changes in compressibility, we plotted Δ compressibility across time (Extended Data Fig. 3-1F). This approach revealed peaks at the light-to-dark transitions in the evenings, consistent with this stimulus eliciting stereotyped behavioral sequences (Burgess and Granato, 2007; Emran et al., 2010).

Next, we used compressibility to assess how our pharmacological and genetic manipulations altered the structure of larval behavior. We found that melatonin decreased day compressibility to night-time levels (Fig. 3B), a difference primarily due to shifts in module distribution (Extended Data Fig. 3-1E). PTZ increased compressibility to a constant day/night value (Fig. 3B). PTZ, however, reduced Δ compressibility (Extended Data Fig. 3-1E), indicating that changes in module distribution, rather than motif usage, are the



Movie 2. Behavioral modules. A video of 96, 6 dpf zebrafish larvae swimming in our rig. The last 1 s of each larva's Δ pixels data is plotted over each well, with each active and inactive bout colored according to its module assignment. This video was filmed at 25 Hz and is played back in real time. [View online]

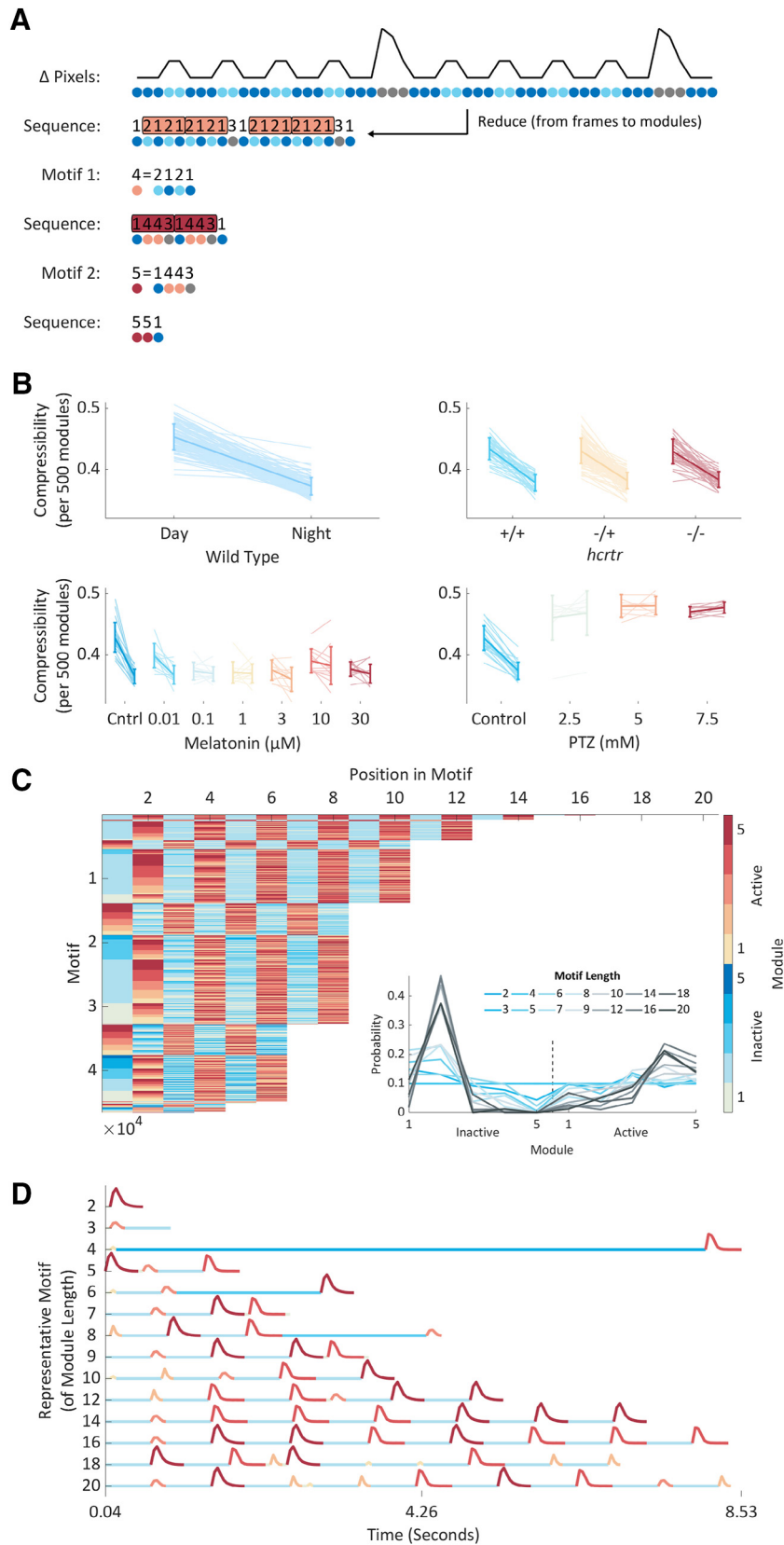


Figure 3. Hierarchical compression reveals structure in zebrafish behavior. **A**, Compression explained using fictive data. Top to bottom, From Δ pixels data (black trace), we classified both active and inactive behaviors into modules (colored circles). From modular behavioral sequences, we identified motifs (sequences of modules) using a compression algorithm. Compression iteratively

continued

identifies motifs (shown as boxes) by replacing them with new symbols until no more motifs can be identified and the sequence is maximally compressed. **B**, Each panel shows how compressibility, calculated from 500 module blocks, varies in different behavioral contexts. Each pale line shows an individual fish's mean compressibility during the day and the night. The darker overlay shows a population day and night mean \pm SD. In the wild-type data, compressibility is higher during the day than the night ($p < 10^{-158}$) and increases from day/night 5–6 ($p < 10^{-4}$), findings consistent across triplicate experiments. Melatonin decreases ($p < 10^{-10}$), while PTZ increases compressibility ($p < 10^{-8}$). There is no effect of *hcrtr* genotype on compressibility. Statistics are two-way or four-way ANOVA. **C**, All 46,554 unique motifs (y -axis) identified by compressing data from all animals. Each motif's module sequence is shown, with the modules colored according to the colormap on the right. Motifs are sorted by length and then sequentially by module. Motifs range in length from 2 to 20 modules long. Inset, For each motif length, the probability of observing each inactive or active module. **D**, Each motif in the library consists of an alternating sequence of Δ pixels changes and pauses (active and inactive modules). A representative motif of each module length is shown with each module colored according to the colormap in **C**. Representative motifs were chosen by determining every motif's distribution of modules and then for each observed module length, selecting the motif closest to the average module distribution (see **C**, inset). Extended Data Figure 3-1 supports Figure 3.

dominant driver of PTZ-induced behavioral changes. Importantly, these drug-induced changes in compressibility do not simply reflect overall activity levels. For example, PTZ-exposed larvae are less active than controls during the day and more active during the night (Extended Data Fig. 1-1E) but have consistently higher compressibility (Fig. 3B). Finally, in *hcrtr* mutants, we found no differences in either compressibility or Δ compressibility, suggesting that *hcrtr* mutant behavior is structured similarly to wild-type animals (Fig. 3B).

Overall, our compressibility results demonstrate that larval zebrafish behavior is structured and that the amount of structure varies with behavioral context. Our Δ compressibility results suggest that such variation predominantly arises from altered module, rather than motif, usage.

To gain insight into the behavioral sequences larvae deploy, we then studied the motifs identified by the compression algorithm. Compression of the real modular sequences identified a mean of 1901 motifs per animal (Extended Data Fig. 3-1C). Interestingly, compression of the real data almost always identified slightly fewer motifs than the shuffled data (Extended Data Fig. 3-1C). This suggests that the motifs identified from the real data were used more frequently than those in the shuffled data and therefore likely reflect enriched behavioral sequences. Merging the motifs identified across all animals generated a library of 46,554 unique behavioral motifs (Fig. 3C). In terms of raw Δ pixels data, each motif represented an approximately repeated pattern of movements and pauses of varying length (Fig. 3D). Motifs in the library ranged from two to 20 modules long with a median length of eight modules and spanned timescales from approximately 0.1 s to 11.3 min with a median length of 3.84 s. In the real data, longer motifs tended to be used more frequently than in the shuffled data (Extended Data Fig. 3-1D), demonstrating the likely behavioral relevance of these sequences. Motifs of different module lengths used distinct subsets of modules (Fig. 3C). For example, motifs comprised of longer module sequences had a lower probability of using long inactive modules. Together, these results reveal the varied timescales at which zebrafish larvae organize their behavior and suggest the presence of structure governing the arrangement of modules into motifs.

Behavioral motif usage is time dependent

The large number of motifs in our library led us to hypothesize that each may be used in specific behavioral contexts. To test this hypothesis, we counted the number of times each larva or set of paired shuffled data used each motif within each time frame (e.g., day or night) and then normalized these counts by calculating the deviation of the real from shuffled counts as well as the deviation of each shuffle from the other shuffles, a metric we termed enrichment/constraint. Overall, we found that enrichment/constraint scores from our real data were more prone to extreme positive (enriched) and negative (constrained) values than the shuffled data (Fig. 4A), suggesting that a minority of behavioral motifs were used more or less frequently than would be expected by chance.

To test whether these extremes occurred in particular contexts, we first compared motif usage between the day and the night in wild-type larvae by generating a matrix of enrichment/constraint scores (Fig. 4B). To distill the most salient motifs from this and other contextual matrices, we used a three-step approach. First, we used the mRMR algorithm (Peng et al., 2005) to rank the motifs from most to least salient. Second, we trained linear discriminant analysis classifiers using 10-fold cross validation as we iteratively increased the number of input motifs from most to least salient (e.g., motif 1, motif 1 and 2, motif 1 – n). Finally, we selected the subset of motifs which achieved the lowest classification error between groups in each context. To determine how accurately these motif subsets could distinguish between behavioral contexts, we compared each classifier's performance to that of a majority class classifier, which performed as well as the ratio of samples between the two contexts. For example, in the day versus night classification, a majority class classifier would have an error rate of 50% (\pm SE of proportion), as each larva contributes an equal number of days and nights to the enrichment/constraint matrix. Additionally, to demonstrate the salience of the motifs selected by the mRMR algorithm, we compared each classifier's performance to a set of 10 classifiers built using the same number of motifs, although randomly selected. For example, for a classifier that achieved its minimal classification error using 50 motifs, we randomly selected 50 motifs from the library and built a classifier. For each comparison we repeated this process 10 times.

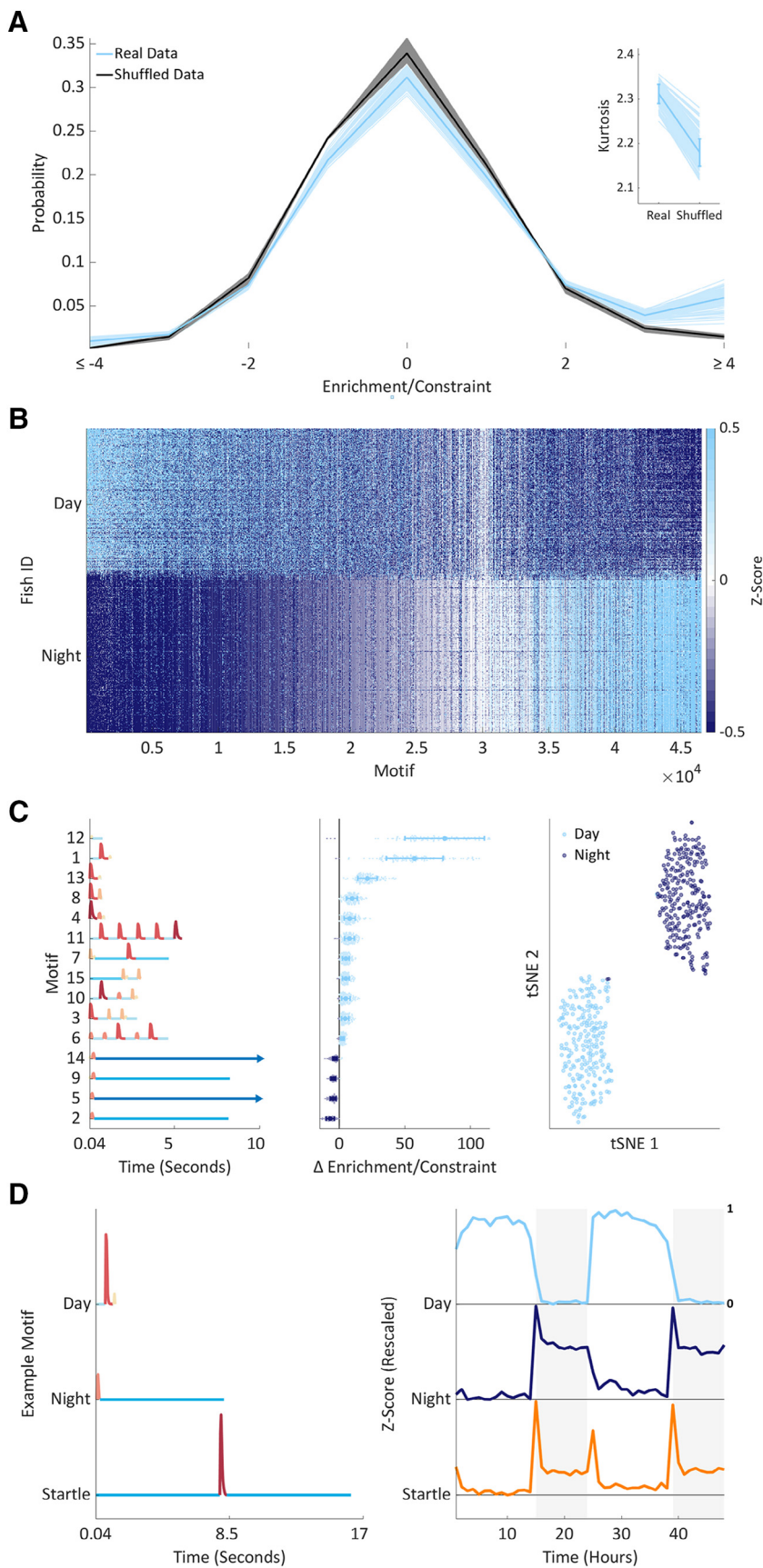


Figure 4. Supervised learning identifies contextual behavioral motifs. **A**, pdfs showing the probability of observing motifs at different

continued

enrichment/constraint scores rounded to whole numbers and summed at values above or below ± 4 for ease of visualization. Each wild-type animal is depicted by a single pale blue (real data) and 10 black (shuffled data) lines; overlaid in bold are mean pdfs. The inset shows that the kurtosis of the real data are higher than the shuffled data ($p < 10^{-271}$; two-way ANOVA, real vs shuffled data, no significant interaction with experimental repeat factor). Each larva is shown as a pale line; overlaid is a population mean and SD. **B**, Enrichment/constraint scores for all 46,554 motifs (x -axis) for each fish during day/night 5 and 6 of development (y -axis). To emphasize structure, motifs are sorted in both axes, first by their average day/night difference (from day to night enriched left to right), then separately day and night by larva. Finally, each motif's enrichment/constraint score is Z-scored to aid visualization. **C**, left, The 15 day/night mRMR motifs module sequences are shown numbered by the order in which they were selected by the algorithm. Motifs are sorted by day minus night enrichment/constraint score (middle). The long pauses at the end of motifs 5 and 14 are cropped at 10 s (arrows). Middle, For each selected motif (y -axis), ordered as in the left panel, each wild-type animal's (124 in total) day minus night enrichment/constraint score (x -axis) is shown as a dot. Values above zero are colored light blue; below zero are dark blue. Overlaid is a population mean and SD per motif. Right, A tSNE embedding of the 15-dimensional motif data (middle) into a two-dimensional space. Each circle represents a single day (light blue) or night (dark blue) sample. **D**, Representative motif temporal dynamics; shown are motifs 1 (day) and 2 (night) from **C**, as well as a startle-like motif. Left, Each motif's module sequence. Right, Each motif's mean enrichment/constraint score each hour, rescaled to 0–1. Extended Data Figure 4-1 supports Figure 4.

Applying this algorithm to wild-type data revealed changes in motif usage across multiple timescales (Extended Data Fig. 4-1B). We found that only 15 motifs were required to classify day-specific and night-specific behavior with only a 0.2% ($\pm 0.63\%$ SD) classification error, compared with a majority class classifier with 50% error and random 15-motif subsets with a mean error of 9.25% (Fig. 4C; Table 1). The day-enriched motifs consisted of high amplitude movements interspersed with short pauses, while the night enriched motifs contained low amplitude movements and long pauses (Fig. 4C). Next, we examined how motif usage changed over time by comparing consecutive days and nights (5–6 dpf). In both day 5 versus day 6 and night 5 versus night 6 comparisons, the classifiers achieved roughly 20% error using 93 and 85 motifs, respectively (Table 1). Thus, motif usage shifted over just 24 h, although these changes were far less prominent than those between the day and night. To study whether motif usage varied at finer timescales, we first divided the day into morning/evening and the night into early/late periods. In each case, the mRMR algorithm performed better than the control classifiers (morning/evening: 33%, early/late night: 36%) although the relatively high classification errors suggest that motif selection did not vary strongly across each day or night (Table 1). Consistent with this conclusion, classifiers attempting to delineate each hour from every other mostly failed to outperform their control classifiers (Table 1). The two notable exceptions were the hour following each lighting transition, where this approach achieved good classification performance (Table 1) and identified startle-like motifs. We interpret these motifs as startle-like as they consist of long pauses interrupted by high-amplitude movements and primarily occur at the lighting transitions (Fig. 4D), stimuli known to elicit startle behaviors in larval zebrafish (Burgess and Granato, 2007; Emran et al., 2010).

Together, these results demonstrate that motif usage varied between the day and the night, but aside from the lighting transitions, was relatively consistent within these periods.

Dose-dependent and dose-specific behavioral motifs

Finally, we hypothesized that behavioral motif usage would vary dose dependently across concentrations of

melatonin and PTZ, providing insight into the mechanisms by which these compounds exert their behavioral effects. Motif dose dependency would suggest a continuously modulated underlying process, which might arise for example if the fraction of bound receptors relates to neuronal activity modulation. Alternatively, motifs enriched at only specific doses would suggest discrete effects on neuronal circuitry.

Applying the mRMR algorithm to our pharmacological data revealed both dose-dependent and dose-specific modulation of motif usage. We found that each melatonin dose could be separated from the others using 40–250 motifs with only 0–2.78% classification error (Fig. 5A; Table 2). Focusing on just the best motif for each comparison, we observed both dose dependency as well as dose specificity. For example, comparing controls to all melatonin-dosed larvae identified a dose-dependent motif that consisted of large magnitude movements and short pauses, whose enrichment/constraint score decreased with increasing melatonin concentration (Fig. 5A). Conversely, the best 10 μM motif, two long pauses broken by a small active bout sequence, showed dose specificity being enriched at only 3 and 10 μM doses (Fig. 5A). When applied to the PTZ data, our approach performed even more accurately, achieving perfect classification (0% error) between all conditions (Fig. 5B; Table 2). Furthermore, in PTZ-dosed larvae we observed enrichment for motifs highly constrained in wild-type larvae, highlighting the usage of motifs beyond the normal wild-type repertoire, such as those corresponding to behavioral seizures (Fig. 5B).

Next, we tested whether our motif subset approach could detect *hcrtr* mutant phenotypes that were not easily captured by other methods. For example, based on human and rodent literature, where loss of hypocretin is associated with narcolepsy (Lin et al., 1999) and prior zebrafish literature (Elbaz et al., 2012), we expected abnormal transitions between active and inactive bouts. We found reasonable performance when discriminating between *hcrtr*^{+/+} and *hcrtr*^{-/-} during both the day ($16.7 \pm 7.5\%$ error with 195 motifs) and night ($12.8 \pm 9.6\%$ error with 53 motifs) but weaker performance when distinguishing between *hcrtr*^{+/+} and *hcrtr*^{-/+}, as expected for a haploinsufficient gene (Extended Data Fig. 4-1C; Table 2). Thus, homozygous loss of *hcrtr* impacts motif usage

Table 1: Wild-type motif classifier performance Download Figure 4-1, TIF file.

Comparison	Motifs (number)	Cv error (%)	Cv error Std	Mc error (%)	Mc EP	RM error (%)	RM error Std
Wild type							
Day/night	15	0.20	0.63	50.0	2.25	9.25	2.50
Day 5/day 6	93	20.2	9.60	50.0	3.18	45.8	4.09
Night 5/night 6	85	19.8	8.09	50.0	3.18	48.3	6.00
Day hours							
● 09-10	102	6.39	1.23	7.14	0.44	7.67	0.26
● 10-11	1	7.37	0.31	7.14	0.44	7.18	0.10
● 11-12	5	7.20	0.23	7.14	0.44	7.15	0.03
● 12-13	9	7.06	0.34	7.14	0.44	7.19	0.08
● 13-14	1	7.14	0.12	7.14	0.44	7.14	0
● 14-15	1	7.14	0.12	7.14	0.44	7.14	0
● 15-16	1	7.11	0.14	7.14	0.44	7.14	0
● 16-17	1	7.09	0.15	7.14	0.44	7.14	0
● 17-18	1	7.14	0.12	7.14	0.44	7.14	0
● 18-19	1	7.14	0.12	7.14	0.44	7.14	0
● 19-20	1	7.14	0.12	7.14	0.44	7.14	0
● 20-21	3	7.11	0.27	7.14	0.44	7.15	0.01
● 21-22	1	7.14	0.12	7.14	0.44	7.14	0
● 22-23	1	7.14	0.12	7.14	0.44	7.15	0.03
Night hours							
● 23-24	23	0.69	0.47	10.0	0.60	6.57	1.23
● 24-01	177	9.84	1.83	10.0	0.60	11.6	0.78
● 01-02	5	9.92	0.51	10.0	0.60	10.0	0.05
● 02-03	88	9.72	1.18	10.0	0.60	10.4	0.31
● 03-04	1	10.0	0.17	10.0	0.60	10.0	0.04
● 04-05	22	9.92	0.47	10.0	0.60	10.1	0.07
● 05-06	1	10.0	0.17	10.0	0.60	10.0	0
● 06-07	1	10.0	0.17	10.0	0.60	10.0	0.02
● 07-08	3	9.84	0.34	10.0	0.60	10.0	0.01
● 08-09	1	10.0	0.17	10.0	0.60	10.0	0.01
Morning/evening	229	33.2	2.32	50.0	0.85	44.7	0.96
Early/late night	26	36.4	2.18	50.0	1.00	43.4	1.55

A table showing the performance of each wild-type motif classifier. Each classifier sought to separate the data shown in the comparison column, e.g., day/night. For the hourly comparisons, each hour was compared with data from all other hours grouped together. For each comparison, 250 motifs were chosen by mRMR, then a smaller number were retained (see motifs column) based on classification error curves (Extended Data Fig. 4-1A). Cv, 10-fold cross validated; Std, SD across the 10 folds; Mc, majority class classifier; EP, SE of proportion; RM, classifiers built from random motif subsets.

enough to allow for successful classification of *hcrtr*^{-/-} mutants, although no single *hcrtr*^{-/-} motifs with large differences in enrichment/constraint scores compared with wild-type siblings were particularly evident.

Collectively, these results demonstrate that behavioral motifs are used context dependently and reveal how motif subsets can parse subtle differences in motif usage between behavioral contexts. However, does motif analysis provide additional discriminatory power over module selection, which also varies between behavioral contexts? To assess this, we compared the performance of each motif classifier to paired module classifiers built from matrices of module probabilities. All of the motif classifiers achieved better performance than their module pairs (Fig. 5C; Table 3), demonstrating both the phenotyping value of the motifs and their importance in the structure of larval behavior.

Discussion

Here, we developed and applied computational tools to describe high-throughput, long-timescale behavioral data in terms of behavioral units (modules) and sequences of modules (motifs) organized across sub-second to day-long timescales.

Low-dimensional representations of behavior

Low-dimensional representations of behavior, such as the Δ pixels metric employed here, result in a loss of information, for example, direction of movement or posture. However, such metrics do facilitate screening approaches and/or long-timescale tracking and in these contexts have provided biological insight into the molecular targets of small molecules (Rihel et al., 2010a) and genetics of aging (Churgin et al., 2017). Our work builds on previous long-timescale studies of behavior by assessing sub-second resolution Δ pixels data across multiple days and nights. This improved resolution enabled the segmentation and parameterization of individual active and inactive bouts from our data, revealing how larvae adapt their behavior across the day/night cycle and how behavior is impacted by small molecules.

Future work should aim to extend our assay by recording more detailed behavioral measures. Indeed, a recent study using centroid tracking in 96-well plates revealed that larvae show a day/night location preference within the well and furthermore uncovered a mutant with a difference in this metric (Thyme et al., 2019), demonstrating that even within the confined space of a 96-well plate, location is an informative metric to record. It is likely that

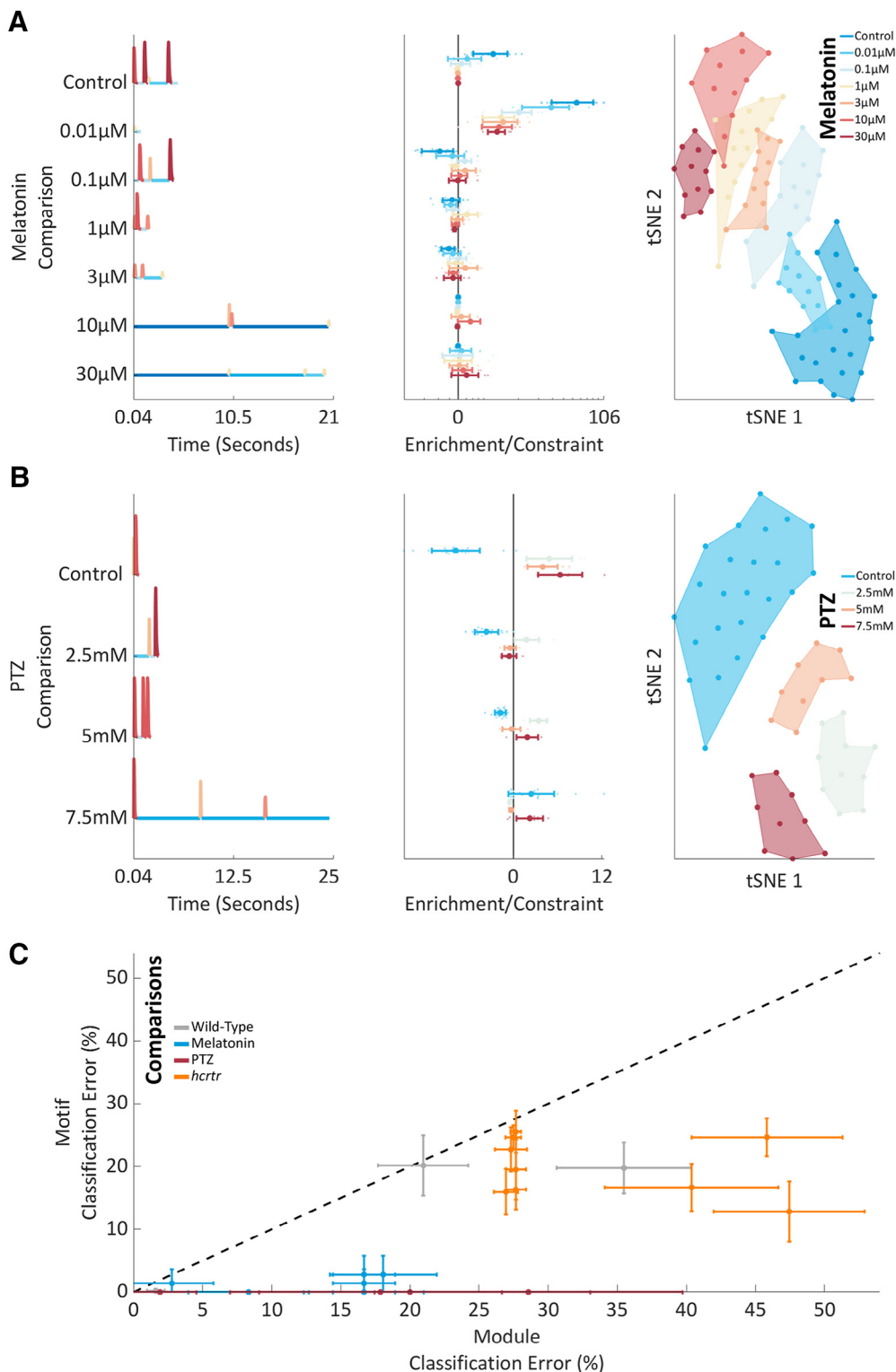


Figure 5. Pharmacological behavioral motifs. **A**, left, Module sequences for the single best motif for each melatonin comparison. Modules are colored as elsewhere. Middle: for each dose’s single best motif, see left panel y-axis for dose, enrichment/constraint scores are shown for every dose on a log x-axis. Each animal is shown as a dot, with a mean \pm std overlaid per dose. Right, A two-dimensional tSNE embedding from a space of 912 unique motifs. Each animal is shown as a single dot overlaid by a shaded boundary encompassing all animals in each condition. **B**, left, Module sequences for the single best motif for each PTZ comparison. To highlight a seizure specific motif, the control motif and corresponding enrichment/constraint score is shown as mRMR motif 2, not 1, for this comparison. Modules are colored as elsewhere. Middle, For each dose’s single best motif, enrichment/constraint scores are shown for every dose on a linear x-axis. Each animal is shown as a dot, with a mean and SD overlaid per dose. Right, A two-

continued

dimensional tSNE embedding from a space of 338 unique motifs. Each animal is shown as a single dot underlaid by a shaded boundary encompassing all animals in each condition. **C**, Each classifier's classification error (%) is shown in terms of modules (x -axis) and motifs (y -axis). Data are shown as mean and SD from 10-fold cross validation. Classifiers are colored by experimental dataset (see Legend). For reference, $y = x$ is shown as a broken black line. Data below this line demonstrates superior performance of the motif classifiers.

Table 2: *hcrtr* and pharmacological classifier performance Download Figure 4-1, TIF file.

Comparison	Motifs (number)	Cv error (%)	Cv error Std	Mc error (%)	Mc EP	RM error (%)	RM error Std
<i>hcrtr</i>							
Day and night							
● WT/Het	173	25.5	6.77	27.7	1.88	39.0	1.93
● WT/Hom	83	24.7	6.07	50.0	2.83	47.8	3.66
● Het/Hom	235	24.7	3.76	27.7	1.88	38.6	1.83
Day							
● WT/Het	80	19.5	9.60	27.7	2.66	37.9	1.08
● WT/Hom	195	16.7	7.50	50.0	4.00	48.7	4.22
● Het/Hom	55	22.7	7.02	27.7	2.66	33.8	2.01
Night							
● WT/Het	79	16.3	6.38	27.7	2.66	37.1	7.17
● WT/Hom	53	12.8	9.58	50.0	4.00	52.3	6.16
● Het/Hom	76	16.0	7.27	27.7	2.66	36.0	5.05
Melatonin (day)							
● Control	40	0	0	25.0	4.42	16.4	3.67
● 0.01 μ M	89	1.39	4.52	16.7	4.39	30.0	15.2
● 0.1 μ M	192	1.39	4.52	16.7	4.39	20.5	18.0
● 1 μ M	132	2.78	6.02	16.7	4.39	29.5	8.67
● 3 μ M	97	0	0	16.7	4.39	48.6	11.2
● 10 μ M	250	2.78	6.02	16.7	4.39	20.0	9.40
● 30 μ M	133	0	0	16.7	4.39	32.2	7.89
PTZ (day)							
● Control	26	0	0	46.2	6.91	15.8	6.43
● 2.5 mM	55	0	0	35.7	9.06	42.1	11.4
● 5 mM	162	0	0	32.1	8.83	34.3	18.0
● 7.5 mM	104	0	0	32.1	8.83	49.9	14.7

A table showing the performance of each classifier. Each classifier sought to separate the data shown in the comparison column, e.g., *hcrtr*^{+/+} (WT) and *hcrtr*^{-/+} (Het). For the pharmacological comparisons, each condition was compared with the rest of the conditions grouped together, aside from the control data which was excluded. For each comparison, 250 motifs were chosen by mRMR, then a smaller number were retained (see motifs column) based on classification error curves (see Extended Data Fig. 4-1A). Cv, 10-fold cross validated; Std, SD across the 10 folds; Mc, majority class classifier; EP, SE of proportion; RM, classifiers built from random motif subsets; WT, *hcrtr*^{+/+}; Het, *hcrtr*^{-/+}; Hom, *hcrtr*^{-/-}.

even more detailed behavioral measures, like eye and tail angles, will yield additional insights, for example enabling the exploration of rapid-eye-movement sleep in zebrafish larvae, as done in reptiles (Shein-Idelson et al., 2016). Such metrics could be extracted by skeletonization or even through the use of an autoencoder applied to the raw video frames from each well (Johnson et al., 2016). Once such high dimensional data had been assigned to modules, our compression and motif enrichment/constraint approach could be applied in the same manner as here.

Modular descriptions of behavior

A key idea in ethology is that behavior consists of stereotyped modules arranged into motifs (Lashley, 1951; Tinbergen, 2010). While early studies described behavior in this manner through manual observations (Dawkins and Dawkins, 1976), recent advances in machine vision and learning have automated these processes (Todd et al., 2017). For example, in zebrafish larvae, recent work used unsupervised learning to uncover a locomotor repertoire

of 13 swim types including slow forward swims and faster escape swims (Marques et al., 2018), although inactive bouts were not considered. From our dataset, we identified five active and five inactive modules, which, respectively, describe swim bouts of different amplitudes (Fig. 2A) and periods of inactivity of varied length (Fig. 2B). Interestingly, all modules were used with reasonably high and similar probability by all wild-type animals (Fig. 2D), demonstrating that these modules represent a set of common larval behaviors. Furthermore, the temporal (Fig. 2E) and pharmacological (Extended Data Fig. 2-2B,C) shifts in these probabilities illustrates that module usage can be flexibly reorganized depending on behavioral context (Wiltshcko et al., 2015).

To discretize our bouts into modules, we first extracted hand-engineered features from each bout (Fig. 1A) and then applied an evidence accumulation-based clustering algorithm (Fred and Jain, 2002, 2005). While our results demonstrate the relevance and utility of these modules in describing larval behavior, it is possible that our approach missed rare bout types. Consequently, future work should

Table 3: Module classifier performance Download Figure 4-1, TIF file.

Comparison	Modules (number)	Cv error (%)	Cv error Std	Mc error (%)	Mc EP
Wild type					
● Day/Night	10	1.61	1.29	50.0	2.25
● Day 5/Day 6	8	21.0	6.53	50.0	3.18
● Night 5/Night 6	1	35.5	9.71	50.0	3.18
hcrtr					
Day and night					
● WT/Het	1	27.7	0.77	27.7	1.88
● WT/Hom	10	45.8	10.9	50.0	2.83
● Het/Hom	8	27.5	1.12	27.7	1.88
Day					
● WT/Het	1	27.7	1.46	27.7	2.66
● WT/Hom	1	40.4	12.5	50.0	4.00
● Het/Hom	3	27.3	2.35	27.7	2.66
Night					
● WT/Het	1	27.7	1.46	27.7	2.66
● WT/Hom	1	47.4	10.9	50.0	4.00
● Het/Hom	10	27.0	1.72	27.7	2.66
Melatonin (day)					
● Control	3	8.33	8.69	25.0	4.42
● 0.01 μ M	10	2.78	6.02	16.7	4.39
● 0.1 μ M	2	16.7	4.52	16.7	4.39
● 1 μ M	1	18.1	7.74	16.7	4.39
● 3 μ M	1	16.8	8.67	16.7	4.39
● 10 μ M	1	16.8	4.52	16.7	4.39
● 30 μ M	1	16.8	4.52	16.7	4.39
PTZ (day)					
● Control	1	1.92	5.27	46.2	6.91
● 2.5 mM	1	17.9	17.6	35.7	9.06
● 5 mM	1	28.6	22.3	32.1	8.83
● 7.5 mM	10	20.0	26.1	32.1	8.83

A table showing the performance of each module classifier. Each classifier sought to separate the data shown in the comparison column, e.g., wild type, day/night. For each comparison, all 10 modules were sequentially chosen by the mRMR algorithm, then a smaller subset was retained (see module column) based on classification error curves. Cv, 10-fold cross validated; Std, SD across the 10 folds; Mc, majority class classifier; EP, SE of proportion.

build on our bout classification by exploring the benefits of including additional features, the use of alternative clustering algorithms and our assumption of stereotypy, i.e., that all bouts can be fit into a module (Berman, 2018). An alternative direction would be to produce a mapping between our active modules and those identified from analysis of larval posture (Marques et al., 2018). Bridging this gap could facilitate behavioral screening approaches, for example, by using data from our set-up to prioritize pharmacological compounds or mutants for postural analysis.

Quantifying structure in behavior

In some contexts, it is beneficial for animals to execute coordinated patterns of behavior. For example, to efficiently search an environment zebrafish larvae will execute organized sequences of left and right turns (Dunn et al., 2016). In other contexts, more random behavior will be advantageous, such as when escaping from a predator (Maye et al., 2007). Quantifying structure in behavior thus provides insight into the overarching strategy being employed in particular contexts. Alterations in behavioral structure can also manifest clinically, for example in Autism Spectrum Disorder, a defining feature of which is increased behavioral stereotypy (American Psychiatric Association, 2013). Consequently, compression would be

a relevant and likely informative metric to record in animal models or even human cases for such conditions.

To quantify structure in larval zebrafish behavior in different contexts, we inputted each larva's modular sequence to a compression algorithm. We found that wild-type behavior was more compressive during the day than the night (Fig. 3B). This echoes recent work in *Drosophila* that revealed higher temporal predictability during the day than the night as well as in females (Fulcher and Jones, 2017). A likely explanation for these findings comes from work in *C. elegans* (Gomez-Marín et al., 2016) that demonstrated that animals who transition slowly between modules, as both zebrafish (Fig. 1B) and *Drosophila* do at night (Geissmann et al., 2019), tend to be less compressive. This may suggest that the underlying mechanisms controlling longer-timescale behaviors are less precise than those controlling fast behavioral sequences.

For future efforts applying compression to behavioral data, there are two avenues left to explore, what compression heuristic to use and how to compress data from multiple animals. Following the work of Gomez-Marín et al. (2016), we defined the best motif at any iteration as the most compressive, which represents a balance between the motif's length and frequency. While this metric generally leads to the best compression (Nevill-Manning and Witten, 2000), alternative measures, such as frequency or length may capture other aspects of behavior. The

second avenue relates to comparisons between animals. Here, each animal was compressed individually, identifying motifs that were later grouped into a common library. While computationally tractable, this approach prevents certain comparisons across animals, for example identifying the most compressive motif across all larvae. This issue could be solved by compressing a single sequence containing all of the animals' modular sequences joined end to end, with spacers to prevent interanimal motifs. Compressing this long sequence would, however, be computationally demanding.

Compressing and merging the identified motifs across all animals generated a library of 46,554 unique motifs (Fig. 3C), each of which described an alternating sequence of movements and pauses (Fig. 3D). Motifs ranged from 0.1 s to 11.3 min in length, revealing the range of timescales at which larval behavior is organized. We cannot, however, rule out the existence of longer timescale motifs in larval behavior as computational demands limited our search to motifs 10 modules long (although the algorithm's hierarchical approach enabled the identification of motifs up to 20 modules long). Thus, future work should aim to extend our approach to explore the full range of timescales at which larval behavior is organized by systematically varying this parameter.

Contextual behavioral motifs

Finally, by distilling salient subsets of motifs from our library, we demonstrated that motif usage was context dependent and highlighted the discriminatory power of motif subsets, which were capable of distinguishing between day/night behavior and even between small changes in compound dose. Comparing motif usage across the day/night cycle identified a set of highly night specific motifs (Fig. 4C), which may represent sleep behaviors. One way in which future studies could address this possibility would be to deprive larvae of these motifs throughout the night, for example, by using a closed-loop paradigm (Geissmann et al., 2019), and observing the impact on larval behavior the following day. In relation to the PTZ data, comparing seizure motifs across epileptogenic compounds and mutants with spontaneous seizures could suggest clues as to their underlying mechanism (Kokel et al., 2010; Rihel et al., 2010a). For example, seizures with similar motif usage patterns may originate in the same brain area or impact awareness in the same manner. This hypothesis could be tested by generating whole-brain activity maps (Randlett et al., 2015) across conditions, with the aim of identifying common and unique neuronal correlates.

Given the amenability of larval zebrafish to high-throughput behavioral screening (Rihel and Ghosh, 2015) future work should leverage our approach to large-scale genetic (Thyme et al., 2019) or pharmacological datasets (Rihel et al., 2010a). Individually, these datasets would provide information on the genetic and molecular basis of behavior across multiple timescales, encompassing processes from sleep to aging. In combination, by identifying mutant and drug-induced phenotypes that cancel each other out (Lamb et al., 2006; Hoffman et al., 2016), these

datasets could be used to identify phenotypic suppressors in genetic disease models, an outcome with potential clinical relevance.

References

- American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders, Ed 5. Arlington, VA: American Psychiatric Association.
- Arthur D, Vassilvitskii S (2007) k-means++: the advantages of careful seeding. Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms, pp 1027–1035. Philadelphia, PA, USA, Society for Industrial and Applied Mathematics.
- Baraban SC, Taylor MR, Castro PA, Baier H (2005) Pentylentetrazole induced changes in zebrafish behavior, neural activity and c-fos expression. *Neuroscience* 131:759–768.
- Barlow IL, Rihel J (2017) Zebrafish sleep: from geneZZZ to neuronZZZ. *Curr Opin Neurobiol* 44:65–71.
- Berman GJ (2018) Measuring behavior across scales. *BMC Biol* 16:23.
- Berman GJ, Bialek W, Shaevitz JW (2016) Predictability and hierarchy in *Drosophila* behavior. *Proc Natl Acad Sci USA* 113:11943–11948.
- Berman GJ, Choi DM, Bialek W, Shaevitz JW (2014) Mapping the stereotyped behaviour of freely moving fruit flies. *J R Soc Interface* 11:20140672.
- Brown AEX, de Bivort B (2018) Ethology as a physical science. *Nat Phys* 14:653–657.
- Burgess HA, Granato M (2007) Modulation of locomotor activity in larval zebrafish during light adaptation. *J Exp Biol* 210:2526–2539.
- Churgin MA, Jung SK, Yu CC, Chen X, Raizen DM, Fang-Yen C (2017) Longitudinal imaging of *Caenorhabditis elegans* in a micro-fabricated device reveals variation in behavioral decline during aging. *Elife* 6:e26652.
- Dawkins R, Dawkins M (1976) Hierarchical organization and postural facilitation: rules for grooming in flies. *Anim Behav* 24:739–755.
- Dunn TW, Mu Y, Narayan S, Randlett O, Naumann EA, Yang C-T, Schier AF, Freeman J, Engert F, Ahrens MB (2016) Brain-wide mapping of neural activity controlling zebrafish exploratory locomotion. *Elife* 5:e12741.
- Elbaz I, Yelin-Bekerman L, Nicenboim J, Vatine G, Appelbaum L (2012) Genetic ablation of hypocretin neurons alters behavioral state transitions in zebrafish. *J Neurosci* 32:12961–12972.
- Emran F, Rihel J, Adolph AR, Dowling JE (2010) Zebrafish larvae lose vision at night. *Proc Natl Acad Sci USA* 107:6034–6039.
- Fred ALN, Jain AK (2002) Data clustering using evidence accumulation', in Object recognition supported by user interaction for service robots. *IEEE Comput. Soc*, pp 276–280.
- Fred ALN, Jain AK (2005) Combining multiple clusterings using evidence accumulation. *IEEE Trans Pattern Anal Mach Intell* 27:835–850.
- Fulcher BD, Jones NS (2017) hctsa: a computational framework for automated time-series phenotyping using massive feature extraction. *Cell Syst* 5:527–531.e3.
- Geissmann Q, Beckwith EJ, Gilestro GF (2019) Most sleep does not serve a vital function: evidence from *Drosophila melanogaster*. *Sci Adv* 5:eaa9253.
- Gomez-Marin A, Stephens GJ, Brown AEX (2016) Hierarchical compression of *Caenorhabditis elegans* locomotion reveals phenotypic differences in the organization of behaviour. *J R Soc Interface* 13:20160466.
- Peng H, Long F, Ding C (2005) Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Mach Intell* 27:1226–1238.
- Hoffman EJ, Turner KJ, Fernandez JM, Cifuentes D, Ghosh M, Ijaz S, Jain RA, Kubo F, Bill BR, Baier H, Granato M, Barresi MJF, Wilson SW, Rihel J, State MW, Giraldez AJ (2016) Estrogens suppress a behavioral phenotype in zebrafish mutants of the autism risk gene, CNTNAP2. *Neuron* 89:725–733.

- Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L, McLaren S, Sealy I, Caccamo M, Churcher C, Scott C, Barrett JC, Koch R, Rauch GJ, White S, Chow W, et al. (2013) The zebrafish reference genome sequence and its relationship to the human genome. *Nature* 496:498–503.
- Johnson MJ, Duvenaud D, Wiltchko AB, Datta SR, Adams RP (2016) Composing graphical models with neural networks for structured representations and fast inference. arXiv 1603.06277.
- Jordan D, Kuehn S, Katifori E, Leibler S (2013) Behavioral diversity in microbes and low-dimensional phenotypic spaces. *Proc Natl Acad Sci USA* 110:14018–14023.
- Kim DH, Kim J, Marques JC, Grama A, Hildebrand DGC, Gu W, Li JM, Robson DN (2017) Pan-neuronal calcium imaging with cellular resolution in freely swimming zebrafish. *Nat Methods* 14:1107–1114.
- Kokel D, Bryan J, Laggner C, White R, Cheung CYJ, Mateus R, Healey D, Kim S, Werdich AA, Haggarty SJ, MacRae CA, Shoichet B, Peterson RT (2010) Rapid behavior-based identification of neuroactive small molecules in the zebrafish. *Nat Chem Biol* 6:231–237.
- Krakauer JW, Ghazanfar AA, Gomez-Marin A, Maclver MA, Poeppel D (2017) Neuroscience needs behavior: correcting a reductionist bias. *Neuron* 93:480–490.
- Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN, Reich M, Hieronymus H, Wei G, Armstrong SA, Haggarty SJ, Clemons PA, Wei R, Carr SA, Lander ES, et al. (2006) The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313:1929–1935.
- Lashley KS (1951) The problem of serial order in behavior. In: *Cerebral mechanisms in behavior; the Hixon Symposium* (Jeffress LA ed), pp 112–146. New York: Wiley.
- Lin L, Faraco J, Li R, Kadotani H, Rogers W, Lin X, Qiu X, de Jong PJ, Nishino S, Mignot E (1999) The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (orexin) receptor 2 gene. *Cell* 98:365–376.
- Marques JC, Lackner S, Felix R, Orger MB (2018) Structure of the zebrafish locomotor repertoire revealed with unsupervised behavioral clustering. *Curr Biol* 28:181–195.e5.
- Maye A, Hsieh C, Sugihara G, Brembs B (2007) Order in spontaneous behavior. *PLoS One* 2:e443.
- Mu Y, Bennett DV, Rubinov M, Narayan S, Yang C-T, Tanimoto M, Mensh BD, Looger LL, Ahrens MB (2019) Glia accumulate evidence that actions are futile and suppress unsuccessful behavior. *Cell* 178:27–43.e19.
- Nevill-Manning CG, Witten IH (2000) On-line and off-line heuristics for inferring hierarchies of repetitions in sequences. *Proc IEEE* 88:1745–1755.
- Oikonomou G, Prober DA (2017) Attacking sleep from a new angle: contributions from zebrafish. *Curr Opin Neurobiol* 44:80–88.
- Prober DA, Rihel J, Onah AA, Sung R-J, Schier AF (2006) Hypocretin/orexin overexpression induces an insomnia-like phenotype in zebrafish. *J Neurosci* 26:13400–13410.
- Proekt A, Banavar JR, Maritan A, Pfaff DW (2012) Scale invariance in the dynamics of spontaneous behavior. *Proc Natl Acad Sci USA* 109:10564–10569.
- Randlett O, Wee CL, Naumann EA, Nnaemeka O, Schoppik D, Fitzgerald JE, Portugues R, Lacoste AMB, Riegler C, Engert F, Schier AF (2015) Whole-brain activity mapping onto a zebrafish brain atlas. *Nat Methods* 12:1039–1046.
- Rihel J, Ghosh M (2015) Zebrafish. In: *Drug discovery and evaluation: pharmacological assays* (Hock FJ, ed), pp 1–102. Berlin, Heidelberg: Springer.
- Rihel J, Prober DA, Arvanites A, Lam K, Zimmerman S, Jang S, Haggarty SJ, Kokel D, Rubin LL, Peterson RT, Schier AF (2010a) Zebrafish behavioral profiling links drugs to biological targets and rest/wake regulation. *Science* 327:348–351.
- Rihel J, Prober DA, Schier AF (2010b) Monitoring sleep and arousal in zebrafish. *Methods Cell Biol* 100:281–294.
- Robie AA, Hirokawa J, Edwards AW, Umayam LA, Lee A, Phillips ML, Card GM, Korff W, Rubin GM, Simpson JH, Reiser MB, Branson K (2017) Mapping the neural substrates of behavior. *Cell* 170:393–406.e28.
- Shein-Idelson M, Ondracek JM, Liaw H-P, Reiter S, Laurent G (2016) Slow waves, sharp waves, ripples, and REM in sleeping dragons. *Science* 352:590–595.
- Stern S, Kirst C, Bargmann CI (2017) Neuromodulatory control of long-term behavioral patterns and individuality across development. *Cell* 171:1649–1662.e10.
- Thyme SB, Pieper LM, Li EH, Pandey S, Wang Y, Morris NS, Sha C, Choi JW, Herrera KJ, Soucy ER, Zimmerman S, Randlett O, Greenwood J, McCarroll SA, Schier AF (2019) Phenotypic landscape of schizophrenia-associated genes defines candidates and their shared functions. *Cell* 177:478–491.e20.
- Tinbergen N (2010) On aims and methods of ethology. *Z Tierpsychol* 20:410–433.
- Todd JG, Kain JS, de Bivort BL (2017) Systematic exploration of unsupervised methods for mapping behavior. *Phys Biol* 14:015002.
- Truett GE, Heeger P, Mynatt RL, Truett AA, Walker JA, Warman ML (2000) Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and tris (HotSHOT). *Biotechniques* 29:52–54.
- Vanwallegheem GC, Ahrens MB, Scott EK (2018) Integrative whole-brain neuroscience in larval zebrafish. *Curr Opin Neurobiol* 50:136–145.
- Vogelstein JT, Park Y, Ohyama T, Kerr RA, Truman JW, Priebe CE, Zlatic M (2014) Discovery of brainwide neural-behavioral maps via multiscale unsupervised structure learning. *Science* 344:386–392.
- Wiltchko AB, Johnson MJ, Iurilli G, Peterson RE, Katon JM, Pashkovski SL, Abreira VE, Adams RP, Datta SR (2015) Mapping sub-second structure in mouse behavior. *Neuron* 88:1121–1135.
- Yokogawa T, Marin W, Faraco J, Pézerson G, Appelbaum L, Zhang J, Rosa F, Mourrain P, Mignot E (2007) Characterization of sleep in zebrafish and insomnia in hypocretin receptor mutants. *PLoS Biol* 5:e277.
- Zhdanova IV, Wang SY, Leclair OU, Danilova NP (2001) Melatonin promotes sleep-like state in zebrafish. *Brain Res* 903:263–268.