

Bioinformatics, Department of Computational Biology, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba, 277-8561, Japan. ⁷Comparative Genomics Laboratory, National Institute of Genetics, 411-8540 Yata 1111, Mishima, Shizuoka, 411-8540, Japan. ⁸Computational Biology Group, IIDMM, University of Cape Town Faculty of Health Sciences, Cape Town, 7925, South Africa. ⁹Department of Biochemistry and Biophysics, Texas A&M University, 328B TAMU, College Station, TX 77843, USA. ¹⁰Department of Biochemistry and Molecular Biology, Egerton University, Post Office Box 536, Njoro, Kenya. ¹¹Department of Biochemistry, Jomo Kenyatta University of Agriculture and Technology (JKUAT), Post Office Box 62000-00200, Nairobi, Kenya. ¹²Department of Biochemistry, Virginia Polytechnic Institute and State University, 309 Fralin Hall, Blacksburg, VA 24061, USA. ¹³Department of Biological Chemistry and Crop Protection, Rothamsted Research, West Common, Harpenden, Herts, AL5 2JQ, UK. ¹⁴Department of Biological Sciences, McMicken College of Arts and Sciences, University of Cincinnati, Cincinnati, OH 45221, USA. ¹⁵Department of Biological Sciences, University of Cape Town, Private Bag, Rondebosch, ZA-7700, South Africa. ¹⁶Department of Biological Sciences, University of Wisconsin–Parkside, 900 Wood Road, Kenosha, WI 53144, USA. ¹⁷Department of Biological Sciences, Wayne State University, 5047 Gullen Mall, Detroit, MI 48202, USA. ¹⁸Department of Biology and Biotechnology, University of Pavia, Via Ferrata 9, Pavia, 27100, Italy. ¹⁹Department of Biology, Baylor University, Waco, TX 76798, USA. ²⁰Department of Biology, KU Leuven, Naamsestraat 59, Leuven, B-3000, Belgium. ²¹Department of Biology, New Mexico State University, Foster Hall 263, Las Cruces, NM 88003, USA. ²²Department of Biology, University of York, Wentworth Way, York, YO10 5DD, UK. ²³Department of Biology, West Virginia University, 53 Campus Drive, 5106 LSB, Morgantown, WV, USA. ²⁴Department of Biomedical Sciences, Institute of Tropical Medicine Antwerp, Nationalestraat 155, Antwerp, B-2000, Belgium. ²⁵Department of Cellular Biology, University of Georgia, 302 Biological Sciences Building, Athens, GA 30602, USA. ²⁶Department of Clinical Research, KEMRI-Wellcome Trust Programme, CGMRC, Post Office Box 230-80108, Kilifi, Kenya. ²⁷Department of Computer and Information Sciences, College of Science and Technology, Covenant University, P.M.B. 1023, Ota, Ogun State, Nigeria. ²⁸Department of Computer Science and Engineering, Department of Biochemistry and Biophysics, Texas A&M University, HRBB 328B TAMU, College Station, TX 77843, USA. ²⁹Department of Entomology, North Carolina State University, Campus Box 7613, Raleigh, NC 27695–7613, USA. ³⁰Department of Entomology, Texas A&M University, 2475 TAMU, College Station, TX 77843, USA. ³¹Department of Entomology, The Ohio State University, 400 Aronoff Laboratory, 318 West 12th Avenue, Columbus, OH 43210, USA. ³²Department of Entomology, University of Arizona, 1140 East South Campus Drive, Forbes 410, Tucson, AZ 85721, USA. ³³Department of Entomology, University of Illinois at Urbana-Champaign, 505 South Goodwin Avenue, Urbana, IL 61801, USA. ³⁴Insect Pest Control Laboratory, Joint FAO/IAEA Programme of Nuclear Techniques in Food and Agriculture, Vienna, 1220, Austria. ³⁵Department of Environmental and Natural Resources Management, University of Patras, 2 Seferi Street, Agrinio, 30100, Greece. ³⁶Department of Epidemiology of Microbial Diseases, Yale School of Public Health, 60 College Street, New Haven, CT 06520, USA. ³⁷Department of Evolution, Ecology, and Organismal Biology, The Ohio State University, 300 Aronoff Laboratory, 318 West 12th Avenue, Columbus, OH 43210, USA. ³⁸Department of Natural Sciences, St. Catharine College, 2735 Bardstown Road., St. Catharine, KY 40061, USA. ³⁹Department of Nutritional Sciences, University of Arizona, Career and Academic Services, College of Agriculture and Life Sciences, Forbes Building, Room 201, Post Office Box 210036, Tucson, AZ 85721–0036, USA. ⁴⁰Department of Nutritional Sciences, University of Arizona, Shantz 405, 1177 East 4th Street, Tucson, AZ 85721–0038, USA. ⁴¹Department of Pathogen Molecular Biology, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, UK. ⁴²Department of Parasitology, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, L3 5QA, UK. ⁴³Department of Vector Biology, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, L3 5QA, UK. ⁴⁴Entomology Section, Onderstepoort Veterinary Institute, Private Bag X5, Onderstepoort, 110, South Africa. ⁴⁵Eck Institute for Global Health, Department of Bio-

logical Sciences, University of Notre Dame, Notre Dame, IN 46556, USA. ⁴⁶Department of Veterinary Tropical Diseases, University of Pretoria, Private Bag X04, Onderstepoort, 110, South Africa. ⁴⁷European Molecular Biology Laboratories, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, Cambridgeshire, CB10 1SA, UK. ⁴⁸Group of Bioinformatics and Modeling, Laboratory of Medical Parasitology, Biotechnology, and Biomolecules, Institut Pasteur de Tunis, 13, Place Pasteur, BP74, Belvédère, Tunis, 1002, Tunisia. ⁴⁹Institut de Recherche pour le Développement (IRD), UMR 177 IRD-CIRAD INTERTRYP, CIRDES Bobo-Dioulasso, Burkina Faso. ⁵⁰Institut de Recherche pour le Développement (IRD), UMR 177 IRD-CIRAD INTERTRYP, LRCT Campus International de Baillarguet, Montpellier, France. ⁵¹Institute of Biological, Environmental, and Rural Sciences, University of Aberystwyth, Old College, King Street, Aberystwyth, Ceredigion, SY23 3FG, UK. ⁵²Institute of Biotechnology Research, Jomo Kenyatta University of Agriculture and Technology (JKUAT), Post Office Box 62000-00200, Nairobi, Kenya. ⁵³Institute of Integrative Biology, The University of Liverpool, Crown Street, Liverpool, L69 7ZB, UK. ⁵⁴Institute of Zoology, Slovak Academy of Sciences, Dúbravská cesta 9, Bratislava, 845 06, Slovakia. ⁵⁵Integrated Database Team, Biological Information Research Center, National Institute of Advanced Industrial Science and Technology, Aomi 2-4-7, Koto-ku, Tokyo, 135-0064, Japan. ⁵⁶Kenya Agricultural Research Institute Trypanosomiasis Research Centre, Post Office Box 362, Kikuyu, 902, Kenya. ⁵⁷Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, 12735 Twinbrook Parkway, Room 2E-32D, Rockville, MD 20852, USA. ⁵⁸Technology Innovation Agency, National Genomics Platform, Post Office Box 30603, Mayville, Durban, 4058, South Africa. ⁵⁹Department of Biochemistry and Sports Science, Makerere University, Post Office Box 7062, Kampala, Uganda. ⁶⁰Bateman Group, Wellcome Trust Sanger Institute, EMBL European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, Cambridgeshire, CB10 1SA, UK. ⁶¹Molecular Biology and Bioinformatics Unit, International Center of Insect Physiology and Ecology, Duduville Campus, Kasarani, Post Office Box 30772-00100, Nairobi, Kenya. ⁶²National Livestock Resources Research Institute (NaLIRRI), Post Office Box 96, Tororo, Uganda. ⁶³National Research Center for Protozoan Diseases, Obihiro University of Agriculture and Veterinary Medicine, Inada-cho, Obihiro, Hokkaido, 080-8555, Japan. ⁶⁴Parasite Genomics Group, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge,

Cambridgeshire, CB10 1SA, UK. ⁶⁵Riddiford Laboratory, Janelia Farm Research Campus, Howard Hughes Medical Institute, 19700 Helix Drive, Ashburn, VA 20147, USA. ⁶⁶RIKEN Center for Integrative Medical Sciences, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan. ⁶⁷South African National Bioinformatics Institute, South African MRC Bioinformatics Unit, University of the Western Cape, 5th Floor Life Sciences Building, Modderdam Road, Bellville 7530, South Africa. ⁶⁸Special Programme for Research and Training in Tropical Diseases (TDR), WHO, Avenue Appia 20, 1211 Geneva 27, Switzerland. ⁶⁹The Genome Institute, Washington University School of Medicine, St. Louis, MO 63110, USA. ⁷⁰Tsetse and Trypanosomiasis Research Institute (TTRI), Majani Mapana, Off Korogwe Road, Post Office Box 1026, Tanga, Tanzania. ⁷¹Vector Health International, Post Office Box 15500, Arusha, Tanzania. ⁷²Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, Cambridgeshire, CB10 1SA, UK. ⁷³WHO Regional Office for Africa, WHO, Cité du Djoué, Post Office Box 06, Brazzaville, Congo. ⁷⁴School of Basic Sciences, Indian Institute of Technology, Mandi 175001, Himachal Pradesh, India. ⁷⁵Department of Parasite, Vector, and Human Biology, KEMRI-Wellcome Trust Programme, CGMRC, Post Office Box 230-80108, Kilifi, Kenya. ⁷⁶Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo, 108-8639, Japan. ⁷⁷Department of Biochemistry and Biotechnology, Kenyatta University, Post Office Box 43844-00100, Nairobi, Kenya. ⁷⁸Department of Biological Sciences, Indian Institute of Science Education and Research, Indore Bypass Road, Bhauri District, Bhopal, Madhya Pradesh, 462066, India. ⁷⁹Faculty of Science, King Abdulaziz University, Jeddah, 21589, SA. ⁸⁰Department of Biostatistics, Harvard School of Public Health, 655 Huntington Ave. Boston, MA 02461. ⁸¹Bioinformatics Research Unit, Agroinformatics Research Center, National Institute of Agrobiology Sciences, 2-1-2, Kannondai, Tsukuba, Ibaraki 305-8602, Japan.

Supplementary Materials

www.sciencemag.org/content/344/6182/380/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S9
Tables S1 to S43
References (39–101)

12 December 2013; accepted 19 March 2014
10.1126/science.1249656

Discovery of Brainwide Neural-Behavioral Maps via Multiscale Unsupervised Structure Learning

Joshua T. Vogelstein,^{1,2*} Youngser Park,^{1*} Tomoko Ohyama,^{3*} Rex A. Kerr,³ James W. Truman,³ Carey E. Priebe,^{1††} Marta Zlatic^{3††}

A single nervous system can generate many distinct motor patterns. Identifying which neurons and circuits control which behaviors has been a laborious piecemeal process, usually for one observer-defined behavior at a time. We present a fundamentally different approach to neuron-behavior mapping. We optogenetically activated 1054 identified neuron lines in *Drosophila* larvae and tracked the behavioral responses from 37,780 animals. Application of multiscale unsupervised structure learning methods to the behavioral data enabled us to identify 29 discrete, statistically distinguishable, observer-unbiased behavioral phenotypes. Mapping the neural lines to the behavior(s) they evoke provides a behavioral reference atlas for neuron subsets covering a large fraction of larval neurons. This atlas is a starting point for connectivity- and activity-mapping studies to further investigate the mechanisms by which neurons mediate diverse behaviors.

Nervous systems can generate a wide range of motor outputs, depending on their incoming sensory inputs and internal state. A comprehensive understanding of how behav-

ioral diversity and selection is achieved requires the identification of neural circuits that mediate many distinct motor patterns in a given nervous system. Mapping a functional circuit for one be-

havior in a given organism is difficult enough, but doing so for many behaviors has been almost impossible. The first step in mapping a circuit that mediates a behavior is to identify neurons whose activity is causally related to the behavior. Such a list of neurons provides a starting point for identifying the connectivity patterns between the relevant neurons. Thus, to map circuits underlying many behaviors, one would need a comprehensive neuron-behavior atlas of the nervous system that would list all neurons causally related with each behavior.

Generating such neuron-behavior maps has been difficult for several reasons. First, the experimental tools to selectively manipulate small sets of neurons while simultaneously observing natural behavior were lacking. Fortunately, recent advances in genetic toolkits allow reasonably selective manipulation of neuron types in genetic model organisms, such as *Drosophila* (1–3). Advances in behavior tracking methods allow high-resolution monitoring of the effect of such manipulations (4, 5). Neural manipulation screens can therefore be coupled with high-resolution monitoring of motor outputs to causally link complex behaviors to correspondingly complex neural circuits.

Second, establishing the causal links between neural manipulations and the resulting time-varying behavioral responses is a daunting computational statistics challenge. Existing supervised machine-learning methods can detect only predetermined

behaviors (6); moreover, they are limited by the speed with which humans can annotate training data sets. An alternative approach uses unsupervised clustering of the multidimensional time series. However, the high-content and high-throughput nature of the time-varying behavior data presents both computational and statistical challenges.

We developed a methodology for data-driven neuron-behavior mapping and applied it to larval *Drosophila*. The nervous system of larval *Drosophila* consists of a well-developed brain and nerve cord containing only about 10,000 neurons, rendering it sufficiently simple to obtain a relatively comprehensive characterization of it. Moreover, there exist more than 1000 genetic GAL4 lines in *Drosophila* larvae with recently characterized sparse neuronal expression patterns that together cover most of the 10,000 neurons in the larval nervous system (<http://flew.janelia.org/cgi-bin/flew.cgi>) (3).

Optogenetic Neural Activation Screen

We designed an optogenetic neural activation screen (see supplementary materials) to obtain a neuron line-behavior atlas of the larval nervous system that would contain causal links between neuron lines and the motor patterns they control. We used 1049 distinct GAL4 lines to selectively target channelrhodopsin-2 (ChR2) (7) to sparse distinct subsets of neurons, with each line activating 2 to ~15 neurons. Because these lines essentially span the entire set of larval neurons, some lines activate sensory and motor neurons as well as many neurons involved in decisions and action selection. We included four positive control lines in the screen that drive expression in nociceptive, mechanosensory, and proprioceptive neurons, previously determined to reliably mediate distinct behaviors (8–10), as well as one negative control line in which no neurons were optogenetically activated (2, 10), for a total of

1054 lines and 37,780 animals tested. In each experiment, we exposed dishes of larvae to 470-nm light stimuli (one exposure of 30 s followed by four exposures of 5 s, with a 30-s interval after the long exposure and a 10-s interval between the short exposures) to optogenetically activate Ch2-expressing neurons; we captured video before, during, and after stimulation (Fig. 1A). The Multi-Worm Tracker (MWT) software (4) tracked time-varying, two-dimensional closed contours of larvae and sketched eight time-varying features that collectively characterize larval shape and motion (Fig. 1B). Streaming and sketching reduced the data complexity by a factor of more than 200,000, enabling a compressive yet expressive representation of the data. These reduced data served as the input into the multiscale unsupervised structure learning methodology to reveal data-driven behavior types (Fig. 1C). Each behavior type was then linked to the subset of lines that mediate them (Fig. 1D).

Discovery of Behavior Types via Multiscale Unsupervised Structure Learning

As a first step, we sought to discover a large, inclusive, and nonpredetermined set of statistically distinguishable behavioral responses performed by the 37,780 animals during the first (30-s) optogenetic activation period. Recently developed methods for multiscale unsupervised structure learning (11–13) can be thought of as generalizations of manifold learning techniques, in that they can learn structures more general than manifolds, such as unions of manifolds. We adopted iterative denoising tree (IDT) methodology (11, 14), which offers demonstrated utility across several domains (15, 16).

The input to IDT is the collection of all 37,780 larval sketches, irrespective of which line generated each sketch (Fig. 2A). IDT consists of five key

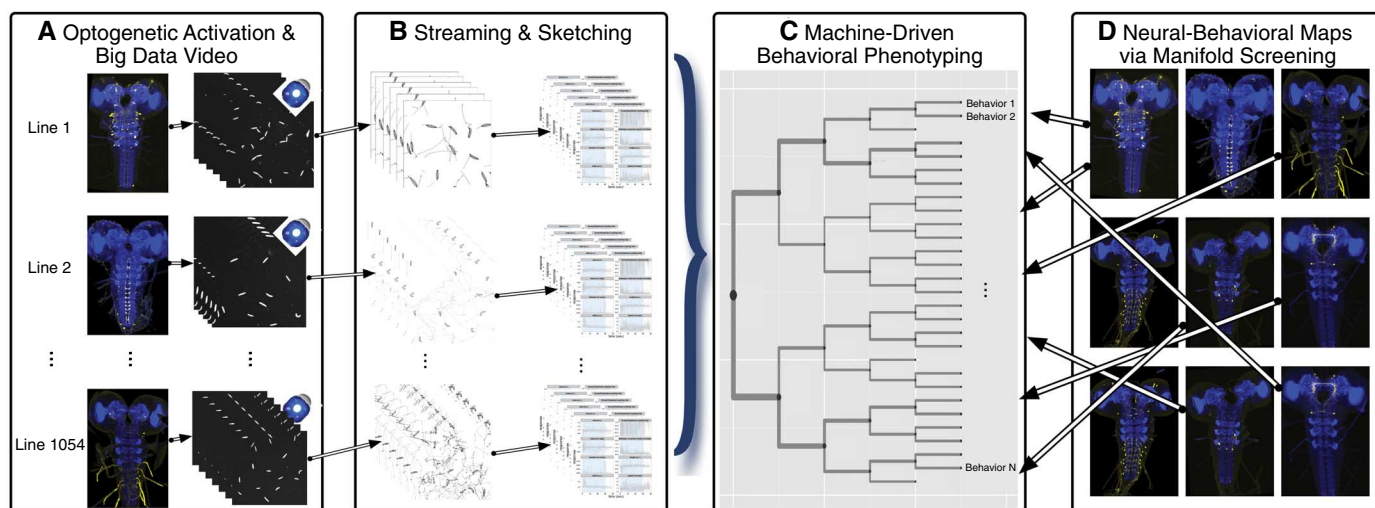


Fig. 1. Experimental design and methodology for obtaining neuron line-behavior maps. (A) Optogenetic activation screen of 1054 lines while digitally recording high-dimensional larval responses. (B) Streaming extracts the contours of each larva from each video frame; sketching extracts eight time-varying features from the contours that characterize the shape and

motion of each animal. (C) Machine-driven behavioral phenotyping learns phenotype categories (called behaviotypes) from the sketches via multiscale unsupervised structure learning. (D) Manifold testing discovers which neuron lines evoke sets of behaviors that are different from negative controls, which facilitates associating each such line with some number of behaviotypes.

¹Whiting School of Engineering, Johns Hopkins University, Baltimore, MD 21218, USA. ²Department of Statistical Science, Duke University, Durham, NC 27708, USA. ³Janelia Farm Research Campus, Ashburn, VA 20147, USA.

*These authors contributed equally to this work.

†These authors contributed equally to this work.

‡Corresponding author. E-mail: cep@jhu.edu (C.E.P.); zlatim@janelia.hhmi.org (M.Z.)

steps collectively resulting in a hierarchical clustering tree. In step one, IDT computes a dissimilarity between all pairs of sketches (Fig. 2B). The dissimilarity choice can have drastic computational and inferential impact; thus, choosing one that captures signal variability, and can be efficiently computed, is key to the success of IDT. This is important because the 37,780 observations yield more than 1.4 billion dissimilarities. We used a smoothed distance between the sketches of the data (17) as a dissimilarity function (see supplementary materials).

In step two, IDT transforms the interpoint dissimilarity matrix into a set of n relatively high-dimensional Euclidean vectors, via an approach such as multidimensional scaling (18). Once each data point is in Euclidean space, an extensive toolkit from classical statistical machine learning is applicable (19).

Step three consists of iterating, once per tree depth, the following substeps: (i) select a subset of the dimensions from each cluster, (ii) cluster all the nodes of the partition tree obtained thus far, and (iii) check for convergence at each resulting cluster. For each cluster at each scale, the number of dimensions to select and the number of sub-clusters to generate are decided in a data-driven adaptive fashion. The final result is a hierarchical tree characterizing families and subfamilies of behavioral responses (Fig. 2C).

To visualize the phenotype categories learned by IDT (or nodes/clusters of the tree), called *behavioypes*, Fig. 2D shows, for each cluster, the time-varying means and standard errors for each of the eight sketched time-varying features. Each behavioype is well separated from at least some other cluster for at least some of the time for at least some features. This provides an intuitive validation of the uniqueness of the behavioypes, thereby demonstrating the efficacy of IDT.

IDT Identified Both Previously Described and Novel Behavioral Sequences

Inspection of the time-varying means of the responses of each behavioype cluster (Fig. 2D) and of the videos of animals at the center of each cluster (movies S1 to S58) revealed that IDT clustered many of the larval behavioral responses into categories similar to those a human expert would have identified. We could label the nodes of the learned tree post hoc (Fig. 2, E and F). For example, the first division revealed by IDT differentiates between slow and fast families. The fast family is subdivided into turn-avoid and escape-crawl sub-families. IDT further distinguished between the right turn-left turn-avoid and the symmetric left turn-right turn-avoid sequence (fig. S1) and between two types of escape crawl that are preceded by more or less hunching and wiggling (fig. S2). The slow family is subdivided into still or backup and turner; the still or backup is further subdivided into still (no movement) and backup (lots of backward crawling) (fig. S3), and the turner is subdivided into turn-turn-turn (continuous turning) and turn-slow-crawl (fig. S4).

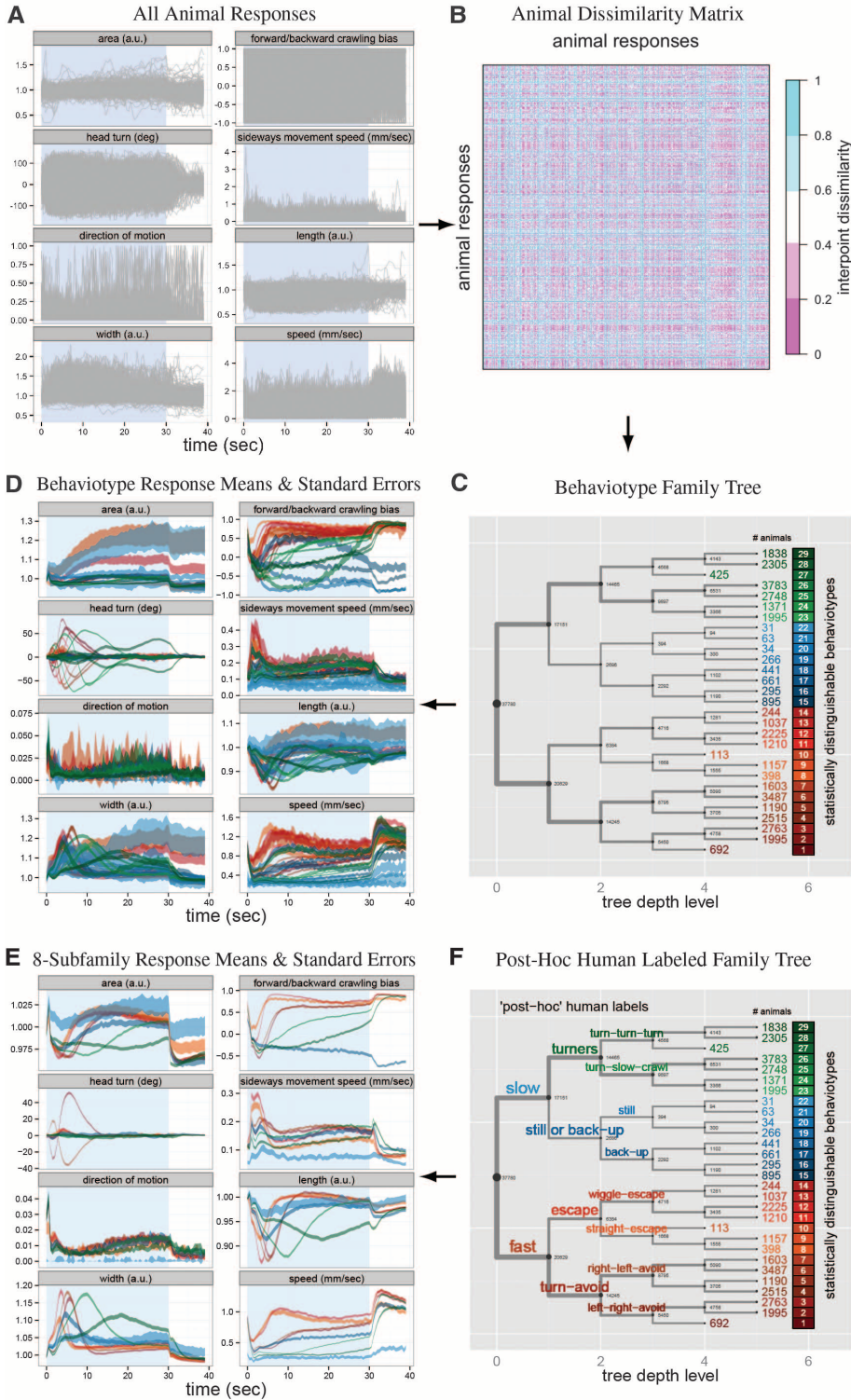


Fig. 2. IDT detected 29 distinguishable behavioypes. (A) The input to IDT is the collection of all 37,780 larval sketches. Structure is not obvious. Blue shading represents the period of photostimulation in all figures. (B) Dissimilarity matrix between all pairs of animals. Only ~1.4 million of the ~1.4 billion pairwise dissimilarities are shown. (C) $K = 29$ distinguishable behavior clusters (behavioypes) were identified using IDT to learn a data-adaptive clustering of the high-content responses. (D) Mean and standard error of the responses of each of the 29 behavioypes identified in (C), using the same color code as in (C). (E) Mean and standard error of the responses of the eight behavioype subfamilies. (F) The same tree as in (C) but with post hoc human labels assigned to the automatically detected behavioype families and subfamilies.

After level three of the behavior type tree, subjective visual inspection of the videos failed to detect differences between related behavior types. Nonetheless, they do have distinct properties for some of the features for some of the time (Fig. 2D). Behavior types 17 and 18 represent animals

with tracking errors in which the MWT streaming and sketching software inverted the front and back of the animals (fig. S6 and movies S33 to S36). IDT assigned animals with this type of tracking error to the two separate behavioral clusters, conveniently isolating such errors.

Many of the optogenetically evoked, automatically detected behavior type subfamilies are similar to previously described larval responses to various natural stimuli (10, 20, 21). IDT also detected behavioral categories not previously documented. This is unsurprising because only

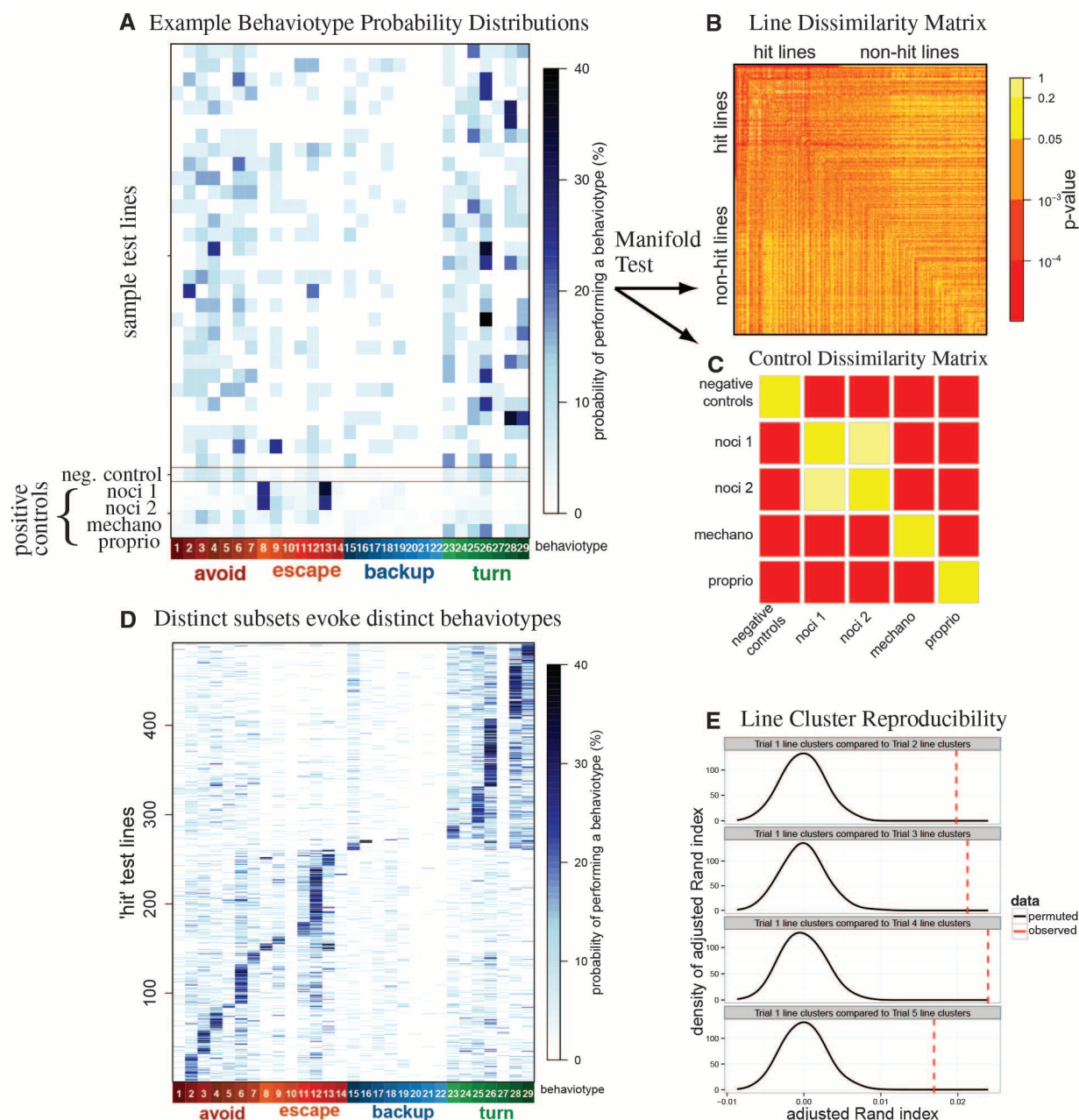


Fig. 3. Learning neuron line-behavior type maps via manifold testing. (A) Behavior type probability vectors for a number of lines. Each row shows the percentage of animals performing a behavior type for optogenetic activation of a particular line: 30 randomly sampled test lines (top), negative controls (pBDPU-ChR2) with no optogenetically activated neurons (middle), and positive controls—nociceptive (ppk-ChR2, noci 1; R38A10-ChR2, noci 2); mechanosensory (Iav-ChR2); and proprioceptive (R11F05-ChR2) neuron lines with known effects on behavior (bottom). (B) Line dissimilarity matrix showing pairwise P values for all 1054 lines computed via the manifold test. The first entry is the negative control. Remaining entries are sorted according to P value for the comparison with the negative control, from lowest to highest; 455 lines (4 positive controls + 451 test

lines) were significantly different from the negative control (hit lines), and 598 lines were not (nonhit lines). (C) The P values between all pairs of known somatosensory neuron lines and negative controls are “correct”: Those that should be significantly different according to previous studies are, and those that should not be are not. This panel is a subset (top left 5×5 submatrix entries) of (B). (D) Behavior type probability distributions for all 451 significantly distinct test neuron lines, sorted according to their maximum probability behavior type. (E) The clusters of lines that bias the probability toward the same behavior types are reproducible, demonstrated by the fact that the mode of the empirical null distribution of the adjusted Rand index (ARI) for reliable clustering (black curve) is lower in all four repeated trials than the observed ARI (all P values < 0.01).

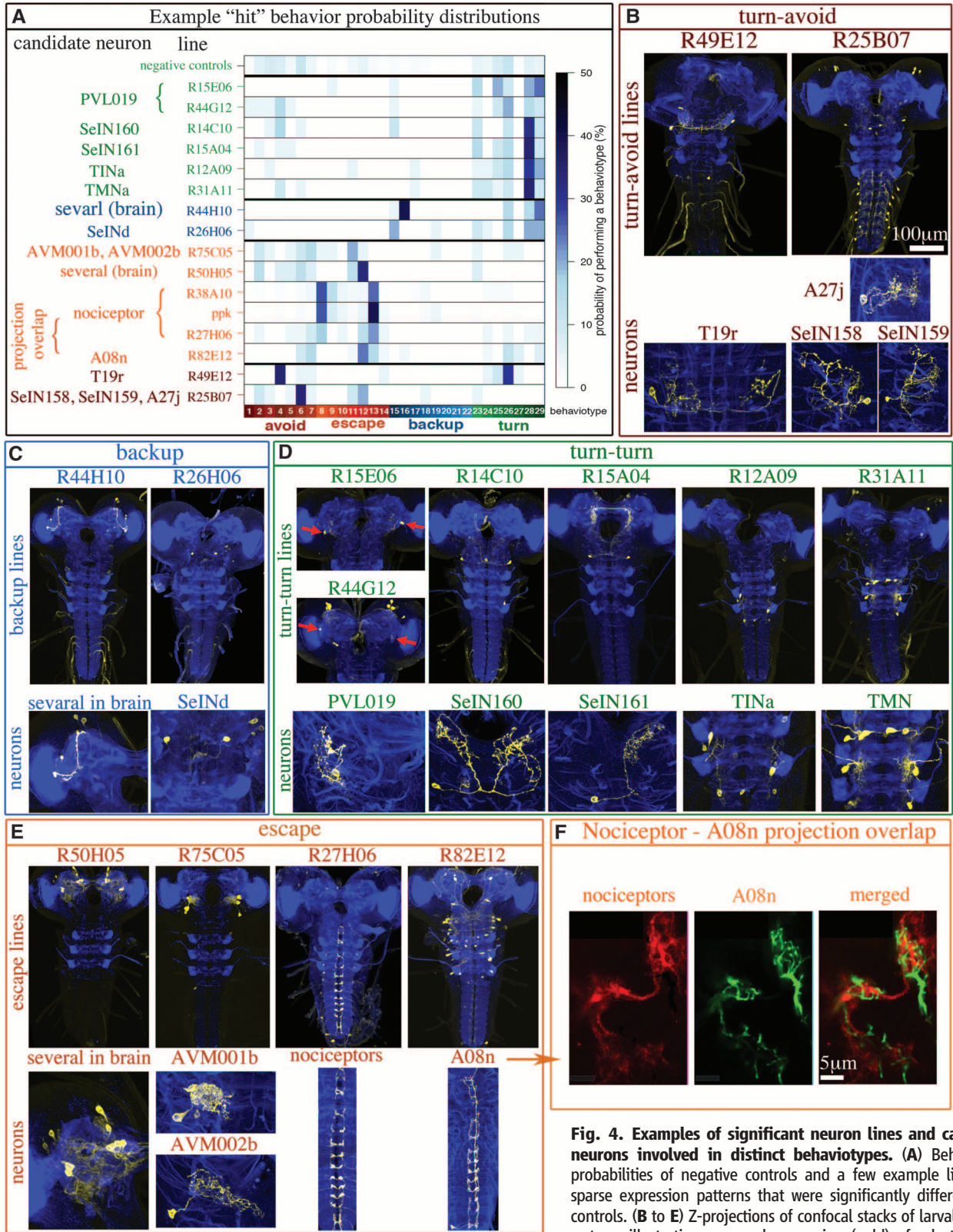


Fig. 4. Examples of significant neuron lines and candidate neurons involved in distinct behaviotypes. (A) Behaviotype probabilities of negative controls and a few example lines with sparse expression patterns that were significantly different from controls. (B to E) Z-projections of confocal stacks of larval nervous systems illustrating neuronal expression (gold) of selected lines.

Neuropil is stained for reference with antibody to N-cadherin (blue). High-resolution images of candidate neurons from these lines are shown beneath the lower-magnification images of the entire nervous system. These lines biased the probability toward (B) turn-avoid, (C) backup, (D) turn-turn, and (E) escape behaviotype subfamilies. In some cases, two or more lines that drive in the same neuron biased the probability toward the same behaviotype. R15E06 and R44G12 both drive in PVL019 (red arrows) and biased the probability toward turn-turn behaviotypes 25, 26, 28, and 29. Ppk, R38A10, and R27H06 drive in the same nociceptor neurons and biased the probability toward escape behaviotype 13. (F) R82E12 projections (green) overlap tightly with nociceptive axon terminals (red) in a single 0.5-µm confocal section, likely forming synaptic contacts. These two lines both biased the probability toward the same escape behaviotype 13 and likely belong to the same anatomical and functional neural circuit.

a relatively small portion of the larval stimulus space has been explored in previous studies.

Biased Probabilistic Relationship Between Neuron Activation and Behaviors

We used the prior results that nociceptive, mechanosensory, and proprioceptive sensory neurons reliably mediate distinct behaviors to assess the validity of this behavior discovery approach. The optogenetic screen included lines that targeted ChR2 expression to the nociceptive [ppk and R38A10 (10, 22)], mechanosensory (iav) (9), and proprioceptive (R11F05) neurons. Activating distinct sensory neurons mostly biased the behavior probability toward distinct subfamilies: (i) Nociceptive stimulation tended to evoke escape behaviors; (ii) mechanosensory stimulation tended to evoke turn-avoid behaviors; and (iii) proprioceptive tended to evoke slow behaviors (Fig. 3A). In contrast, activating the same nociceptive neurons with two distinct GAL4 lines evoked similar behavior probability distributions. The response profiles of the negative controls (pBDPU-ChR2) (2, 10) that did not express ChR2 likely represented larval reactions to blue light alone and appeared different from those of animals with optogenetically activated nociceptive, mechanosensory, or proprioceptive neurons (Fig. 3A). Finally, analysis of the 1049 previously unexplored neuronal test lines (Fig. 3A; top rows show 30 examples) demonstrated that distinct lines bias the probability toward distinct behaviors. Optogenetic stimulation of different animals of the same line (that is, activating the “same” neurons in different animals) did not always evoke the same behavior; rather, it biased the probability toward a few possible behaviors.

We also applied IDT jointly to the responses from the four additional identical repeats of the 5-s optogenetic activation stimulus of the same organisms and analyzed the behavior distributions of the same individuals on different trials (fig. S5A). We found that even repeated activation of the same neurons in the same individual did not always evoke the same behavior; rather, it biased the probability toward a few possible behaviors (fig. S5A). However, the responses of individual animals were significantly more similar to each other than to the responses of distinct individuals (fig. S5B).

Screening Neuron Lines via Manifold Testing

Each neuron line is characterized by the empirical probability of larvae performing each behavior and is encoded by a $\hat{K} = 29$ -dimensional vector with nonnegative entries that sum to unity. This encoding enables direct testing of each pair of lines by choosing an appropriate test statistic and applying a standard test. However, the choice of test statistic for these multivariate probability vectors is not obvious; moreover, existing tests do not sufficiently address the multiple dependent hypothesis-testing problem. We therefore

devised the following manifold test (see supplementary materials).

The main idea of the manifold test is to jointly and nonlinearly embed each experiment into a lower-dimensional representation. The test first computes the Hellinger distance between all 1054^2 pairs of experiments and then uses multidimensional scaling (MDS) (18) to obtain a low-dimensional Euclidean embedding. In this space, one can easily compute the difference between any pair of trials via an “out-of-sample” extension to this embedding methodology (23). Moreover, via bootstrap, the test computes the significance of those differences (24). Then it adjusts the P values to account for multiple correlated hypotheses (25) and other batch effects (26). Using this manifold test, we identified a large fraction of lines (4 positive controls and 451 test lines) as being significantly different (hit lines) from the negative control line (Fig. 3B; see supplementary data set 1 for the list of all such lines).

The positive control lines with different neurons activated (nociceptive, mechanosensory, and proprioceptive) were all significantly distinct from the negative controls and from each other (Fig. 3C). The two positive control lines with the same nociceptive neurons activated (ppk and R38A10) were not significantly different from one another (Fig. 3C). The known somatosensory neuron controls lend additional credence to this manifold test.

A Neuron Line–Behavior Atlas

Given the list of lines significantly distinct from the negative controls, we desired to identify which lines bias the probability toward which behaviors, thereby generating a neuron line–behavior atlas. Many lines biased the probability of behavior primarily to one behavior, with the remaining probability distributed to a few related behaviors (Fig. 3D). We confirmed that the line clusterings are reliable by applying IDT independently to the four additional repeats of the 5-s optogenetic activation stimulus of the same organisms, and then comparing how similar the line clusters from each of these additional trials were to the first trial. The identified line clusters were indeed reproducible ($P < 0.001$ via an adjusted Rand index permutation test, Fig. 3E).

For most behaviors, activation of at least one line biased the probability toward that behavior significantly more than the negative controls (see supplementary data set 1 for significant line–behavior probability distribution numbers and P values). Collectively, these significance results constitute a reference atlas that associates each neural line with a set of behaviors it mediates (and vice versa, associating each behavior with a set of neural lines that mediates it). Images of neuronal expression patterns of all of these lines are available from <http://fweb.janelia.org/cgi-bin/fweb.cgi>.

Figure 4 shows examples of behavior probability vectors (Fig. 4A) and neuronal expression patterns of a few significant lines that bias the probability toward turn-avoid (Fig. 4B), backup

(Fig. 4C), turn-turn (Fig. 4D), or escape subfamilies (Fig. 4E). We analyzed with higher resolution the projections of individual neurons from some of these lines using a single-cell flip-out method (27) and gave names to the identified neurons. A number of the significant lines only drove expression in single pairs of larval neurons, yet they still biased the probability toward specific behaviors. For example, activation of the pair of PVL019 neurons in the basal posterior region of the larval brain (with the line R15E06) or of the pair of SeIN161 neurons in the subesophageal ganglion (with the line R15A04) evoked turn-turn behaviors in a large fraction of animals. This indicates that single pairs of cells can exert drastic control over behavior, consistent with findings in other systems (28, 29).

In some cases, we found that lines that contribute to the same behaviors contain neurons with anatomically overlapping projections that may constitute presynaptic and postsynaptic partners. For example, the motor neurons in R31A11 and the interneurons in R12A09 both strongly evoked turn-turn-turn behavior 28, and their projections occupy the same motor region of the nerve cord and likely overlap. Similarly, R27H06 and R82E12 (Fig. 4E) biased the probability toward escape behavior 13 and drive expression in the nociceptive neurons, and in a class of ascending projection neurons (A08n), respectively. Double labeling of the nociceptive and the A08n neurons confirmed that their arbors do indeed tightly overlap, likely forming synaptic contacts (Fig. 4F).

Discussion

An important step toward understanding the structure and function of neural circuits is to identify comprehensive lists of neurons whose activation is causally related to a comprehensive set of motor patterns. This requires statistical approaches for identifying structure in high-dimensional behavior data.

Methods for systematic, automated, and unsupervised clustering are rarely applied to high-dimensional behavioral data. A recent study used unsupervised learning to detect behavioral motifs defined as sequences of four “eigenworm” positions in mutant and wild-type freely moving *Caenorhabditis elegans* in the absence of any stimulation (30). Here, we developed methods for clustering entire behavioral response sequences during a stimulation period and made use of discovered behavioral clusters to detect significantly distinct experiments from a large-scale screen, providing statistical validation that the learned clusters are meaningful. We applied these methods to discover neurons causally related to behaviors, providing a proof of principle that it is possible to classify neurons in terms of their activation of motor outputs in a fully automated way.

The collection of neuron line–behavior relationships collectively constitutes an atlas. Starting with a list of lines whose activation is sufficient to evoke each behavior, the relevant

individual neurons from each line can readily be identified. Whereas some lines drive expression in a single pair of neurons, most drive in the range of two to five candidate neuron types. In cases where lines drive in more than one neuron, intersectional strategies can be used to target individual neurons and test the effect of their activation on behavior (2).

This reference atlas provides a valuable starting point for understanding how distinct behaviors are selected and controlled. Large-scale connectomics (31–33) and functional brain imaging methods (34, 35) will soon provide similarly comprehensive views of the structure of neural circuits and of the activity patterns within those circuits. However, a connectome by itself does not carry information about which neurons mediate which behaviors. Similarly, a brain-activity map alone shows the flow of information through the network, but does not reveal causal relationships between neurons and behavior. Together, the neuron-behavior map, the neuron-activity map, and the connectome complement one another, laying the groundwork for a brainwide understanding of the principles by which brains generate behavior.

The statistical methods presented here are generally applicable to discovery of scientifically meaningful structure from big data—a pressing problem in the information age.

References and Notes

1. B. D. Pfeiffer *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 9715–9720 (2008).
2. B. D. Pfeiffer *et al.*, *Genetics* **186**, 735–755 (2010).
3. A. Jenett *et al.*, *Cell Rep.* **2**, 991–1001 (2012).

4. N. A. Swierczek, A. C. Giles, C. H. Rankin, R. A. Kerr, *Nat. Methods* **8**, 592–598 (2011).
5. K. Branson, A. A. Robie, J. Bender, P. Perona, M. H. Dickinson, *Nat. Methods* **6**, 451–457 (2009).
6. M. Kabra, A. A. Robie, M. Rivera-Alba, S. Branson, K. Branson, *Nat. Methods* **10**, 64–67 (2013).
7. C. Schroll *et al.*, *Curr. Biol.* **16**, 1741–1747 (2006).
8. R. Y. Hwang *et al.*, *Curr. Biol.* **17**, 2105–2116 (2007).
9. C. W.-H. Wu *et al.*, *Neuron* **70**, 229–243 (2011).
10. T. Ohya *et al.*, *PLOS ONE* **8**, e71706 (2013).
11. C. E. Priebe, D. J. Marchette, D. M. Healy, *Mod. Signal Process.* **46**, 223 (2003).
12. W. K. Allard, G. Chen, M. Maggioni, *Appl. Comput. Harmon. Anal.* **32**, 435–462 (2012).
13. P. Bendich, D. Cohen-Steiner, H. Edelsbrunner, J. Harer, D. Morozov, in *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)* (IEEE, Providence, RI, 2007), pp. 536–546.
14. K. E. Giles, M. W. Trosset, D. J. Marchette, C. E. Priebe, *Comput. Stat.* **23**, 497–517 (2008).
15. C. E. Priebe, D. J. Marchette, D. M. Healy, *IEEE Trans. Pattern Anal. Mach. Intell.* **26**, 699–708 (2004).
16. D. Karakos, S. Khudanpur, J. Eisner, C. E. Priebe, in *Proceedings of the 2005 IEEE International Conference on Acoustics Speech and Signal Processing ICASSP* (IEEE, Philadelphia, 2005), vol. 5, pp. 1081–1084.
17. C. E. Priebe, *IEEE Trans. Pattern Anal. Mach. Intell.* **23**, 404–413 (2001).
18. I. Borg, P. J. F. Groenen, *Modern Multidimensional Scaling: Theory and Applications* (Springer, New York, 2010).
19. T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, New York, 2001).
20. E. A. Kane *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, E3868–E3877 (2013).
21. M. Kernan, D. Cowan, C. Zuker, *Neuron* **12**, 1195–1206 (1994).
22. J. A. Ainsley *et al.*, *Curr. Biol.* **13**, 1557–1563 (2003).
23. M. Tang, Y. Park, C. E. Priebe, <http://arxiv.org/abs/1305.4893> (2013).
24. P. I. Good, *Permutation, Parametric, and Bootstrap Tests of Hypotheses* (Springer, New York, 2010).
25. D. Yekutieli, Y. Benjamini, *Ann. Stat.* **29**, 1165–1188 (2001).
26. R. A. Fisher, *Statistical Methods for Research Workers* (Oliver and Boyd, Edinburgh, 1925).
27. G. Struhl, K. Basler, *Cell* **72**, 527–540 (1993).
28. I. Kupfermann, K. R. Weiss, *Behav. Brain Sci.* **1**, 3–10 (1978).
29. B. Hedwig, *J. Neurophysiol.* **83**, 712–722 (2000).
30. A. E. X. Brown, E. I. Yemini, L. J. Grundy, T. Jucikas, W. R. Schafer, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 791–796 (2013).
31. D. D. Bock *et al.*, *Nature* **471**, 177–182 (2011).
32. M. Helmstaedter, *Nat. Methods* **10**, 501–507 (2013).
33. S.-Y. Takemura *et al.*, *Nature* **500**, 175–181 (2013).
34. M. B. Ahrens, M. B. Orger, D. N. Robson, J. M. Li, P. J. Keller, *Nat. Methods* **10**, 413–420 (2013).
35. T. Schrödel, R. Prevedel, A. Aumayr, M. Zimmer, A. Vaziri, *Nat. Methods* **10**, 1013–1020 (2013).

Acknowledgments: We thank G. M. Rubin, B. D. Pfeiffer, A. Nern, and B. Condron for fly stocks; B. D. Mensh for exceptionally helpful comments on the manuscript; A. Cardona, J. H. Simpson, and K. Branson for helpful discussions; C. Sullivan and A. Mondal for help with editing; H. Li and Fly Light Project Team at Janelia HHMI for images of neuronal lines; Janelia Fly Core for setting up the fly crosses for the activation screen; and Janelia Scientific Computing for help with data processing and storage, especially E. Trautman, R. Svirkas, and D. Olbris. Supported by the Larval Olympiad Project and Janelia HHMI, the XDATA program of the Defense Advanced Research Projects Agency administered through Air Force Research Laboratory contract FA8750-12-2-0303, and a National Security Science and Engineering Faculty Fellowship. All raw data, data derivatives, and code are freely available from <http://openconnecto.me/> behavotypes.

Supplementary Materials

www.sciencemag.org/content/344/6182/386/suppl/DC1
Materials and Methods
Figs. S1 to S6
Movies S1 to S58
Supplementary Data Sets 1 and 2
References (36–40)

31 December 2013; accepted 17 March 2014
Published online 27 March 2014;
10.1126/science.1250298

REPORTS

A Dual-Catalysis Approach to Enantioselective [2 + 2] Photocycloadditions Using Visible Light

Juana Du,* Kazimer L. Skubi,* Danielle M. Schultz,* Tehshik P. Yoon†

In contrast to the wealth of catalytic systems that are available to control the stereochemistry of thermally promoted cycloadditions, few similarly effective methods exist for the stereocontrol of photochemical cycloadditions. A major unsolved challenge in the design of enantioselective catalytic photocycloaddition reactions has been the difficulty of controlling racemic background reactions that occur by direct photoexcitation of substrates while unbound to catalyst. Here, we describe a strategy for eliminating the racemic background reaction in asymmetric [2 + 2] photocycloadditions of α,β -unsaturated ketones to the corresponding cyclobutanes by using a dual-catalyst system consisting of a visible light-absorbing transition-metal photocatalyst and a stereocontrolling Lewis acid cocatalyst. The independence of these two catalysts enables broader scope, greater stereochemical flexibility, and better efficiency than previously reported methods for enantioselective photochemical cycloadditions.

Modern stereoselective synthesis enables the construction of a vast array of organic molecules with precise control over their three-dimensional structure (1, 2), which

is important in a variety of fields ranging from drug discovery to materials engineering. Photochemical reactions could have a substantial impact on these fields by affording direct access to

certain structural motifs that are otherwise difficult to construct (3, 4). For example, the most straightforward methods for the construction of cyclobutanes and other strained four-membered rings are photochemical [2 + 2] cycloaddition reactions. The stereochemical control of photocycloadditions, however, remains much more challenging than the stereocontrol of analogous non-photochemical reactions (5, 6) despite the chemistry community's sustained interest in photochemical stereoreduction over the last century (7, 8).

Although many strategies using covalent chiral auxiliaries (9, 10) or noncovalent chiral controllers (11, 12) have been used to dictate absolute stereochemistry in photochemical cycloaddition reactions, the development of methods that utilize substoichiometric stereodifferentiating chiral catalysts has proven a more formidable challenge.

Department of Chemistry, University of Wisconsin–Madison, 1101 University Avenue, Madison, WI 53706, USA.

*These authors contributed equally to this work.

†Corresponding author. E-mail: tyoon@chem.wisc.edu

Discovery of Brainwide Neural-Behavioral Maps via Multiscale Unsupervised Structure Learning

Joshua T. Vogelstein, Youngser Park, Tomoko Ohyama, Rex A. Kerr, James W. Truman, Carey E. Priebe and Marta Zlatic

Science **344** (6182), 386-392.
DOI: 10.1126/science.1250298 originally published online March 27, 2014

Optogenetic Insights

Mapping functional neural circuits for many behaviors has been almost impossible, so **Vogelstein *et al.*** (p. 386, published online 27 March; see the Perspective by **O'Leary and Marder**) developed a broadly applicable optogenetic method for neuron-behavior mapping and used it to phenotype larval *Drosophila* and thus developed a reference atlas. As optogenetic experiments become routine in certain fields of neuroscience research, creating even more specialized tools is imperative (see the Perspective by **Hayashi**). By engineering channelrhodopsin, **Wietek *et al.*** (p. 409, published online 27 March) and **Berndt *et al.*** (p. 420) created two different light-gated anion channels to block action potential generation during synaptic stimulation or depolarizing current injections. These new tools not only improve understanding of channelrhodopsins but also provide a way to silence cells.

ARTICLE TOOLS

<http://science.sciencemag.org/content/344/6182/386>

SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2014/03/26/science.1250298.DC1>

RELATED CONTENT

<http://science.sciencemag.org/content/sci/344/6182/372.full>

REFERENCES

This article cites 32 articles, 6 of which you can access for free
<http://science.sciencemag.org/content/344/6182/386#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2014, American Association for the Advancement of Science