

A compressed sensing framework for efficient dissection of neural circuits

Jeffrey B. Lee^{1,8}, Abdullah Yonar ^{1,8*}, Timothy Hallacy², Ching-Han Shen^{3,4}, Josselin Milloz^{3,4}, Jagan Srinivasan⁵, Askin Kocabas ^{3,6} and Sharad Ramanathan ^{1,2,3,4,7*}

A fundamental question in neuroscience is how neural networks generate behavior. The lack of genetic tools and unique promoters to functionally manipulate specific neuronal subtypes makes it challenging to determine the roles of individual subtypes in behavior. We describe a compressed sensing-based framework in combination with non-specific genetic tools to infer candidate neurons controlling behaviors with fewer measurements than previously thought possible. We tested this framework by inferring interneuron subtypes regulating the speed of locomotion of the nematode *Caenorhabditis elegans*. We developed a real-time stabilization microscope for accurate long-term, high-magnification imaging and targeted perturbation of neural activity in freely moving animals to validate our inferences. We show that a circuit of three interconnected interneuron subtypes, RMG, AVB and SIA control different aspects of locomotion speed as the animal navigates its environment. Our work suggests that compressed sensing approaches can be used to identify key nodes in complex biological networks.

Many studies in nematodes, fruitflies, zebrafish, and mice have revealed that the number of key neurons regulating specific behaviors is a small fraction of the total number of neurons in the nervous system^{1–10}. Conventional methods for the identification of this small fraction (Fig. 1a,b) require the interrogation of the different subtypes of neurons in the neural network one at a time¹¹. However, genetic and molecular approaches to implement this search are challenging owing to the lack of unique genetic markers and promoters specific to only one neural subtype both in *Caenorhabditis elegans*¹² and in other animals^{13–15}. Even if specific genetic tools did exist, the measurements would have to be made one neuron subtype at a time. Compressed sensing techniques from statistics^{16,17} have been successful in both efficient sampling and accurate inference of important sparse features in data^{18–21}. Thus, compressed sensing methods could provide an alternative strategy for rapidly prioritizing the contribution of different neurons to a given behavior, enabling a focused analysis of key neurons. Indeed, we show that an experimental framework based on compressed sensing enables the identification of key sets of neurons controlling the speed of locomotion of *C. elegans*, despite the lack of promoters to target individual neuronal subtypes.

The first principle of compressed sensing is to make measurements on an incoherent basis where the data are not sparse²². For example, if the data are sparse in the time domain, measurements should be made in the frequency domain in which basis the same data are no longer sparse. In the context of neural circuits, this means that perturbations of neural activity should not target one neuron subtype at a time to match measured responses to a small number of key neurons. Instead, groups of neuron subtypes should be perturbed simultaneously (over a total of at least $\sim \log[N]$ groups, where N is the total number of neuron subtypes in the network). Following theorems by Candes and Tao^{16,22,23}, these groups need not be carefully chosen, and the method will work for most

arbitrary choices, even with very different numbers of neuron subtypes in each group. The second principle of compressed sensing is that the data from as few as $\sim n \log[N]$ such measurements can be analyzed using L1-norm regularization to find the small number n of key neuron subtypes driving the behavior ($n \ll N$). This is much less than the N measurements needed using a conventional approach that uses unique promoters to address one neuronal subtype at a time.

Results

Compressed sensing framework to identify essential interneurons controlling the speed of locomotion in *C. elegans*. The lack of specific promoters and mutations that affect individual interneuron subtypes in *C. elegans* has hindered efforts to systematically uncover the roles of these neurons in different behaviors. We therefore used *C. elegans* as a system in which to test a framework based on compressed sensing to systematically identify the key interneuron subtypes controlling the speed of locomotion. We first designed an incoherent measurement matrix by leveraging non-specific promoters. We selected 27 promoters, primarily expressed in interneurons, with expression patterns that have been characterized (Supplementary Table 1). Each of these promoters drives expression in 2 to 15 neurons. Together, these promoters were active in 88 of the 118 neuron types in the animal, including 56 types of interneurons, 22 types of sensory neurons, and 10 types of motor neurons (Fig. 1c and Supplementary Table 1)^{12,24,25}. These 27 promoters generate an incoherent 27×88 dimensional measurement matrix M , where each row of this matrix corresponds to the promoter identity and each column to the neuron type. If promoter i drives expression in neuron type j , $M_{ij}=1$, and else $M_{ij}=0$ (Fig. 1c). To implement the measurement matrix experimentally, we constructed transgenic *C. elegans* lines in a *lite-1* mutant background²⁶, with each line driving the expression of archaerhodopsin-3 (Arch)^{27,28} under the

¹School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA. ²Biophysics Program, Harvard University, Cambridge, MA, USA.

³FAS Quantitative Biology Initiative, Center for Brain Science, Harvard University, Cambridge, MA, USA. ⁴Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA, USA. ⁵Biology and Biotechnology, Worcester Polytechnic Institute, Worcester, MA, USA. ⁶Department of Physics, Koç University, Sarıyer, Istanbul, Turkey. ⁷Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA. ⁸These authors contributed equally: Jeffrey B. Lee, Abdullah Yonar. *e-mail: yonar@g.harvard.edu; sharad@post.harvard.edu

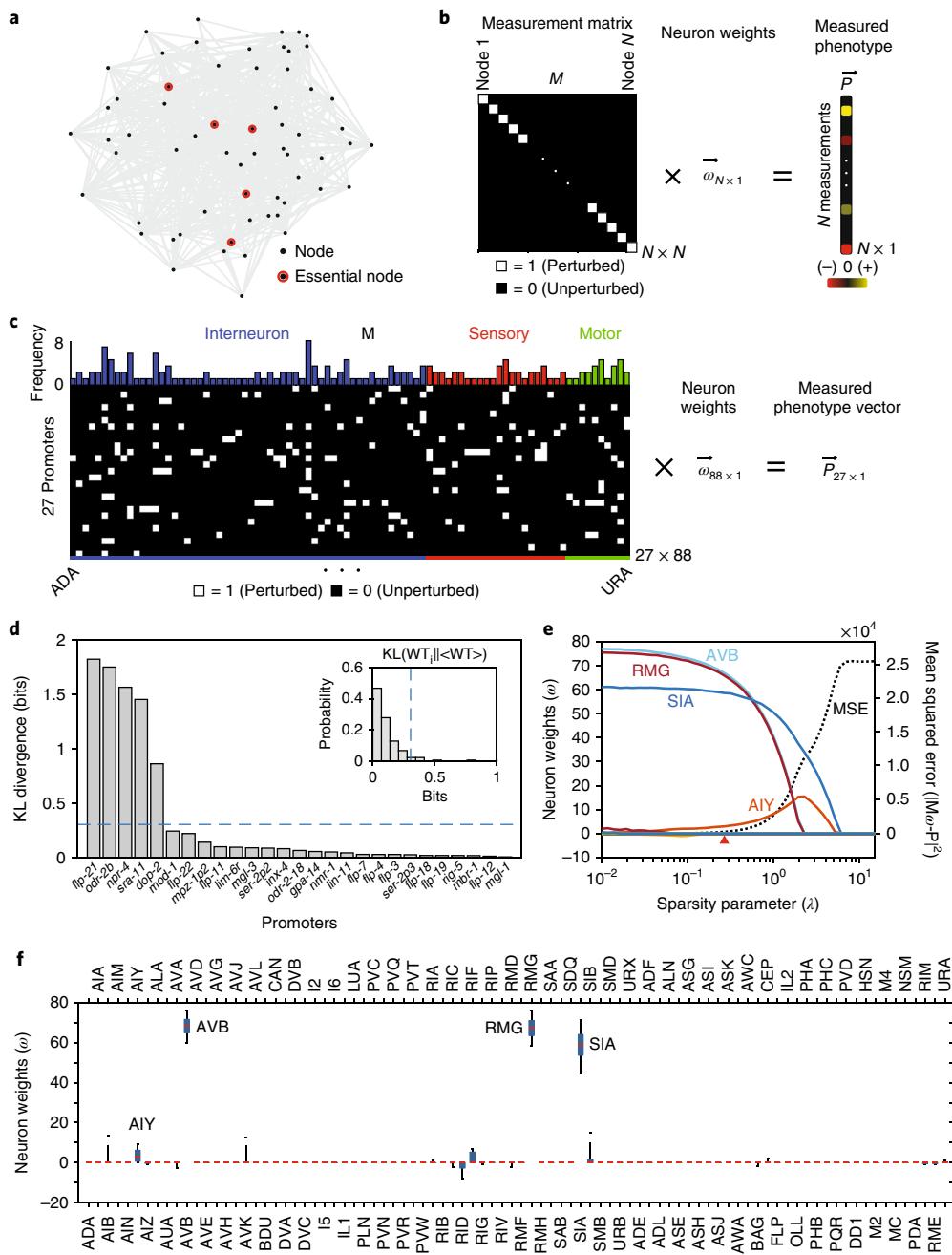


Fig. 1 | Compressed sensing framework enables identification of essential neurons controlling speed of locomotion in *C. elegans*. **a**, A network of nodes (black dots) with a small number of essential nodes for a phenotype (red), and synaptic connections (gray lines). A node can either be an individual neuron or neuron subtypes. **b**, Identifying essential neurons by perturbing nodes one at a time is equivalent to solving a matrix equation with an $N \times N$ diagonal measurement matrix (diagonal entries of one for the perturbed neuron, zero otherwise) multiplying unknown neuron weight vector $\vec{\omega}$ set equal to measured phenotype vector \vec{P} . The number of measurements N linearly increases with the size of the network. **c**, The 27×88 measurement matrix M generated from 27 transgenic lines expressing archaerhodopsin-3 under the control of non-specific promoters. Rows are promoters, columns are neuron types. Matrix has an entry of 1 (white) if the promoter drives expression in that neuron type, else 0 (black). Bars above the measurement matrix show the number of times each neuron appears in the matrix (interneurons, blue; sensory neurons, green; motor neurons, red). Finding key neurons from 88 neuron types using just 27 measurements requires solving an underdetermined set of linear equations $M\vec{\omega} = \vec{P}$. **d**, KL divergence between speed distributions of control and Arch lines placed in descending order. Inset, variation of KL divergence in wild-type replicates. Blue dashed lines, 95th percentile of the distribution. **e**, Median neuron weights from 10,000 lasso regression solutions through bootstrapping for sparsity parameter spanning three orders of magnitude. Red arrow, sparsity parameter picked for **f**. Dashed line, mean squared error $\chi^2 = (M\vec{\omega} - \vec{P})^2$ of optimized solution for given sparsity parameters. **f**, Neuron weights distribution for each neuron types for sparsity parameter shown in **e**. Red line, median; box, 25 and 75; whiskers, 5th and 95th percentiles. The neurons arranged in the same order as in **c**.

control of one of the 27 promoters. Although most of our lines showed expression patterns in a defined set of neurons, consistent with the literature, a few promoters (Supplementary Fig. 1) were substantially different. We used our observed expression patterns in these instances to correct the measurement matrix.

To measure the locomotor phenotypes of the 27 lines, we grew animals on bacteria with all-*trans*-retinal (ATR) and then placed them 1.5 cm away from the center of a bacterial lawn approximately 0.3 mm in radius on a standard agarose plate while being exposed to 5 mW mm⁻² of 525 nm (green) light to inhibit all neurons expressing archaerhodopsin-3 (Supplementary Fig. 2a,b). Using a multi-particle tracking approach, we measured the speed of the transgenic animals as they performed chemotaxis (Supplementary Fig. 2b,c and Supplementary Fig. 3). Of the 27 transgenic lines, five (*Prom_{fpl-2}*, *Prom_{npr-4}*, *Prom_{odr-2b}*, *Prom_{sra-11}*, *Prom_{dop-2}*) showed speed distributions that were significantly different from those of control animals based on the 95th percentile of the variation in the wild-type Kullback–Leibler (KL) divergence (Fig. 1d and Methods). Thirteen lines showed changes in reversal frequency of the animals or a lower fraction of animals arriving on the bacterial lawn than in the control group (Supplementary Fig. 4), despite not showing any change in speed phenotype. Thus, while the optogenetic perturbations are effective in many of these different lines, the ability to control speed is not broadly distributed among all of the neurons of the nervous system.

Next, we used these 27 distinct sets of measurements to find the contribution of each of the 88 neuron types to the observed phenotype. We formulated the problem as an underdetermined set of equations $\vec{M}\vec{\omega} = \vec{P}$, where M is the measurement matrix, $\vec{\omega}$ is the vector of the neuron weights, and \vec{P} is the phenotype vector. The measured fractional reduction in the mean speeds of the 27 lines compared to the wild type gave us the phenotype vector \vec{P} (Supplementary Fig. 2d). To solve this set of equations, we imposed a sparsity constraint that only a small fraction of all the neuron subtypes must be essential for controlling speed. As we did not know how many neurons were key to controlling speed, we could not constrain the sparse solution to contain a specific number of neurons. We therefore used Lasso regression^{29,30}, imposing sparsity on the solution by constraining the sum of the absolute values of the individual weights of the neurons (that is, the L1 norm of $\vec{\omega}$). Minimizing the sum of the mean squared error ($\chi^2 = (\vec{M}\vec{\omega} - \vec{P})^2$) and the L1 norm of the weight vector ($\|\vec{\omega}\|_1$) multiplied by a Lagrange multiplier (λ) gives $(\vec{M}\vec{\omega} = \vec{P}) + \lambda\|\vec{\omega}\|_1$, which imposes a sparsity constraint on the solution. Tuning the sparsity parameter λ changes the sparsity of the solution: increasing the value of λ leads to fewer neurons being used to fit the data. Thus, using Lasso we obtained solutions for the neuron weights with λ spanning over three orders of magnitude. We also calculated the errors in the estimated neuron weights. Over a wide range of values of the sparsity parameter, the chi-squared error of the fit of the weight vectors to $\vec{M}\vec{\omega} = \vec{P}$ remained stable, and our analysis yielded three key interneurons with large weights: AVB, RMG, and SIA (Fig. 1e). AIY appeared in the solution but had lower weights. Although the weights of AIY increased for large values of λ , in this regime, the fit of the inferred weight vectors to $\vec{M}\vec{\omega} = \vec{P}$ deteriorated rapidly (Fig. 1e). Owing to error propagation, AIB, AIZ, AVA, AVK, RIA, RIC, RID, RIF, RMD, SIB, BAG, CEP, IL2, RIM, and URA appeared with low weights in a small fraction of the solutions (Fig. 1f).

We tested the robustness of the results. Using either an L1-norm minimization alone or a nonlinear phenotype vector produced results consistent with those above: a core set of three neurons (AVB, RMG, and SIA) along with a few other neurons with low weights (Supplementary Fig. 5a-d). We further tested our results for robustness to corruption of the original measurement matrix to account for potential variations in the levels of archaerhodopsin expression (Supplementary Figs. 6,7) and the sensitivity of our

inference to the chosen promoters by changing the number of promoters from 22 to 32 (Supplementary Fig. 8). Next, we determined the accuracy of our results by estimating the false-positive and false-negative discovery rates (Supplementary Fig. 9). The number of false-negative interneurons using our measurement matrix with 27 promoters was on average less than one for up to 6 essential neurons (Supplementary Fig. 9a,f). Finally, we computationally tested the efficacy of our method to analyze nonlinear deep-learning networks. This suggested that in such networks³¹ our method has the potential to identify the key neurons (Supplementary Fig. 10). This is consistent with the efficacy of Lasso regression in the presence of nonlinear distortion³².

Stabilization microscope for long-duration calcium imaging and optogenetic manipulation of freely moving animals. To validate the solutions inferred using Lasso, we sought to perturb and accurately image neural activity in freely behaving animals. Imaging neural activity in a moving animal at high magnification is challenging, despite recent progress^{33–36}. To make such imaging possible, animal movement was restricted in these studies^{33–36} (at least by a glass coverslip for accurate z tracking), potentially affecting their speed of movement. Furthermore, these studies used light levels of at least 14 mW mm⁻² to track the animal for 5 min, by which time the effects of phototoxicity are evident³⁵. To overcome these limitations, we built a real-time image-stabilization microscope with the capacity to perform imaging at high magnifications and at less than a tenth of the light power levels than previously possible^{33–35} to image animals for 1 hour without any sign of phototoxicity. Our stabilization microscope is capable of accurately imaging fluorescence signals from both the soma and processes of multiple neurons in different z -planes as the animal moves freely on an agar plate over 10 cm (Fig. 2a). We achieved this level of performance by stabilizing a marker neuron (AWC^{ON}, expressing mKOOrange; Fig. 2b, top right panel) within our field of view in x , y , and z to micrometer accuracy (Fig. 2b, top right panel, and Supplementary Video 1) and its angular movement to within 45° as the animal crawled freely (Fig. 2c, $\sigma=22.3^\circ$, and Supplementary Video 2) through a 7 ms control loop. We achieved z tracking without a coverslip in a freely moving animal by imaging the fluorescent marker neuron simultaneously on two focal planes (Fig. 2b, left panel, and Supplementary Video 3). Rotational stabilization was achieved using a dove prism (Supplementary Video 4) mounted on a rotation stage. Having stabilized the marker neuron accurately, we used a tunable liquid lens to simultaneously image other neurons expressing GCaMP6s³⁷, for which the fluorescence changes based on the calcium levels in the cell, at different z planes (Supplementary Fig. 11 and Supplementary Video 5). We calibrated the accuracy and performance of this microscope to image from individual neurons (Fig. 2d–f and Supplementary Fig. 12a). Using a DLP projector, we could target light on a specific individual neuron in an animal in which multiple neurons expressed Arch, allowing for specific inhibition (Fig. 2f, Supplementary Fig. 12b, and Supplementary Video 6). The ability to simultaneously monitor and perturb neural activity with micron precision in animals moving freely (several centimeters) on an agar plate enabled us to uncover the roles of individual neuron subtypes in controlling speed (Supplementary Note 1).

Targeted inhibition of single neurons SIA, RMG, and AVB reduces the speed of locomotion. Using our microscope, we tested our inference that SIA, RMG, and AVB control speed. We selectively inhibited SIA in a *Prom_{dop-2}::Arch* line using targeted illumination with a randomized duty cycle to accurately assess the effects of the inhibition³⁸. We observed a consistent decrease in speed every time we illuminated SIA (Fig. 3a). To quantify this effect, we normalized the speed during light-on intervals using the speed during light-off intervals as a baseline (Fig. 3b). Although control worms fed no

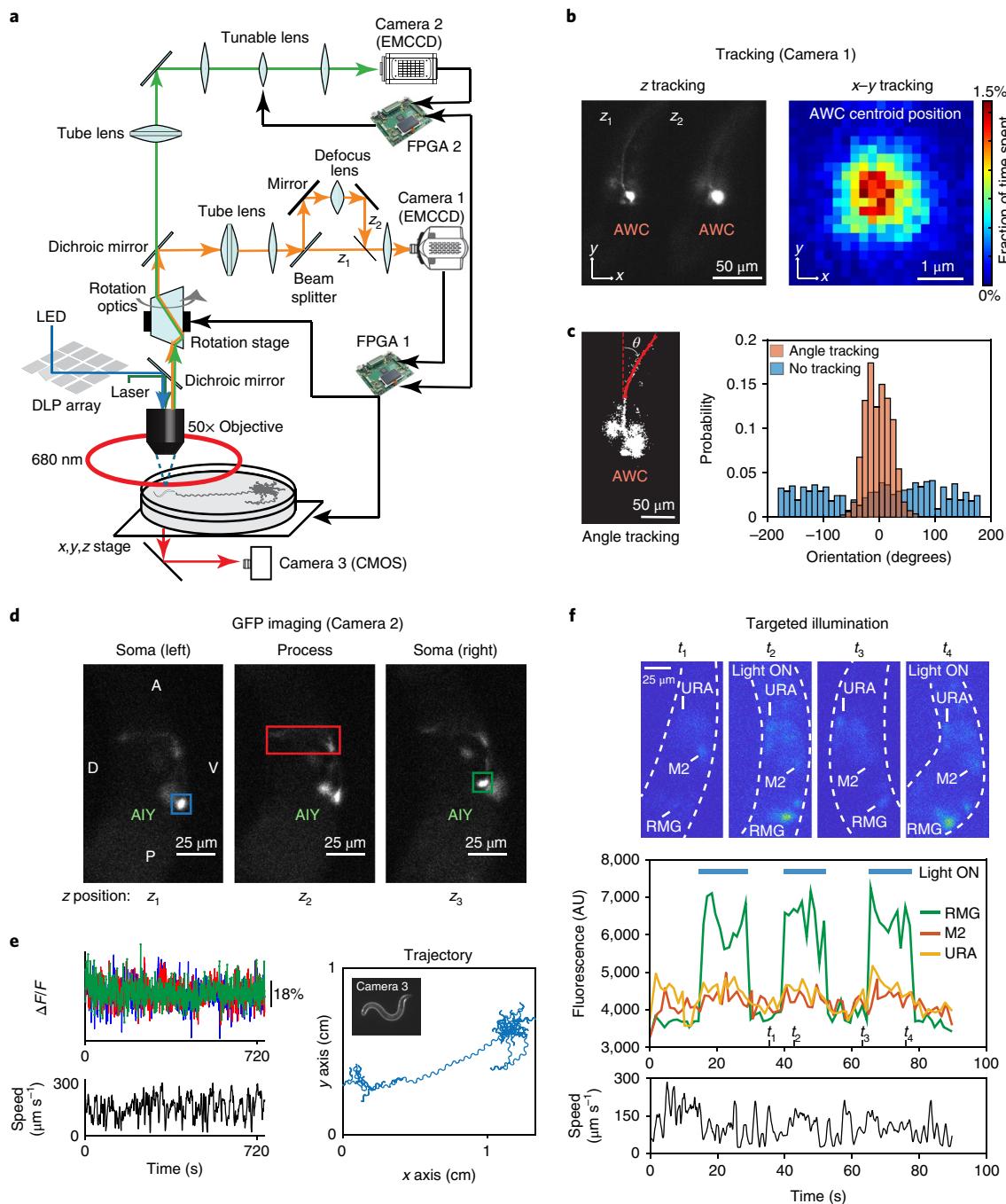


Fig. 2 | Stabilization microscope for long-time calcium imaging and optogenetic manipulation of freely moving animals. **a**, Schematic of tracking and image stabilization microscope (orange, mKO emission; light green, GCaMP emission; blue, GCaMP excitation; dark green, mKO excitation). Data from Camera 1 is processed by FPGA1 to control x,y,z and rotation stages for simultaneous tracking. Camera 2 acquires GCaMP imaging data via FPGA2, which also controls the tunable lens for the z scan. Camera 3 is used to capture animal behavior at low magnification. A DLP mirror array, controlled by a PC, is used to target specific neurons. **b**, Left, images of marker neuron AWC^{ON} on two z planes in a freely moving animal for z tracking. Right, heat map of the percentage of time spent by the center of mass of a marker neuron at each pixel for animal in **a** shows 1 μm x and y tracking accuracy. **c**, Left, AWC^{ON} marker neuron thresholded image (from Camera 1) to track rotation angle (θ). Right, histogram of θ with (orange) and without (blue) stabilization using rotation optics. **d**, AIY::GFP images (Camera 2) at different z positions with a z scan using a liquid lens (AIYL, blue; AIY process, red; AIYR, green; A, anterior; P, posterior; D, dorsal; V, ventral). **e**, Quantification of noise. Top left, GFP intensity (AIYL, blue; AIYR, green; AIY process, red) from freely moving animal shows 18% fluctuations. Bottom left, speed of an animal moving freely on an agar plate during the experiment. Right, trajectory of the animal. Inset image from Camera 3. **f**, Control for targeted illumination of the neuron of interest. Top, *Pifl-21::GCaMP6s* images (Camera 2) at four time points with targeted light on or off. Middle, GCaMP6s intensities from RMG, M2, and URA neurons. Blue bar, the patterned light on. Bottom, speed of an animal during targeted illumination. AU, arbitrary units.

ATR did not show speed reduction, selective SIA inhibition resulted in a significant reduction in speed ($32\%, P=2.23 \times 10^{-5}$, two-sample two-sided *t*-test) (Fig. 3b). Similarly, targeted inhibition of RMG

(*Prom_{fpl-21}::Arch*) and AVB (*Prom_{sra-11}::Arch*) led to a significant reduction in speed ($59\%, P=5.36 \times 10^{-11}$, and $33\%, P=1.26 \times 10^{-6}$, respectively, two-sample two-sided *t*-test) compared to the no ATR

controls (Fig. 3c–f and Supplementary Fig. 13a), validating the inference methods.

In addition to certain sensory neurons (Supplementary Note 2), the interneurons AIY and RIB have been implicated as being important for speed control³⁹. In our study, AIY was assigned a low weight through compressed sensing, and RIB was not identified. Furthermore, although the command interneuron AVB has been shown to control forward movement and is identified above, the command interneurons PVC^{40,41}, as well as AVA, AVE, and AVD, which control reversals, were not. To test the validity of our findings we inhibited these neurons individually using our microscope. Inhibition of AIY or RIB did not result in a speed change (Fig. 3g). Moreover, although the inhibition of command interneurons AVA, AVD, and AVE led to a suppression of reversals, this inhibition did not affect speed distribution (Supplementary Fig. 13b–d)⁴². In addition, we selectively inhibited the activity of 10 neural subtypes that had low or no weights in our compressed sensing solution. Of the neurons that we tested (AIB, RIM + RIC, IL2, CEP, BAG, URX, AVE, RIM, MI, and PVQ), only URX showed a small but significant change in speed with inhibition (17% increase, $P=0.01$, two-sample two-sided t -test).

Calcium imaging reveals distinct roles of SIA, RMG, and AVB in controlling speed. To further understand how these neurons control speed during foraging behavior, we used our microscope to measure calcium activity in different neurons as freely moving animals searched for food on an agar plate. Calcium activity in URX as well as in the other neurons (AIY, CEP, IL2, BAG, and RIM) showed no significant correlation with speed (Fig. 4a and Supplementary Fig. 14, P values for significance are listed in Fig. 4a). Again, calcium activity in RIB, which was implicated in the control of speed³⁹, did not show any correlation with speed (Fig. 4a). Similarly, although the calcium activity in the command interneuron AVA correlated with reversals, it did not correlate with speed (Fig. 4a and Supplementary Fig. 13d). Thus, interneurons that were previously implicated in speed regulation do not show any contributions to speed in the context of chemotaxis. Thus, they are not false negatives of the compressed sensing approach.

In contrast to the neurons above, activity patterns in the neuron types RMG, SIA, and AVB, which we identified through our analysis as speed-controlling neurons, showed a significant correlation with the speed of animals (Fig. 4a and Supplementary Fig. 12; $P=0.0103$, $P=0.0003$, $P=0.0159$, respectively, one-sample two-tailed t -test). These results, together with the results from the targeted inhibition of these neurons (Fig. 3), validate our compressed sensing solutions.

We next investigated the activity patterns of SIA, RMG, and AVB to understand their role in controlling speed. SIA calcium dynamics in animals freely searching for food (Fig. 4b, Supplementary Fig. 15a,b and Supplementary Video 7) showed an autocorrelation time of approximately 50 s (Fig. 4c and Supplementary Fig. 15c). Furthermore, the correlation between SIA activity and speed increased with the window size of the low-pass filter beyond 1 min, plateauing at window sizes over 3 min (Fig. 4d and Supplementary Fig. 15d). The slow modulation in SIA activity and the frequency-dependent cross-correlation analysis suggests that SIA controls speed modulations on the timescale of minutes (Supplementary Fig. 15b). In contrast, correlations of RMG calcium activity with speed arose solely due to pausing events (Fig. 4e,f, Supplementary Fig. 15e and Supplementary Video 8); when pausing events were removed computationally, correlations between RMG activity and speed vanished (Fig. 4g). During pausing, RMG activity decreased, but then increased again as the animal began moving, which in conjunction with our inhibition experiments suggests that RMG neurons control whether the animal pauses or moves (Supplementary Fig. 13a). Consistent with the role of AVB as a forward command interneuron^{40,41}, we observed that AVB activity decreases whenever the animal makes a reversal and increases

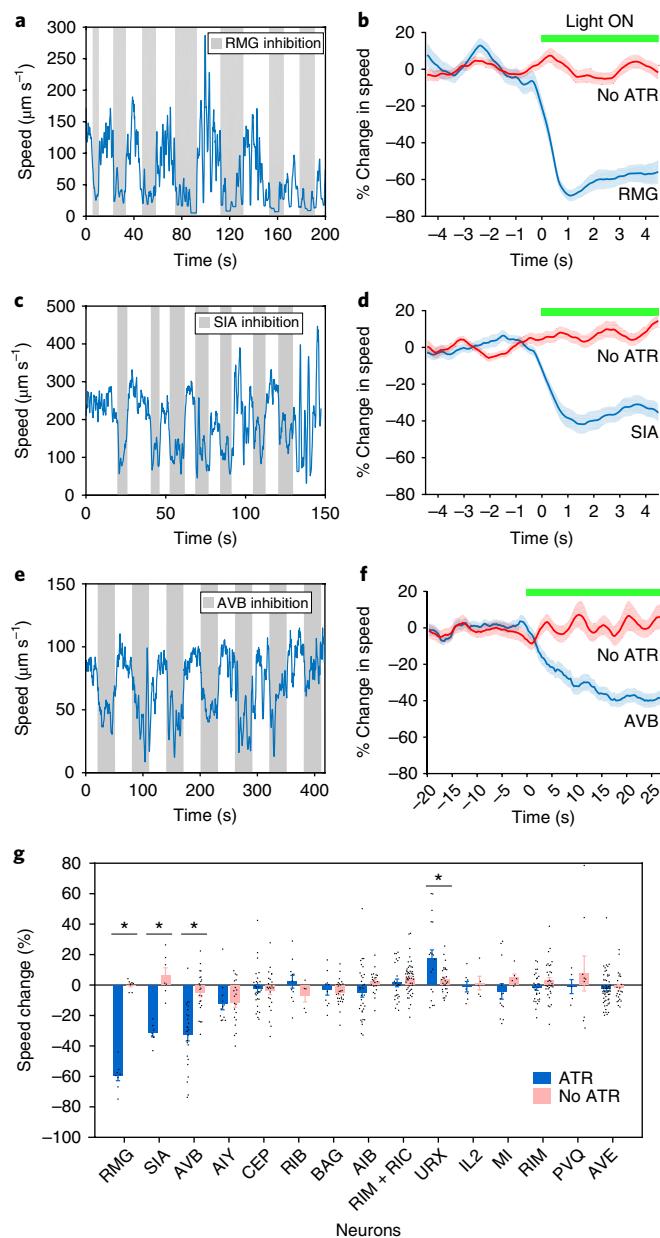


Fig. 3 | Targeted inhibition of single neurons SIA, RMG, and AVB reduces the speed of locomotion. **a**, Time trace of the speed of an animal with targeted RMG inhibition using on-off 540 nm light cycle. Gray bars, light on. **b**, Percentage speed change upon optical inhibition. Speed trace normalized to the light-off speed as light switches from off to on, inhibiting RMG ($n=71$ ATR+, 86 ATR on-off light pulses on $n=5$ animals each). Green bars, light on. **c,d**, SIA ($n=47$ ATR+, 54 ATR on-off light pulses on $n=5$ animals each). **e,f**, AVB ($n=153$ ATR+, 122 ATR on-off light pulse on $n=10$ animals each). **b,d,f**, Data mean (solid line) \pm s.e.m. (shaded area). **g**, Percentage speed change upon targeted inhibition of specified neurons (all n values can be found in Supplementary Table 3). Blue bars, animals fed with ATR; pink bars, no ATR controls. Data shown as mean \pm s.e.m. (*RMG, $P=5.36 \times 10^{-11}$; SIA, $P=2.23 \times 10^{-5}$; AVB, $P=1.26 \times 10^{-6}$; URX, $P=0.01$; two-sample, two-sided t -test). P values for other neurons are greater than 0.05 (see Supplementary Table 3).

during the onset of transition from backward to forward motion (Fig. 4h,i, Supplementary Fig. 15f and Supplementary Video 9). In addition to these dynamics, we also observed that AVB activity decreases during pausing (Fig. 4h and Supplementary Fig. 15c,d).

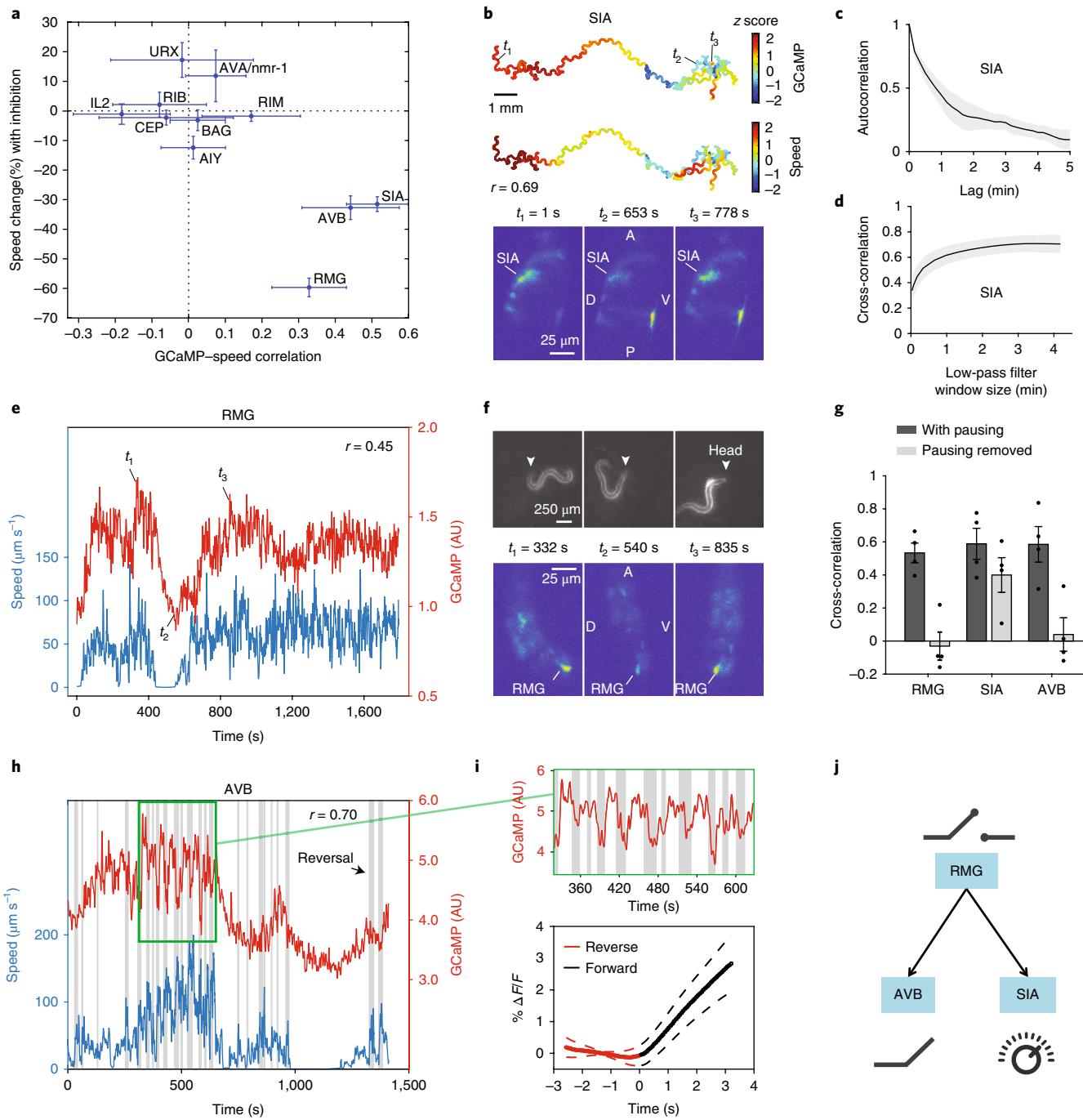


Fig. 4 | Calcium imaging of SIA, RMG, and AVB reveals distinct roles in controlling speed. **a**, Cross-correlation between calcium imaging and speed versus the speed change upon inhibition for neuron types. RMG ($n=10, P=0.0103$), SIA ($n=9, P=0.0003$), AVB ($n=7, P=0.0159$), AIY ($n=7, P=0.89$), CEP ($n=5, P=0.76$), RIM ($n=9, P=0.24$), RIB ($n=7, P=0.56$), IL2 ($n=5, P=0.24$), URX ($n=4, P=0.93$), BAG ($n=5, P=0.76$), AVA ($n=5, P=0.42$) where n is the number of animals imaged for each neuron type. One-sample two-tailed t -test. AVA/nmr-1 data point represents GCaMP imaging from AVA and inhibition of all neurons in Pnmr-1::Arch line. **b**, Trajectory of a worm color coded by SIA calcium activity (top) and the speed of the animal (bottom). High-magnification images (from camera 2 in Fig. 2) are shown at three time points marked on the trajectory. A, anterior; P, posterior; D, dorsal; V, ventral. **c**, Autocorrelation function for SIA activity shows the correlation time of approximately 1 min ($n=9$ animals). **d**, Cross-correlation between speed and SIA activity with low-pass filter ($n=9$). **e**, Time trace of RMG activity (red) and speed (blue) for a typical animal (representative of $n=4$ animals). **f**, RMG intensity images and the corresponding images of the animal at three different time points from the experiment shown in **e**. **g**, Cross-correlation between RMG, SIA, AVB ($n=4$ for each), and speed, with pausing and after pausing regions were removed. **h**, Trace of AVB GCaMP6s (red) and speed (blue) for a typical freely behaving animal (representative of $n=7$ animals). AVB activity correlates both with pausing events and forward motion onset. Gray bars indicate reversal events. **i**, Top, zoom-in to show the increase in AVB activity with forward motion initiation. Bottom, quantification of the increase in AVB activity with forward motion onset over reversal cycles ($n=63$ reversals). Time zero is the end of reversal and beginning of forward motion. **a,c,d,g,i**, Data are shown as mean \pm s.e.m. **j**, RMG-SIA-AVB circuit controls speed. RMG synapses onto both AVB and SIA. RMG acts as switch controlling pausing, AVB acts as a rectifier controlling forward motion, and SIA modulates the speed continuously.

Discussion

As *C. elegans* searches for food in its environment, the spatial profiles of odors are translated into temporally changing signals by ciliated sensory neurons exposed at the nose tip. These signals are then processed by the interneurons of the animal to determine exploration strategies. The translation of the spatial information into a time series of odor signals is determined by the speed of the animal. Here, we identified the three interneuron types, RMG, SIA and AVB, that control the nematode's speed. These three neuron types form a circuit with RMG synapsing onto both AVB and SIA²⁴ (Fig. 4j). RMG, an inter-motor neuron that synapses predominantly onto both head motor neurons and head muscles, has been implicated previously in the control of how active the animal is in different oxygen concentrations^{43,44}, and has been termed a 'hub' interneuron involved in signal integration⁴⁵. SIA neurons were originally identified as interneurons²⁴ but they also innervate neck muscle groups¹¹. In this circuit of RMG, SIA, and AVB, our data together suggest that RMG acts as a switch to determine whether the animal moves or not, AVB acts as a rectifier, to determine whether the animal moves forward or not, and SIA modulates the speed continuously over timescales of a minute (Fig. 4j). Our work raises the question of how inputs into this circuit, including from dopaminergic neurons, which are upstream of RMG and SIA, have a role in modulating circuit activity and the resulting speed modulations during chemotaxis.

Statistical analyses of the measurement matrix through corruption and changes in size enable the robustness, and false-positive and false-negative rates of the inferences to be determined. Such analyses also enable us to identify specific promoters that can be added to the measurement matrix to improve the quality and accuracy of the inferences. These statistical analyses on our experimental measurement matrix suggest that we could have missed one interneuron subtype in our set of key neurons controlling speed. Future experiments activating neurons using channel rhodopsin and analyzing the data using our framework will help identify additional neurons we could have missed.

In conclusion, a compressed sensing-based framework that exploits non-specific genetic tools, in conjunction with the microscope and the set of archaerhodopsin lines developed in this work should enable rapid and comprehensive understanding of the neural circuits that drive the behavior of *C. elegans*. Similar experimental methods based on compressed sensing have the potential to discover the key nodes that control a phenotype in complex biological networks, including the nervous systems of higher organisms as well as gene regulatory networks.

URLs. CVX package (<http://cvxr.com/cvx>); MNIST (<http://yann.lecun.com/exdb/mnist/>)

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41592-018-0233-6>.

Received: 18 February 2018; Accepted: 31 October 2018;

Published online: 20 December 2018

References

- Trent, C., Tsuing, N. & Horvitz, H. R. Egg-laying defective mutants of the nematode *Caenorhabditis elegans*. *Genetics* **104**, 619–647 (1983).
- Beverly, M., Anbil, S. & Sengupta, P. Degeneracy and neuromodulation among thermosensory neurons contribute to robust thermosensory behaviors in *Caenorhabditis elegans*. *J. Neurosci.* **31**, 11718–11727 (2011).
- Luo, L. et al. Bidirectional thermotaxis in *Caenorhabditis elegans* is mediated by distinct sensorimotor strategies driven by the AFD thermosensory neurons. *Proc. Natl Acad. Sci. USA* **111**, 2776–2781 (2014).
- Zimmer, M. et al. Neurons detect increases and decreases in oxygen levels using distinct guanylate cyclases. *Neuron* **61**, 865–879 (2009).
- Zaslaver, A. et al. Hierarchical sparse coding in the sensory system of *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **112**, 1185–1189 (2015).
- White, J. Q. & Jorgensen, E. M. Sensation in a single neuron pair represses male behavior in hermaphrodites. *Neuron* **75**, 593–600 (2012).
- Tran, D. H., Meissner, G. W., French, R. L. & Baker, B. S. A small subset of fruitless subesophageal neurons modulate early courtship in *Drosophila*. *PLoS ONE* **9**, e95472 (2014).
- Huang, K.-H., Ahrens, M. B., Dunn, T. W. & Engert, F. Spinal projection neurons control turning behaviors in zebrafish. *Curr. Biol.* **23**, 1566–1573 (2013).
- Capelli, P., Pivetta, C., Soledad Esposito, M. & Arber, S. Locomotor speed control circuits in the caudal brainstem. *Nature* **551**, 373–377 (2017).
- Milo, R. et al. Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827 (2002).
- Gray, J. M., Hill, J. J. & Bargmann, C. I. A circuit for navigation in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **102**, 3184–3191 (2005).
- Hobert, O., Glenewinkel, L. & White, J. Revisiting neuronal cell type classification in *Caenorhabditis elegans*. *Curr. Biol.* **26**, R1197–R1203 (2016).
- Hawrylycz, M. J. et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012).
- Lein, E. S. et al. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).
- Murphy, D. K., Herman, A. M. & Arenkiel, B. R. Dissecting inhibitory brain circuits with genetically-targeted technologies. *Front. Neural Circuits* **8**, 124 (2014).
- Candès, E. J., Romberg, J. K. & Tao, T. Stable signal recovery from incomplete and inaccurate measurements. *Commun. Pure Appl. Math.* **59**, 1207–1223 (2006).
- Donoho, D. L. Compressed sensing. *IEEE Trans. Inf. Theory* **52**, 1289–1306 (2006).
- Lustig, M., Donoho, D. L., Santos, J. M. & Pauly, J. M. Compressed sensing MRI. *IEEE Signal Process. Mag.* **25**, 72–82 (2008).
- Satat, G., Tancik, M. & Raskar, R. Lensless Imaging With Compressive Ultrafast Sensing. *IEEE Trans. Comput. Imaging* **3**, 398–407 (2017).
- Shental, N., Amir, A. & Zuk, O. Identification of rare alleles and their carriers using compressed se(que)nsing. *Nucleic Acids Res.* **38**, e179 (2010).
- Amir, A. & Zuk, O. Bacterial community reconstruction using compressed sensing. *J. Comput. Biol.* **18**, 1723–1741 (2011).
- Candes, E. J. & Wakin, M. B. An introduction to compressive sampling. *IEEE Signal Process. Mag.* **25**, 21–30 (2008).
- Candes, E. J. & Tao, T. Decoding by linear programming. *IEEE Trans. Inf. Theory* **51**, 4203–4215 (2005).
- White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Phil. Trans. R. Soc. Lond. B* **314**, 1–340 (1986).
- Xu, M. et al. Computer assisted assembly of connectomes from electron micrographs: application to *Caenorhabditis elegans*. *PLoS ONE* **8**, e54050 (2013).
- Edwards, S. L. et al. A novel molecular solution for ultraviolet light detection in *Caenorhabditis elegans*. *PLoS Biol.* **6**, e198 (2008).
- Chow, B. Y. et al. High-performance genetically targetable optical neural silencing by light-driven proton pumps. *Nature* **463**, 98–102 (2010).
- Husson, S. J. et al. Microbial light-activatable proton pumps as neuronal inhibitors to functionally dissect neuronal networks in *C. elegans*. *PLoS ONE* **7**, e40937 (2012).
- Tibshirani, R. Regression shrinkage and selection via the lasso: a retrospective. *J. R. Stat. Soc. Series B Stat. Methodol.* **73**, 273–282 (2011).
- Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Series B Stat. Methodol.* **58**, 267–288 (1996).
- Hu, T. & Chklovskii, D.B. Reconstruction of Sparse Circuits Using Multi-neuronal Excitation (RESCUME). *Adv. Neural Inf. Process. Syst. NIPS* **22**, (2009).
- Thrampoulidis, C., Abbasi, E. & Hassibi, B. in *Advances in Neural Information Processing Systems* **28** (eds. Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M. & Garnett, R.) 3420–3428 (Curran Associates, 2015).
- Shipley, F. B., Clark, C. M., Alkema, M. J. & Leifer, A. M. Simultaneous optogenetic manipulation and calcium imaging in freely moving *C. elegans*. *Front. Neural Circuits* **8**, 28 (2014).
- Piggott, B. J., Liu, J., Feng, Z., Wescott, S. A. & Xu, X. Z. S. The neural circuits and synaptic mechanisms underlying motor initiation in *C. elegans*. *Cell* **147**, 922–933 (2011).
- Nguyen, J. P. et al. Whole-brain calcium imaging with cellular resolution in freely behaving *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **113**, E1074–E1081 (2016).
- Faumont, S. et al. An image-free opto-mechanical system for creating virtual environments and imaging neuronal activity in freely moving *Caenorhabditis elegans*. *PLoS ONE* **6**, e24666 (2011).
- Chen, T.-W. et al. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295–300 (2013).
- Jazayeri, M. & Afraz, A. Navigating the neural space in search of the neural code. *Neuron* **93**, 1003–1014 (2017).

39. Li, Z., Liu, J., Zheng, M. & Xu, X. Z. S. Encoding of both analog- and digital-like behavioral outputs by one *C. elegans* interneuron. *Cell* **159**, 751–765 (2014).
40. Kawano, T. et al. An imbalancing act: gap junctions reduce the backward motor circuit activity to bias *C. elegans* for forward locomotion. *Neuron* **72**, 572–586 (2011).
41. Chalfie, M. et al. The neural circuit for touch sensitivity in *Caenorhabditis elegans*. *J. Neurosci.* **5**, 956–964 (1985).
42. Roberts, W. M. et al. A stochastic neuronal model predicts random search behaviors at multiple spatial scales in *C. elegans*. *eLife* **5**, e12572 (2016).
43. Busch, K. E. et al. Tonic signaling from O₂ sensors sets neural circuit activity and behavioral state. *Nat. Neurosci.* **15**, 581–591 (2012).
44. Laurent, P. et al. Decoding a neural circuit controlling global animal state in *C. elegans*. *eLife* **4**, e04241 (2015).
45. Macosko, E. Z. et al. A hub-and-spoke circuit drives pheromone attraction and social behaviour in *C. elegans*. *Nature* **458**, 1171–1175 (2009).

Acknowledgements

We thank N. Ringstad and members of the Ramanathan Laboratory for useful discussions and comments. This work was supported by a US National Institutes of Health (NIH) Directors Pioneer award (DP1OD008197, to S.R.) and a grant from the NIH (R01DC016058, to J.S.). We also acknowledge support from the NSF-Simons Center for Mathematical and Statistical Analysis of Biology at Harvard University, and from NSF grant number DMS-1764269, and the Harvard FAS Quantitative Biology Initiative (to S.R.). A.K. is supported by Harvard University - Koç University Visiting Scholar Program.

Author contributions

J.B.L. designed and performed experiments, analyzed data, developed the tracking microscope, and wrote the manuscript. A.Y. generated transgenic lines, designed and performed experiments, analyzed data and performed simulations, developed the structural illumination for the microscope, and wrote the manuscript. T.H. performed and analyzed experiments. C.-H.S. generated transgenic lines, and performed and analyzed experiments. J.M. generated transgenic lines, and performed and analyzed some of the Arch inhibition experiments. J.S. supervised the annotation of the expression patterns of different promoters in the different neuronal subtypes. A.K. developed the tracking microscope, and performed and analyzed experiments. S.R. designed and supervised the project, and wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41592-018-0233-6>.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to A.Y. or S.R.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2018

Methods

C. elegans strains. Strains were grown and maintained under standard conditions unless indicated otherwise. All experiments were carried out using *lite-1*(ce314)-mutant animals, which have lower sensitivity to blue light than the wild type. All transgenic Arch lines were generated by injecting fused promoter::Archaeorhodopsin-3-tagRFP or GFP constructs. The details of all transgenic lines used in this study can be found in Supplementary Table 1 and Supplementary Table 2.

Promoter expression, neuron identification, and the measurement matrix.

The neuronal expression pattern of the promoter::Arch constructs was verified using high-magnification (63 \times) fluorescence and differential interference contrast (DIC) microscopy (Zeiss Cell Observer). Only lines showing a phenotype or expression patterns that were obviously different from that in the literature under a lower magnification (15 \times) fluorescence microscope were examined (Zeiss SteREO Discovery, v12). In most instances, a combination of fluorescence–DIC co-localization and neuron process morphology was sufficient to determine the expression pattern. We imaged a minimum of 10 animals for each line and claimed that the line expressed those neurons reliably if they were observed in at least 75% of the animals. Promoters and lines that showed a higher degree of mosaicism were discarded. The measured expression patterns were used to generate the measurement matrix M .

Behavioral assay with optogenetic perturbations and image processing. All promoters were fused to Archaeorhodopsin-3 by fusion PCR and injected into *C. elegans* using standard protocols⁴⁶ to create transgenic animals. Transgenic animals were fed ATR (1 mM), a cofactor required for rhodopsin activity, for 12 h or more before the behavioral assay. To minimize the effect of the mosaicism in the experiments, only animals that expressed Arch in all of the identified neurons in each behavioral experiment were selected carefully using a fluorescent microscope. Only young adults from each transgenic line were used in the behavioral assay. Between 5 and 50 animals at a time were placed 1.5 cm away from the center of the bacteria lawn (approximately 3 mm in radius). The same number of worms per run was used for corresponding controls every time. The green light (5 mW mm⁻²) from a ScopeLED G250 that was used to activate Arch covered >70% of the plate area with less than 15% variation in intensity, as verified with a Thorlabs S120C photodiode. Worms can behave normally for more than 1 h under a 5.5 mW mm⁻² green light⁴⁶. Therefore, we used 5 mW mm⁻² as the safest known high power level in our silencing experiments. A video camera was used to record the movements of the worms at 3–5 frames per s (fps)⁴⁷. Only data from approximately 70% of the plate area centered on the bacterial lawn was processed and analyzed. For each set of experiments performed on a day, a parallel set of positive and negative control experiments were performed under identical conditions. As positive controls, experiments were performed on animals with pan-neuronal expression of Archeorhodopsin. Under illumination, these animals are paralyzed, thus validating our set up as well as the quality of ATR. As negative controls, animals from the same line that were not fed ATR (and hence were unaffected by green light) and *lite-1* animals were used. The same LED illumination was used for the control and the experiments.

Custom MATLAB software with elements from ‘The Parallel Worm Tracker’⁴⁸ was used to segment and track the position of individual nematodes across camera frames. We tracked worms until they entered the bacterial lawn. To estimate sub-pixel and hence sub-resolution-scale displacements, we binned the real space into discrete pixels based on pixel size. Thus, a nonzero displacement was only observed if the animal moved more than two times the pixel width, the distance between diagonally adjacent pixels. Such analysis is likely to be more accurate than arbitrary time averaging, which is typically used but does not take spatial resolution into account. Under roughly 1 \times magnification, each animal was typically around 50–100 pixels in size. Speed was calculated by taking the time derivative of the centroid position.

Data analysis for behavior screening. The speed of animals was measured until they reached the bacterial lawn. Speed naturally fluctuates during the movement of wild-type animals, and so to determine whether certain neurons affect speed, it is necessary to measure the distribution of the speed over the course of the behavioral experiment. Therefore, for each transgenic line and corresponding control, speed data from at least 10 animals were analyzed to obtain the speed distribution for that line. To determine whether the transgenic line showed significant changes in speed, we measured the KL divergence between the speed distribution of this line and the control experiment. In parallel, we also measured the KL divergence between the 131 distinct control experiments in our dataset to obtain the distribution of the intra-control KL divergence values. As long as the KL divergence between the speed distribution of the transgenic line and the control was above the 95th percentile of the intra-control KL divergence values, we assumed that the transgenic line showed a speed phenotype.

To infer key neurons robustly, we tested two different phenotype vectors based on KL divergence: binary and continuous. For the continuous phenotype vector, mean speed reductions were calculated with respect to control. Phenotypes for promoters in which experimental and control lines are indistinguishable based on

KL-divergence were taken to have no speed change. Promoters not distinguishable from control were taken to have no speed change (we carried out further tests to check that this assumption did not affect the results). Using the phenotype vector \vec{P} , and the measurement matrix M , both the neuron weights $\vec{\omega}$ and confidence in these weights were calculated for each of 88 neuron types. To calculate the confidence in the weights, errors in the measured mean speed changes were propagated onto inferred neuron weights with bootstrapping. We sampled the phenotype vector 10,000 times from normal distribution with a mean as fractional mean speed changes and s.d. as the standard error of mean speed changes. We then used these phenotype vectors to calculate the weight distribution for each of the 88 neuron types. For these analyses, Lasso and LassoCV packages from the Python machine learning library (scikit-learn v0.19.2) were used with the default tolerance value. Other optimization methods were also tested, such as MATLAB lasso, convex optimization package (CVX package), and MATLAB *linprog* (L1-norm minimization: $\min ||\vec{\omega}||_1$ subject to $M\vec{\omega} = \vec{P}$). These gave consistent results.

Artificial neural networks. To test whether key neurons could be inferred in a nonlinear neural network, we trained a fully connected feed-forward neural network using deep learning algorithms to recognize handwritten digits (which is a standard task for deep neural networks)⁴⁹. This network classified digits as 0 to 9 from 28 \times 28 grayscale images (MNIST). The network consisted of the input layer of 784 neurons, 4 hidden layers each with 100 neurons, and the output layer consist of 10 neurons signaling the output class of the digits (Supplementary Fig. 10a). Each neuron in the network was extremely nonlinear with a sigmoidal activation function (Supplementary Fig. 10b). We trained this network on a training dataset (60,000 images) using categorical cross entropy as a loss function. During training to produce networks with a small number of key neurons and to promote sparse representation of the data, we added L1 constraint to activations of neurons. After training, when the network was shown examples of handwritten digits from the training dataset, a fraction of units (neurons) in the network showed activity for individual digit images, whereas with white noise data the activation was broad (Supplementary Fig. 10c). For simplicity, we focused on a single hidden (intermediate) layer of the trained network in our analysis. Exploiting our computational access to the neurons in this network, we inhibited each neuron in a layer one at a time to measure the change in accuracy of handwriting recognition on a test dataset consisting of 10,000 images. These measurements quantified the weights of individual neurons in handwriting recognition and constituted the ground truth for the network (Supplementary Fig. 10d). We then tested whether compressed sensing could recover these weights and, in particular, identify neurons that have an essential role in digit recognition. To do so, we built a measurement matrix by generating 50 groups each containing 5 randomly chosen neurons. As in our *C. elegans* experiments, we inhibited all the neurons in a row of the measurement matrix and measured the phenotype (change in accuracy of handwriting recognition on the test dataset) (Supplementary Fig. 10e,f). Using Lasso, with this measurement matrix we could correctly identify the key neurons essential for digit recognition. (Supplementary Fig. 10g–i). We also determined how error rate and accuracy of inference changes with the size (number of rows) of the measurement matrix. We changed the number of rows (equivalent to the number of independent measurements using different groups) from 25 to 60 to determine how performance of the method changes with the size of the measurement matrix (Supplementary Fig. 10j). These tests suggested that compressed sensing can be used to find essential neurons in this nonlinear artificial neural network. This work was implemented using Keras v2.1.3 Python deep learning library with TensorFlow backend.

Statistical analysis of the measurement matrix. To evaluate the performance of our measurement matrix, we performed simulations. In these simulations we first assigned a randomly chosen set of n neurons (with n between 1 and 11 out of the total of 88 subtypes covered by the measurement matrix) as essential for behavior. We assigned weights ($\vec{\omega}_{\text{real}}$) to these essential neurons such that when each of these essential neurons was inhibited in our simulation, the phenotype changed positively or negatively by at least 30% of the wild-type value. We assigned 0 weights to non-essential neurons. We first assumed that when multiple neurons were inhibited simultaneously, the effects of these neurons on the phenotype added linearly. We used our measurement matrix to calculate the phenotype vector $\vec{P} = M\vec{\omega}_{\text{real}}$. We then hid the identities of the essential neurons and attempted to infer these identities from this phenotype vector \vec{P} using compressed sensing. We used Lasso to infer the weights $\vec{\omega}_{\text{inferred}}$ of individual neurons. We compared $\vec{\omega}_{\text{inferred}}$ against $\vec{\omega}_{\text{real}}$ to obtain false-positive and false-negative rates. By performing this computational test thousand times for different randomly chosen sets of essential neurons, we determined the average false-positive and false-negative rates (Supplementary Fig. 9a). We then ascertained that the false-positive and false-negative rates were robust to how the phenotype vector was evaluated by using a nonlinear function such that the combined phenotype of multiple neurons did not add linearly. We did so by evaluating the phenotype $\vec{P} = \text{sgn}(M\vec{\omega}_{\text{real}})$ where $\text{sgn}(x) = \{-1 \text{ if } x < 0; 0 \text{ if } x=0; 1 \text{ if } x > 0\}$ and repeated the calculation of the false-positive and false-negative rates (Supplementary Fig. 9b). We further evaluated the false-positive and false-negative rates as the measurement matrix size changed

from 22 to 32 rows (promoters) (Supplementary Fig. 9c–e), and evaluated how these rates depended on the total number of promoter constructs used in the experiments (Supplementary Fig. 9f).

Promoter removals and additions. We tested the robustness of the solution (RMG, SIA, and AVB) to the removal of promoters from and additions of promoters to the original set, and thus to changes in the measurement matrix. We used mean speed changes as our phenotype vector as before, but we did not carry out bootstrapping analysis with the uncertainty on the measurements, for simplicity. First, we randomly removed from 1 to 5 promoters from our measurement matrix to create new measurement matrices. We next obtained the identities of the key neurons for a range of sparsity parameters for each of these new matrices (Supplementary Fig. 6a). In all instances, the inference of AVB, SIA, and RMG was robust to the removal of promoters (Supplementary Fig. 6b). Nevertheless, upon removal of promoter *mgl-1* (*Pmgl-1*), the solution contains AIA and the weight of AVB decreases (Supplementary Fig. 6b). This is because only two promoters drive expression in AIA in our measurement matrix: *Pmgl-1* with no speed change and *Psra-11* with significant speed change. Therefore, removal of *Pmgl-1* introduces AIA as a neuron that can control speed. Once *Pmgl-1* and the fact that it does not show a speed phenotype are taken into account, AIA is not contained in the sparse solution (Supplementary Fig. 6c). The sensitivity of the inference to perturbations of the measurement matrix enables us to determine the neurons whose inferences are most sensitive to the identity of the chosen promoters.

To test the robustness upon promoter additions, we generated five more transgenic lines expressing Arch in multiple neurons and measured the speed phenotype in each of these lines. We added these five new promoters (*Podr-2(16)*, *Popt-3*, *Pntr-2*, *Psams-5*, *Ptdc-1*) to the original set to increase the size of our matrix to 32 promoters. The identity of the key inferred neurons remained unchanged with these new behavioral data and the expanded measurement matrix (Supplementary Fig. 7).

Arch efficiency simulations. Although the Arch lines were imaged at high magnification to validate their expression pattern and only animals showing clear expression patterns of the fluorescent tag were picked for behavioral assays, errors in our inference due to variations in the levels of archaerhodopsin expression are possible. We further tested the robustness of our methods to corruption of the measurement matrix. Our original measurement matrix had an entry of 1 if a promoter drove expression in a neuron type, and 0 otherwise. We corrupted 10% of the entries of our measurement matrix at random computationally, such that the 1s were changed with equal probability to a number between 0 and 0.5 (corresponding to randomly reducing or abrogating the expression of a subset of the neurons in a subset of the promoters). We thus generated 1,000 distinct corrupted measurement matrices. We then used these corrupted measurement matrices to infer the key neurons from behavioral experiments to determine the statistical robustness of our inferences to reduction and abrogation of expression levels in subsets of neurons. In almost every instance, the solutions (RMG, SIA, and AVB) were robust. There were two exceptions: when the corruption of the measurement matrix (Supplementary Fig. 8a) led to the abrogation of expression in RMG by the promoter *Pflp21* or in AVB by the promoter *Psra-11*, RMG and AVB were missed as key neurons, respectively. This is because in our entire set of promoters, only *Pflp21* drives expression in RMG, and only *Psra-11* drives expression in AVB. Such computational perturbations of expression of individual promoters in SIA did not affect the inference because multiple promoters in our Arch lines show expression in SIA (Supplementary Fig. 8a,b).

Tracking (stabilization) microscope. Real-time tracking and image stabilization. Tracking and image stabilization are achieved through a combination of fast acquisition, computation, and mechanical feedback. Our system is able to image, process, and feedback to *x*, *y*, *z* coordinates and rotational controls, all within 7 ms. To stabilize worm movement, motion in the *x*, *y*, *z* coordinates, and rotation (*r*) resulting from head bending, must all be tracked simultaneously. In brief, a green 543 nm laser (<5 mW mm⁻²) is used to excite the mKusabira-Orange (mKO)-tagged marker neuron, AWC^{ON}. Fluorescence is collected by a 50×/0.55-NA (numerical aperture) (Nikon 50× LU Plan ELWD 0.55-NA) objective and passed through a dove prism mounted on a rotation stage. The dove prism uses internal reflections to rotate the image. The rotated beam is then split into two adjacent images that focus at slightly different *z* depths on an EMCCD (electron multiplying charge-coupled device); one is in focus, the other slightly out of focus (Fig. 2b). This step, colloquially referred to as dual-view, was accomplished by inserting a lens in one of the otherwise optically equal length paths and is a modern interpretation of a classic technique used by H. Berg to track swimming *E. coli*⁵⁰. The pixels from the camera are then read serially onto the FPGA (field programmable gate array) and processed immediately as they enter.

To facilitate fast and reliable real-time processing, FPGAs were used. FPGAs are parallel-processing CPUs (central processing units) that enable high-speed synchronization of multiple simultaneous processes, and this was required for our tracking system. Positional *x*, *y*, *z*, *r* information was simultaneously extracted from raw images in parallel, without any one process delaying another. The *x*,

y, *z* position was calculated from the centroid of the cell body in the focused image. The *z* position was measured by calculating the difference in intensity between the focused and out-of-focus images. The angular orientation, *θ*, was calculated from the marker neuron's process, which extends from the cell body to the animal's nose tip. The displacement from the target position was then passed to *x*, *y*, *z*, *θ* stages (*x*, *y*, piezo + servo; *z*, stepper; *θ*, servo) to mechanically move the worm back to the target location. This was all done at 100 fps and less than 10 ms latency. As a result, we could stabilize the >250 μm s⁻¹ movements of the marker neuron cell body to 1 μm; less than its radius. Head bending motion artifacts were also reduced twofold (Fig. 2b). Tracking and stabilization with 1-micron accuracy could be maintained over an hour and over centimeters of travel.

Dove prism for image rotation. We used a dove prism (Thorlabs PS992-A) mounted in a rotation stage to rotate the image and compensate for bending of the worm's head, as rapid computation by the FPGA was only possible using addition and subtraction operations, and hence there was insufficient time to computationally rotate the image. Dove prisms have been used in astronomy and machine vision as 'beam rotators'⁵¹. A challenge of aligning the dove prism is that its rotation axis must be collinear with that of the rotation stage. Small lateral or angular deviations result in image nutation, or 'rocking'. Fortunately, there are some established alignment procedures^{52,53}. We assembled an alignment apparatus using a combination of off-the-shelf stages (ThorLabs CXY1 and KS1). Then, we aligned the dove prism by minimizing the rotational displacement. More precisely, we measured the trajectory of a point-like fluorescent particle in our field of view as we rotated the dove prism and iteratively adjusted translations orthogonal to the optical axis as well as the azimuth (tilt) to minimize the path total length and asymmetry, as done previously in the literature^{52,53}.

***z* tracking from two focal planes.** We tracked the *z* position by comparing the 'fuzziness' between two images of the marker neuron taken at different focal planes. The technique was first used by Berg in 1971 (ref. ⁵⁰) to track swimming *E. coli*. The fuzziness of an object, as measured by its intensity, does not give the absolute position as the fuzziness is symmetric above and below the focal plane. Comparing this to the image from a different focal plane, however, enables the exact *z* position to be calculated as the difference in intensity between the two images is monotonic near either focal plane. Berg explains, as the object moves toward one focal plane, "its image sharpens on [that focal plane] and fuzzes out more on [the other plane]"⁵⁰.

***x*, *y*, *z*, *r* stages.** To achieve precision and speed, as well as range, we stacked fast but small *x*, *y* piezo stages (PI) on long travel servomotor stages (Newport LTA-HS). The micron-scale position was read from the encoded stages onto an FPGA (National Instruments PXIe-7966R) as an analog voltage. A non-encoded linear stepper motorized stage (Thorlabs ZFS13B, Sparkfun EasyDriver) was used to control the *z* position. The dove prism used for rotating the image was placed in a servo-controlled rotation stage (Newport URB100C) that was set to track a voltage and corresponding angle using its own internal PID feedback. The servo stages were all driven by a stage controller (Newport XPS Controller) and custom-written.tcl scripts. All tracking movement signals were processed and sent by an FPGA. Custom-built digital-to-analog and analog-to-digital converters (Harvard University Electronic Instrument Design Lab) were used wherever necessary.

LabVIEW FPGA for hardware control and real-time processing. FPGAs were used to control both tracking and imaging. Processes on FPGAs can run in parallel, as opposed to sequentially on a CPU. This means that each process loop runs exactly at its specified rate no matter the loading on itself or other loops. We used a National Instruments 1483 Camera Link Adaptor Module to interface the camera (Andor iXon Ultra 897) to the FPGA. The different stages have different sampling rates and because the FPGA can communicate with each of them independently, no process delays another.

FPGAs were programmed with The LabVIEW FPGA platform without using low-level hardware language. LabVIEW FPGA compiles code written in the standard LabVIEW interface into VHDL (Hardware Description Language). It has the additional benefit of a PC interface that we used to not only save images and data, but also to monitor the experiment in real time and manually control the stages when necessary. Although data were saved on the PC when rates were variable, we minimized timing uncertainty by bundling timestamps from the FPGA clock with the data. All GCaMP (a genetically encoded calcium indicator), tracking, and wide-field images were synchronized using these timestamps.

The algorithms used to process images and track the worm are simple and fast. In the *x*, *y*, *z* tracking loop, two focal plane images of the marker neuron are first thresholded to remove all but the soma. The centroid of the soma from the in-focus image was calculated to obtain *x* and *y* positions. The difference in power, sum of all pixel intensities, between the in- and out-of-focus images was used as a measurement for *z* position. In the tracking loop, a different threshold was used to remove background noise but keep the cell process. The angular orientation of the process was determined by first fitting an approximate spline. The process was first split into roughly eight segments by binning pixel rows roughly perpendicular to the process, calculating the centroid in each of them, and then drawing a line

between centroids from adjacent bins. The angular orientations of these segments were then averaged to obtain the process' orientation. All of these position calculations occurred as pixels are serially read from the camera and completed by the time the last pixel in the image was read. This setup ensured the calculations did not add any latency, and their speeds were limited only by the camera's data transfer rate.

GCaMP calcium imaging. Tracking and image stabilization enabled the calcium signal to be imaged as if the worm were immobilized. A blue LED or DLP (digital light processing) projector (Logic PD DLP LightCommander) was used to provide wide-field illumination for GCaMP6s imaging. An exposure time of 30 ms was used, which enabled the use of blue light levels around only 0.5–1 mW mm⁻², several times less than the reported dangerous level⁵⁵. This low blue light level is one of the reasons that we have been able to image freely moving worms at high spatiotemporal resolution for an order of magnitude longer than other methods. The sampling rate was 15.625 fps. Stabilizing the image also enabled us to perform volumetric z scans as can be done on some commercial microscopes equipped with autofocus systems, such as the Zeiss Definite Focus and Nikon Perfect Focus, except that those systems track the coverslip surface rather than the sample. These commercial systems can compensate for temperature related z drift but cannot do volumetric scanning in a moving sample. Rather than changing the objective to sample distance to vary the focal plane, a liquid lens with an adjustable focal length (Optotune EL-10-30-C) was placed in a 1× relay system between the tube lens and GCaMP imaging camera (Hamamatsu ImagEM X2). An FPGA (National Instruments PXIe-7961) was used to acquire images from the camera and control the liquid lens. The focal length of the tunable lens is voltage controlled and can be varied between 100 and 200 mm. In this configuration, the tunable lens can scan through the entire approximately 50 μm width of the worm without any variance in image quality or magnification⁵⁴. Including the 30 ms exposure times, the entire worm (15 z-slices) could be scanned through in 1 s.

GFP control to characterize motion artifacts and sampling noise. We used GFP as a control to characterize remaining motion artifacts and noise in the microscope. A specific promoter (*Ptx-3*) was used to drive GFP expression in a left-right pair of interneurons: AIYR and AIYL. Movement artifacts would have the same effect on measuring GFP as on GCaMP6s. Similarly, readout and dark noise from the camera would be the same, as would shot noise given that the GFP and GCaMP6s intensities are comparable. GFP control shows that the error in the measurement, including all remaining motion artifacts and camera noise, was less than 18% as measured by the s.d. (Fig. 2e). Smoothed GFP intensity time series over 3 s had a s.d. of less than 10% indicating that much of the noise we observe is at frequencies faster than 0.3 Hz.

Structured illumination. The DLP projector was used for spatially structured illumination for targeted photoactivation. A series of approximately 75 mm relay lenses were used to map the image of the DLP to the sample plane and appropriate excitation filters were used depending on the application. The DLP was controlled using a combination of software written in LabVIEW and its native software. In our configuration, the DLP is limited to refreshing at 30 fps through HDMI, but with more hardware-level control, this could be increased to 60+ fps. See Supplementary Note 1 for the detailed catalog of the components of the microscope and the critical notes about the components.

Targeted inhibition of neurons. Animals expressing multiple neurons were fed with ATR overnight before tracking with our stabilization microscope. After tracking and stabilization, we located the position of the neuron of interest on the camera. A pattern of light was then projected from the DLP projector to selectively target the neuron of interest. To ensure that only the neuron of interest remained illuminated as the worm changed orientation, we used a custom LabView script that computationally rotated the projected image in synchrony with the dove prism. This was made possible by aligning the center of the rotation of the projected pattern to the center of rotation of the dove prism. By following the entire process on camera, we were able to confirm the selective illumination of the neuron (Fig. 2f).

Targeted inhibition experiments were performed by pulsing structured illumination patterns while recording the animals' speed. Speed changes due to inhibition (Fig. 3) were found by smoothing the speed traces with an approximately 1 s low-pass filter and then normalizing speed during pulses (light ON periods) with the speed between pulses (light OFF periods). If the OFF period average speed was less than 20 μm s⁻¹, we discarded those cycles as it generates artifacts and introduces bias to the distribution. The results are not sensitive to a 20 μm s⁻¹ threshold. Each track (data points) had five cycles on average. Transgenic lines used in targeted inhibition experiments and single-neuron-expressing lines related to Fig. 3 are listed in Supplementary Table 2.

Image analysis of calcium imaging data. GCaMP6s intensity information was extracted using custom software written in MATLAB. Regions of interest and

segmentation corresponding to different neurons in x,y are obtained from the maximum intensity projection. As a result of image stabilization and tracking, each neuron stays localized to a specific region of interest throughout the time series (Supplementary Fig. 11a,b). Within a region of interest that resolves neurons in x,y (Supplementary Fig. 11b), neurons at different z positions could be identified from the histogram of the position of each slice containing the maximum intensity pixel within each z-stack, as the intensity of a neuron increases towards the focused plane. The best focal plane was taken as that which maximizes intensity in the region of interest for a specific neuron (Supplementary Fig. 11c,d).

Calcium activity and speed correlation analysis. Cross correlation between calcium activity and speed of animals were obtained at different frequencies by applying low-pass, high-pass or band-pass filters to both GCaMP6s and speed to find the best significant correlation between two signals using the Pearson correlation. GCaMP6s data were randomly shuffled 100 times and the correlation coefficient was calculated. The correlation with the shuffled data was used to measure whether cross correlation between the actual calcium activity and speed was significant. For example, SIA experiments in Supplementary Fig. 15a,b show significant correlation between calcium activity and speed of locomotion even when the smoothing window sizes are up to 3 min long (Supplementary Fig. 15g,h) while the significance of the correlation of other neurons deteriorated with the smoothing window size (Supplementary Fig. 15i,j).

Statistical information and reproducibility. Statistical tests used in this study are indicated in the corresponding figure legends. Other randomization and bootstrapping methods applied to data were explained in detail in the Methods section. Details on the statistical test used in Fig. 3 can be found in the Supplementary Table 3. Experiments in Fig. 2b,c were repeated independently with similar results more than 50 times. Experiments in Fig. 2d,e were repeated three times independently with similar results. Targeted illumination control experiments as in Fig. 2f were repeated three times independently with similar results. Experiments in Fig. 3a–f were repeated independently with similar results 5, 5, 5, 10 and 10 times, respectively. Experiments in Fig. 4b were repeated nine times independently with similar results. Experiments in Fig. 4e,f were repeated four times independently with similar results. Experiments in Fig. 4h,i were repeated seven times independently with similar results. Experiments in Supplementary Videos 7,8,9 were repeated independently with similar results nine, four and seven times, respectively.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Code availability. Custom codes implementing the compressed sensing analysis and the statistical analysis of the measurement matrix are available at the Code Ocean platform^{55,56}.

Data availability

The data that support the findings presented in this study are available from the corresponding authors upon reasonable request.

References

46. Mello, C. & Fire, A. DNA transformation. *Methods Cell Biol.* **48**, 451–482 (1995).
47. Baek, J.-H., Cosman, P., Feng, Z., Silver, J. & Schafer, W. R. Using machine vision to analyze and classify *Caenorhabditis elegans* behavioral phenotypes quantitatively. *J. Neurosci. Methods* **118**, 9–21 (2002).
48. Ramot, D., Johnson, B. E. Jr, Berry, T. L. Jr, Carnell, L. & Goodman, M. B. The Parallel Worm Tracker: a platform for measuring average speed and drug-induced paralysis in nematodes. *PLoS ONE* **3**, e2208 (2008).
49. Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2324 (1998).
50. Berg, H. C. How to track bacteria. *Rev. Sci. Instrum.* **42**, 868–871 (1971).
51. Shin, C. W., Inokuchi, S. & Kim, K. I. Retina-like visual sensor for fast tracking and navigation robots. *Mach. Vis. Appl.* **10**, 1–8 (1997).
52. Mao, W. Error and adjustment of reflecting prisms. *Opt. Eng.* **36**, 3367–3372 (1997).
53. Sullivan, D. L. Alignment of rotational prisms. *Appl. Opt.* **11**, 2028–2032 (1972).
54. David Giese, J., Ford, T. N. & Mertz, J. Fast volumetric phase-gradient imaging in thick samples. *Opt. Express* **22**, 1152–1162 (2014).
55. Lee, J.B. et al. A compressed sensing framework for efficient dissection of neural circuits. *Code Ocean* <https://doi.org/10.24433/CO.020095e2-4067-4d44-9ea4-30d55309dda9>
56. Lee, J.B. et al. A compressed sensing framework for efficient dissection of neural circuits. *Code Ocean* <https://doi.org/10.24433/CO.3bc0d23b-b316-474d-9ab2-076e1deac88a>

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.

For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

► Experimental design

1. Sample size

Describe how sample size was determined.

No statistical methods are used to predetermine sample size. For speed measurements, the sample size were large enough to obtain accurate average speed distributions (corresponding to an hour long measurement) for each line. For targeted neuron inhibition and GCaMP imaging experiments, our sample sizes were similar to or larger than those reported in previous publications in the field.

2. Data exclusions

Describe any data exclusions.

In targeted neuron inhibition with light pulsing experiments, If off-light baseline average speed was less than 20um/s, we discarded those cycles since it generates artifacts and introduce bias as also described in the Methods section.

3. Replication

Describe the measures taken to verify the reproducibility of the experimental findings.

The number of replicates and experiments performed were reported in either corresponding figure legends or statistical information and reproducibility section in the Methods. Experiments by different team members for optogenetic inhibition and calcium imaging were pooled. All attempts at replication were successful.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

In optogenetic experiments, genetically identical worms randomly selected as either experimental group fed with ATR or control group not fed with ATR.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

No blinding was performed. The data acquisition and analysis pipeline was identical for all experimental and control animals and the analyses did not involve human scoring.

Note: all *in vivo* studies must report how sample size was determined and whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a Confirmed

- The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
- A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- A statement indicating how many times each experiment was replicated
- The statistical test(s) used and whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of any assumptions or corrections, such as an adjustment for multiple comparisons
- Test values indicating whether an effect is present
Provide confidence intervals or give results of significance tests (e.g. P values) as exact values whenever appropriate and with effect sizes noted.
- A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
- Clearly defined error bars in all relevant figure captions (with explicit mention of central tendency and variation)

See the web collection on [statistics for biologists](#) for further resources and guidance.

► Software

Policy information about [availability of computer code](#)

7. Software

Describe the software used to analyze the data in this study.

Python 2.7 - scikit-learn v0.19.2 (Lasso and LassoCV)
 Keras v2.1.3 (Python 3.5)
 Matlab (2015a, 2017a)
 CVX - <http://cvxr.com/cvx>
 Labview 2014
 Custom Software for Compressed Sensing analysis and its statistical analysis that produce exact figures in the paper are available from CodeOcean platform (DOI: 10.24433/CO.020095e2-4067-4d44-9ea4-30d55309dda9 and DOI:10.24433/CO.3bc0d23b-b316-474d-9ab2-076e1deac88a)

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). [Nature Methods guidance for providing algorithms and software for publication](#) provides further information on this topic.

► Materials and reagents

Policy information about [availability of materials](#)

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.

All unique materials are readily available from the authors

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies were used.

10. Eukaryotic cell lines

- a. State the source of each eukaryotic cell line used.
- b. Describe the method of cell line authentication used.
- c. Report whether the cell lines were tested for mycoplasma contamination.
- d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

No eukaryotic cell lines were used.

► Animals and human research participants

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study.

C. elegans (nematode)

Policy information about [studies involving human research participants](#)

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

This study did not involve human research participants.