Megapixel camera arrays for high-resolution animal tracking in multiwell plates

Ida Barlow[1,2]*, Luigi Feriani[1,2]*, Eleni Minga[1,2], Adam McDermott-Rouse[1,2], Thomas O'Brien[1,2], Ziwei Liu[1,2,3], Maximilian Hofbauer[4], John R. Stowers[4], Erik C. Andersen[5], Siyu Serena Ding[1,2,6], André E.X. Brown[1,2‡]
*These authors contributed equally
[1]Institute of Clinical Sciences, Imperial College London, London, UK; [2]MRC London Institute of Medical Sciences, London, UK; [3]Present address: Department of Medical Biochemistry and Biophysics, Karolinska Institute, Stockholm, Sweden; [4]Loopbio GmbH, Vienna, Austria; [5]Department of Molecular Biosciences, Northwestern University, Evanston, IL, USA; [6]Present address: Max Planck Institute of Animal Behavior, Konstanz, Germany
[‡] correspondences to: andre.brown@imperial.ac.uk

Abstract

Tracking small laboratory animals such as flies, fish, and worms is used for phenotyping in neuroscience, genetics, disease modelling, and drug discovery. Current imaging systems are limited either in spatial resolution or throughput. A system capable of imaging a large number of animals with sufficient resolution to estimate their pose would enable a new class of experiments where detailed behavioural differences are quantified but at a scale where hundreds of treatments can be tested simultaneously. Here we report a new imaging system consisting of an array of six 12-megapixel cameras that can simultaneously record from all the wells of a 96-well plate with a resolution of 80 pixels/mm at 25 frames per second. We show that this resolution is sufficient to estimate the pose of nematode worms including head identification and to extract high-dimensional phenotypic fingerprints. We illustrate the potential application of the system in three domains with a study of behavioural variability across wild isolates, experiments on the escape response and sensitisation of worms to repeated blue light stimulation, and the utility of this stimulus for phenotyping disease models. Because the system is compatible with standard multiwell plates, it makes computational ethological approaches accessible in existing high-throughput pipelines and greatly increases the scale of possible phenotypic screening experiments in *C. elegans*.

Introduction

Recording and quantifying animal behaviour is a core method in neuroscience, behavioural genetics, disease modelling, and psychiatric drug discovery. Both the scale of behaviour experiments and the information that can be extracted from them have increased dramatically and led to new classes of experiments that were not previously possible[1–4]. In terms of phenotypic content, a key parameter is the resolution of the recording. If there is sufficient spatial and temporal resolution, then body parts can be identified and tracked, the animal's pose can be estimated, and the full suite of computational ethology methods can be applied to analyse any behaviour of interest. Maintaining this resolution was the first key design constraint in our approach.

Automation of behaviour recording is also important for achieving scale. We wanted to use off-the-shelf multiwell plates so that any behaviour screening pipeline would still be compatible with existing pipeline elements such as liquid and plate-handling robots as well as small animal sorting machines. Because behaviour occurs over time, a standard plate-scanning approach in which each well of a multiwell plate is imaged in turn using a motorised stage limits throughput regardless of scan speed because each well must be recorded long enough to observe the behaviour of interest. We therefore wanted a system that would simultaneously record all of the wells of a multiwell plate. Moreover, mechanical arrangements with moving parts introduce higher maintenance costs and have a higher risk of failure compared to a static camera system.

Our solution to simultaneously image a large area with high resolution was to use an array of machine vision cameras that are small enough to be arranged in close proximity to one another with partially overlapping fields of view at a resolution sufficient to track small animal pose. Here we present the design of a six 12-megapixel camera array that uses a near-infrared light panel for illumination, a set of high intensity blue LEDs for photostimulation, and the associated open-source software for automatically identifying wells and keeping track of metadata. The software is fully integrated into our existing Tierpsy Tracker software[5], including a graphical user interface for reviewing tracking data, joining trajectories, and annotating problematic wells to discard from analysis as well as a neural network for distinguishing worms from non-worm objects. We demonstrate the potential of the new tracking system in neuroscience, disease modelling, and genetics.

Results

Megapixel camera array design

Based on our previous work with single-worm tracking, we set a target resolution of at least 75 pixels/mm and a recording rate of 25 frames per second in order to accurately estimate worm pose and identify the head from the tail which requires the measurement of the small head swinging behaviour that is often referred to as 'foraging' in the worm community. These constraints require a total of 8100 x 5400 pixels, or about 44 megapixels with a 3:2 aspect ratio, to cover a standard 96-well plate. Single cameras with this resolution that can record at 25 frames per second are not available commercially. We therefore considered arrays of cameras and found that six Basler acA4024 cameras (Basler AG, Ahrensburg, Germany) in a 3x2 array equipped with Fujinon HF3520-12M lenses (Fujifilm Holding Corporation, Tokyo, Japan) was an optimal solution: This combination of lenses and cameras allowed mounting the cameras in close proximity, whereas cameras with higher resolution or larger sensors would have required significantly larger lenses. Imaging a multiwell plate with

multiple cameras significantly reduces the blind spots caused by vertical separators between wells, compared with using a single camera with a conventional lens. A similar effect could be obtained by using a single camera with a telecentric lens, but the multi-camera approach remains a more compact and cost-effective way of achieving the required resolution. To provide uniform illumination whilst mitigating light-avoidance response, we used a dedicated light system using 850 nm LEDs (Loopbio GmbH, Vienna, Austria). To avoid edge effects, it is important that the light panel is larger than the area to be imaged. Blue light stimulation is provided by a custom LED array (Marine Breeding Systems, St. Gallen, Switzerland) using four Luminus CBT-90 TE light-emitting diodes (bin J101, 456 nm peak wavelength, 10.3W peak radiometric flux each). The camera lenses are equipped with long pass filters (Schneider-Kreuznach IF 092, Schneider-Kreuznach, Germany, and Midopt LP610, Midwest Optical Systems Inc, Palatine, IL, USA) to block the photostimulation light while allowing the bright field IR signal to reach the camera sensors. A schematic of the imaging system is shown in **Error! Reference source not found.**A-C. To further increase throughput, we built five units of the camera arrays which can be operated in parallel (Kastl-High-Res, LoopbBio GmbH, Vienna, Austria).

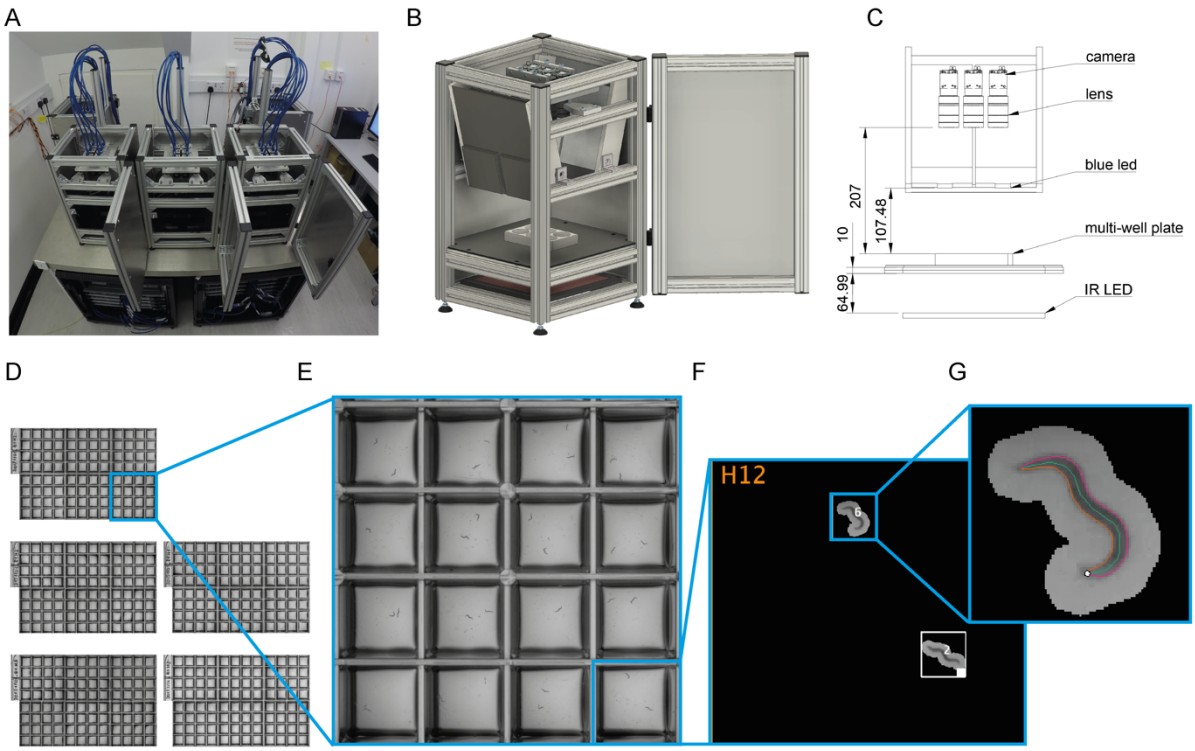


Figure 1: **Schematic of megapixel camera arrays**. **A**) Five identical camera arrays were mounted on an air supported table. The associated workstations to run the Motif software were arranged in the two server racks underneath. **B**) 3D schematic drawing of a single imaging unit (Kastl – Highres). The six cameras were mounted on a plate that is connected to the rig frame by three spring-loaded screws, and can be moved along the vertical axis. This allows for changing the focal plane of all six cameras at once. One of the imaging unit's side panels is omitted from this view. **C**) Technical drawing of an imaging unit annotated with dimensions in millimetres. **D**) Using five identical units, 480 wells can be recorded simultaneously. Zooming in to the **E**) camera, **F**) well, **G**) and worm level shows that this system achieves enough resolution to precisely track the nematodes. Each square well measures 8 x 8 mm.

Choice of suitable multiwell plate design

Because the fields of view of the six cameras partially overlap, the imaging system provides flexibility in selecting a multiwell plate with any number of wells. For our purposes, 96-well plates with square wells provided a good balance between imaging area and number of wells (**Error! Reference source not found.**D, E). Plates with smaller numbers of wells would reduce imaging throughput while 96-well plates with circular wells would reduce the area available for worms to behave and increase shadowing around the well edges (**Supplementary Figure 1**A). Using square well plates (Whatman 96 well plate with flat bottom, GE Healthcare, Chicago, IL, USA) significantly increases the efficiency of the system: in our tests, in plates with circular wells only 21% of the imaging area is available for capturing useful data, while the rest is outside of any well or lost in shadows. For square wells, 43% is available for behaviour. This fraction can be further increased by using custom plates with thinner wall dividers and shallower wells to reduce the shadowing (**Supplementary Figure 1**B), but this comes at an increased cost of manufacture.

The output of the combined system is 30 videos tiling across the five imaged multiwell plates corresponding to 480 simultaneous behaviour assays (**Error! Reference source not found.**D). Expanding the image to the level of a single well (**Error! Reference source not found.**F) and a single animal (**Error! Reference source not found.**G) shows that the resolution is sufficient to estimate the pose and identify the head of single worms, which can reveal detailed trajectory differences between individuals that are the basis for quantitative behavioural phenotyping.

High throughput imaging

Due to the high amount of raw image data produced by USB3 cameras at full bandwidth (6 cameras recording at 25 fps produce approximately 6.5TB/hour of raw footage), live compression on a dedicated system was required. To achieve this, a total of 10 Motif Recording Units (Motif - Video Recording System, LoopbBio GmbH, Vienna, Austria) equipped with Nvidia Quadro P2000 GPUs (Nvidia Corporation, Santa Clara, CA, USA), each recording from 3 cameras, was required. The two recording units with cameras from the same system were set up in a parent-child configuration.

The Motif software acquires and compresses images on the fly and stores them in the open imgstore format (https://github.com/loopbio/imgstore) along with timestamps and frame numbers for each individual frame, as well as continuous and synchronised recordings of environmental data (for each unit this was: outside temperature and inside temperature, humidity, and light level). Recording the time and frame number for each image allows precise timekeeping over a long recording duration as it removes temporal drift due to skipped or dropped frames and due to differences in camera clocks. For good scientific data management practice, additional metadata for each recording is saved with the imgstore. This includes the camera serial numbers, camera and system settings and any user defined data.

A single workstation manages all imaging experiments on all units across the whole system, from video collection to data transfer, by accessing the Motif user interfaces using the web-browser of the parent machines in the camera arrays. Given the large number of high-resolution cameras, the control workstation was connected to a large monitor (we use a 43-inch 4k monitor) to facilitate camera focussing and sample positioning.

In addition to providing a web accessible user interface, Motif allows complete control of the camera arrays and arbitrarily complex scheduling of data acquisition and photostimulation

programmatically, via an API (https://github.com/loopbio/python-motifapi). This allows us to run imaging experiments on all camera arrays by executing a single Python script on the monitoring workstation. Encoding the parameters of experiments in a script improves reproducibility by making parameters consistent over time by default.

Updates to Tierpsy Tracker, and companion software, for multiwell imaging format

In our camera array system, each camera records multiple wells which complicates metadata handling since there is no longer a one-to-one correspondence between a video file and a particular experimental condition. We have updated Tierpsy Tracker[5] to handle videos with multiple wells: it can automatically identify wells from the video (**Error! Reference source not found.**F), and return results on a well-by-well basis. In the Viewer, the user can see the names and boundaries of the wells, and have the option of marking any well as "bad" if necessary. This flag is propagated to the final tracking results so that the contribution of "bad" wells can be filtered out for downstream analysis.

To keep track of the experimental conditions of each well we have developed an open source module in Python to automatically handle experimental metadata (github.com/Tierpsy/tierpsy-tools-python). For each experiment, the user specifies, in a series of csv files, the worms and compounds (if applicable) that were added to each well. This can include information on how a COPAS worm sorter was used to dispense different strains in the wells of the imaging plates, which compound source plate was replicated onto each imaging plate, or any column shuffling performed by a liquid handling robot. These tables are then combined to create a mapping between each well in an imaging plate (identified by a unique ID) and an experimental condition. For each imaging run, the user needs to log the camera array used for each imaging plate at the time of the experiment. This information is then mapped to video file names to create a final metadata table suitable for subsequent analysis (see Material and Methods and *Supplementary Figure* **3** for more details).

Another key software improvement we incorporated is a convolutional neural network (CNN) to exclude non-worm objects from subsequent analysis. While we previously used contrast-based segmentation and size-based filtering for worm detection in our analysis[5], introducing the CNN into Tierpsy Tracker improves the quality of the tracking data and the subsequent analysis results as well as the speed of the analysis because fewer objects are analysed (see Materials and Methods for more details).

Rapid assessment of natural variation in behaviour

We tracked the behaviour of N2 and wild isolates of the divergent set in the CeNDR strain catalogue with our system to detect natural variation in behaviour[6]. To further increase the dimensionality of the behavioral phenotypes, we included a blue light stimulation protocol using a set of four bright blue light LEDs. Each tracking experiment is divided into three parts: 1) a 5-minute pre-stimulus recording, 2) a 6-minute stimulus recording with three 10-second blue light pulses starting at 60, 160, and 260 seconds, and 3) a 5-minute post-stimulus recording. Blue light can elicit an escape response in worms, thus expanding the range of observable behaviours[7,8]. Programmable blue light stimulation is reproducible, compatible with high throughput assays, and is also useful for optogenetic stimulation.

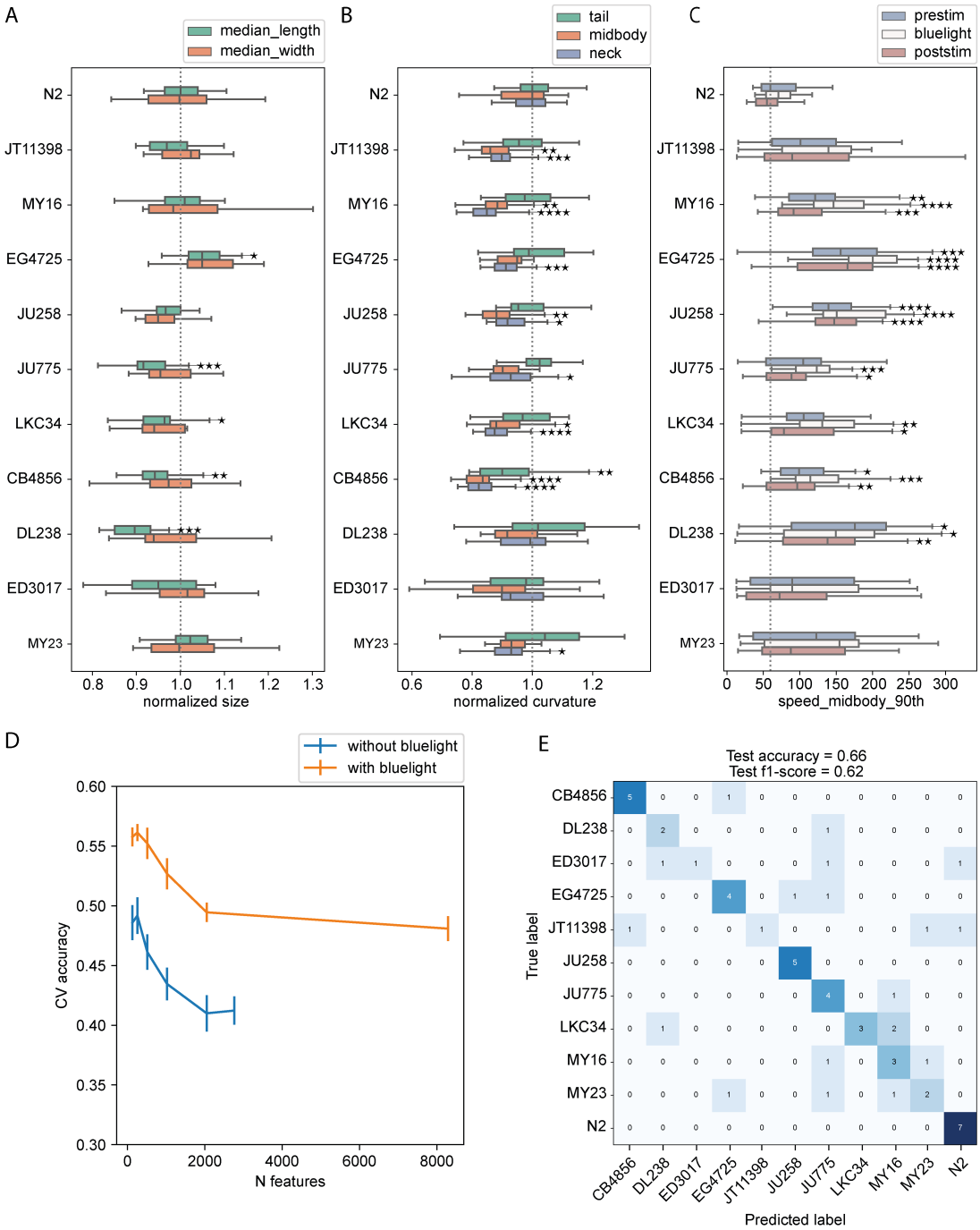


Figure 2: Natural variation in behaviour. A-C) Examples of features describing morphology, movement, and posture. Stars indicate statistically significant differences between N2 and wild isolates at a false discovery rate of 1% using Kruskal-Wallis tests and correcting for multiple comparisons with the Benjamin-Yekutieli method. A) Morphological differences were detected between strains. The length and the midbody width varied in a non-uniform way among strains. B) Adequate resolution enabled detailed characterisation of the worm posture and the detection of differences among strains in multiple dimensions. The curvature at different parts of the body varied in a non-uniform way among strains. The neck curvature showed more significant differences. The parts of the body are defined following the conventions adopted in Tierpsy Tracker[9]. C) The speed of wild isolates was on average higher than the speed of N2 worms. The response of wild isolates to blue light stimulus varied; some strains (e.g. EG4725) were more sensitive to blue light compared to N2, while others showed less obvious escape response (e.g. DL238). This provided additional dimensions to the behavioural phenotype. D) Using the quantitative behavioural phenotypes, strains were classified with significantly higher accuracy than random. Combining features from different blue light conditions increased the dimensionality of the data and the classification accuracy between strains. E) Worm strains were predicted in a held-out test set with 66% accuracy which is higher than random (9%).

We tracked on average 20 wells per strain. Given the high throughput achieved with our new system, this experiment can be performed within a few hours. The recordings of the camera array maintain enough resolution to extract the full set of Tierpsy features[9], which describe in detail the morphology, movement, and posture of the worms, including subdivision by motion mode (forward, backward, and paused) and body part. We extract a set of 3076 summary features per well for each recording period (pre-stimulus, blue light, and post-stimulus), resulting in a total of 9228 features for each well. This allows us to detect fine differences in the morphology, posture and movement of the worms which varies in a non-uniform way among wild isolates (**Error! Reference source not found.**A-C). The neck curvature of wild isolates tends to show more significant differences to N2 worms than the curvature of other parts of the body, which might be related to differences in foraging behaviour between N2 and wild isolates (**Error! Reference source not found.**B). However, not all strains show the same curvature pattern across the body indicating natural variation in posture. All the wild isolates move on average faster than N2 worms but their response to blue light varies (with some being more and others less sensitive to blue light), showing that the blue light stimulus increases the dimensionality of behavioural differences (**Error! Reference source not found.**C).

To assess how well we can predict the worm strain based on its behavioural fingerprint, we estimated the classification accuracy using a random forest classifier. We first split the data into a training/tuning set and a held-out test set. We used the training set to select features using recursive feature elimination (RFE) and tune the hyperparameters of the model. **Error! Reference source not found.**D shows the highest cross-validation accuracy achieved for different sizes of selected feature sets. The accuracy improves when we select features increasing the samples-to-features ratio, as this helps control overfitting and, in parallel, reduces the correlation between features. Combining features from different blue light conditions (blue curve) increases the dimensionality of the data and therefore increases the classification accuracy. Using the best performing features and hyperparameters, we trained a classifier with the entire training/tuning set and used it to make predictions in the test set. The test accuracy we achieved is 66% which is significantly higher that random (9%).

Worms show a rich temporal response to aversive blue light stimulation

Having established that blue light stimulation can be leveraged to improve classification accuracy, we moved to further investigated the response elicited by blue light in N2 and CB4856 at a higher temporal resolution.

We performed imaging with blue light stimulation on both the N2 and CB4856 strains, and observed different behavioural responses between these two strains. We automatically extracted a set of 3076 previously defined features[9] with a time resolution of 10 seconds using Tierpsy Tracker[5], and used these to construct the behaviour phenotype space. N2 and CB4856 have well known behavioural diffrences[10–15] and are expected to occupy different regions of the phenotype space. Principal component analysis (PCA) shows that application of blue light stimulation moves the strains from their already distinct positions in the plane defined by the first two principal components (PCs) to new positions, indicating detectable responses to the stimulus in both strains (**Error! Reference source not found.**A). Blue light-induced displacement through phenotype space led to better separation between the two strains (**Error! Reference source not found.**A, right), confirming that the addition of the stimulus can reveal further behavioural differences between two strains already known to be distinct.

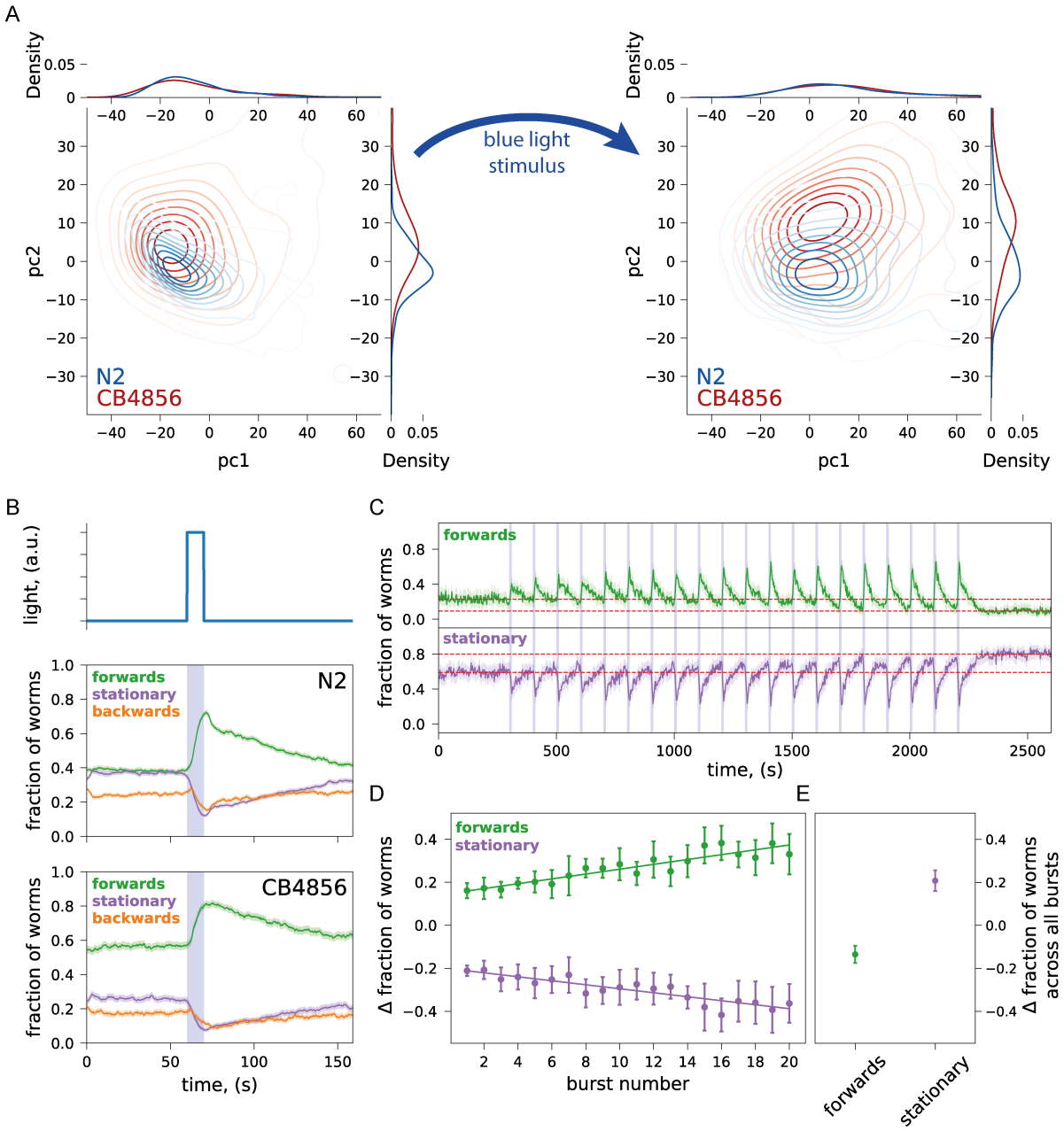


Figure 3: Escape response to photostimulation. **A**) PCA plots of N2 and CB4856 in the 10 seconds immediately before (left) and immediately after (right) a 10 second stimulus showing detectable behavioural responses: both strains moved to new, better separated positions in the phenotype space as a result of stimulation. **B**) Photostimulation with blue light elicited similar escape responses in both N2 and CB4856 strains, with the fraction of worms moving forwards increasing during the stimulus and decreasing after the stimulus. However, post-stimulus recovery appears to occur at two timescales for N2 but not for CB4856. **C**) Repeated photostimulation triggered increasing aversive response in N2, also leaving a higher fraction of worms stationary after serial stimulation than before (vertical separation between the two dashed lines to contrast the before and the after levels). **D**) The fraction of worms triggered to move forwards by each stimulus increased throughout the stimulation series, with the opposite trend true for worms initiating movement after pausing. Each data point was obtained by taking the difference in a 10 second window just before and just after the end of each stimulus. **E**) After repeated photostimulation, a larger fraction of the population than before was stationary. This was quantified by taking the difference of the population fractions in each motion mode between the final 5 minutes and the initial 5 minutes of the experiment (red dashed lines in **C**).

The *C. elegans* escape response is characterised by a combination of increased forward locomotion and decreased spontaneous reversals[16]. The differences in blue light-induced escape response between the N2 and CB4856 strains can thus also be seen by simply examining the fraction of the worm population moving forwards, moving backwards, or remaining stationary, without needing to use a large number of features. For both strains, a single dose of photostimulation triggers a sharp and steady increase in the fraction of worms moving forwards, followed by a relaxation towards the pre-stimulus level once the stimulus ceases. The fraction of worms moving backwards has a very slight increase at the beginning of the stimulus, and then decreases without increasing again until the stimulus is over. Finally, the fraction of stationary worms declines rapidly during the stimulus and is restored after the stimulus ends (**Error! Reference source not found.**B). However, while in CB4856 the rate at which the population fractions return to the pre-stimulus levels is steady, N2 shows a sharp initial decline in forward-moving worms (and a corresponding sharp increase in stationary worms and backward-moving worms) before assuming steady relaxation. Repeated photostimulation (twenty pulses of 10 s on, 90 s off) of N2 worms causes sensitisation, as light pulses trigger a progressively increasing fraction of worms to move forwards (**Error! Reference source not found.**C). Meanwhile, between light pulses worms recover to a progressively decreasing baseline level of forward locomotion (and conversely, progressively increasing stationary fraction), possibly due to worms becoming tired from increased activities during the pulse. This reduced forward locomotion fraction persists in the absence of photostimulation, with no obvious return towards pre-stimulus levels over a 6.5 minute period after the final pulse (**Error! Reference source not found.**C, E). The combined effects of sensitisation and tiring leads to a roughly linearly increasing response over multiple light pulses, as illustrated by taking the difference between the fraction of worms moving forward before and after stimulation (**Error! Reference source not found.**D).

Photostimulation can thus better distinguish between worm strains using existing pre-defined feature sets, as well as create new features for quantifying the rich details of the escape behaviour. The addition of these new and interpretable features increases the dimensionality of the worm behavioural phenotypic space, which can be useful for many phenotyping applications.

Blue light stimulation reveals differential behavioural responses in ALS disease models

A previous study generated several Amyotrophic Lateral Sclerosis (ALS) disease model strains that carry patient amino acid changes in the *C. elegans sod-1* gene[17]. This study found that the disease model strains have no obvious behavioural defects unless they are exposed to oxidative stress by overnight treatment with paraquat.

We phenotyped these ALS disease model strains on our system and saw similar results. PCA of a pre-defined set of 256 Tierpsy features[9] under standard imaging conditions (five minutes of spontaneous behaviour) does not show clear differences between the strains (**Error! Reference source not found.**A). Adding blue light pulses (three 10-second blue light pulses over six minutes) leads to better separation between the strains in PC space (**Error! Reference source not found.**B). Although the SOD-1(+) wild-type control strain (blue) and the SOD-1(A4V) mutant disease strain (orange) clearly separate into their own clusters, SOD-1(H71Y), SOD-1(G85R) and SOD-1(0) null strains cluster together, suggesting that their overall responses to blue light are similar to each other. The clustering of SOD-1(H71Y), SOD-1(G85R), and SOD-1(0) strains upon blue light stimulation is consistent with the previous finding that all three strains have loss of *sod-1* function in glutamatergic neurons. By contrast, the SOD-1(A4V) strain has overexpression of *sod-1* in cholinergic

neurons without affecting glutamatergic neurons[17], and this disease strain forms its own cluster in the blue light PC space (**Error! Reference source not found.**B).

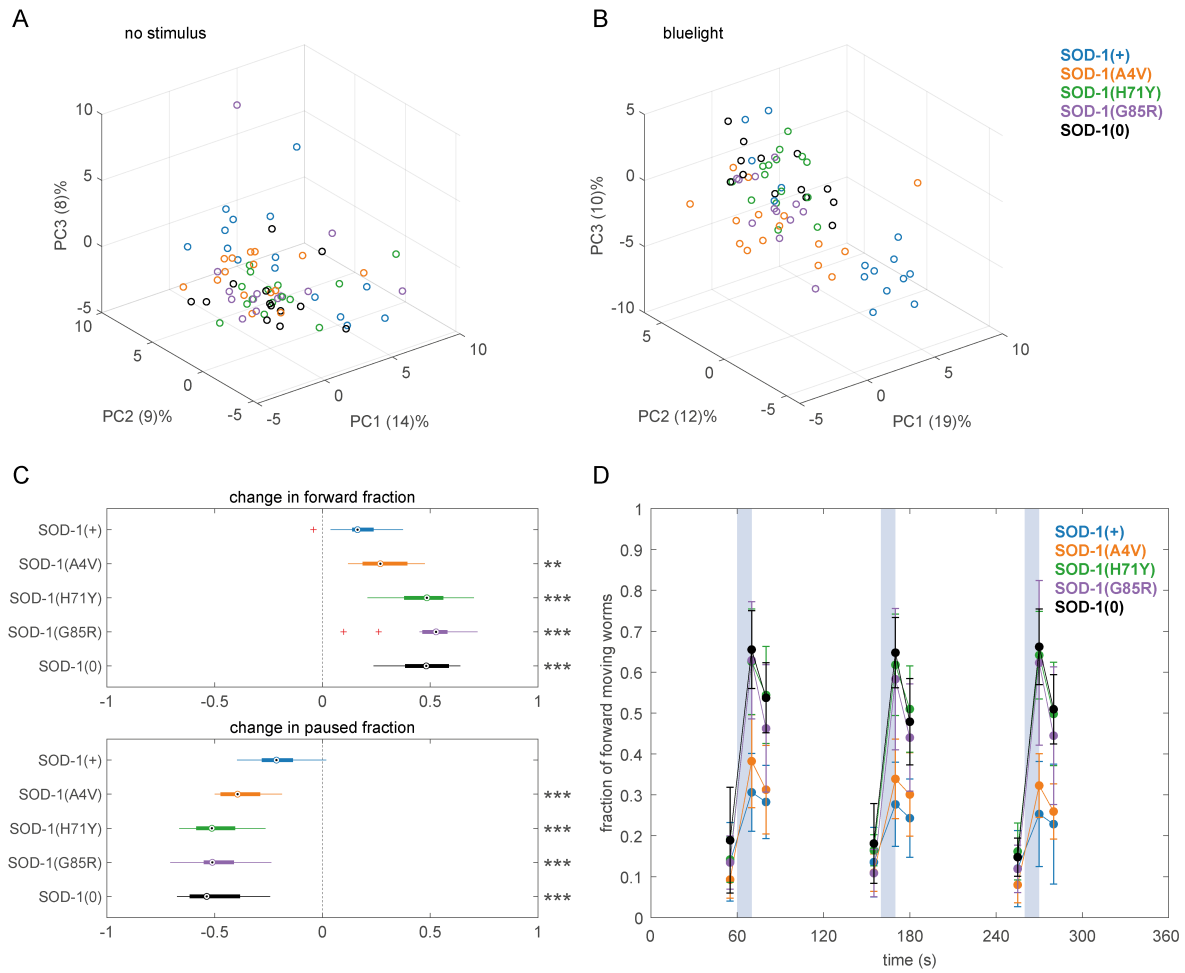


Figure 4: Blue light stimulation elicits different responses amongst ALS disease models. **A-B**) Principal component analysis of 256 extracted behavioural features from (**A**) standard or (**B**) blue light imaging conditions. Features were extracted by Tierpsy Tracker. Each datapoint represents one plate average of the strain, with up to 12 independent wells for each strain in every 96-well plate. Each well contained an average of three worms. The time window represented in (**B**) is also shown in (**D**). **C**) Changes in the overall fraction of forward (top) or paused (bottom) locomotion upon blue light stimulation. The difference was calculated by subtracting the average feature values over the t = 50-60 second pre-stimulus window from those over the t = 65-75 second first blue light pulse window (these correspond to the first and the second time points in (**D**), respectively). Plate averages were used to generate the plot for each strain. Two sample t-test against SOD-1(+) control strain: ** $p<0.01$; *** $p<0.001$. **D**) Overall fraction of forward locomotion under blue-light imaging conditions. Three 10 second blue light pulses (blue shading) started at t = 60, 160, 260 seconds, and feature values were calculated using 10 second windows centred around 5 seconds before, 10 seconds after, and 20 seconds after the beginning of each blue light pulse. Plate averages were used to generate the plot for each strain.

Behaviourally, the previous study reported an increased escape response from noxious paraquat in the three *sod-1* loss-of-function strains caused by increased neuromuscular junction function[17]. Our results using blue light as a noxious stimulus reproduce the paraquat-induced differential escape responses. Upon blue light stimulation, SOD-1(H71Y), SOD-1(G85R), and SOD-1(0) strains show significantly bigger increases in forward locomotion compared to the SOD-1(+) control strain and the SOD-1(A4V) disease strain (**Error! Reference source not found.**C, top). This increase in forward movement appears to be

primarily at the expense of stationary (**Error! Reference source not found.**C, bottom) rather than backwards locomotion (**Supplementary Figure 2**A). Nevertheless, a closer look at reversal frequencies at a finer temporal resolution reveals decreased reversals in the three *sod-1* loss-of-function strains but not the other two strains (**Supplementary Figure 2**B). Finally, the level of blue light-induced escape response is sustained across three pulses of light stimulation for all five strains with no obvious habituation effect (**Error! Reference source not found.**D and **Supplementary Figure 2**B).

Discussion

We have developed a megapixel camera array system to enable high throughput, high content imaging of worms in standard multiwell plates. By partially overlapping the fields of view of six cameras, we can image an entire 96-well plate at spatial and temporal resolutions that are sufficient for tracking *C. elegans* and extracting high-dimensional phenotypic fingerprints. We have added features to Tierpsy Tracker to make it compatible with the multiwell imaging format, so that each well is detected and analysed separately. We incorporated strong blue LED lights into the camera array system to provide precise and repeatable photostimulation and found that this leads to better separation between wild isolates and ALS disease model strains, in the latter case revealing phenotypes that could not be detected in standard unstimulated assays. Repeated blue light stimulation also revealed a novel sensitisation phenotype in N2 worms, in marked contrast to our initial expectation of habituation, as reported in previous experiments on repeated mechanical stimuli which are used to study learning in worms[18,19].

Our imaging hardware and analysis software are designed to support high throughput phenotypic screening, as the multiwell format allows for a large number of experiments to be conducted simultaneously. Furthermore, our experimental pipeline uses liquid handling robots for dispensing agar, food, drugs, and worms, in order to streamline the workflow for large-scale phenotypic screening. On a typical eight-hour imaging day, a single experimenter can operate five runs on all five camera array units, thus collecting imaging data from 2400 independent wells in a 96-well plate format. Typical post-acquisition processing time for this volume of data (assuming the standard 16 minute video length at 25 fps, three worms per well) is 50-85 hours using a MacPro (Processor: 2.7 GHz 12-Core Intel Xeon E5; Memory: 64GB 1866 MHz DDR3) to go from raw video data to fully extracted behavioural features. Processing time increases significantly with object number and depends on the quality of the video (good contrast, lack of debris, etc.).

A main strength of our camera array system is its scalability. Screening throughput can be readily expanded with additional imaging units, as the system is modular and each camera array has a relatively small physical footprint. Motif software enables on-the-fly compression of raw videos during acquisition, thereby keeping data volume manageable. Post-acquisition analysis is easily parallelised since videos can be analysed independently and processing time can be decreased linearly by allocating more computational cores to the task (e.g. by using a high-performance cluster).

The megapixel camera arrays we describe here represent a natural progression in worm tracking hardware where advances in the past have come from multiplexing to increase throughput[20] and increasing resolution to get more information from multi-worm trackers[18]. Our new system has a throughput that is approaching what has been achieved for screens with simpler endpoints such as survival or gross motility, but with a resolution that enables the full suite of computational ethology tools to be brought to bear on phenotyping. We

anticipate this will open new directions in large scale behaviour quantification with applications in genetics, disease modelling, and drug screening.

Materials and Methods

Worm strains

C. elegans strains used in this work are listed in Supplementary Table 1. Worms are cultured on Nematode Growth Medium (NGM) agar at 20 ºC and fed with *E. coli* OP50 following standard procedures[18].

Standard phenotyping assay

The standard phenotyping assay was used for most experiments in this work unless otherwise noted (detailed protocol: https://dx.doi.org/10.17504/protocols.io.bsicncaw). See Supplementary Table 2 for the detailed protocols used to collect the data shown in each figure panel.

Briefly, Day 1 adult worms were obtained by bleach-synchronisation (detailed protocol: https://dx.doi.org/10.17504/protocols.io.2bzgap6) and used for all imaging experiments. Imaging plates were prepared by filling 96 well plates with 200 µL of low peptone (0.013% Difco Bacto) NGM agar per well using an Integra VIAFILL reagent dispenser (INTEGRA Biosciences Ltd, UK) (detailed protocol: http://dx.doi.org/10.17504/protocols.io.bmxbk7in), and stored at 4 ºC until use. On the day before imaging, plates were placed in a LEEC BC2 drying cabinet (LEEC Ltd, Nottingham, UK) to lose 3-5% weight (2-3 hours). Each plate was then seeded with 5 µL per well of 1:10 diluted OP50 using VIAFILL, and stored at room temperature overnight.

On imaging day, synchronised Day 1 adults were washed in M9 (detailed protocol: https://dx.doi.org/10.17504/protocols.io.bfqbjmsn) and dispensed into imaging plate wells using COPAS 500 Flow Pilot worm sorter (detailed protocol: https://dx.doi.org/10.17504/protocols.io.bfc9jiz6). Three worms were placed into each well unless noted otherwise. Plates were returned to a 20 ºC incubator for 1 hour to dry following liquid handling, and then placed onto the multi-camera tracker for 0.5 hour to acclimatise prior to image acquisition.

Drug experiments

Drug experiments followed standard phenotyping assay workflow, but with a few modifications. Detailed protocol of the below can be found at http://dx.doi.org/10.17504/protocols.io.bs6znhf6.

Briefly, imaging plates were prepared with drugs the day before imaging and stored in the dark overnight at 4 ºC. Using COPAS 500 Flow Pilot, three worms were dispensed into each well (two worms per well for the first round of imaging, featured in **Error! Reference source not found.**A, B). Following liquid handling, plates were kept in a 20 ºC incubator for an extra three hours to allow drug exposure (total drug exposure time was thus four hours).

Image acquisition

All videos were acquired at 25 fps on the trackers in a temperature-controlled room at 20 ºC, with a shutter time of 25 ms, and 12.4 µm px^{-1} resolution. For all experiments unless

otherwise noted, three sequential videos were taken, run in series by a script: a 5-minute pre-stimulus video, a 6-minute blue light recording with 10-second 100% intensity blue light pulses at the 60, 160, and 260 seconds mark, and a 5-minute post-stimulus recording. The timing of recordings and photostimulation was controlled using Loopbio's API for Motif software [https://github.com/loopbio/python-motifapi] in a script.

For the serial blue light stimulation experiments, the plates were continuously imaged for 43 minutes and 20 seconds in the following stimulation pattern: 5 minutes off, 20 x (10 s on, 90 s off), 5 minutes off.

Image processing and quality control

Segmentation, tracking, and pose estimation over time was performed using Tierpsy Tracker. Each video was checked using Tierpsy Tracker's Viewer, and wells with visible contamination, agar damage, or excess liquid (from worm sorter, so that worms swim rather than crawl) were marked as bad and excluded from the analysis.

Convolutional neural network to exclude non-worm objects

We improved Tierpsy tracking by incorporating a CNN classifier after segmentation to exclude non-worm objects from being analysed and skewing the results.

In the video compression step at the beginning of the Tierpsy analysis pipeline, a segmentation algorithm detects putative worm objects according to a set of user-defined parameters. The pixels in the frame that are further away than a threshold from any of the putative worms are set to 0, creating a "Masked Video". The objects selected by the masking algorithm are tracked throughout the video, but now if only they pass the filtering step powered by a CNN classifier.

The classifier was trained on a dataset of 43,561 grey-scale "masked" images measuring 80x80 pixels each, collected across several imaging systems in our lab. All images were manually annotated and objects were marked as either "worm" or "non-worm" by two independent researchers, so a consensus could be sought. The annotated dataset was split into training, validation, and test sets containing 80%, 10%, and 10% of the images, respectively, while keeping the classes balanced in each set. All images were pre-processed in two steps. First, the background pixels set to 0 by the masking algorithm were shifted to the top 95 percentile of the grey values in the unmasked area. This prevents the artificial edge between the masked and non-masked area from disproportionately influencing the classifier. Second, all pixel values were scaled to the range of 0 to 1 by min-max normalisation, to reduce the influence of variable illumination and contrast in different imaging setups.

The architecture of the CNN is a shallower adaptation of VGG16[21], featuring eight convolution layers with 3x3 filter size and stride 1, each followed by a rectified linear activation unit, four max-pool layers (filter size 2x2, stride 2) applied every two convolution layers, and a fully connected layer. Batch normalisation is applied to the third and seventh convolution layer to accelerate training by reducing internal covariate shift[22], and a Dropout layer is added before the fully connected layer to prevent overfitting[23]. In total, the CNN has about 1.78 million trainable parameters.

The CNN classifier was implemented in PyTorch 1.6, and was trained with the cross entropy loss function and the Adam optimisation algorithm[24] at a learning rate of 10^{-4}. It achieved an accuracy of 97.68% and F1 score of 97.98% as measured on the independent test set.

To improve performance at the inference step, we apply the CNN to a subset (one image per second) of all the images featuring the same putative worm object. This yields, per snapshot, the probability of the object to be a valid worm. If the median of this probability over time is higher than 0.5, the object is classified as a valid worm.

Video processing with multiple wells

Using multiwell plates for imaging significantly increased the experimental throughput, but also introduced challenges for data analysis as each video output contains 16 separate wells. Further software engineering was thus warranted to process multiwell videos, so that wells are detected and analysed separately.

To achieve this, we implemented an algorithm in Tierpsy Tracker that automatically detects multiple wells in a field of view and stores the coordinates of well boundaries. Briefly, we created a template that approximates the appearances of a well in the video, and replicated it on a lattice to simulate the grid of wells. The overall dimensions of the lattice are defined in Tierpsy's configuration file, but the lattice spacing parameters were chosen, via SciPy's differential evolution routine[25], to minimise the differences between the video's first frame (or its static background, if Tierpsy was instructed to calculate it) and the simulated grid of wells.

Automatic extraction of behavioural features was then performed on a per-worm basis, before worms were sorted into their respective wells based on their (x, y) coordinates in order to obtain well-averaged behavioural features.

Data provenance

Tracking multiwell plates complicates the handling of metadata, since there isn't a unique mapping between videos and experimental conditions. When well shuffling is performed using the liquid handling robot, the well contents in the imaging plate also needs to be tracked. To handle experimental metadata for imaging with the camera arrays, the records that need to be compiled manually during the experiments was standardised and an open source module in Python (https://github.com/Tierpsy/tierpsy-tools-python/hydra) was developed to combine the experimental records to create a full metadata table with the experimental conditions for each well (*Supplementary Figure* **3**).

The experimental records are typically compiled in the form of csv files. In each tracking day, the experimenter needs to record: i) information about the media type and the bacterial food present on the imaging plates, and the worm strains that were dispensed into the wells of the plates (this is recorded in a summarized way in the *wormsorter.csv* file), ii) information about the experimental runs, including the unique IDs of the imaging plates, the instrument name where each plate was imaged, and the environmental conditions (*manual_metadata.csv*), iii) if applicable, information about the contents of the compound source plates (*sourceplate.csv*) and the mapping between imaging plates and source plates (if the liquid handling robot was used for column shuffling, this mapping will be recorded automatically in the *robotlog.csv*; if there was no shuffling, this will be recorded in *imaging2source.csv*).

Using the functions in the *hydra* module, firstly a plate metadata table is created to contain all the well-specific experimental conditions for every well of each unique imaging plate,

including the compound contents if applicable. Then, the information about the experimental runs is merged with the plate metadata to create a final metadata table with the complete experimental conditions for every recording of every well. At this stage, the video filenames are also matched to the sample based on the camera array instrument ID.

Motion mode analysis of timeseries

Tierpsy Tracker[5] was used to detect the motion mode (forwards, backwards, stationary) of each worm over time. To calculate the fraction of worms in each motion mode over time (**Error! Reference source not found.**B), the number of worms in each motion mode for each time point in each well was counted and then divided by the total number of tracked worms at each time point in each well. This gave the fraction of worms in each motion mode, at each time point, for each well, so that an average could be taken across all wells. The 95% confidence interval for the average was obtained by nonparametric bootstrap (n = 1000 resamplings, with replacement).

For the longer experiments in **Error! Reference source not found.**C-E, the motion mode detected by Tierpsy Tracker for each worm over time was first down-sampled to 0.5 Hz by dividing the video into non-overlapping two seconds intervals and taking the prevalent motion mode in each interval. The fraction of the worm population in each motion mode over time was calculated by counting the number of worms in each motion mode and then dividing by the total number of worms detected at each time point. The 95% confidence interval was calculated via nonparametric bootstrap by the seaborn Python library.

Classification of wild isolates

For the classification of the divergent set we used a random forest classifier as implemented in scikit-learn[26]. For feature selection we used recursive feature elimination with a random forest estimator (RFE), as implemented in scikit-learn[26]. We started by splitting the data randomly in a training/tuning set and a test set, with 20% of the data from each strain assigned to the test set. We used the training/tuning set for feature selection. We tried specific candidate feature set sizes $\{2^i, \text{for } i=7:11\}$. For each size, we performed cross-validation and: i) used each training fold to select N features and train a classifier with the selected features; ii) used each test fold to estimate the classification accuracy. We repeated the process 20 times to get statistical estimates of the mean cross-validation accuracy for each size and selected the best performing size N_{best}. We then selected N_{best} features using the entire training/tuning set and used this set for downstream analysis. At a second stage, we tuned the hyperparameters of the random forest classifier using grid search with cross-validation as implemented in scikit-learn[26] with the grid shown in Table 1. The best performing parameters are reported in Table 1. Finally, we trained a classifier on the entire training set using the selected features and hyperparameters and used it to make predictions on the test set.

Table 1: Tested parameter grid for random forest classifier

Parameter	Values	Selected
n_estimators	200:200:2000	1000
max_features	‘auto’, ‘sqrt’	‘sqrt’
max_depth	10:10:110, None	50
min_samples_split	2, 5, 10	2
min_samples_leaf	1, 2, 4	2

Acknowledgements

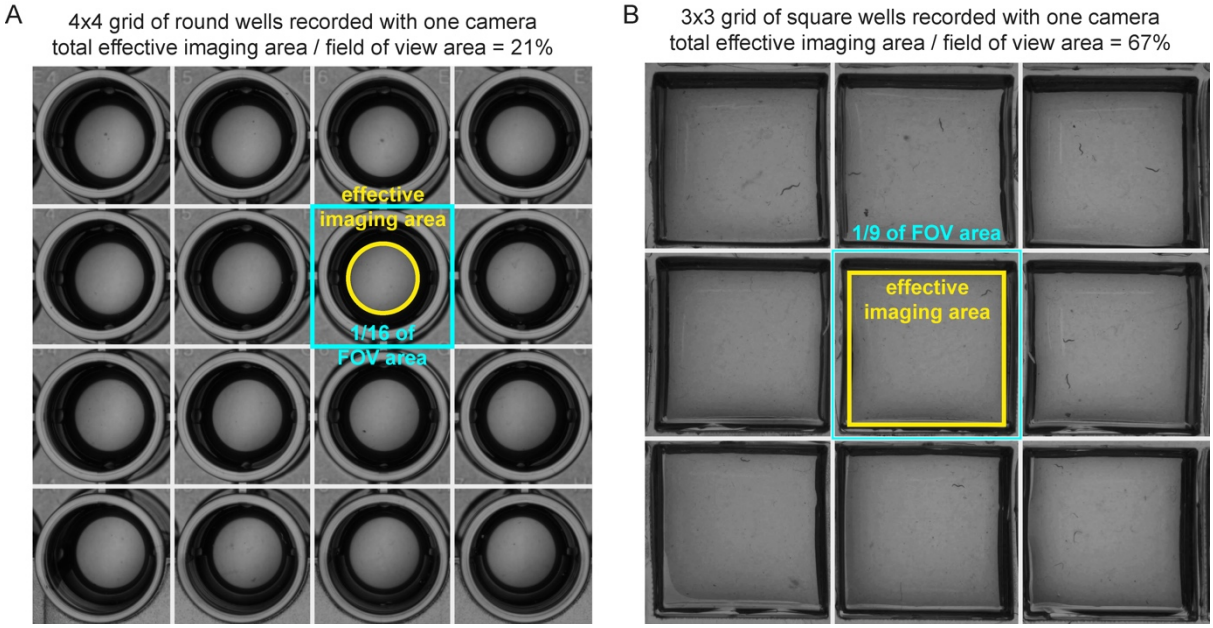
This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 714853) and was supported by the Medical Research Council through grant MC-A658-5TY30. AMR was supported by a BBSRC CASE studentship part-funded by Syngenta.

References

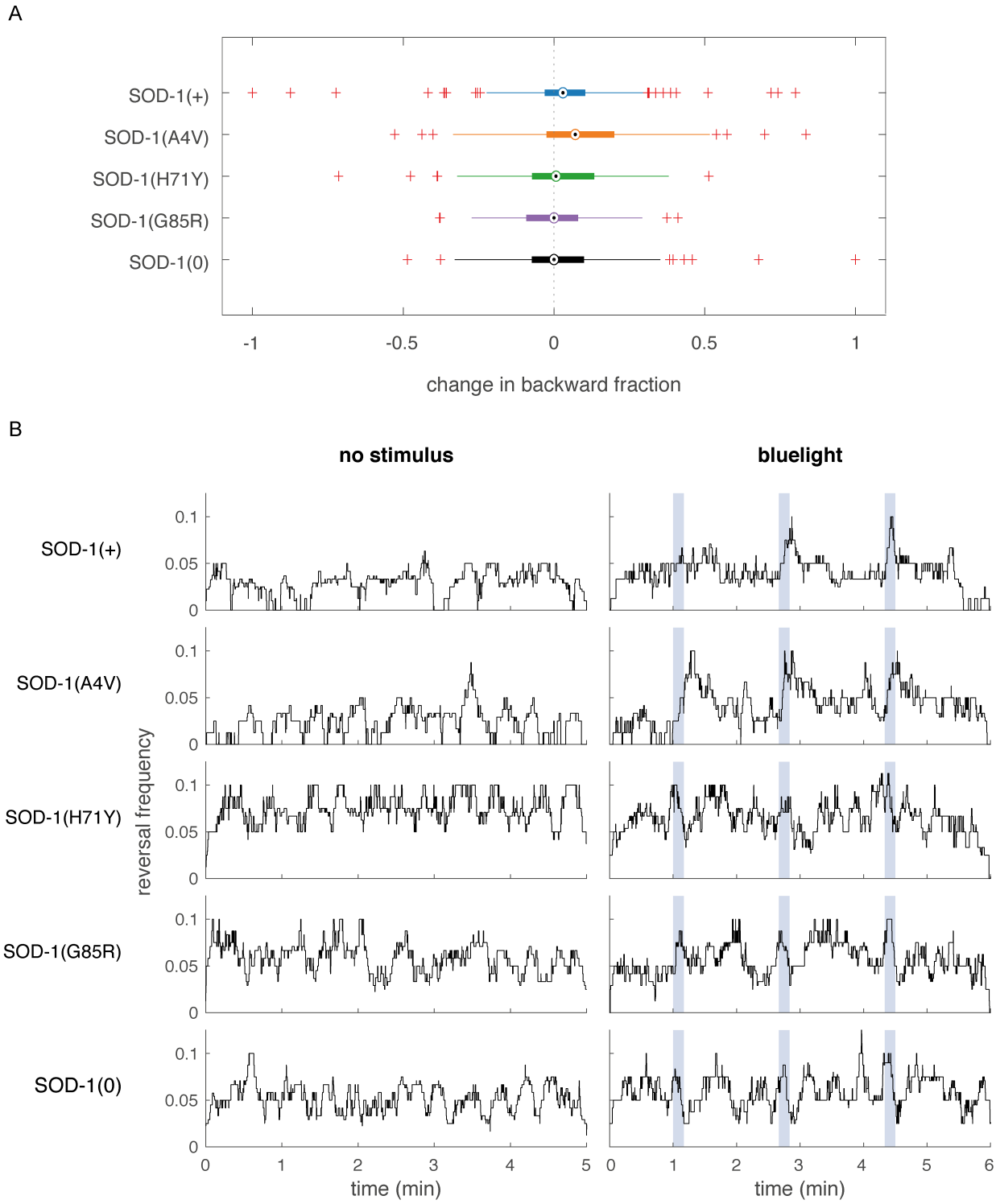
1. Gomez-Marin, A., Paton, J. J., Kampff, A. R., Costa, R. M. & Mainen, Z. F. Big behavioral data: psychology, ethology and the foundations of neuroscience. *Nat. Neurosci.* **17**, 1455–1462 (2014).
2. Anderson, D. J. & Perona, P. Toward a Science of Computational Ethology. *Neuron* **84**, 18–31 (2014).
3. Berman, G. J. Measuring behavior across scales. *BMC Biol.* **16**, 23 (2018).
4. Brown, A. E. X. & de Bivort, B. Ethology as a physical science. *Nat. Phys.* (2018) doi:10.1038/s41567-018-0093-0.
5. Javer, A. *et al.* An open-source platform for analyzing and sharing worm-behavior data. *Nat. Methods* **15**, 645–646 (2018).
6. Cook, D. E., Zdraljevic, S., Roberts, J. P. & Andersen, E. C. CeNDR, the *Caenorhabditis elegans* natural diversity resource. *Nucleic Acids Res.* gkw893 (2016) doi:10.1093/nar/gkw893.
7. Edwards, S. L. *et al.* A Novel Molecular Solution for Ultraviolet Light Detection in Caenorhabditis elegans. *PLoS Biol.* **6**, e198 (2008).
8. Churgin, M. A. *et al.* Longitudinal imaging of Caenorhabditis elegans in a microfabricated device reveals variation in behavioral decline during aging. *eLife* **6**, (2017).
9. Javer, A., Ripoll-Sánchez, L. & Brown, A. E. X. Powerful and interpretable behavioural features for quantitative phenotyping of Caenorhabditis elegans. *Philos. Trans. R. Soc. B Biol. Sci.* **373**, 20170375 (2018).
10. de Bono, M. & Bargmann, C. I. Natural variation in a neuropeptide Y receptor homolog modifies social behavior and food response in C. elegans. *Cell* **94**, 679–689 (1998).
11. Andersen, E. C., Bloom, J. S., Gerke, J. P. & Kruglyak, L. A Variant in the Neuropeptide Receptor npr-1 is a Major Determinant of Caenorhabditis elegans Growth and Physiology. *PLoS Genet.* **10**, e1004156 (2014).
12. Gomez-Marin, A., Stephens, G. J. & Brown, A. E. X. Hierarchical compression of *Caenorhabditis elegans* locomotion reveals phenotypic differences in the organization of behaviour. *J. R. Soc. Interface* **13**, 20160466 (2016).
13. Schwarz, R. F., Branicky, R., Grundy, L. J., Schafer, W. R. & Brown, A. E. X. Changes in Postural Syntax Characterize Sensory Modulation and Natural Variation of C. elegans Locomotion. *PLOS Comput. Biol.* **11**, e1004322 (2015).
14. Thompson, O. A. *et al.* Remarkably Divergent Regions Punctuate the Genome Assembly of the Caenorhabditis elegans Hawaiian Strain CB4856. *Genetics* (2015) doi:10.1534/genetics.115.175950.
15. Vergara, I. A. *et al.* Genome-wide variations in a natural isolate of the nematode Caenorhabditis elegans. *BMC Genomics* **15**, 1–25 (2014).
16. Schild, L. C. & Glauser, D. A. Dynamic switching between escape and avoidance regimes reduces Caenorhabditis elegans exposure to noxious heat. *Nat. Commun.* **4**, 2198 (2013).

17. Baskoylu, S. N. *et al.* Single copy/knock-in models of ALS SOD1 in C. elegans suggest loss and gain of function have different contributions to cholinergic and glutamatergic neurodegeneration. *PLOS Genet.* **14**, e1007682 (2018).
18. Swierczek, N. A., Giles, A. C., Rankin, C. H. & Kerr, R. A. High-throughput behavioral analysis in C. elegans. *Nat. Methods* **8**, 592–598 (2011).
19. Ardiel, E. L., Yu, A. J., Giles, A. C. & Rankin, C. H. Habituation as an adaptive shift in response strategy mediated by neuropeptides. *Npj Sci. Learn.* **2**, (2017).
20. Yemini, E., Jucikas, T., Grundy, L. J., Brown, A. E. X. & Schafer, W. R. A database of Caenorhabditis elegans behavioral phenotypes. *Nat. Methods* **10**, 877–879 (2013).
21. Simonyan, K. & Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. in *International Conference on Learning Representations* (2015).
22. Ioffe, S. & Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. in *International conference on machine learning* 448–456 (PMLR, 2015).
23. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
24. Kingman, D. & Ba, J. Adam: A Method for Stochastic Optimization. Conference paper. in *3rd International Conference for Learning Representations* (2015).
25. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
26. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

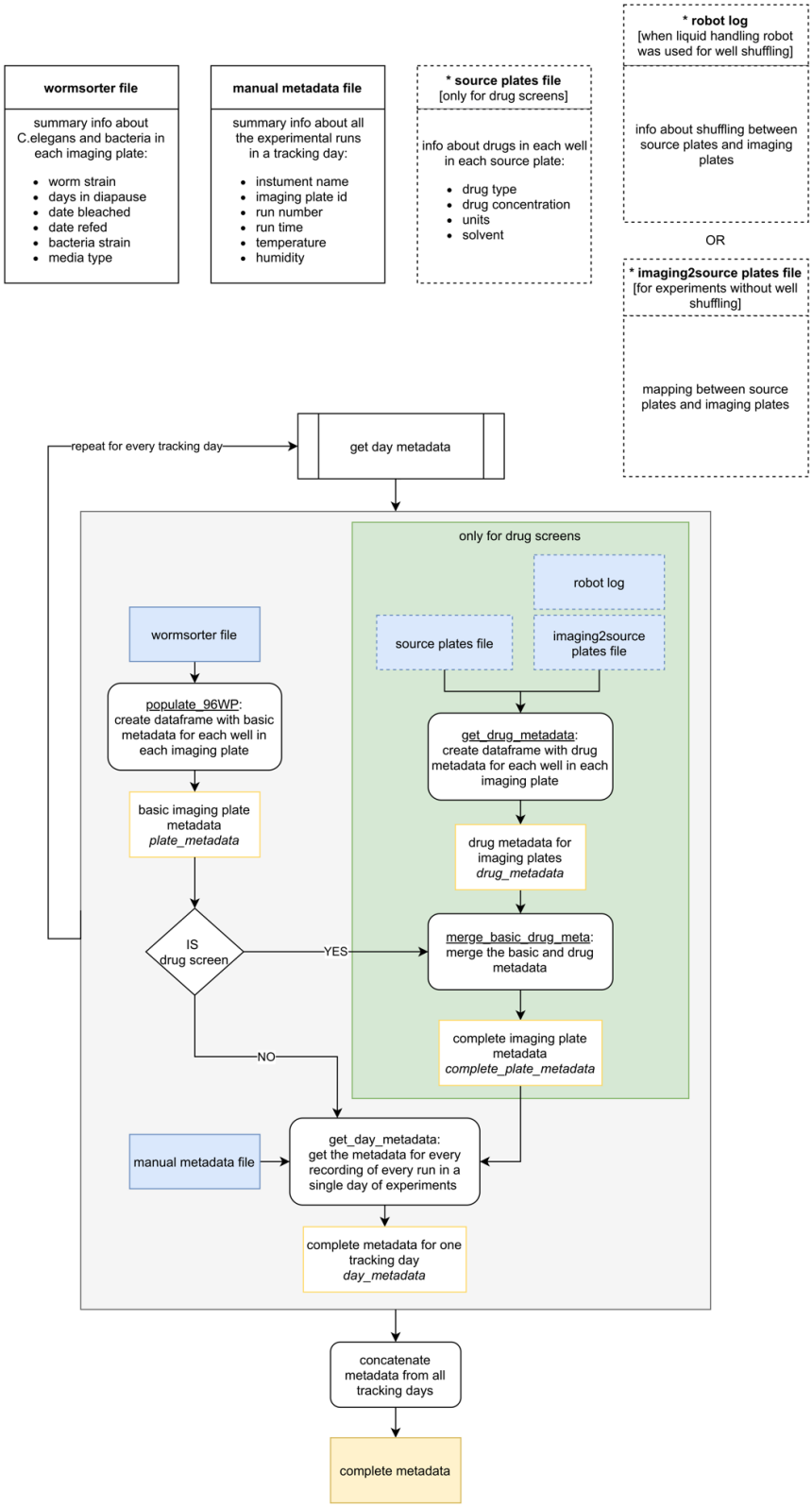
Supplementary figures



***Supplementary Figure 1: Alternative plate designs investigated**. **A**) Round wells are less optimal than square wells as a large part of the field of view is lost to the area between wells, thus reducing the effective area for behaviour. Round wells further lose more effective area due to the shadows cast by the separators. **B**) A prototype custom made 54 multiwell plate. Custom plates with shallow wells and thin separators are the best option to maximise effective imaging area, but this comes with increased manufacturing costs and undermines interoperability with established automated liquid handling instruments.*



Supplementary Figure 2: Backward locomotion in ALS disease models. A) *Changes in the overall fraction of backward locomotion upon blue light stimulation. The difference was calculated by subtracting the average feature values over the t = 50-60 second pre-stimulus window from those over the t = 65-75 second first blue light pulse window (these correspond to the first and the second time points in Fig.3D, respectively. Two sample t-test against SOD-1(+) control strain: ns not significant; * p<0.05.* ***B)*** *Reversal frequency per worm per second, calculated for each frame at 25 fps. A reversal was detected when a worm changes its motion state from forward or paused to backwards. A 10 second sliding window was used to calculate reversal frequency, then the data was smoothed over a 3 second window. A random sample of 20 independent experiments was used to generate the left and the right column plots for each strain.*



*Supplementary Figure **3**: Flow chart of the full metadata compilation based on standardised experimental records when imaging with the Kastl - HighRes. As mentioned in the methods section "Data provenance", experiments with multiwell plates pose challenges in terms of keeping track of data provenance. Our pipeline to handle the automatic creation of metadata can flexibly accommodate different experimental needs (objects marked with * and surrounded by dashed boxes are optional and only present in some experimental designs).*

Supplementary tables

Supplementary Table 1: List of strains in this study

Strain name	Alias	Genotype	Source	Study
N2		Bristol wild-type reference strain	*Caenorhabditis elegans* Natural Diversity Resource	Natural Variation, Sensitisation
CB4856		Wild isolate	*Caenorhabditis elegans* Natural Diversity Resource	Natural Variation
CX11314		Wild isolate	*Caenorhabditis elegans* Natural Diversity Resource	Natural Variation
DL238		Wild isolate	*Caenorhabditis elegans* Natural Diversity Resource	Natural Variation
ED3017		Wild isolate	*Caenorhabditis elegans* Natural Diversity Resource	Natural Variation
EG4725		Wild isolate	*Caenorhabditis elegans* Natural Diversity Resource	Natural Variation
JT11398		Wild isolate	*Caenorhabditis elegans* Natural Diversity Resource	Natural Variation
JU258		Wild isolate	*Caenorhabditis elegans* Natural Diversity Resource	Natural Variation
JU775		Wild isolate	*Caenorhabditis elegans* Natural Diversity Resource	Natural Variation
LKC34		Wild isolate	*Caenorhabditis elegans* Natural Diversity Resource	Natural Variation
MY16		Wild isolate	*Caenorhabditis elegans* Natural Diversity Resource	Natural Variation
MY23		Wild isolate	*Caenorhabditis elegans* Natural Diversity Resource	Natural Variation
HA2427	SOD-1(+)	sod-1(tm776) II; unc-119(ed3) III; rtSi001[sod-1p::sod-1(WT); Cbr-unc-119(+)] IV	Andersen Lab (Northwestern University)	ALS
HA2464	SOD-1(A4V)	sod-1(tm776) II; unc-119(+) III; rtSi008 [sod-1p::sod-1A4V^{M} ::sod-1 3'UTR + Cbr-unc-119(+)] IV	Andersen Lab (Northwestern University)	ALS
HA2425	SOD-1(H71Y)	sod-1(tm776) II; unc-119(+) III; rtSi007 [sod-1p::sod-1H71Y^{M} ::sod-1 3'UTR + Cbr-unc-119(+)] IV	Andersen Lab (Northwestern University)	ALS
HA2426	SOD-1(G85R)	sod-1(tm776) II; unc-119(+) III; rtSi006 [sod-1p::sod-1G85R^{M} ::sod-1 3'UTR + Cbr-unc-119(+)] IV	Andersen Lab (Northwestern University)	ALS
HA2622	SOD-1(0)	sod-1(tm776) II; unc-119(+) III; rtSi026 [Cbr-unc-119(+)] IV	Andersen Lab (Northwestern University)	ALS

Supplementary Table 2: detailed protocols

Figure panel	Protocol	Number of worms per well	Acclimatisation time (hours)
Error! Reference source not found.	https://dx.doi.org/10.17504/protocols.io.bn5zmg76	3	4.5
Error! Reference source not found.A-B	https://dx.doi.org/10.17504/protocols.io.9vqh65w	2	4.5
Error! Reference source not found.C - E	https://dx.doi.org/10.17504/protocols.io.bsicncaw	3	1.5
Error! Reference source not found., **Supplementary Figure 2**	https://dx.doi.org/10.17504/protocols.io.bsicncaw	3	1.5

Note: Acclimatisation time is defined here as time of imaging – (start time on COPAS worm sorting + end time on COPAS worm sorting) / 2.