An image analysis toolbox for high-throughput *C. elegans* assays

Carolina Wählby[1,2], Lee Kamentsky[1], Zihan H Liu[1], Tammy Riklin-Raviv[3], Annie L Conery[4], Eyleen J O'Rourke[4], Katherine L Sokolnicki[1], Orane Visvikis[5], Vebjorn Ljosa[1], Javier E Irazoqui[5], Polina Golland[3], Gary Ruvkun[4], Frederick M Ausubel[4] & Anne E Carpenter[1]

We present a toolbox for high-throughput screening of image-based *Caenorhabditis elegans* phenotypes. The image analysis algorithms measure morphological phenotypes in individual worms and are effective for a variety of assays and imaging systems. This WormToolbox is available through the open-source CellProfiler project and enables objective scoring of whole-worm high-throughput image-based assays of *C. elegans* for the study of diverse biological pathways that are relevant to human disease.

Phenotypic assays of intact organisms allow for the study of biological pathways and diseases that cannot be reduced to biochemical or cell-based assays. *C. elegans* is a useful model organism for studying biological processes: they share many molecular and physiological homologies with humans[1], high-throughput instrumentation and reagent libraries exist for sample preparation and imaging in *C. elegans*[2], and deviations from wild type are often readily apparent because the worms are visually transparent and follow a stereotypic developmental pattern[3]. Large-scale chemical and RNA interference (RNAi) screens using *C. elegans* are widespread[4–6] and can be used to probe complex processes such as metabolism, infection and behavior; however, to date, the analysis of such experiments has largely been manual, subjective and onerous.

Much progress has been made in automating the analysis of particular types of *C. elegans* experiments, such as those involving low-throughput, high-resolution, three-dimensional or time-lapse images or those involving images of embryos[7–11]. However, there is still a strong need to automate the analysis of high-throughput, static images of adult worms in liquid culture, which is a common screening output. For most assays, the density of worms per microplate well causes the worms to touch or cluster, which has limited automated analyses to population-averaged measurements[12,13], hiding population heterogeneity and prohibiting measurements on individual worms.

An alternative to microscopy is flow systems adapted for worms (for example, COPAS, Union Biometrica) that measure length, optical density and fluorescence emission at transverse slices along the length of individual worms. However, image-based screens have several benefits: they allow detection of more complex phenotypes by a two-dimensional analysis of shape and signal patterns, and they do not require resuspension of worms in additional liquid before analysis, allowing for the use of smaller sample volumes and closed culture conditions, a key factor when screening large libraries of small molecules and RNAi clones and when using pathogenic microbes. In addition, image-based screening allows for the visual confirmation of results, the images form a permanent record that can be rescreened for additional phenotypes, and low-throughput experiments require no more equipment than a microscope and a digital camera.

To improve *C. elegans* phenotype scoring from images of adult worms in liquid, we developed an image-analysis toolbox that can detect individual worms regardless of crossing or clustering. This method can measure hundreds of phenotypes related to shape, biomarker intensity and staining patterns in relation to the anatomy of the worms.

A typical workflow starts with bright-field images (**Fig. 1a**). We preprocess these images to compensate for illumination variations, detect well edges and make the image binary (**Fig. 1b**). The next step, and the major challenge of this method, is 'untangling', meaning the detection of individual worms among clustered worms and debris. To address this issue, we first construct a model of the variability in worm size and shape from a representative set of training worms. This model is then used to untangle and identify individual worms (**Fig. 1c,d**). Many measurements, such as size, shape, intensity, texture and spot counts, can then be made on a per-worm basis using all image channels available, as is common for cell-based assays[14]. Many phenotypes, such as spot area per worm, can be scored directly using such measurements; more complex phenotypes, such as subtle or complex changes in protein expression patterns, can be scored using a combination of measurements and machine learning[15]. If a reporter signal location is of interest, we map each worm to a low-resolution atlas, allowing for a quantification of the signal that is correlated to the worm's anatomy.

We evaluated the untangling performance using images from our prior work[8] in which 15 worms were placed in each well of a 384-well plate. Approximately 1,500 worms from 100 wells were
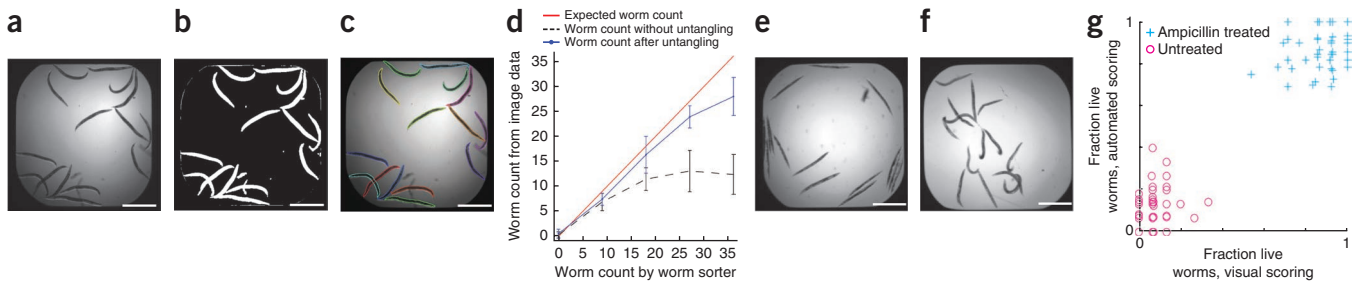

Figure 1 | Workflow and performance of the WormToolbox. (**a,b**) Starting from a bright-field image (**a**), we removed variations in illumination and separated objects from the background (**b**). (**c**) We created a worm model from non-touching training worms and untangled individual worms from clusters and debris using the model. Each worm is shown with a uniquely colored outline. (**d**) Automated detection of adult worms in the presence of eggs and progeny in images from a 384-well plate, with each well containing a predetermined percentage of adults and counted with or without untangling. The number of adult worms counted using the image analysis is plotted against the known number of adult worms sorted into each well. Error bars, s.d. (*n* = 16 images). (**e,f**) The images show the morphology of worms infected with *E. faecalis* without (**e**; worms are dead) and with (**f**; worms are alive) treatment with ampicillin. (**g**) The automated viability scoring based on the shape and texture measurements from individual worms compared to the visual scoring of positive (ampicillin treated) and negative (untreated) control wells. Scale bars, 1 mm.

manually delineated, revealing that 46% of the worms were clustered or were touching other worms (**Supplementary Fig. 1**). Compared to the manual delineation, only 51% of the worms were correctly detected with automated foreground-background segmentation followed by connected component labeling. When applying the untangling algorithms of the WormToolbox, the performance increased to 81%, which was sufficient for the assays presented here. The major remaining source of error in worm delineation was poor image contrast close to the well edges; the performance improved to 94% when we manually corrected the foreground-background segmentation, which decoupled the errors caused by untangling from the errors in the initial segmentation. We also tested the performance of the untangling in relation to the size of the training set and found that performance plateaued when using a worm model constructed from 50 randomly selected training worms. This means that training can be done on a relatively small number of worms representing the phenotypic variation of a given experiment (**Supplementary Fig. 2**).

We then evaluated the toolbox on data from a different laboratory and imaging system[13]. The challenge was to detect individual adult worms that were partly clustered and mixed with eggs and progeny. We trained the worm model on stage L4 and adult worms only and observed that untangling improved the accuracy of finding individual adult worms as compared to using thresholding and size sorting alone (**Fig. 1d** and **Supplementary Fig. 3**). The model efficiently excluded smaller larvae (stages L1, L2 and L3) and eggs, and the performance of the model was relatively robust in the presence of up to sixfold more progeny than adults (**Supplementary Fig. 4**). We also evaluated the performance of worm untangling as the number of worms per well increased. The wells contained L1, L3 or adult worms at increasing concentrations, and we created a separate worm model for each developmental stage. The performance of the model was higher for the slightly smaller L3 worms, as more space between these worms led to less clustering, but untangling became unstable when the worms were so small (L1) that the image resolution only allowed for a few pixels per worm (**Supplementary Figs. 5** and **6**).

In the second assay, we evaluated the toolbox for scoring viability, which can be read out as a morphological phenotype in the bright-field images alone without the need for a viability stain. Worms in liquid culture tend to be curved and evenly opaque

when alive but become rod-shaped and textured when they die (**Fig. 1e,f**). We untangled high-throughput images of worms that we infected with *Enterococcus faecalis* and either mock-treated with DMSO or treated with ampicillin[12]. After making shape, intensity and texture measurements of each untangled worm, we manually selected 150 live and 150 dead training examples from one 384-well plate. We then used the machine learning of CellProfiler Analyst[15] (**Supplementary Fig. 7**) to identify a combination of measurements that discriminated between live and dead worms. Finally, we applied the classifier to 1,500 worms from a different 384-well plate and verified that it distinguished between live and dead worms as well as humans can (**Fig. 1g**). To evaluate the performance of the viability scoring on more heterogeneous data from a real high-throughput experiment, we selected 1,766 random images and 200 hits from a 37,200-compound screen[12] and compared automated scoring with visual scoring based on bright-field images (**Supplementary Fig. 8**). We achieved an accuracy of 97% and a precision of 83%, indicating that morphology-based viability screening could be a feasible alternative to the viability stains (SYTOX) used in the original screen.

In the third assay, we evaluated how well the toolbox could differentiate between a positive and a negative control from an RNAi screen for regulators of fat accumulation[16]. The positive control downregulates *daf-2*, and the negative control was an empty vector. We compared two different approaches for pattern quantification: per-well measurements (using the basic functionality of CellProfiler), where no effort was made to assign fatty regions to individual worms, yielded a false discovery rate of 22.2% (**Supplementary Fig. 9**); and per-worm measurements (using the untangling functionality of the WormToolbox) yielded a false discovery rate of 4.5% (**Supplementary Fig. 10**). The per-worm measurements were superior because they captured the heterogeneity of the population, which was lost in the population averages from per-well measurements.

Finally, we evaluated the ability of the toolbox to detect worms with a change in the location of GFP expression (**Fig. 2a**). We used a *C. elegans* strain in which GFP expression (enhanced GFP from the Fire set of *C. elegans* vectors) in the intestine is under the control of a promoter that responds to *Staphylococcus aureus* infection[17]. A pharyngeal stain (expressing monomeric (m)Cherry) served as an internal control. We could not score the assay using simple approaches, such as measuring the total

Figure 2 | Scoring fluorescence signal distribution with the WormToolbox. (**a**) The images show *clec-60∷GFP* expression (green) in wild-type worms (WT, top) and in *pmk-1* (*km25*) mutants (*pmk-1*, bottom). The pharynx is marked with *myo-2∷mCherry* (red). (**b**) All the worms in **a** are juxtaposed after automated detection and digital straightening. (**c**) A low-resolution worm atlas enables feature measurements in relation to the worm's anatomy. T_1of6 through T_6of6 indicate the different transversal segments of the worms. (**d**) Comparison of the s.d. of GFP fluorescence intensity in transversal segment T_2of6 of the worms in 48 wells containing either positive (*pmk-1* mutant, blue stars) or negative (wild type, magenta diamonds) controls. Scale bars, 1 mm.
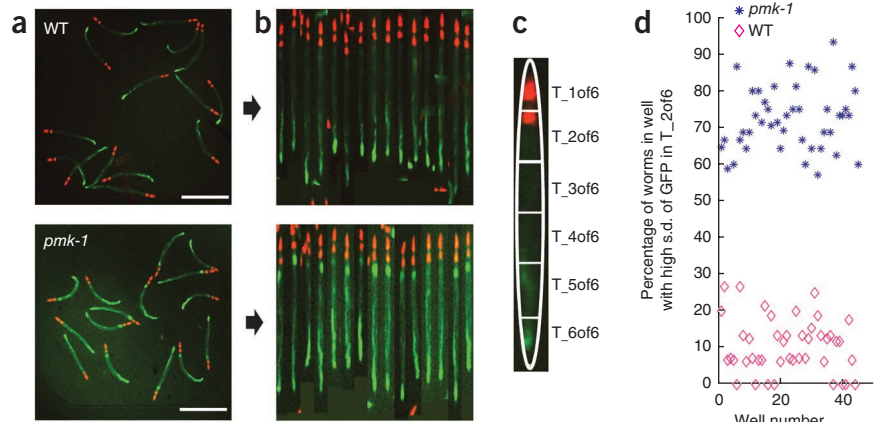
intensity of GFP expression per well or per worm or counting the number of green fluorescent spots (**Supplementary Fig. 11**). However, using worm straightening and our atlas mapping (**Fig. 2b,c**), we were able to quantitatively detect elevated expression of *clec-60∷GFP* in the anterior intestine (**Fig. 2d**) and separate positive and negative controls with a Z'-factor of 0.21. Here we focused on the location of the signal along the length of the worm, but we can also discern the asymmetric signal distribution across the width of the worm.

The WormToolbox is, to our knowledge, the first system to automatically, quantitatively and objectively score a variety of phenotypes in individual *C. elegans* in static, high-throughput images. We implemented the toolbox as modules for the free and open-source CellProfiler[14,18] software. CellProfiler is designed with an emphasis on ease of use, is compatible with cluster computing and is flexible to new assays that are developed by the scientific community. Training the worm model takes less than an hour, and once an image analysis pipeline is set up for an assay, a typical analysis takes 10–30 s per image, or much less if a computing cluster is available. It is also worth noting that we were able to reuse the worm model created for the second assay in both the third and the fourth assays, which consisted of images captured over several years and on different microscope systems.

The performance of the WormToolbox depends on the contrast between the worms and the surrounding background, making it sensitive to large variations in background illumination and to the worm-like tracks sometimes formed when growing worms on agar medium. The WormToolbox can handle images of worms on agar in large plates, but further optimization is needed for 384-well plates. In liquid culture, the untangling can handle up to 20 adult worms per well in a 384-well format and is designed to detect worms in the range of size and shapes of the training worms that were used to create the worm model. It is probable that unexpected phenotypes will be discarded as debris, but wells with a low fraction of correctly detected worms can be flagged for visual examination. In future work we will extend the WormToolbox by adding more worm-specific measurements based on their unique anatomy and incorporating better methods for mixed worms at various stages of development.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

AUTHOR CONTRIBUTIONS
A.E.C. and E.J.O. conceived of the idea for the study. C.W., L.K., Z.H.L., P.G., V.L. and T.R.-R. designed and implemented the algorithms of the WormToolbox. A.L.C., E.J.O. and O.V. developed sample assays and collected image data. J.E.I., G.R. and F.M.A. designed and supervised screens. C.W. and K.L.S. developed analysis pipelines and evaluated results, with input from E.J.O. and A.L.C., C.W., L.K., K.L.S., E.J.O. and A.E.C. wrote the manuscript.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Harris, T.W. *et al. Nucleic Acids Res.* **32**, D411–D417 (2004).
2. O'Rourke, E.J., Conery, A.L. & Moy, T.I. *Methods Mol. Biol.* **486**, 57–75 (2009).
3. Sulston, J.E. & Horvitz, H.R. *Dev. Biol.* **56**, 110–156 (1977).
4. Artal-Sanz, M., de Jong, L. & Tavernarakis, N. *Biotechnol. J.* **1**, 1405–1418 (2006).
5. Kamath, R.S. & Ahringer, J. *Methods* **30**, 313–321 (2003).
6. Labbé, J.C. & Roy, R. *Clin. Genet.* **69**, 306–314 (2006).
7. Buckingham, S.D. & Sattelle, D.B. *Invert. Neurosci.* **8**, 121–131 (2008).
8. Long, F., Peng, H., Liu, X., Kim, S.K. & Myers, E. *Nat. Methods* **6**, 667–672 (2009).
9. Murray, J.I. *et al. Nat. Methods* **5**, 703–709 (2008).
10. Sönnichsen, B. *et al. Nature* **434**, 462–469 (2005).
11. Green, R.A. *et al. Cell* **145**, 470–482 (2011).
12. Moy, T.I. *et al. ACS Chem. Biol.* **4**, 527–533 (2009).
13. Gosai, S.J. *et al. PLoS ONE* **5**, e15460 (2010).
14. Carpenter, A.E. *et al. Genome Biol.* **7**, R100 (2006).
15. Jones, T.R. *et al. Proc. Natl. Acad. Sci. USA* **106**, 1826–1831 (2009).
16. O'Rourke, E.J., Soukas, A.A., Carr, C.E. & Ruvkun, G. *Cell Metab.* **10**, 430–435 (2009).
17. Irazoqui, J.E., Urbach, J.M. & Ausubel, F.M. *Nat. Rev. Immunol.* **10**, 47–58 (2010).
18. Kamentsky, L. *et al. Bioinformatics* **27**, 1179–1180 (2011).

ONLINE METHODS

Software. The WormToolbox is available as modules for CellProfiler[14,18]. The open-source code of the CellProfiler WormToolbox algorithms described here is available as **Supplementary Software 1** and for download at http://www.cellprofiler.org/. Example pipelines for worm model training, worm untangling and feature extraction are available as **Supplementary Software 2** and on the CellProfiler website. Compiled versions of the code and updates are available at http://www.cellprofiler.org/. Below we describe the steps of the workflow, as well as our four sample assays. Instructions for how to get started using the WormToolbox are provided in **Supplementary Methods 1**. Image data and manual delineations are available from the Broad Bioimage Benchmark Collection (http://www.broadinstitute.org/bbbc/).

Compensating for uneven illumination. Uneven illumination often distorts bright-field microscopy images of the multi-well plates typically used in high-throughput chemical and genetic screens, making foreground-background intensity thresholding difficult. Our new approach for approximating background illumination and well-edge position is based on the convexity of both the well and the illumination field (**Supplementary Fig. 12**). The algorithm is as follows: choose 256 evenly spaced intensity levels between the minimum and maximum intensity for the image. Starting from the lowest intensity, for each intensity, find all pixels with equal or higher intensity. Find the convex hull that encloses those pixels, set the pixels of the output image within the convex hull to the current intensity and continue to the next intensity level. If the well edges are dark and the well has a convex shape, this approach removes the well edge and compensates for uneven illumination without the need for any input parameters, making it robust to the variations that are often present in high-throughput experiments. The final result is thresholded using Otsu's[19] method, resulting in a binary image that serves as the input for the worm untangling step.

Worm detection. After illumination correction and thresholding, we created a mathematical description of each worm cluster (**Supplementary Fig. 13**). We reduced each binary object to its morphological skeleton and let each segment of the skeleton represent a worm segment and each branch point represent a point where worms touched or intersected. This way, the segments and branch points comprising the worm cluster can be described as a mathematical graph, and untangling becomes a search for paths through the graph that probably represent complete worms. More precisely, we searched for the ensemble of paths through a cluster that best represented the true worms as compared to a worm model, limiting worm overlap and maximizing cluster coverage. Our advancements made here compared to our previous work[20,21] are described in the next three sections.

Worm model construction and shape cost. The worm model was created from a comprehensive set of non-touching training worms, essentially as in our prior work[20], but here with the shape descriptor based on angles rather than on spatial coordinates. We sampled equidistant control points along the morphological skeleton of each training worm using cubic spline interpolation. Each of the control points other than the first and last was at the vertex of an angle formed by the lines from its predecessor and successor.

These angles and the length of the path formed a feature vector that functioned as our shape descriptor. Worm width, length and area were also extracted, and we made the training set symmetric by mirroring all samples along the *x* and *y* axes. The shape cost of a path potentially representing a worm in a cluster was given by the dot product of the feature vector describing the path and a cross-correlation matrix derived from the training data. Note that we were only looking at the overall body shape when training the model, which does not vary as much as other features of the worms, such as fluorescent markers or bright-field stains and texture. The training worms should represent the worms in the dataset, and the variation in size and shape of the training worms must be within certain limits, for example, a variation in length of a factor of two might cause the untangling step to divide some long worms in half or exclude some short worms as debris. Any worms that deviate in shape and posture from that expected by the model are discarded as debris. It is therefore feasible to flag wells with few detected worms compared to foreground pixels so they can be screened visually (or with an improved worm model) to detect unexpected body shape or size phenotypes (or failures in worm detection resulting from large amounts of debris or other problems). It is also worth noting that because of similarity in worm size and shape, we were able to reuse the worm model created for the second assay in both the third and fourth assays, which consisted of images captured over several years and on different microscope systems.

Preprocessing of cluster skeletons. Artifacts appear when two or more adjacent worms form regions wider than the worm width, resulting in a skeleton not centered on a true worm. In the two-worm case (**Supplementary Fig. 14**), the skeleton is composed of the two segments that enter the area where the worms touch, the two segments that leave the area and a single segment running the length of the area where the worms touch. To improve the alignment of the segmentation result with the true worms, we introduced a preprocessing step: touching areas were defined by a circular structuring element whose diameter was the maximum width of a worm. All skeleton ends adjacent to the area were connected with new paths, and the best paths were selected by the path search described below. To improve worm detection in the cases where two worms touched end to end without producing a branch point, we added branch points at an average worm's length starting from each endpoint of every skeleton segment longer than the longest training worm.

Worm untangling by path search. Once the skeleton had been preprocessed, we considered the combined cost of different ensembles of paths representing worms. Conceptually, the algorithm is composed of three steps: enumeration of paths, calculation of the costs of individual paths and calculation of the costs of ensembles of paths. The first step is to generate all paths whose lengths are between the minimum and maximum acceptable lengths, as defined by the worm model, and discard a path if the shape-cost function exceeds the maximum acceptable cost. There are three parts to the cost of a particular ensemble of paths: the sum of the costs of the individual paths in the ensemble, a penalty cost that is proportional to the length of all segments that are shared by the paths in the ensemble and a penalty cost that is proportional to the length of all segments that do not appear in

any path in the ensemble. Details on the algorithm are described in **Supplementary Methods 2** and the open source of the code in **Supplementary Software 1** (and at http://www.cellprofiler.org/).

Straightening and atlas-based feature extraction. To extract the reporter signal location, worms were transformed to a straight shape by resampling the image data along lines perpendicular to the central axis of the worm, similar to the method previously described[22]. However, here we made the processing much faster by reusing the spline function describing the path through the worm during untangling. If a head- or tail-specific marker was available, the worms were flipped so that they all had an intensity distribution skewed in the same direction. Our low-resolution worm atlas consists of a user-defined number of transversal and longitudinal segments. The intensity mean and s.d. were extracted from each subsegment in any number of image channels.

Evaluation of worm detection. We evaluated the performance of the untangling on 100 bright-field images from our previously published high-throughput experiment[12]. Each image of a well from a 384-well plate contained approximately 15 worms. Ground truth was created by manually delineating all worms and saving them as individual binary masks to enable the evaluation of worm detection, clustering and overlap (provided through http://www.broadinstitute.org/bbbc/). In this dataset, 46% of the worms touched or overlapped, with most of the worms located in clusters of two (**Supplementary Fig. 1**). We calculated the accuracy, precision, recall and F-factor for the individual worms, and a threshold on an F-factor of 0.8 yielded 81% correctly segmented worms by automated foreground-background segmentation followed by untangling and 94% correct segmentation if the foreground-background segmentation was manually corrected before untangling. If the worms were defined by conventional intensity thresholding and connected component labeling, only 51% of the worms were correctly segmented. The performance was generally higher for smaller clusters of worms (**Supplementary Fig. 2**). The performance was also affected by the size of the training set and plateaued at about 50 training worms. The processing time was reduced by about tenfold compared to our previously published implementation[17], and all the steps combined, including image preprocessing, worm untangling and straightening, typically took less than 10 s.

Assay 1: finding individual adult worms in the presence of eggs and progeny. Images were kindly provided by Gosai *et al.*[13] Briefly, *C. elegans* worms were cultured at 22 °C on nematode growth medium plates seeded with *Escherichia coli* strain OP50. Next, 36 worms, containing a predetermined percentage (0%, 25%, 50%, 75% or 100%) of adult worms, were dispensed into each well of a 384-well plate. Images were acquired on the ArrayScan VTI HCS Reader (Cellomics, ThermoFisher) fitted with a 2.5× objective and a 0.63× coupler. We compensated for uneven background illumination using the convex-hull approach and identified objects by automated intensity thresholding. We constructed a worm model from the non-touching, adult worms, which we identified based on their areas and maximum widths. We then untangled all worms (**Supplementary Fig. 3**) and counted the number of adult worms per image (**Fig. 1d**). Next, we tested the limit of adult worm detection at increasing

concentrations of progeny (**Supplementary Fig. 4**) and concluded that the untangling was stable to a concentration of about sixfold more progeny than adults. We then found the limits on worms per well using a COPAS (Union Biometrica) worm sorter to seed 1–96 L1, L3 or adult worms per well in a 384-well plate in four replicates (**Supplementary Fig. 5**). For the adult worms, the untangling worked well until the concentration reached about 20 worms per well, whereas for L3 worms, the limit was reached at a concentration of about 30 worms per well, which is to be expected, as these worms are smaller and are therefore farther apart in the well. For the smaller L1 worms, the image resolution only allowed a few pixels per worm, making the model-based worm untangling unstable at this stage, particularly in the presence of small bubbles, which can be confused with small worms at this resolution. To explore the modes of failure of the assay, we examined a subset of segmentation results visually and found that the initial illumination correction and intensity thresholding had a large effect on the resulting segmentation (**Supplementary Fig. 6**).

Assay 2: live or dead scoring based on bright-field morphology. We cultured *C. elegans* (*glp-4* (*bn2*); *sek-1* (*km4*) mutant) worms on plates seeded with *E. coli* strain HB101. We infected sterile adult worms by pipetting them onto a lawn of *E. faecalis* and incubating them for 15 h at 15 °C. Using a COPAS (Union Biometrica) worm sorter, we transferred 15 of the infected worms into each well of a 384-well plate. As a positive control, we added 21 μg/ml ampicillin to 192 of the wells and mock treated 192 of the wells with an equal volume of DMSO[12]. We captured bright-field transmitted-light images showing the entire well using a Molecular Devices Discovery-1 microscope with a transmitted light module, a 2× low-magnification objective and MetaXpress software. We manually delineated a total of 60 worms from the positive and negative control wells and used them to construct a worm model. After untangling, we measured worm shape, intensity and texture. Using CellProfiler Analyst[15], we trained a classifier to distinguish the live and dead phenotypes based on 150 training examples. The classifier used eight shape, intensity and texture features (**Supplementary Fig. 7**). We applied the classifier to a set of images that did not include the training examples, classifying each worm as alive or dead. We then scored wells by the fraction of live worms and compared the results to the majority vote of three *C. elegans* specialists that scored the wells by visual inspection (**Fig. 1e**). We also compared automated and visual scoring of images randomly selected from a full-scale high-throughput screening experiment (**Supplementary Fig. 8**).

Assay 3: fat accumulation scoring based on staining pattern. We treated worms with either an empty vector (L4440 (ref. 23), Addgene) or RNAi against the insulin receptor (*daf-2*) according to standard procedures. We stained the worms with the fat-specific stain oil red O. Using an upgraded Axioscope microscope (Zeiss) with automated hardware (Biovision Inc.) and Surveyor software (Objective Imaging Ltd.), we acquired six bright-field color images per well. The original images were 2,782 × 3,091 pixels, but we scaled them to 690 × 765 pixels to speed the analysis. Before detecting and untangling the worms, we combined the color channels into a single gray-scale image. We used the same worm model as that for assay 2 and achieved satisfactory

segmentation results without any further adjustments. We defined fat regions by intensity thresholding and quantified fat patterns by measuring the extent of the fatty regions (**Supplementary Fig. 9**). We then compared per-well, per-cluster and per-worm measurements (**Supplementary Fig. 10**) and found the last to be ideal for the assay.

Assay 4: reporter pattern detection by worm straightening and atlas mapping. We used a transgenic strain of *C. elegans* that expresses GFP from the promoter of the gene *clec-60* and *myo-2*::mCherry for labeling the pharynx. We cultured the worms on nematode growth medium plates seeded with *E. coli* strain OP50 at 15–20 °C according to standard procedures and sorted 15 worms into each well of a 384-well plate. Forty-eight wells received wild-type worms (L4440 (ref. 23), Addgene) expressing *clec-60*::*GFP*, and 48 wells received *pmk-1* (*km25*) mutant worms. Using a Discovery-1 microscope (Molecular Devices), we acquired bright-field images, as well as fluorescence images, at two wavelengths (corresponding to GFP and mCherry). The images were 696 × 520 pixels[17]. We tested three approaches to phenotype scoring (**Supplementary Fig. 11**). First, we defined spots of GFP signal by intensity thresholding and likewise approximated worm count by thresholding the intensities in the image channel showing the pharynx (mCherry). Because large variations in GFP expression and touching worm heads led us to underestimate the number of worms, this approach was not successful. Instead, we untangled the worms using the same worm model as that for assay 2 and achieved satisfactory segmentation results without

any further adjustments. The mean and s.d. of the GFP expression in the individual worms was also insufficient to separate the two phenotypes, so we continued the analysis by straightening the worms, aligning them to the low-resolution worm atlas and measuring the mean and s.d. of the GFP expression from each of six transversal subsegments that were spread evenly along the length of the worm. Instead of examining each measurement separately, we trained CellProfiler Analyst software[15] to distinguish the phenotypes by presenting it with examples of mutant and wild-type worms. The resulting classifier relied primarily on the difference in s.d. of the GFP fluorescence in transversal segment number two (the segment second from the head, referred to as T_2of6) to distinguish between wild-type and mutant worms. Based on this method, we labeled as mutant worms those with an s.d. of GFP expression in segment T_2of6 greater than 0.4. We then scored each well by the percentage of mutant worms and achieved a Z'-factor of 0.21.

A list of the general aspects of the assay design and error handling for the image-based screening is presented in **Supplementary Table 1**, and details regarding image settings are given in the **Supplementary Note**.

19. Otsu, N. *IEEE Trans. Syst. Man Cybern.* **9**, 62–66 (1979).
20. Riklin-Raviv, T. *et al. Int. Conf. Med. Image Comput. Comput. Assist. Interv.* **13**, 634–641 (2010).
21. Wählby, C. *et al. Proc. IEEE Int. Symp. Biomed. Imaging* 552–555 (2010).
22. Peng, H., Long, F., Liu, X., Kim, S.K. & Myers, E.W. *Bioinformatics* **24**, 234–242 (2008).
23. Timmons, L. & Fire, A. *Nature* **395**, 854 (1998).