HHS Public Access

Author manuscript

Nat Methods. Author manuscript; available in PMC 2013 April 01.

Published in final edited form as:

Nat Methods. 2012 October ; 9(10): 977–980. doi:10.1038/nmeth.2141.

Autonomous screening implicates new genes in synaptogenesis of *C. elegans*

Matthew M. Crane[1,2], **Jeffrey N. Stirman**[1,3], **Chan-Yen Ou**[4], **Peri T. Kurshan**[4], **James M. Rehg**[5], **Kang Shen**[4], and **Hang Lu**[1,3,*]

[1]Interdisciplinary Program in Bioengineering, Georgia Institute of Technology, Atlanta, GA, USA

[3]School of Chemical & Biomolecular Engineering, Georgia Institute of Technology, Atlanta, GA, USA

[4]Department of Biology, Howard Hughes Medical Institute, Stanford University, California, USA

[5]School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA, USA

Abstract

Morphometric studies in multicellular organisms are mostly performed manually because of the complexity of multidimensional features and lack of appropriate tools for handling these organisms. Here we present an integrated system to autonomously (i.e. without human supervision) identify and sort mutants with altered subcellular traits in real-time. We performed self-directed screens of synapse formation 100× faster and found both novel genes and phenotypic classes previously unidentified in extensive manual screens.

Microscopy is a powerful tool widely used to investigate cells, tissues, and organisms. Combining imaging and perturbations (e.g. mutations, siRNA treatment, or drugs) allows us to probe functions, explore mechanisms, and remediate dysfunctions[1]. Recent technological developments have made automated microscopy and sample handling for single-cell samples routine, even allowing automated identification of cells[2]. This has enabled large-scale imaging and drug interaction studies in cultured cells[3–5]. Statistical methods have been applied to high-content quantitative phenotypical descriptors of cell lines to identify subtle phenotypes and the underlying networks[5–7]. In contrast, although advances in sample handling, largely through the use of a modified FACS system[8] and multiwell-plate methods, have increased imaging and screening throughput, they are low-resolution technologies, and thus unable to distinguish many phenotypes of interest. A primary obstacle is that automated screening requires equipment that can robustly handle large sample numbers and a system for extracting and understanding data from high-content images. Recently, microfluidic approaches have improved the throughput of high-resolution imaging of *Caenorhabditis*

*corresponding author hang.lu@gatech.edu.

[2]Current address: SynthSys, University of Edinburgh, Edinburgh UK

Author Contributions: M.M.C and J.N.S designed and built the external system components. M.M.C wrote the software, designed the microfluidic device and performed the screens. C.Y.O performed the complementation tests. P.T.K. performed the mapping and provided valuable reagents. M.M.C, J.N.S, J.M.R., K.S., and H.L. designed the experiments and prepared the manuscript.

elegans (*C. elegans*)[9–11], but sorting has been based on user input[10] or simplistic criteria, such as the local intensity[9].

Here we present a system for performing autonomous screens based on high-content quantitative features in *C. elegans*. The system combines a microfluidic device, computer-vision tools, and a statistical framework to classify animals. The microfluidic device allows animals to be imaged and sorted rapidly (Fig. 1a). The acquired images are processed to identify the fluorescently-labeled objects using a two-stage computer-vision algorithm (Fig. 1b–d), and quantitative phenotypical descriptors are extracted and used to predict whether the animal has a pattern of interest (Fig. 1e–f). Finally, the microfluidic device sorts the identified mutants. Using this integrated system, we performed an autonomous forward screen for new classes of mutants affecting synaptogenesis.

Our automated microfluidic system (Fig. 1a and Supplementary Fig. 1) is optimized to simplify fabrication, minimize possible failures, and increase throughput while image at high magnification. It uses a simple "load, image, and sort" routine[9,10] that reliably manipulates animals and allows extended automated operation, even with the large variations in animal size resulted from mutagenesis (Supplementary Fig. 2). When loaded, the animal is transiently (~10sec) cooled to ~3°C. This rapidly immobilizes the animal for high-resolution image acquisition[9] without having to use feedback from the images to control mechanical immobilization mechanisms. In order for fully automated operation, we also developed extensive external system-level components and error handling that collectively minimizes situations requiring human intervention (Supplementary Note 1 and Supplementary Fig. 3).

To automate the decision-making and sorting based on morphometric criteria, we developed a computational framework that identifies specific fluorescent objects of interest such as the synapses or neuron soma, extracts quantitative phenotypical descriptors from these objects, and classifies the animal based on the descriptors (Fig. 1b,c). In contrast to most published computer-vision methods applied to biological analysis which are run offline[5,6,12], real-time processing is critical to allow sorting decisions in forward genetic screens. We therefore designed the algorithm to balance two competing requirements: high accuracy to maximize enrichment, which usually correlates with computational time, and minimal computational burden to maximize throughput (Supplementary Note 2).

We applied the method to identification of an EGFP-tagged synaptic vesicle marker in the well-characterized motor neuron DA9[13]. This is challenging because fluorescently labeled synapses are small, have limited numbers of fluorophores, and autofluorescent fat granules often have similar size and appearance to synapses. To automatically extract specific quantitative phenotypical descriptors requires the ability to distinguish relevant fluorescent structures from the irrelevant. A low false positive rate of fat granules labeled as synapses is important as a single synapse located far from stereotyped synapse positions could indicate an interesting mutant (Supplementary Note 2). Furthermore, to identify mutants, we need to detect subtle changes in synapse locations, size, and intensity because synaptogenesis is both complex and regulated[13,14].

To minimize errors, we designed a computer vision framework to identify the fluorescent objects of interest to operate in two stages (Fig. 1c): first, the program identifies pixels associated with probable synapses using features based on the local neighborhood surrounding each pixel; second, it uses these probable synapses to extract features based on the relative positions of potential synapses to each other and within the image. The second stage features were designed using *a priori* knowledge about the synaptic patterns, for example, that synapses are more likely to cluster near one another than to be randomly located in the worm body. Support Vector Machine (SVM) classifiers were trained to evaluate these features and identify synapses (Supplementary Notes 2,3 and Supplementary Figs. 4–7). This framework was designed so identification took less time than acquisition of the *z*-stack (~4 sec). The total processing time, including imaging and sorting, per imaged animal was ~10 seconds (Supplementary Table 1). This is a two-orders-of-magnitude saving in time compared to manual screening method with quantitation, which requires picking, mounting, imaging, manually quantifying phenotypical features using software such as ImageJ, classifying, and rescuing the animals (~20 minutes). The computer vision framework results in overall pixel accuracy of >99.9% when estimated by 5-fold cross-validation, which is exceptionally good given the difficulty (Supplementary Fig. 8). Once synapses are identified from each image, we use the overall synaptic pattern to extract 30 quantitative descriptors of the phenotypes such as the average size, shape, number and location of the synapses (Fig. 1f and Supplementary Note 3). These features are used to classify an animal as wild-type or mutant with high accuracy and limited bias (Supplementary Fig. 9).

We used two types of classifiers to screen and classify animals, as wild-type or mutant: a discriminative classifier and an outlier-detection classifier (Fig. 2a,b). The discriminative screen used both wild-type animals and a known mutant (*lin-44*$^{-/-}$)[13] to train a classifier to maximize the differences between the phenotypical descriptors of both populations. To train the classifier, ~130 wild-type and ~80 *lin-44*$^{-/-}$ mutants were imaged and quantitatively phenotyped using the system. Using two populations allowed the classifier to remove irrelevant descriptors that provide limited discriminative power. In some cases this method can provide a reduced false-positive rate and focus on identification of mutants with phenotypes similar to the known mutant (Fig. 2a). Although useful, this type of screen is less likely to find new classes of phenotypes, which in many problems would be more interesting. To screen more broadly for novel mutant classes, we used an outlier-detection scheme, where ~400 wild-type animals were imaged and used to model the wild-type population. The wild-type phenospace was modeled as a multi-variate Gaussian, and mutagenized animals with a cumulative distribution function <0.1% of wild-type were sorted as mutants (Fig. 2b). This allows screening for completely novel patterns, with the potential trade-off of a higher false positive rate.

To compare the discriminative powers of each screening method, and to obtain quantitative metrics on screening performance, we performed two small-scale screens. The discriminative classifier was used to perform a ~1,000 haploid genome screen, while a ~1,500 haploid genome screen was done with the outlier detection approach (Supplementary Note 4). Both screens resulted in similar false positive (~70%) and false negative (~10%)

rates (Online Methods), which was due to optimal setting of the decision boundary and selection of highly relevant phenotypical descriptors. Previous manual screens resulted in a discovery rate of about one mutant per few thousand haploid genomes. In contrast, our screens resulted in a discovery rate of one mutant per several hundred animals, and included phenotypes similar to those previously published[13–15] as well as novel ones. This higher discovery rate implies a potential ability to identify more genes affecting the phenotype of interest when compared with current methods.

To demonstrate our ability to perform large-scale experiments, we screened ~20,000 haploid genomes (Online Methods), and identified nearly 60 mutants with altered phenotypical features. The majority of this screening was performed with an outlier detection approach. From all screens, we obtained several phenotypic classes of mutants, and a subset of animals has been further characterized, including one mutant that exhibited a strikingly novel phenotype (Fig. 2c,d and Supplementary Table 2). Interestingly, complementation tests of some of the mutants with dendritic RAB-3 localization against genes in the CDK-5 pathway[15] revealed two novel genes (*a117* and *a102*) and two alleles of previously implicated genes (*a107* and *a109*) (Fig. 2e). Moreover, allele *a085* has a striking novel morphological phenotype: axons contain enlarged spine-like protrusions filled with synaptic markers (Fig. 2f). The allele is mapped to a small region on chromosome IV. This demonstrates that our system is not only capable of finding alleles of previously manually identified genes, but also has the ability to identify novel genes and phenotypes missed by previous manual screens.

Our system is both camera- and microscope-independent, and can be easily adapted to other laboratories and other model networks. We demonstrated here for the first time autonomously screens in *C. elegans* for subtle subcellular changes using quantitative phenotypical descriptors. We screened at a sustained rate of over 100 haploid genomes per hour. Moreover, we identified novel phenotypes too subtle to be reliably assessed by eye. This allowed the discovery of weak alleles, as well as peripheral players in specific pathways. Applied to additional forward genetic, *RNAi*, or drug screens, this method will not only allow faster screens, but with a higher efficiency to find more genes and interactions than previously possible.

C. elegans culture

C. elegans were cultured according to established methods[1]. Mutagenesis was performed on age-synchronized L4 animals using ethyl methanesulfonate (EMS) according to standard protocols[2]. The overall scheme is shown in SFig 1. Briefly, when the animals reached the L4 stage, a large number of synchronized wild-type animals carrying the *wyIs85* transgene were suspended in a buffer solution and 20mMEMS (Sigma-Aldrich) for a period of four hours. Following the incubation, the animals were then rinsed with the M9 buffer solution and placed on plates. These P0 animals were allowed to grow to adults, and once the F1 progeny from these animals became adults, the F2 offspring were age-synchronized. F2 embryos were obtained by bleaching F1 adults using a solution containing about 1% NaOCl and 0.1 MNaOH, washed in M9 buffer, and cultured for 24 hours. Animals were then transferred and cultured on Nematode Growth Medium (NGM) plates seeded with *E. coli* OP50 until L4

stage. The strains used in this project included XA7810: N2, *wyIs85* (Pitr-1 *pB::gfp::rab-3*); XA7812: *lin-44(n1792)I, wyIs85*(Pitr-1 *pB::gfp::rab-3*).

For imaging and screening, animals were washed and suspended in M9 solution containing 0.01 volume% Triton X100 (Sigma-Aldrich). This prevented the animals from adhering to the tubing during injection. After washing animals off of the plates, animals were allowed to settle, the supernatant removed and additional M9 added. Animals were screened under a compound microscope using a 40× (NA=1.4) oil objective using the microfluidic chips. Sorting decisions were made based on differences in the reporter expression pattern or intensity; potential animals of interest were sorted into the mutant outlet and were collected directly from tubing connected to the mutant outlet with M9 solution containing 0.01% Triton X100. Animals were subsequently transferred to individual plates for culture and further examination. SNIP-SNP mapping and complementation tests were performed using standard protocols.

Materials and Equipment

The equipment utilized during our experiments includes the following: Peltier cooler (PJT-5 30 mm square and PJT-6 40 mm square peltier coolers), copper heat exchanger (custom designed and machined), peristaltic pump for coolant: 400 F/A (Watson Marlow), digital I/O Card (Pacdrive from Ultimarc), solenoid Valves and Manifold (3-Way 10mm Solenoid Valve, 188 series, ASCO Valve), microscope (Leica DM4500) with lens (63X Oil, NA=1.4). and camera (Infinity 3-1, Lumenera).

Automation, Handling of Multicellular Organisms, and System Design

Although manipulation and imaging of multicellular model organisms are an important part of modern biology, most methods are still labor intensive and time-consuming. Recent work, largely through multiwell plate methods, and a modified flow cytometer (COPAS), has made significant improvements in sample handling and low-resolution imaging of *C. elegans*, and *D. rerio* and *D. melanogaster* embryos. Recently, microfluidic approaches have improved the throughput of high-resolution imaging of *Caenorhabditis elegans*[3–6], but sorting has been based on user input [4,7] or simplistic criteria, such as the local intensity[3]. To date, all *C. elegans* screens, both automated and manual, have required human intervention to set the sorting boundaries[4,8,9], and the few automated screens have only been performed at very low resolution[8,10,11]. Thin glass capillary tubes have recently been used to image and then ablate axons in *D. rerio*, but image processing and ablation was performed manually using user input[12]. Similarly, microfluidic traps have been employed to align *D. melanogaster* embryos to enable highly controlled imaging conditions[13].

Automation of information extraction has largely been focused on post-processing of data acquired during an experiment. Within multicellular model organisms, this has focused on understanding animal behavior and interactions, and with lineage tracking of cell fates. Social interaction between flies has been analyzed using computer vision to extract behavioral information[14,15]. Similar work has also been extended to mammalian models with mice[16]. Tracing of cell lineages from early to late embryogenesis has allowed tracking

of *C. elegans* embryos through many divisions[17–20]. A digital atlas of *C. elegans* cell locations, and methods to identify them, has recently been created[21]. Performing image processing to identify quantitative information in near real time during a screen, however, requires a framework that places a significantly higher value on speed and throughput. The framework provided in this work can be extended to tackling these problems for real-time screening of other model organisms.

Included on the website is software used during screening. Because of differences in equipment this is primarily intended as a guide for development of customized code for the end user. The code contains functions used to operate the microfluidic device, as well as that to extract relevant information from the fluorescent images. The optimized feature extraction and phenotypical parameters can be modified to focus on identification of other small, fluorescent objects. More information can be found in a readme available with the software.

The completed microfluidic device is shown in SFig 4A. The device is a standard two-layer PDMS device[22,23], similarly to our previous work[3,4,24], fabricated using a rapid prototyping method and thermal bonding between layers (SFig 4B). Rather than using conventional fullclosure valves that rely on a curved cross-section of the flow-layer, this device uses a rectangular flow layer. Although valves incapable of fully closing are typically considered drawbacks in microfluidic devices, in this case they could be used advantageously to simplify the design. Because the valves always allow a small amount of flow (Supplementary Fig. 2C), the device was designed using partial closure valves to position animals in the imaging channels, instead of suction channels[3].

In addition to the development of a microfluidic device for screening, external systemic components were required to allow for automated sorting. This required a closed loop control system and the development of specific hardware to interface with and control the on-chip components. Creating external components that would allow computerized control of on-chip components was necessary (Supplementary Fig. 1).

Because we needed to run the system for a very long period during the screen, a comprehensive framework was developed for handling errors robustly (Supplementary Fig. 3). This includes an external macroscale control components to be integrated with the microfluidic device to allow closed-loop control, an error handling routines to reduce the need for operator intervention for the closed loop control, and optimization of the operational sequence to minimize the amount of time spent per animal. The entire system was coded within Matlab. External calls were to the machine learning libraries and a dll for valve control operation. Due to potential stability issues with Matlab operation, however, it could be advisable for future efforts to focus on the open microscopy environment and code in a more robust language such as Java.

Genetic Screening

The screening methodology was optimized to maximize the number of independent F2 animals screened, and to minimize the likelihood of screening clonal siblings. This was done by pooling F2 animals from a number of F1 mothers, and then screening a subset of this

group. Due to the screening methodology of the number of haploid genomes cannot be directly calculated, but several assumptions have to be made regarding the sampling from F2 population. These assumptions are derived from an earlier paper[25].

Following mutagenesis, approximately 25 P0 animals were placed on each large NGM plate. Each P0 animal has around 60 progeny. This results in approximately 1,500 F1 animals per plate, or 3,000 haploid genomes. Given that F1 animals are A/B, we can assume that half of the animals are homozygous (A/A or B/B), and the naïve assumption is that for every 2 F2 animals screened, we have screened 1 haploid genome. On average, 500 F2 animals were screened from each of these populations, resulting in 250 haploid genomes. Given that we were sampling from a larger population of animals, there is the possibility that an F2 animal could be homozygous for the same genome present in an animal that was previously screened. The probability of this occurring was calculated as $1-\left(\frac{2999}{3000}\right)^{250}=0.08$. Given there were 3,000 haploid genomes present, we screened 0.08 * 3,000 = 240. This yields a ratio of $\frac{240}{500}=0.48$, implying that because we have screened a total of ~40,000 animals, that corresponds to nearly 20,000 haploid genomes.

Following screening, the animals identified were cloned and then verified by reimaging. Once the identified mutants have been cloned and decontaminated, it is necessary to verify that the animals are probable mutants. Furthermore, because genetic validation by sequencing is such a slow and time consuming process, it is important to visually confirm that each of the animals to be mapped or sequenced indeed have an altered synaptic expression. Validation was done using standard worm protocols using a worm-slide and sodium azide for immobilization. A small population (~15–20) of animals of each genotype was imaged, and these images were used to determine whether the animals were actual mutants, and thus the accuracy of the screening. Some of the mutants were phenotyped with larger numbers, and were characterized in greater detail to predict the potential pathway that was affected (Supplementary Table 2). Each of the genotypes classified as mutants of interest were phenotyped. For the purpose of these results, animals that were sorted as mutants but failed to produce any offspring that could be used to verify whether the animals were correctly or incorrectly sorted, were removed from the results.

To determine the accuracy of the screening protocol, images of all processed animals were collected and analyzed to determine whether animals were correctly or incorrectly sorted as wild-type animals. This information was used to determine the accuracy of the system during screening. The false negative rate was determined by comparing the manual scores to the computer scores to determine the accuracy of the algorithm. The screen using the discriminative classifier resulted in a false positive rate of 69%, a false negative rate of 9%, and a theoretical enrichment of 6,000%. The screen using the outlier detection approach resulted in a false positive rate of 74%, a false negative rate of 7%, and an enrichment of 6,000%. The marginally higher false negative rate could be a result of the discriminative classifier not recognizing some mutant phenotypes, or merely the sample size. The false positive rates of these classifiers are similar to or better than the manual methods due to animal-to-animal variations and the manual method’s subjectivity in scoring the phenotypes.

These false positive animals also do not pose an issue in these rare-sort problems since it is a very small number of candidate mutants that need to be characterized in the following steps.

False positives were calculated using the following equation where animals sterile animals were removed from the denominator: $FP=\frac{wild-type\ sorted\ as\ mutants}{Animals\ sorted\ as\ mutants}$.

The false negative rate was calculated using the following equation:

$$FN=\frac{mutants\ sorted\ as\ wildtype}{mutants\ sorted\ as\ mutants+mutants\ sorted\ as\ wild-type}.$$

The upper bound of the enrichment was calculated using the equation:

$$Enrichment=\frac{Mutants\ Correctly\ Sorted/Number\ of\ animals\ sorted\ as\ mutants}{Mutants\ in\ starting\ population/Animals\ in\ starting\ population}.$$

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment

We thank funding from the US National Institutes of Health, the US National Science Foundation, the Alfred P. Sloan Foundation, and the Howard Hughes Medical Institute.

References

1. Jorgensen EM, Mango SE. The art and design of genetic screens: Caenorhabditis elegans. Nat Rev Genet. 2002; 3:356–369. [PubMed: 11988761]
2. Huang K, Murphy RF. Boosting accuracy of automated classification of fluorescence microscope images for location proteomics. Bmc Bioinformatics. 2004; 5
3. Whitehurst AW, et al. Synthetic lethal screen identification of chemosensitizer loci in cancer cells. Nature. 2007; 446:815–819. [PubMed: 17429401]
4. Jones TR, et al. Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106:1826–1831. [PubMed: 19188593]
5. Collinet C, et al. Systems survey of endocytosis by multiparametric image analysis. Nature. 2010; 464:243–249. [PubMed: 20190736]
6. Bakal C, Aach J, Church G, Perrimon N. Quantitative morphological signatures define local signaling networks regulating cell morphology. Science. 2007; 316:1753–1756. [PubMed: 17588932]
7. Perlman ZE, et al. Multidimensional drug profiling by automated microscopy. Science. 2004; 306:1194–1198. [PubMed: 15539606]
8. Doitsidou M, Flames N, Lee AC, Boyanov A, Hobert O. Automated screening for mutants affecting specification in C. elegans. Nature Methods. 2008; 5:869–872. [PubMed: 18758453]
9. Chung K, Crane MM, Lu H. Automated on-chip rapid microscopy, phenotyping and sorting of C. elegans. Nat Meth. 2008; 5:637–643.
10. Crane MM, Chung K, Lu H. Computer-enhanced high-throughput genetic screens of C. elegans in a microfluidic system. Lab on a Chip. 2009; 9:38–40. [PubMed: 19209332]

11. Rohde CB, Zeng F, Gonzalez-Rubio R, Angel M, Yanik MF. Microfluidic system for onchip high-throughput whole-animal sorting and screening at subcellular resolution. PNAS. 2007; 104:13891–13895. [PubMed: 17715055]
12. Murray J, et al. Automated analysis of embryonic gene expression with cellular resolution in C. elegans. Nature Methods. 2008; 5:703–709. [PubMed: 18587405]
13. Klassen MP, Shen K. Wnt Signaling Positions Neuromuscular Connectivity by Inhibiting Synapse Formation in C. elegans. Cell. 2007; 130:704–716. [PubMed: 17719547]
14. Poon VY, Klassen MP, Shen K. UNC-6/netrin and its receptor UNC-5 locally exclude presynaptic components from dendrites. Nature. 2008; 455:669–U668. [PubMed: 18776887]
15. Ou C-Y, et al. Two Cyclin-Dependent Kinase Pathways Are Essential for Polarized Trafficking of Presynaptic Components. Cell. 2010; 141:846–858. [PubMed: 20510931]
1. Brenner S. Genetics of Caenorhabditis-Elegans. Genetics. 1974; 77:71–94. [PubMed: 4366476]
2. Wood, WB. The Nematode Caenorhabditis elegans. Cold Spring Harbor Laboratory Press; 1988.
3. Chung K, Crane MM, Lu H. Automated on-chip rapid microscopy, phenotyping and sorting of C. elegans. Nat Meth. 2008; 5:637–643.
4. Crane MM, Chung K, Lu H. Computer-enhanced high-throughput genetic screens of C. elegans in a microfluidic system. Lab on a Chip. 2009; 9:38–40. [PubMed: 19209332]
5. Doitsidou M, Flames N, Lee AC, Boyanov A, Hobert O. Automated screening for mutants affecting specification in C. elegans. Nature Methods. 2008; 5:869–872. [PubMed: 18758453]
6. Rohde CB, Zeng F, Gonzalez-Rubio R, Angel M, Yanik MF. Microfluidic system for on-chip high-throughput whole-animal sorting and screening at subcellular resolution. PNAS. 2007; 104:13891–13895. [PubMed: 17715055]
7. Pardo-Martin C, et al. High-throughput in vivo vertebrate screening. Nat Meth. 2010; 7:634–636.
8. Doitsidou M, Flames N, Lee AC, Boyanov A, Hobert O. Automated screening for mutants affecting dopaminergic-neuron specification in C. elegans. Nat Meth. 2008; 5:869–872.
9. Jorgensen EM, Mango SE. The art and design of genetic screens: Caenorhabditis elegans. Nat Rev Genet. 2002; 3:356–369. [PubMed: 11988761]
10. Gosai SJ, et al. Automated High-Content Live Animal Drug Screening Using C. elegans Expressing the Aggregation Prone Serpin α1-antitrypsin Z. PLoS ONE. 2010; 5:e15460. [PubMed: 21103396]
11. O'Rourke EJ, Conery AL, Moy TI. Whole-animal high-throughput screens: the C. elegans model. Methods Mol Biol. 2009; 486:57–75. [PubMed: 19347616]
12. Pardo-Martin C, et al. High-throughput in vivo vertebrate screening. Nature methods. 2010; 7:634–636. [PubMed: 20639868]
13. Chung K, et al. A microfluidic array for large-scale ordering and orientation of embryos. 2011; 8
14. Branson K, Robie AA, Bender J, Perona P, Dickinson MH. High-throughput ethomics in large groups of Drosophila. Nature Methods. 2009; 6:451–U477. [PubMed: 19412169]
15. Dankert H, Wang LM, Hoopfer ED, Anderson DJ, Perona P. Automated monitoring and analysis of social behavior in Drosophila. Nature Methods. 2009; 6:297–303. [PubMed: 19270697]
16. Chaumont FD, Coura R, Serreau P. Computerized video analysis of social interactions in mice. Nature Methods. 2012; 9
17. Bao Z, et al. Automated cell lineage tracing in Caenorhabditis elegans. PNAS. 2006; 103:2707–2712. [PubMed: 16477039]
18. Boyle T, Bao Z, Murray J, Araya C, Waterston R. AceTree: a tool for visual analysis of Caenorhabditis elegans embryogenesis. Bmc Bioinformatics. 2006; 7:275. [PubMed: 16740163]
19. Murray J, et al. Automated analysis of embryonic gene expression with cellular resolution in C. elegans. Nature Methods. 2008; 5:703–709. [PubMed: 18587405]
20. Murray J, Bao Z, Boyle T, Waterston R. The lineaging of fluorescently-labeled Caenorhabditis elegans embryos with StarryNite and AceTree. Nature Protocols. 2006; 1:1468–1476. [PubMed: 17406437]
21. Long F, Peng H, Liu X, Kim SK, Myers E. A 3D digital atlas of C. elegans and its application to single-cell analyses. Nature methods. 2009; 6:667–672. [PubMed: 19684595]

22. Unger MA, Chou HP, Thorsen T, Scherer A, Quake SR. Monolithic microfabricated valves and pumps by multilayer soft lithography. Science. 2000; 288:113–116. [PubMed: 10753110]
23. Duffy DC, McDonald JC, Schueller OJA, Whitesides GM. Rapid prototyping of microfluidic systems in poly(dimethylsiloxane). Analytical Chemistry. 1998; 70:4974–4984. [PubMed: 21644679]
24. Chung K, Lu H. Automated high-throughput cell microsurgery on-chip. Lab on a Chip. 2009; 9:2764–2766. [PubMed: 19967110]
25. Ellis RE, Jacobson DM, Horvitz HR. Genes Required for the Engulfment of Cell Corpses During Programmed Cell Death in Caenorhabditis elegans. Genetics. 1991; 129:79–94. [PubMed: 1936965]

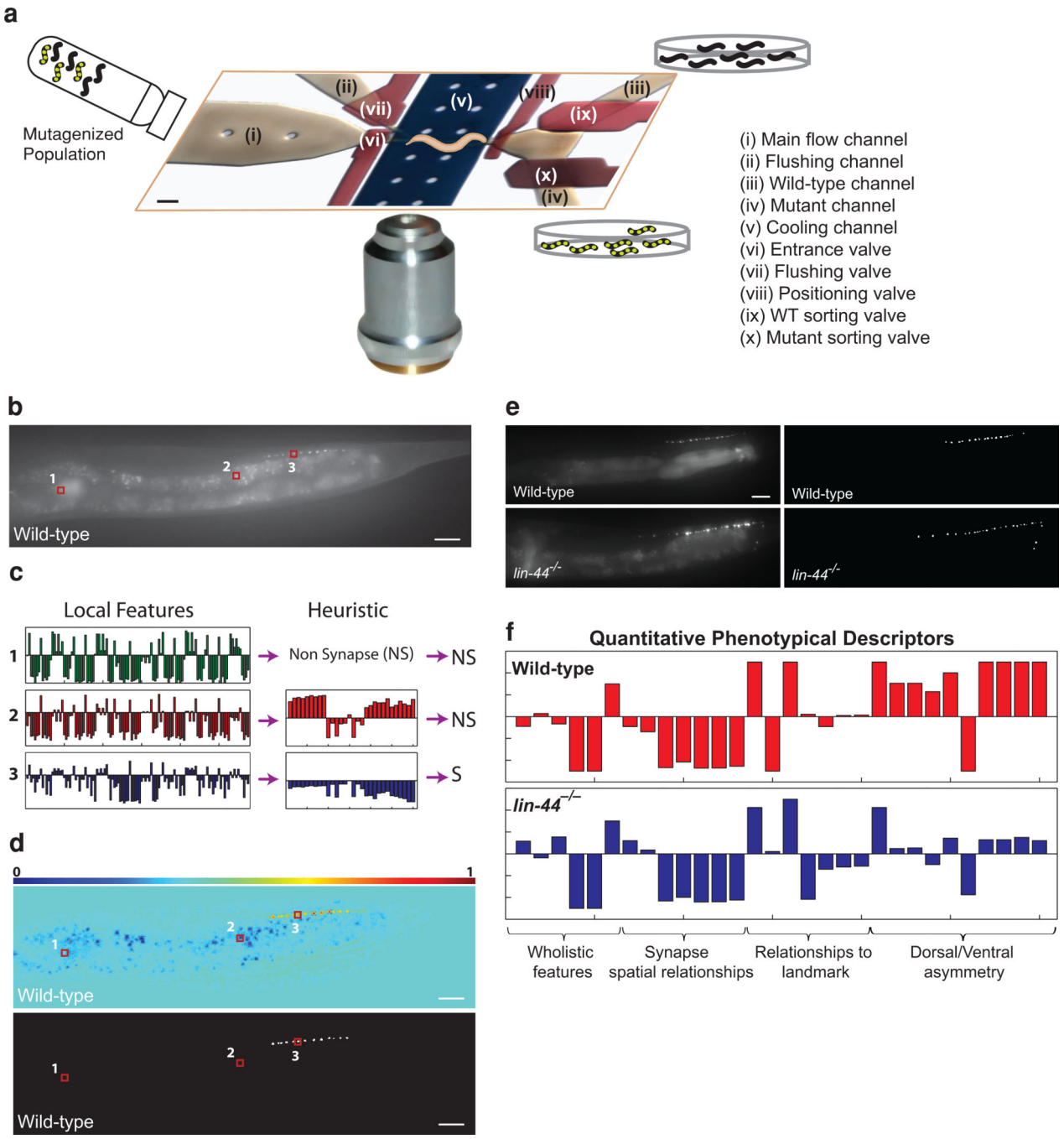
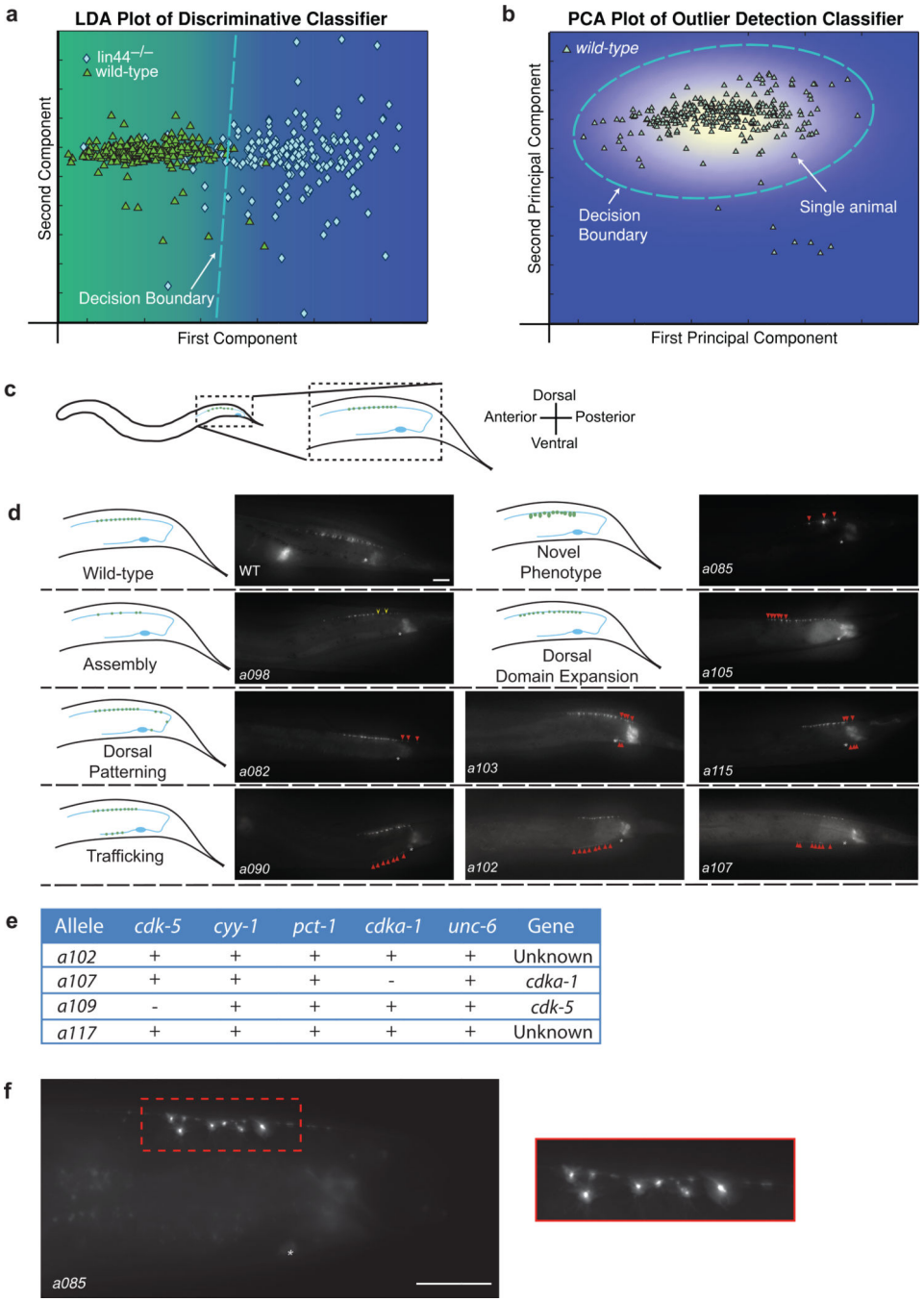


Figure 1.

Integrated system for autonomous screening of fluorescent reporters. (a) Schematic of the microfluidic device used to automate sample handling of a mutagenized *C. elegans* population and allow imaging and sorting (scale bar 150 μm). (b–d) The computer vision framework to identify the fluorescent reporter in a low signal-to-noise environment: (b) Maximum projection of a representative wild-type animal acquired in the device (scale bar 20 μm). (c) Computer vision framework applied to identify the objects of interest (synapses). For each pixel in (b) local features and used to predict the probability that a pixel is a

synapse; for high probability locations the spatial relationship between potential synapses is used to distinguish between autofluorescence and the signal of interest. (d) The probability that each pixel of the wild-type animal is a synapse (top), and the locations of the identified synapses (bottom)(scale bar 20 μm). (e–h) Statistical framework for quantitative phenotyping and autonomous decision making during screening. (e) Representative images of wild-type and *lin-44*$^{-/-}$ mutants acquired in the device, and the resulting identified synapse locations (scale bar 20 μm). (f) Quantitative phenotypical descriptors extracted from the representative images. These descriptors are used to train the classifier for performing autonomous screens and predicting whether an animal is a mutant.



e

Allele	*cdk-5*	*cyy-1*	*pct-1*	*cdka-1*	*unc-6*	Gene
a102	+	+	+	+	+	Unknown
a107	+	+	+	-	+	*cdka-1*
a109	-	+	+	+	+	*cdk-5*
a117	+	+	+	+	+	Unknown

Figure 2.

Autonomous screens for mutants containing altered synaptic patterns. (a) LDA projection of the phenotypical descriptors from wild-type and *lin-44*$^{-/-}$ animals used to train the discriminative classifier. Classification during screening was performed in the original high-dimensional space using an RBF-kernel SVM. (b) PCA projection of the phenotypical descriptors from wild-type animals used for the outlier detection screen, and showing a representative decision boundary. Classification during screening was performed in the original high-dimensional space. (c) Schematic location of the DA9 neuron within *C.*

elegans. (d) (left) Schematics of the phenotypical classes identified during autonomous screening, including both previously identified and novel phenotypical classes. (right) Representative images of alleles with phenotypes falling into the appropriate categories. (e) Candidate mutants of a single phenocluster, trafficking, were selected for further investigation. Complementarity tests were performed between these new alleles and genes known to cause a similar dendritic puncta phenotype. Two alleles complemented with genes known to act within this pathway while two alleles failed to complement with any of the known genes, suggesting the discovery of additional players within the pathway. (f) A mutant (*a085*) showing a novel phenotype of enlarged spine-like protrusions (scale bars 20 μm).