PNAS
www.pnas.org

Supplementary information for

Systematic phenomics analysis of Autism-associated genes reveals parallel networks underlying reversible impairments in habituation

Troy A. McDiarmid, Manuel Belmadani, Joseph Liang, Fabian Meili, Eleanor A. Mathews, Gregory P. Mullen, Ardalan Hendi, Wan-Rong Wong, James B. Rand, Kota Mizumoto, Kurt Haas, Paul Pavlidis, and Catharine H. Rankin

Corresponding author: Catharine H. Rankin
Email: crankin@psych.ubc.ca

This PDF file includes:

www.pnas.org/cgi/doi/10.1073/pnas.1912049116

Materials and Methods

Ortholog identification and strain selection

C. elegans orthologs of human ASD-associated genes were identified by querying OrthoList using Ensembl gene IDs, as previously described (1). **Dataset S1 and S2** describe all Orthology relationships used in this study. During the course of this project The Alliance of Genome Resources (https://www.alliancegenome.org/) created a web tool that allows for identification of the "best" matched Ortholog, defined as the ortholog predicted in the queried species by the highest number of gold-standard algorithms. The vast majority of the *C. elegans* orthologs predicted by OrthoList (85%) are also defined as the best ortholog by this new tool (**Dataset S1**). OrthoList was also updated to OrthoList2 during the course of this work (2). Again there is large agreement between our orthology predictions and those made using the recently updated OrthoList2 (95% of OrthoList relationships are supported by OrthoList2; **Dataset S1**, note that 48 of the 135 strains were predicted later than the rest using the updated OrthoList2 and Alliance of Genome Resources tools). In situations where multiple human ASD-associated genes share a single *C. elegans* ortholog predicted by OrthoList the single *C. elegans* gene was used to study the larger vertebrate family, as has been done previously (3–5). For example, *nlg-1* and *shn-1* are the sole *C. elegans* orthologs of all vertebrate neuroligin and shank family proteins, respectively (**Dataset S1**). Note that throughout the manuscript the "•" symbol is used to represent the human-to-*C. elegans* orthology relationship of interest, e.g. *GAPVD1•rme-6(b1014)*.

Strains harboring mutations in orthologs of ASD-associated genes were identified using WormBase and ordered from the *Caenorhabditis* Genetics Center (CGC), the National BioResource Project of Japan (NBRP), or received from a collaborator following a formal request. Strains carrying putative null and loss-of-function alleles were prioritized (6). In many cases multiple strains harboring distinct loss-of-function alleles were characterized. Where such alleles were not available, or where null alleles are known to result in lethality or fecundity defects severe enough to impede high-throughput characterization in our assay a strain carrying a known or predicted deleterious missense mutation was characterized instead. The complete list of all strains and alleles assessed in the initial large-scale characterization are described in **Dataset S1**. It is important to note that in *C. elegans* a single, heavily characterized wild-type strain (N2) is used across laboratories.

Animals

Strains were maintained on NGM (nematode growth medium) plates seeded with *Escherichia coli* strain OP50 according to standard experimental procedures. 96h post-hatch hermaphrodite animals were used for all experiments.

Microbe strains

The *E. coli* OP50 strain was used as a food source for *C. elegans*.

Behavioral assays

For the mechanosensory habituation paradigm animals were cultured on NGM seeded with *E. coli* (OP50) and age synchronized for behavioral tracking as described previously (7–9). Animals were synchronized for behavioral testing on Petri plates containing NGM seeded with 50 µl of OP50 liquid culture 12-24 hours before use. Five gravid adults were picked to plates and allowed to lay eggs for 3-4 hours before removal. For all Multi-Worm Tracker experiments 3-6 plates (20-100 worms/plate) were run for each strain on each testing day. The animals were maintained in a 20°C incubator for 96 hours.

Our behavioral paradigm (**Fig. S1*E***) consisted of a 5-minute period to recover from being placed on the tracker followed by a 5 min baseline period from which we computed multiple measures of morphology and baseline locomotion. Beginning at 10 minutes we administered 30 mechanosensory stimuli to the Petri plate holding the animals at a 10 second interstimulus interval (ISI) using an automated push solenoid (**Fig. S1*A* and *E***). Animals respond to a mechanosensory stimulus by emitting a reversal response (crawling backwards) allowing us to assess multiple measures of naïve sensitivity (e.g. reversal likelihood, duration, etc.) (**Fig. S1 *E-I***). With repeated stimulation there is a decrease in the likelihood of a reversal, as well as the duration, speed, and distance of reversals (habituation). Following habituation training, we allowed a 5 minute recovery period after which we administered a 31st stimulus to gauge spontaneous recovery from short-term habituation - an assay of short-term memory retention (**Fig. S1 *E-I***).

Standard sensory integration, spontaneous reversal, octanol avoidance, and thermotaxis

assays were conducted as previously described (10, 11). Briefly, in order to assess thermotaxis, thermal gradients were established by placing a vial of frozen glacial acetic acid (16.7°C) in the center of inverted 10 cm unseeded plates kept in a 25°C incubator (12). Approximately 50 worms were transferred to each thermal gradient plate and allowed to move freely on the gradient for 30 minutes, after which their positions were scored on an overlay of concentric circles demarking eight equal areas on the plate. Octanol chemotaxis plates were prepared by placing 2 µl of 0.1% octanol on one side of a 10 cm plate (side 'A') and 2 µl of diluent on the other side of the plate (side 'B'). Approximately 50 worms were transferred to the center of each assay plate and allowed to move freely for 20 minutes and then scored for location. Only worms within 2 cm of the test or control spots were scored, and therefore worms in neutral areas were disregarded. A chemotaxis index (C.I.) was calculated using the formula C.I.= (A–B)/(A+B). A positive C.I. value (up to 1.00) indicates attraction, a negative value (down to −1.00) indicates repulsion, and a value near 0.00 indicates neutrality. Spontaneous reversals were measured by transferring worms to unseeded plates, waiting two minutes for equilibration, and visually scoring for changes in direction for 10 minutes. To test the integration of two conflicting (attractive and repulsive) sensory stimuli a repellent barrier consisting of 50 µl of 0.5 mM cupric acetate was pipetted in a line across the middle of a 10 cm plate (13). After approximately 16 hours, a chemo-attractant (2 µl of 0.1% diacetyl in ethanol) was placed on one side of the barrier, and approximately 75 worms were transferred to the plate on the opposite side of the barrier. The worms were allowed to move freely on the plate for 30 minutes, at which point the fraction of worms that crossed the barrier was scored.

Multi-Worm Tracker behavioral analysis and statistics

Multi-Worm Tracker software (version 1.2.0.2) was used for stimulus delivery and image acquisition (14). Phenotypic quantification with Choreography software (version 1.3.0_r103552) used "--shadowless", "--minimum-move-body 2", and "--minimum-time 20" filters to restrict the analysis to animals that moved at least 2 body lengths and were tracked for at least 20 s. Standard choreography output commands were used to output morphology and baseline locomotion features (14). A complete description of the morphology, baseline locomotion, sensory, and habituation learning features can be found in the Multi-Worm Tracker user guide (https://sourceforge.net/projects/mwt/) (14) and ***SI Appendix,* Table S1.** The MeasureReversal plugin was used to identify reversals occurring within 1 s (dt = 1) of the

mechanosensory stimulus onset. Comparisons of "final response" comprised the average of the final three stimuli. Custom R scripts organized and summarized Choreography output files. No blinding was necessary because the Multi-Worm Tracker scores behavior objectively. For the initial large-scale characterization (**Figs. 1-3**), features were pooled across plate replicates for each mutant strain and means were compared to the mean of the wild-type distribution with an unpaired t-test and Benjamini-Hochberg control of the false discovery rate at 0.1. Final figures were generated using the ggplot2 package in R (15). For targeted confirmation and follow-up analyses (**Figs. 4-6**) responses were pooled across plates and compared across strains using binomial logistic regression for habituation or one-way ANOVA for additional behavioral assays (**Fig. S9**) with Tukey's honestly significant difference (HSD) criterion as previously described (9, 16). Each targeted confirmation and follow up experiment was independently replicated at least twice. Alpha values of 0.001 or 0.0001 were used to determine significance for Logistic regression and one-way ANOVA statistical tests respectively. Final figures were generated using the ggplot2 package in R. For all Multi-Worm Tracker experiments 3-6 plates (20-100 worms/plate) were run for each strain on each testing day. Sample sizes for each behavioral assay were chosen to be either equal to or greater than sample sizes reported in the literature that were sufficient to detect biologically relevant differences.

All raw and processed data, and the results of all statistical tests can be found at (https://doi.org/10.5683/SP2/FJWIL8), all analysis code is freely available at (https://github.com/PavlidisLab/McDiarmid-etal-2019_Multi-Worm-Tracker-analysis).

Clustering analyses

T-statistic's were used as numerical scores to represent the difference between wild-type and mutant animals for each phenotypic feature; this created a numerical profile of phenotypic features for further analysis. All clustering analyses were performed in R. Correlation distributions were visualized using ggplot2 (15). Average-linkage hierarchical clustering was performed with pvclust (17) using correlation as the distance measure, and 50,000 rounds of bootstrapping. t-SNE clustering was performed using the Rtsne package (18) with hyperparameters: initial principal component analysis = TRUE, perplexity = 10, and theta = 0 for the final 2D and 3D visualizations. The dendrogram and heat maps were visualized with pheatmap (19) for the correlation matrix heat map, iheatmapr (20) for the phenomic profile

heat maps, and pvclust for the dendrograms.

CRISPR-Cas9 missense variant lines

The *chd-7* and *nlg-1* missense variant lines were generated using CRISPR-Cas9 genome editing as previously described (21). The following missense variants strains were characterized:

PS7293 *chd-7(sy861*[G1225S]) *I*
PS7317 *chd-7(sy855*[L1487R]) *I*
PS7316 *chd-7(sy854*[R2624Q]) *I*
PS7318 *chd-7(sy859*[L1220P]) *I*
PS7267 *chd-7(sy1049*[P253L]) *I*
PS7711 *nlg-1(sy959*[G59R]) *X*
PS7713 *nlg-1(sy961*[A268T]) *X*

List of strains generated in this work

The following strains were created for this work either via standard genetic crosses for double mutants, or via microinjection of plasmid DNA for the generation of extrachromosomal array or CRISPR-Cas9 genome engineered transgenic lines:

VG870-871 *chd-7(gk306) I; nlg-1(ok259) X*
VG872-873 *chd-7(gk306) I; rme-6(b1014) X*
RM3707 *nlg-1(ok259) X* - 6x outcrossed
RM3249 *pha-1(e2123ts) III*; *nlg-1(ok259) X*
RM3540 *pha-1(e2123ts) III; nlg-1(ok259) X; mdEx1035[Pnlg-1::NLG-1(R62W)::YFP; pBX]*
RM3516-17 *pha-1(e2123ts) III; nlg-1(ok259) X; mdEx1016-1017[Pnlg-1::NLG-1(V397M)::YFP; pBX]*
RM3536 *pha-1(e2123ts) III; nlg-1(ok259) X; mdEx1033[Pnlg-1::NLG-1(R430C)::YFP; pBX]*
RM3537 *pha-1(e2123ts) III; nlg-1(ok259) X; mdEx1034[Pnlg-1::NLG-1(R714C)::YFP; pBX]*
VG880 *nlg-1(yv15[nlg-1p::nlg-1::AID::GFP:: + LoxP pmyo-2::GFP::unc-54 UTR prps-27::NeoR::unc-54 UTR LoxP + nlg-1 3' UTR])*
VG881 *nlg-1(yv16[nlg-1p::nlg-1::AID::GFP:: + LoxP + nlg-1 3' UTR])*

VG890-891 nlg-1(yv16[nlg-1p::nlg-1::AID::GFP:: + LoxP + nlg-1 3' UTR]); ieSi57[eft-3p::TIR1::mRuby::unc-54 3' UTR + cbr-unc-119(+)] II; unc-119(ed3) III

Strain and plasmid generation

The neuroligin missense variant plasmids were constructed by performing standard site-directed mutagenesis on our previously described *nlg-1* YFP functional fusion protein construct derived from the yk497a9 cDNA (10).

The Moerman lab guide selection tool (http://genome.sfu.ca/crispr/) was used to identify the *nlg-1* targeting sgRNA (22). The *nlg-1* sgRNA sequence: TCACCAACGTGTCCACGTCA was cloned into the *pU6::klp-12* sgRNA vector (obtained from Calarco lab) using site-directed mutagenesis and used for all editing experiments. The *nlg-1::AID::GFP::nlg-1* upstream and *nlg-1 3' UTR* downstream homology arms were synthesized by IDT and cloned into the loxP_myo2_neoR repair construct (obtained from Calarco lab) using Gibson Assembly.

C. elegans wild-type N2 strain was used for all CRISPR-Cas9 editing experiments. Genome edits were created as previously described (23). In brief, plasmids encoding sgRNA, Cas9 co-transformation markers pCFJ90 and pCFJ104 (Jorgensen lab, Addgene) and the selection cassette flanked by homology arms (~500 bp) containing *AID::GFP* were injected into wild-type animals. Animals containing the desired insertions were identified by G418 resistance, loss of extrachromosomal array markers, and uniform dim fluorescence of the inserted GFP.

Genotype confirmation

Correct insertion of the *GFP::AID sequence* was confirmed by amplifying the two regions spanning the upstream and downstream insertion borders using PCR followed by Sanger sequencing (primer binding locations depicted in **Fig. 6*A***). The genotyping strategy is essentially as described for deletion allele generation via DMS cassette insertion in Norris et al. (2015) (23, 24).

The forward and reverse primers used to amplify the upstream insertion region were GAAGTTTCCAAATGGTCGTAGAAC (located within the *nlg-1* genomic locus) and CGAGAAGCATTGAACACCATAAC (located within GFP in the selection cassette) respectively.

The forward and reverse primers used to amplify the downstream insertion region were TTCCTCGTGCTTTACGGTATCG (located within the Neomycin resistance gene) and GGTAGCTTGATTCGCCTTCTAT (located downstream of the *nlg-1* genomic locus) respectively.

Following cassette excision via injection of cre-recombinase the *nlg-1* genomic locus (GAAGTTTCCAAATGGTCGTAGAAC)
and *nlg-1* downstream
(GGTAGCTTGATTCGCCTTCTAT)
primers were used to amplify and confirm error free insertion of the *AID::GFP sequence* at the *nlg-1* locus via Sanger sequencing (**Fig. 6*A***).

Auxin administration

Auxin administration was performed as previously described (25). Auxin treatment was performed by transferring animals to bacteria-seeded plates containing auxin. Auxin indole-3-acetic acid (IAA) (Thermo Fisher, Alfa Aesar™ #A1055614) was dissolved in ethanol to create a 400 mM stock solution. The stock solution was stored at 4°C for up to one month. Auxin was diluted into the molten NGM agar (cooled to ~50°C before Auxin addition) before pouring plates. Auxin plates were seeded with 50 µl of OP50 liquid culture 12-24 hours before use. For continuous exposure groups, animals were age synchronized (as described above) for behavioral testing on auxin plates and tested at 96 hours old. For developmental exposure animals were age synchronized and reared on auxin plates for 48 hours before being transferred to standard OP50 seeded NGM plates, and then tested 48 hours later (96 hours old). For adult auxin treatment groups animals were age synchronized on standard OP50 seeded NGM plates and reared for 48 hours before being transferred to auxin plates and then tested 48 hours later (96 hours old).

Confocal imaging

Adult animals were anesthetised on glass microscope slides in 5 mM Levamisole and 150mM BDM (2,3-butanedione monoxime) dissolved in M9 buffer and covered with a 1.5 coverslip. A Leica SP8 white light laser confocal microscope and 63× oil immersion lens was used for imaging. Step size was 0.3 µm. GFP was excited using a 488 nm wavelength laser with emitted light collected through a 493–582 nm bandpass filter. YFP was excited using a 506 nm wavelength laser with emitted light collected through a 511-600 nm bandpass filter. mRuby was excited using a 587 nm wavelength laser with light collected through a 592-779 nm bandpass filter. Final figures were generated using ImageJ (National Institutes of Health, Bethesda, MD). For **Fig. S8** NLG-1::AID::GFP fusion proteins were captured in live *C. elegans* using a Zeiss LSM800 confocal microscope (Carl Zeiss, Germany). Worms were immobilized on 2% agarose pad using a mixture of 7.5 mM levamisole (Sigma-Aldrich) and 0.225M BDM (2,3-butanedione monoxime) (Sigma-Aldrich). Images were analyzed with Zen software (Carl Zeiss). The Dorsal nerve cord was highlighted using 'Segmented Line', Edit, Selection, Straighten. The dorsal nerve cord of each individual channel (Pflp-13::mCherry::rab-3 and NLG-1::AID::GFP) were straightened and cropped using ImageJ (NIH, Bethesda, MD).

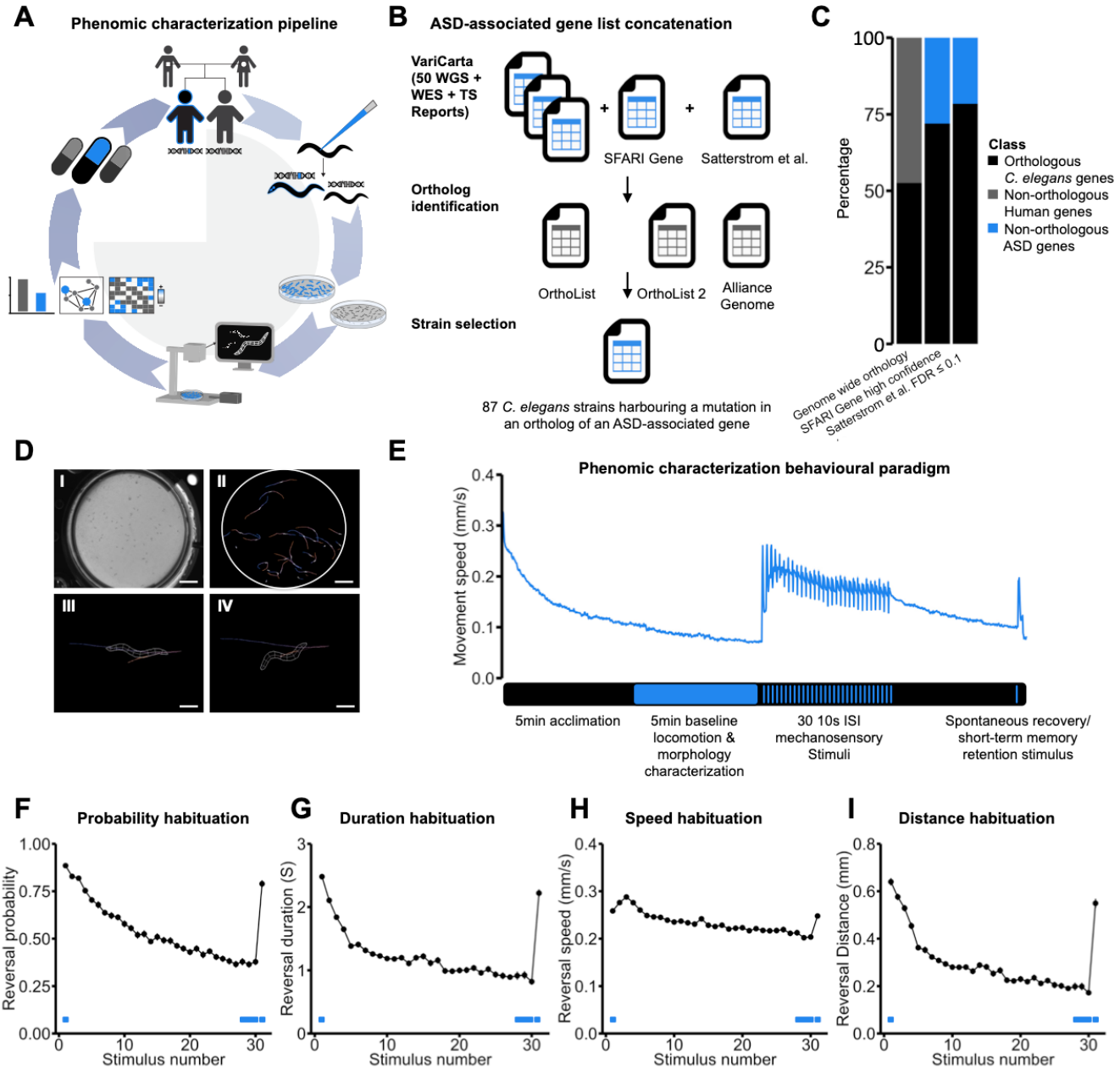

Fig. S1. ASD-associated gene ortholog identification and phenomic characterization pipeline. (A) Animals carrying mutations in orthologs of ASD-associated genes are systematically engineered using CRISPR-Cas9 or ordered from stock centers, large populations of are grown and their phenotypic profiles are characterized using The MWT. Genotype-phenotype relationships are used to cluster strains based on phenotypic similarity, establish variant functional assays, and test phenotypic reversibility. These data can serve to facilitate precision medicine drug screens in future work. (B) Databases used during risk gene list concatenation and ortholog identification. Whole-Genome Sequencing (WGS), Whole-

Exome Sequencing (WES), Targeted Sequencing (TS). (C) A greater proportion of high-confidence ASD-associated genes have *C. elegans* orthologs (blue, 72% 18/25 SFARI Gene, 78% 80/102 Satterstrom et al. 2019) compared to all human genes (gray, 53% 10,678/23,010). (D) The MWT delivers stimuli and performs image acquisition, object selection, and parameter selection in real time while choreography software extracts phenotypic information offline. (panels) I) Petri plate of *C. elegans* II) *C. elegans* selected for analysis III) Digital representation illustrating phenotypic detail. An example behavior scored by the MWT: the *C. elegans* response to a mechanosensory tap to the side of the plate is brief backwards locomotion (from III to IV). Scale bars: 1cm, 1cm, 0.25mm, 0.25mm from I-IV. (E) Phenomic characterization paradigm plotted alongside a single phenotype, absolute movement speed. Following a 5min acclimation phase a further 5min period is recorded from which multiple measures of morphology and baseline locomotion are extracted. Beginning at 10min 30 mechanosensory stimuli are delivered at a 10s ISI to which the animals initially reverse and then habituate, allowing for assessment of multiple measures of initial sensitivity and learning. This is followed by a 5min recovery before administering a 31st stimulus to assess spontaneous recovery from habituation/short-term memory retention. F-I) Multiple measures of habituation of the same reversal responses exhibit different extents and rates of response decrement. Data shown as mean±s.e.m using number of plates as n.

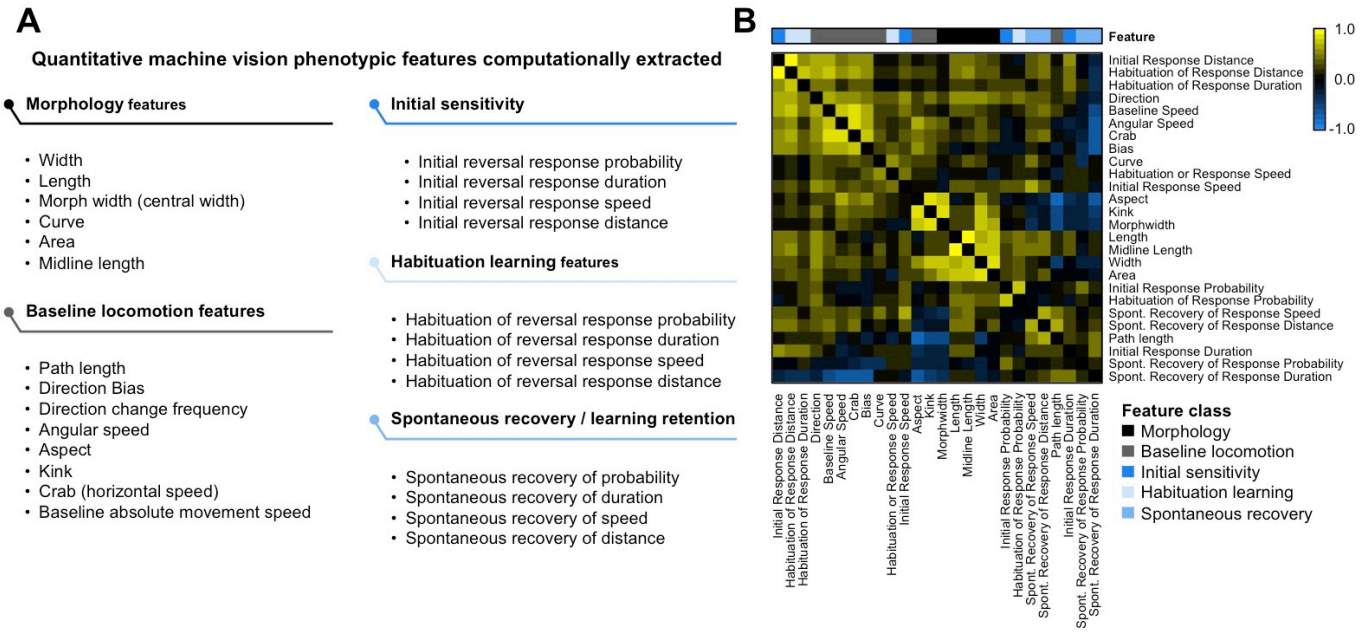

Fig. S2. Phenotypic features display varied moderate correlations that correspond to feature subclasses. (A) List of quantitative machine vision phenotypic features computationally extracted and their predefined subclasses. (B) Hierarchically clustered correlation matrix illustrating varied moderate correlations between features. Pearson's r is shown.

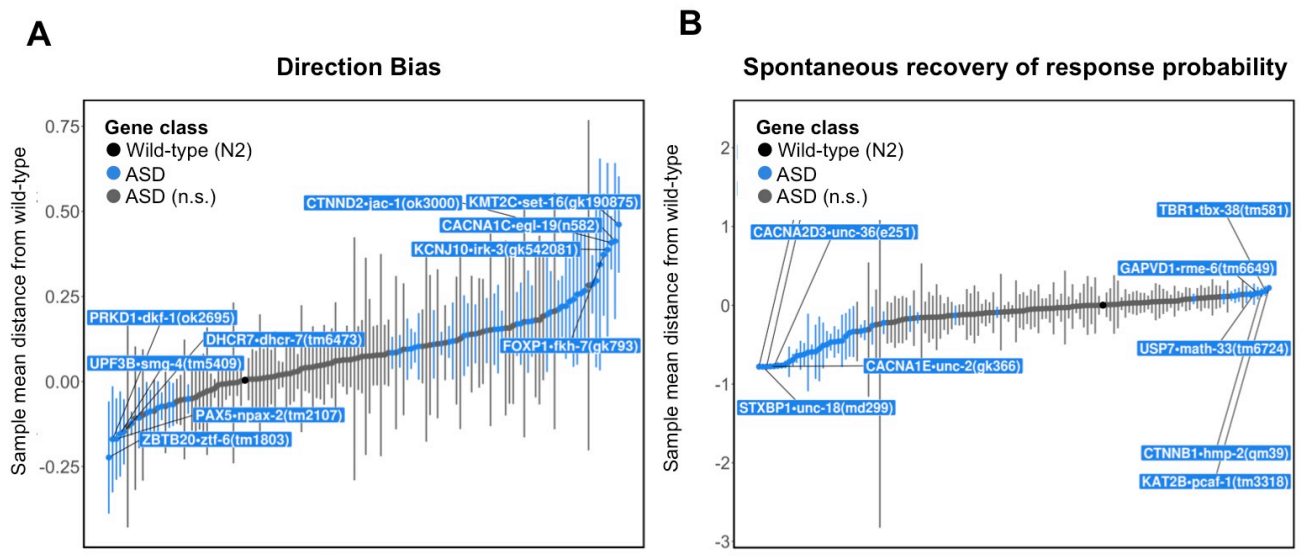

Fig. S3. Quantitative phenotypic profiles enable rapid reverse genetic screens to identify shared functions among ASD-associated genes. All plots illustrate the sample mean distance of each genotype group from wild-type. Strains outside the 95% confidence interval of the wild-type distribution are labeled and colored blue. Only a maximum of five strains are labeled in each direction per feature to prevent overplotting. (A) Movement direction bias by genotype. (B) Spontaneous recovery of response probability by genotype. Distinct partially overlapping sets of ASD-associated gene orthologs alter initial sensitivity, habituation learning, and spontaneous recovery/memory retention, indicating genetically dissociable underlying mechanisms. Error bars represent 95% confidence intervals.

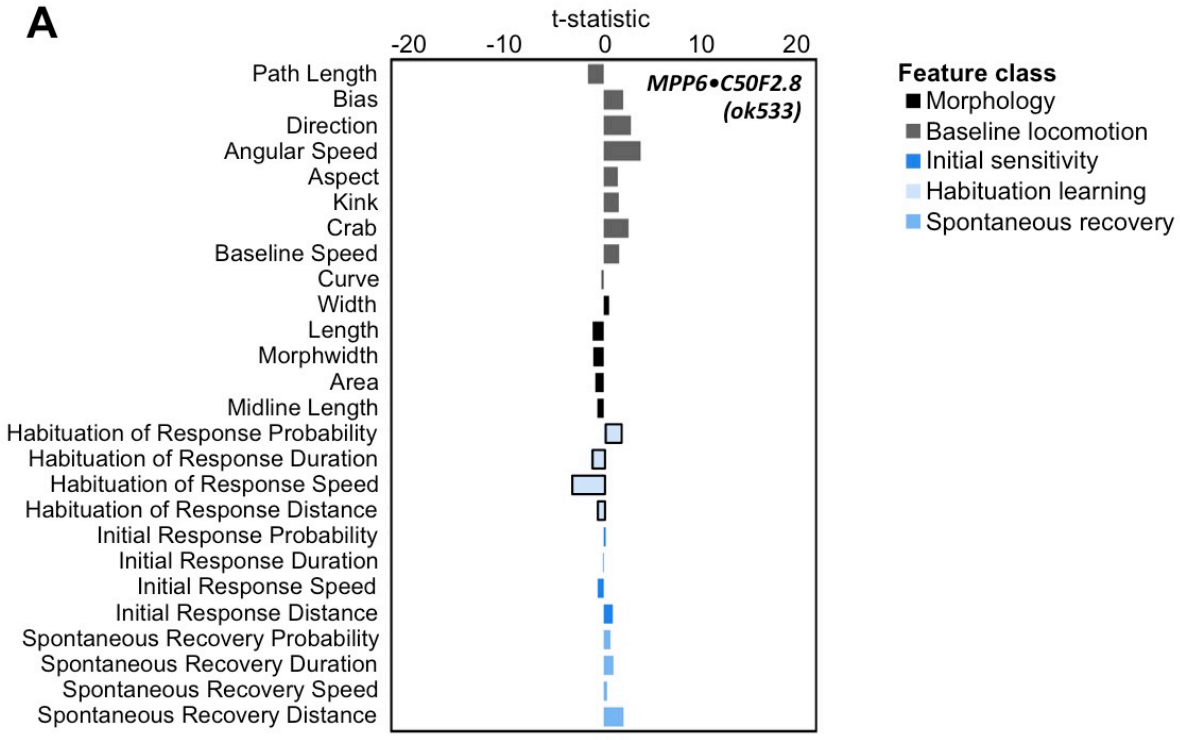

Fig. S4. Phenotypic profiles of strains with mutations in ASD-associated genes define shared and unique functions and phenotypic modularity. Bars represent directional t-statistics for each phenotypic feature listed across the y-axis. Color coding reflects feature classification. (A) Phenotypic profile for *MPP6•C50F2.8(ok533)* exhibiting relatively minor phenotypic alterations.

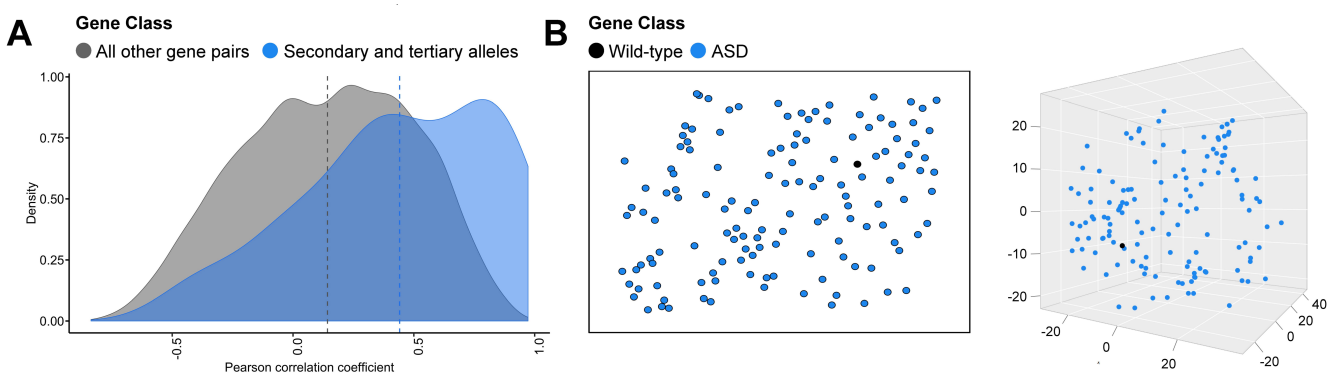

Fig. S5. Phenotypic correlations and t-SNE clustering of 135 strains harboring mutations in ASD-associated gene orthologs. (A) Density plot illustrating the distribution of overall phenotypic profile Pearson correlation coefficients between independent alleles in the same gene (blue) and all other possible gene pairs (gray). (B) 2D (left) and 3D (right) t-SNE plots illustrating the distance between the 135 strains harboring mutations in ASD-associated gene orthologs (blue) and wild-type (black).

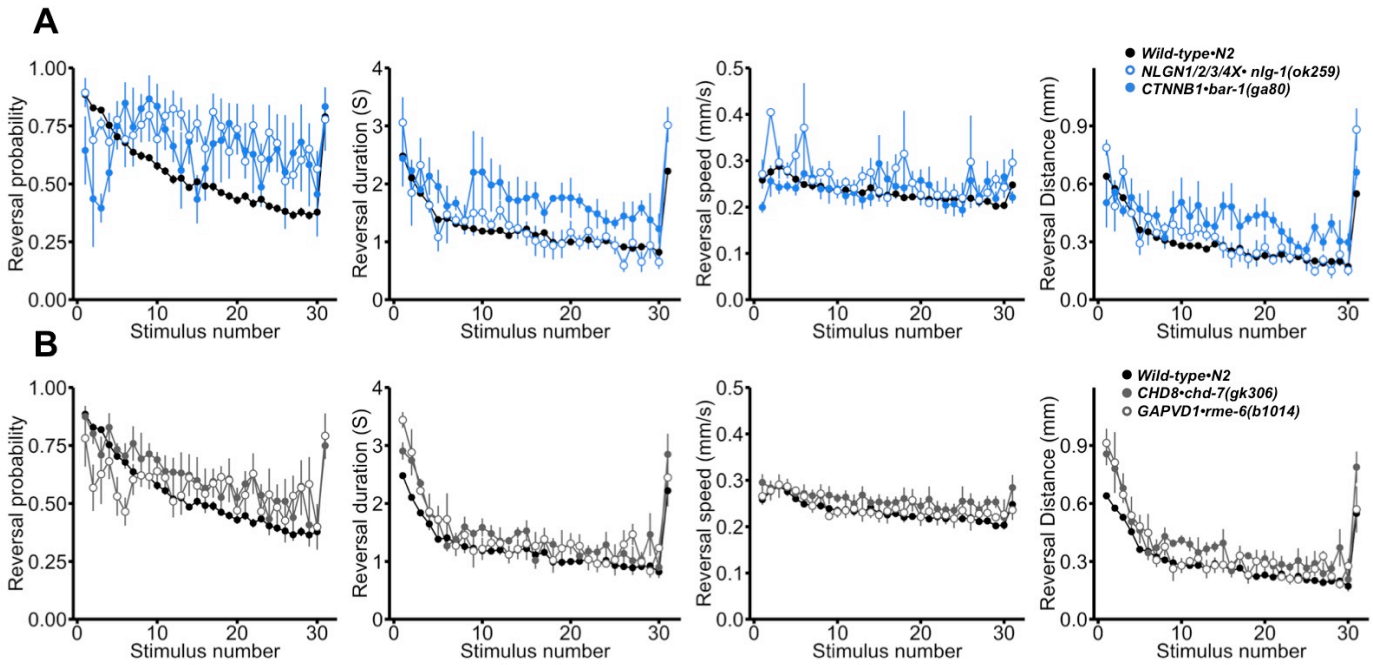

Fig. S6. Sensory and learning phenotypic profiles of cluster members. (A) Sensory and learning phenotypic profile for NLGN1/2/3/4X•nlg-1(ok259) and CTNNB1•bar-1(ga80) mutants. Each point represents a reversal response. Data are shown as mean±s.e.m using the number of plates as n. (B) Sensory and learning phenotypic profile for CHD8•chd-7(gk306) and GAPVD1•rme-6(b1014) mutants. Each point represents a reversal response. Data are shown as mean±s.e.m using the number of plates as n.

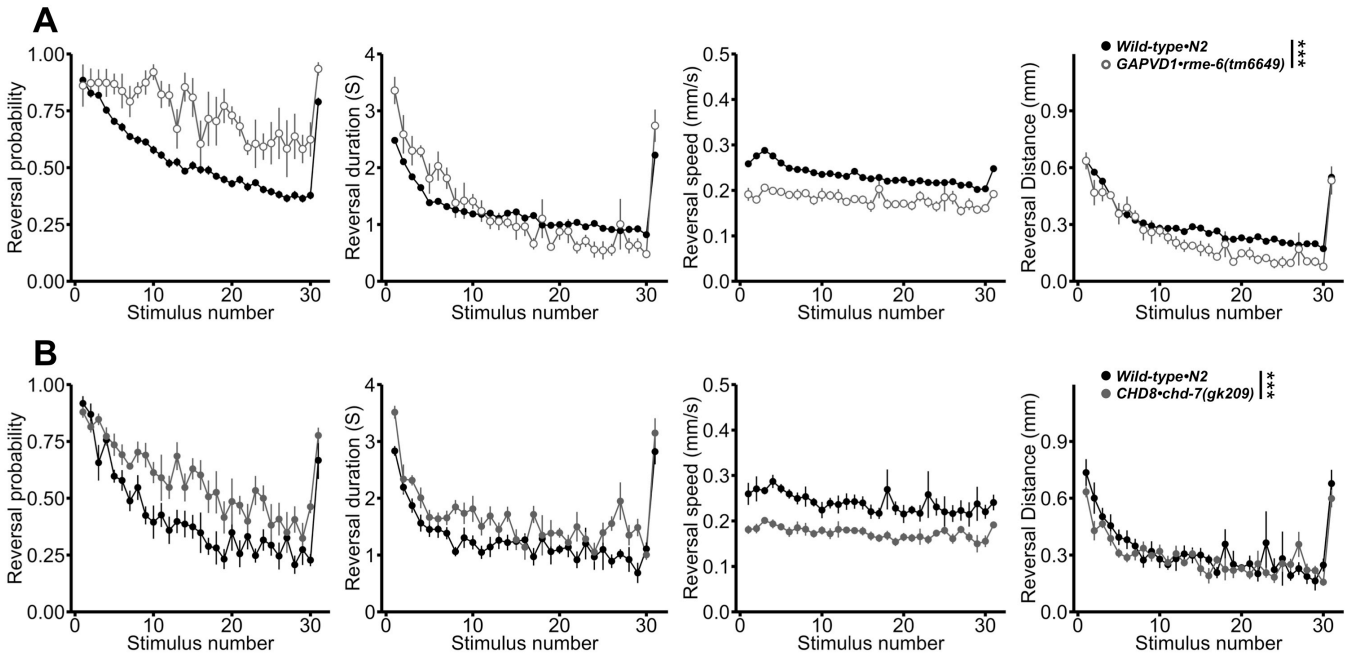

Fig. S7. Second alleles of *GAPVD1•rme-6(tm6649)* and *CHD8•chd-7(gk209)* also display increased initial reversal response duration and impaired habituation of response probability. (A) Sensory and learning phenotypic profile for *GAPVD1•rme-6(tm6659)* and (B) *CHD8•chd-7(gk209)* mutants. Data are shown as mean±s.e.m using the number of plates as n. ***p<0.001, binomial logistic regression followed by Tukey's HSD criterion was used to determine significance of the habituated level (proportion reversing at tap 30) for each pair of strains, n.s., not significant.

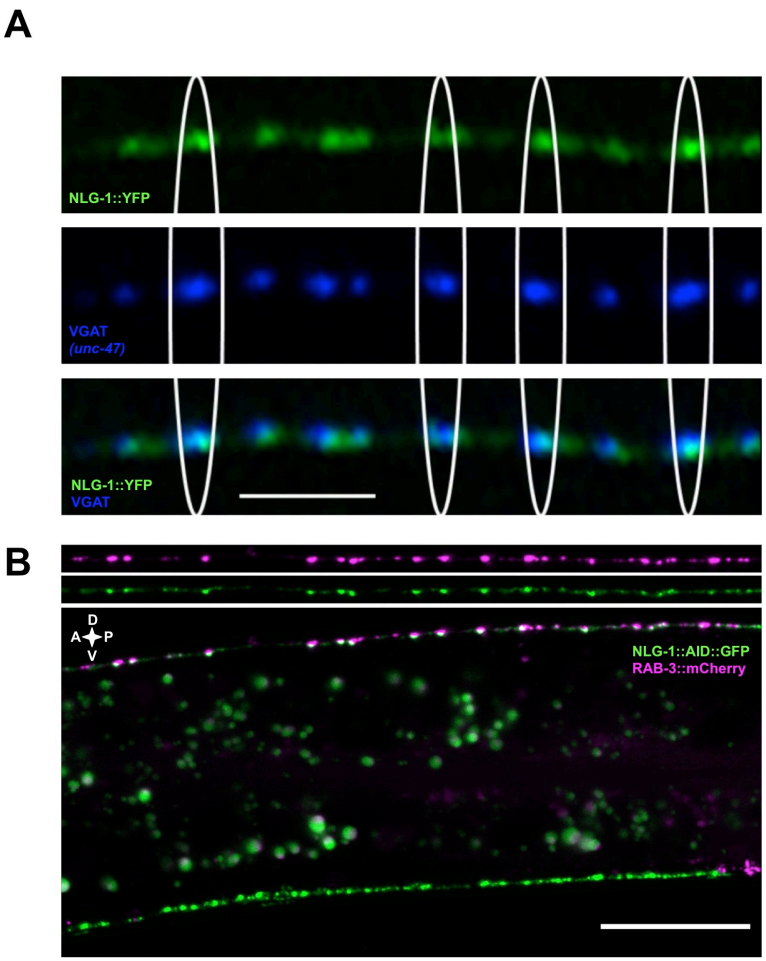

Fig. S8. Synaptic localization of neuroligin transgenes. (A) Immunofluorescent staining of *nlg-1(ok259)* animals expressing a NLG-1::YFP transgene driven by a muscle-specific promoter, illustrating apposition of the post-synaptic NLG-1::YFP with the presynaptic vesicular GABA transporter (VGAT, a marker of GABAergic synapses). Shown is the dorsal nerve cord of a young adult stained with antibodies to GFP, which also recognize YFP (top, green), VGAT (middle, blue), and the merged image (bottom), scale bar = 1 µm. (B) The NLG-1::AID::GFP endogenous transgene is apposed to a RAB-3::mCherry synaptic vesicle marker driven by an *flp-13* promoter in DD class GABAergic neurons (DD5 output region is shown), scale bar = 0.02 mm.

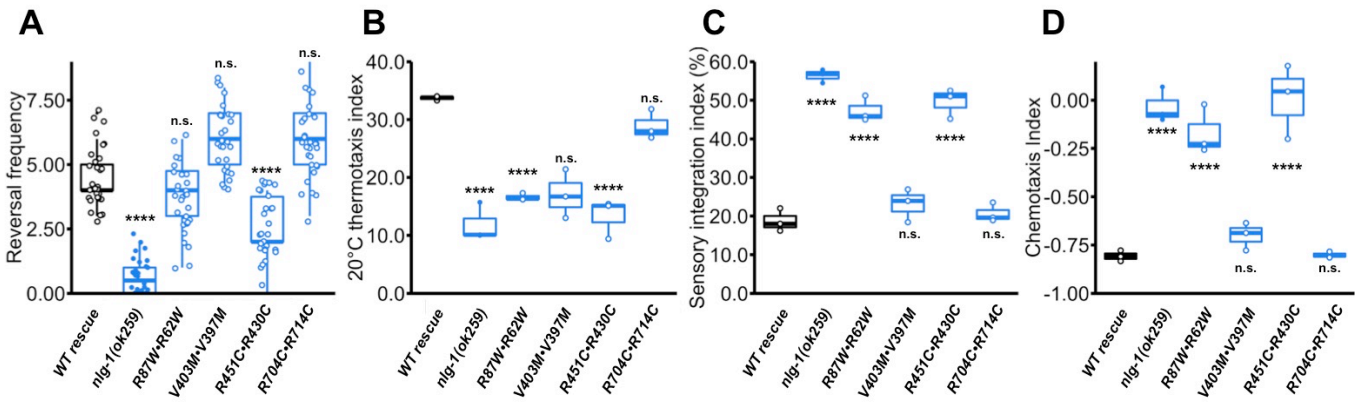

Fig. S9. Additional functional assays for ASD-associated neuroligin missense variants. (A-D) Each ASD-associated neuroligin variant was scored for its ability to rescue spontaneous reversal frequency (A), thermotaxis (B), sensory integration (C), and octanol avoidance chemotaxis (D) in the nlg-1(ok259) deletion mutant, again revealing varied partial loss-of-function. Dots represent individual worm reversal frequencies (A) or plate means (B-D), horizontal lines represent median of replicates. (A-D) ****p<0.0001, one-way ANOVA followed by Tukey's HSD criterion was used to determine significance, n.s., not significant.

Captions for Datasets S1 and S2

Dataset S1. Genes phenotyped. Contains the complete list of ASD risk gene orthologs phenotyped, orthology prediction software results, the specific alleles phenotyped, the outcross status of phenotyped strains, predicted molecular function of the genes from Satterstrom et al., and measures of ASD-association confidence from SFARI Gene and Satterstrom et al.

Dataset S2. Satterstrom et al. ASD risk gene to *Caenorhabditis elegans* orthology. Contains predicted *C. elegans* orthologs for ASD risk genes reported by Satterstrom et al., whether there is an assay-suitable ortholog deletion or severe missense allele available, and whether the ortholog has been characterized in the present work.

Table S1. Phenotypic feature descriptions.

Feature	Description
Width	The distance spanned perpendicular to the major axis, in mm.
Length	The distance spanned by objects along their major axis (defined to be the axis of a least squares fit), in mm.
Morph Width	The mean width of the central 60% of the object, measured perpendicular to the skeleton. This is only available if the object has a skeleton.
Curve	The average angle, in radians, of points along the object's skeleton taken two apart. This measure is only available if skeletons exist.
Area	The area in square millimeters of objects
Midline Length	The length of the object, in mm, measured along the skeleton; this is only available if skeletons were collected (or outlines were collected, and skeletons were generated with a plugin).
Path Length	The distance that the animal has traveled. This is both signed (forwards is positive, backwards is negative) based on the results of the “bias” output and cumulative. If movement direction cannot be computed, then the length starts again from zero.
Bias	The fraction of objects moving in the dominant movement direction as opposed to the other direction. The dominant direction is labeled 1, the opposite 1, and stationary objects are 0. This is usually inaccurate unless the segment (S) option is used.
Direction	The frequency of changes in direction. Forward motion is 1; motion that backtracks is 1 at the point of turning. This metric is somewhat unreliable unless segment (S) is used.
Angular Speed	The angular speed in radians/second of objects; this is calculated over the same interval as speed, but reports the greatest difference in angle between primary axes over that time
Aspect	The ratio of the width to length

Kink	The angle in radians between the line from the first to third point of the skeleton and the fourth through last points; or the angle between first and fourth to last and third to last to last, whichever is greater. This is similar to the MWT endwiggle measure.
Crab	The speed perpendicular to the length of the body, in mm/sec, averaged as specified by s. Note that the speed along the length of the body is sqrt(speed*speed – crab*crab).
Baseline Speed	The speed in mm/second of objects. This speed is averaged over the duration given in the s parameter; it is equal to the distance between the most widely spaced pair of points over that period of time (half ahead of and half behind the current timepoint) divided by that duration.
Initial Reversal Response Probability	The proportion of animals reversing to the first mechanosensory stimulus.
Initial Reversal Response Duration	The duration of reversal responses to the first mechanosensory stimulus
Initial Reversal Response Speed	The speed of reversal responses to the first mechanosensory stimulus
Initial Reversal Response Distance	The distance (mm) of reversal responses to the first mechanosensory stimulus
Habituation of Response Probability	The difference between the initial reversal response probability and the final habituated level of response probability (average of 28-30th responses).
Habituation of Response Duration	The difference between the initial reversal response duration and the final habituated level of response duration (average of 28-30th responses).
Habituation of Response Speed	The difference between the initial reversal response speed and the final habituated level of response speed (average of 28-30th responses).
Habituation of Response Distance	The difference between the initial reversal response distance and the final habituated level of response distance (average of 28-30th responses).

Interactive Supplementary Figure Markdown Captions

We have included portions of our supplemental Material as R Markdown .html documents to allow for interactive exploration that is necessary for readers' full understanding and to facilitate future use. These can be downloaded from the article page at the PNAS journal website, opened, and viewed by any Internet browser, similar to a supplemental dataset or video file. The captions provided below also appear in the R Markdown documents.

Dataset S3. Reverse genetic screens. All plots illustrate the sample mean distance of each genotype group from wild-type. Strains outside the 95% confidence interval of the wild-type distribution are labeled and colored blue. Only a maximum of ten strains are labeled in each direction per feature to prevent overplotting.

Dataset S4. Phenotypic profiles. For all plots bars represent directional t-statistics from an unpaired t-test comparing the indicated mutant to wild-type for each phenotypic feature listed across the x-axis. Color coding reflects feature classification.

Dataset S5. Phenomic heat maps. Phenomic heat maps summarizing the phenotypic profiles of 135 strains harboring a mutation in an ortholog of an ASD-associated gene. Cells represent directional t-statistics from comparisons to wild-type controls. T-statistics are shown unclipped and at various clippings (t clipped at ±10, ±20, etc.). On select indicated heat maps, only cells significant at FDR < 0.1 are colored for ease of interpretation. The heat maps are interactive allowing for more detailed inspection of selected observations. Absolute t-statistics values are clipped at 3.0, 10.0 and 20.0 in the last three figures.

Dataset S6. Multiple independent allele confirmation phenomic heat maps. Phenomic heat maps summarizing the phenotypic profiles of multiple strains harboring independent mutations in an ortholog of an ASD-associated gene. Cells represent directional t-statistics from comparisons to wild-type controls. T-statistics are shown unclipped and at various clippings (t clipped at ±10, ±20, etc.). On select indicated heat maps, only cells significant at FDR < 0.1 and with directionally selective genotype-phenotype relationships observed in both strains are colored for ease of interpretation. The heat maps are interactive allowing for more detailed inspection of selected observations. Absolute t-statistics values are clipped at 3.0, 10.0 and 20.0 in the last three figures.

Dataset S7. Reverse genetic screens colored by ASD-association confidence. All plots illustrate the sample mean distance of each genotype group from wild-type. Individual genotype groups are colored by the current evidence for ASD-association of the corresponding gene according to Satterstrom et al. (2019) (see main text for citation). The Satterstrom et al. categories denote evidence for ASD-association from strongest to weakest: FWER ≤ 0.05, FDR ≤ 0.05, FDR ≤ 0.1, None (no categorization). Strains outside the 95% confidence interval of the wild-type distribution are labeled. Only a maximum of five strains are labeled in each direction per feature to prevent overplotting.

Dataset S8. Reverse genetic screens colored by ASD & NDD or ASD predominant categorization. All plots illustrate the sample mean distance of each genotype group from wild-type. Individual genotype groups are colored by ASD & NDD (neurodevelopmental disorders) or ASD predominant categorization of the corresponding gene according to Satterstrom et al. (2019) (see main text for citation). None = no categorization. Strains outside the 95% confidence interval of the wild-type distribution are labeled. Only a maximum of five strains are labeled in each direction per feature to prevent overplotting.

Dataset S9. Pvclust dendrograms. Dendrograms depict hierarchical clustering of strains based on similarity in their phenotypic profiles. T-statistics were used as a numerical score to represent the difference between wild-type and mutant animals for each phenotypic feature; this created a numerical profile of phenotypic features for further analysis. Average-linkage hierarchical clustering was performed with pvclust using correlation as the distance measure, and 50,000 rounds of bootstrapping. Clustering was performed on all features as well as the morphology features only and sensory and learning features only.

References

1. D. D. Shaye, I. Greenwald, OrthoList: A Compendium of C. elegans Genes with Human Orthologs. *PLoS One* **6**, e20085 (2011).
2. W. Kim, R. S. Underwood, I. Greenwald, D. D. Shaye, OrthoList 2: A New Comparative Genomic Analysis of Human and *Caenorhabditis elegans* Genes. *Genetics* **210**, 445–461 (2018).
3. Z. Hu, *et al.*, Neurexin and neuroligin mediate retrograde synaptic inhibition in C. elegans. *Science* **337**, 980–4 (2012).
4. X.-J. Tong, *et al.*, A network of autism linked genes stabilizes two pools of synaptic GABA $_A$ receptors. *Elife* **4**, e09648 (2015).
5. E. Pym, *et al.*, Shank is a dose-dependent regulator of Cav1 calcium current and CREB target expression. *Elife* **6** (2017).
6. C. elegans Deletion Mutant Consortium, Large-Scale Screening for Targeted Knockouts in the *Caenorhabditis elegans* Genome. *G3: Genes|Genomes|Genetics* **2**, 1415–1425 (2012).
7. S. Brenner, The genetics of Caenorhabditis elegans. *Genetics* **77**, 71–94 (1974).
8. E. L. Ardiel, *et al.*, Insights into the roles of CMK-1 and OGT-1 in interstimulus interval-dependent habituation in *Caenorhabditis elegans*. *Proc. R. Soc. B Biol. Sci.* **285**, 20182084 (2018).
9. C. M. Loucks, *et al.*, EFHC1, implicated in juvenile myoclonic epilepsy, functions at the cilium and synapse to modulate dopamine signaling. *Elife* **8** (2019).
10. J. W. Hunter, *et al.*, Neuroligin-deficient mutants of C. elegans have sensory processing deficits and are hypersensitive to oxidative stress and mercury toxicity. *Dis. Model. Mech.* **3**, 366–76 (2010).
11. J. W. Hunter, "Of Autism and Worms: Neuroligin mutants and synaptic function in C. elegans," University of Oklahoma, Oklahoma City. (2011).
12. E. M. Hedgecock, R. L. Russell, Normal and mutant thermotaxis in the nematode Caenorhabditis elegans. *Proc. Natl. Acad. Sci. U. S. A.* **72**, 4061–5 (1975).
13. T. Ishihara, *et al.*, HEN-1, a Secretory Protein with an LDL Receptor Motif, Regulates Sensory Integration and Learning in Caenorhabditis elegans. *Cell* **109**, 639–649 (2002).
14. N. A. Swierczek, A. C. Giles, C. H. Rankin, R. A. Kerr, High-throughput behavioral analysis in C. elegans. *Nat. Methods* **8**, 592–8 (2011).
15. H. Wickham, "Introduction" in *Ggplot2*, (Springer New York, 2009), pp. 1–7.
16. E. L. Ardiel, A. J. Yu, A. C. Giles, C. H. Rankin, Habituation as an adaptive shift in response strategy mediated by neuropeptides. *npj Sci. Learn.* **2** (2017).
17. R. Suzuki, H. Shimodaira, Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540–2 (2006).
18. L. van der Maaten, J. Krijthe, "Package 'Rtsne'. T-Distributed Stochastic Neighbor Embedding using a Barnes-Hut Implementation" (2018) (March 20, 2019).
19. R. Kolde, "Package 'pheatmap'" (2019) (March 20, 2019).
20. A. N. Schep, S. K. Kummerfeld, iheatmapr: Interactive complex heatmaps in R https:/doi.org/10.21105/joss.00359 (March 20, 2019).
21. W.-R. Wong, *et al.*, Autism-associated missense genetic variants impact locomotion and neurodevelopment in Caenorhabditis elegans. *Hum. Mol. Genet.* **28**, 2271–2281 (2019).
22. V. Au, *et al.*, CRISPR/Cas9 Methodology for the Generation of Knockout Deletions in *Caenorhabditis elegans*. *G3: Genes|Genomes|Genetics*, g3.200778.2018 (2018).
23. A. D. Norris, H.-M. Kim, M. P. Colaiácovo, J. A. Calarco, Efficient Genome Editing in Caenorhabditis elegans with a Toolkit of Dual-Marker Selection Cassettes. *Genetics* **201** (2015).
24. D. J. Dickinson, B. Goldstein, CRISPR-Based Methods for Caenorhabditis elegans Genome Engineering. *Genetics* **202**, 885–901 (2016).
25. L. Zhang, J. D. Ward, Z. Cheng, A. F. Dernburg, The auxin-inducible degradation (AID) system enables versatile conditional protein depletion in C. elegans. *Development* **142**, 4374–4384 (2015).