# Natural variation in the *irld* gene family affects insulin/IGF signaling and starvation resistance

Amy K. Webster[1], Rojin Chitrakar[1], Maya Powell[1], Jingxian Chen[1], Kinsey Fisher[1], Robyn Tanny[2], Lewis Stevens[2]†, Kathryn Evans[2], Igor Antoshechkin[3], Erik C. Andersen[2], L. Ryan Baugh[1,4]*

[1]Department of Biology, Duke University, Durham, NC 27708, USA.

[2]Department of Molecular Biosciences, Northwestern University, Evanston, IL, USA.

[3]Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA.

[4]Center for Genomic and Computational Biology, Duke University, Durham, NC 27708, USA.

*Corresponding author. Email: ryan.baugh@duke.edu

†Current address: Tree of Life, Wellcome Sanger Institute, Cambridge, CB10 1SA, UK.

**Abstract:**

Starvation resistance is a fundamental, disease-relevant trait, but the genetic basis of its natural variation is unknown. We developed a synthetic population-sequencing approach to measure starvation resistance for many wild *C. elegans* strains simultaneously. We identified three quantitative trait loci with variants in 16 insulin/EGF receptor-like domain (*irld*) family members. We show that four *irld* genes affect starvation resistance by regulating insulin/IGF signaling. We propose that IRLD proteins bind insulin-like peptides to modify signaling in the sensory nervous system thereby affecting organismal physiology. This work demonstrates efficacy of using population sequencing to dissect a complex trait, identifies *irld* genes that regulate insulin/IGF signaling, and shows that an expanded gene family modifies a deeply conserved signaling pathway to affect a fitness-proximal trait.

## INTRODUCTION

Enduring periods of starvation is a near-ubiquitous feature of animal life, making starvation resistance a fitness-proximal trait. However, the genetic basis of natural variation in starvation resistance is unclear, though understanding if and how critical conserved pathways are affected will inform therapeutic strategies. Insulin signaling is reduced during starvation, and evolutionary biologists, geneticists, and clinicians have all noted the parallel roles of insulin signaling in metabolic syndrome and starvation resistance (SOETERS AND SOETERS 2012; FLIER 2019). The nematode *C. elegans* is frequently starved in the wild and has robust starvation responses (SCHULENBURG AND FELIX 2017; BAUGH AND HU 2020). Larvae that hatch in the absence of food arrest development and can survive starvation for weeks (BAUGH 2013). In addition to causing mortality, extended starvation impacts growth and reproductive success (JOBSON *et al.* 2015; JORDAN *et al.* 2019), and these effects can be uncoupled (ROUX *et al.* 2016; BAUGH AND HU 2020). Reduction of insulin/IGF signaling activates the forkhead transcription factor DAF-16/FoxO (LIN *et al.* 1997; OGG *et al.* 1997), which promotes starvation resistance (MUNOZ AND RIDDLE 2003; BAUGH AND STERNBERG 2006; HIBSHMAN *et al.* 2017). Hundreds of genome-sequenced, wild *C. elegans* strains are available for genetic analysis of natural variation (COOK *et al.* 2017). By capturing targeted sequences, molecular inversion probe sequencing (MIP-seq) enables extremely deep sequencing of polymorphic loci (CANTSILIERIS *et al.* 2017; MOK *et al.* 2017), providing sensitive and precise estimation of strain frequencies in a complex pool as a means of high-throughput phenotyping. Here, we used MIP-seq to phenotype starvation resistance in a pool of genetically diverse wild *C. elegans* strains, and we identified multiple members of the *irld* gene family as modifiers of insulin/IGF signaling (IIS) and starvation resistance.

## RESULTS AND DISCUSSION

### MIP-seq effectively measures starvation resistance for pooled *C. elegans* strains

We chose 100 genetically diverse, wild *C. elegans* strains from around the world including the laboratory reference N2 to assay for starvation resistance (Fig 1A-B). We designed four molecular inversion probes per strain, with each targeting a single-nucleotide variant (SNV) specific to the strain (Fig 1C). To pilot MIP-seq, we pooled the molecular inversion probes and prepared sequencing libraries from an equimolar mix of genomic DNA from each strain (Fig 1D). We determined the frequency of strain-specific reads for each

molecular inversion probe, and we censored probes that produced frequencies substantially different than the expected value of 0.01, leaving three or four reliable probes for 85% of strains (Fig 1E, Data S1). We also mixed genomic DNA at different concentrations to prepare a standard curve. MIP-seq accurately measured individual strain frequencies over three orders of magnitude (Fig 1F), and greater sequencing depth can expand the dynamic range if necessary.

We pooled the 100 diverse strains and subjected them to starvation. The pool was sampled multiple times throughout starvation (Fig 2A). These samples were then fed, given time to recover and then reproduce for one day, and DNA was isolated for MIP-seq. By incorporating recovery and early fecundity, this sampling scheme integrates effects of starvation on mortality, developmental rate, and reproductive success, each of which are important for starvation resistance. DNA was also isolated directly from starved larvae at the onset of starvation as a 'baseline' sample to normalize differences in pool composition in each of five biological replicates (Fig S1A,B). Principal component analysis (PCA) on strain frequencies at all time points showed the first principal component correlates with duration of starvation, especially across the first three time points (Fig 2B, Fig S1B). Substantial mortality occurred by day 17 (Fig S1C), and day 17 recovery cultures thus produced relatively few progeny. Consequently, differences in strain frequencies were actually smaller at day 17 than 13, but trends were stable (Fig S2).

We developed two metrics to quantify starvation resistance for each strain. "PC1" is the value of the first principal component for each strain from the strain-level PCA, and "Slope" is a measure of how much a strain increases or decreases in frequency over time across days 1, 9, and 13 (Data S1). These two metrics are highly correlated but also show some differences (Fig 2C). As validation, strains' Slope values are correlated with starvation-resistance values produced from a population-sequencing approach with less power (Fig S1D) (WEBSTER *et al.* 2019). In addition, it is modestly correlated with the latitude from which strains were collected, suggesting possible adaptation to starvation or other correlated traits based on location (Fig S3). We used Slope to order strains from most resistant to sensitive, revealing natural variation in starvation resistance (Fig 2D, Fig S2).

As validation, starvation-resistant strains DL238 and EG4725 survived starvation significantly longer than sensitive strains MY2147 and NIC526 in a standard assay (Fig 2E). Survival times of MY2147 and

NIC526 are comparable to N2, suggesting that these differences are driven by increased resistance of DL238 and EG4725. Notably, differences in starvation survival among wild strains are relatively small compared to many published mutants in the N2 reference background (BAUGH AND HU 2020), suggesting mutants that die quickly during starvation are likely purged from the population. On the other hand, we find modest evidence of a trade-off between starvation resistance and growth rate after one day of starvation (control condition) (Fig S4), potentially limiting starvation resistance among wild strains. Finally, DL238 and EG4725 recovered from starvation better than MY2147 but not NIC526 (Fig 2F), demonstrating that differences in starvation survival and/or recovery may underlie variation in resistance, as expected given our recovery-based sampling scheme. These results validate the MIP-seq approach and reveal the extent of natural variation in starvation resistance.

**Natural variation in *irld* gene family members affects starvation resistance**

We used Slope and PC1 as trait values to perform genome-wide association (GWA) analyses (COOK *et al.* 2017). GWA identified quantitative trait loci (QTL) on the right arm of chromosome IV and on the left and right arms of chromosome V (Fig 3A-B, Data S2). We confirmed each QTL affected starvation resistance by generating near-isogenic lines and performing a starvation-recovery assay (Fig S5). These QTL include enrichments for several large gene families, including serpentine receptors, C-type lectins, and nuclear hormone receptors (Fig 3C). We used gene name prefixes to identify additional enriched families, and we found that nearly a quarter of *irld* gene family members contain variants in the associated QTL (Fig 3D).

In addition to the enrichment of *irld* genes as a group, *irld-39* was the strongest individual candidate gene based on association of a predicted high-impact variant with starvation resistance (Fig 3E, Data S2). *irld-52* also contains a high-impact variant associated with starvation resistance (Fig 3F). *irld-11* and *irld-57* overlap with 'hyper-divergent' regions containing numerous variants in some starvation-resistant strains (LEE *et al.* 2021), and hyper-divergence at these loci is associated with starvation resistance (Fig 3G-H). Because *irld-39* and *irld-52* contain singular, putative high-impact variants associated with starvation resistance, we used genome editing to generate these specific variants in the starvation-sensitive MY2147 and the laboratory-adapted N2 backgrounds (Fig 3I). Since *irld-11* and *irld-57* contain so many candidate variants, we deleted these genes in MY2147 and N2 (Fig 3I). In all four cases, putative loss-of-function alleles improved starvation recovery in the MY2147 background but not N2 (Fig 3J,K). These results show that multiple types of variants in

different *irld* family members improve starvation recovery, identifying four individual genes from one family that affect starvation resistance in wild strains.

Function of the *irld* family has been largely unexplored. Sequence homology suggests that IRLD proteins contain extracellular L domains, like insulin receptors, but lack a tyrosine kinase domain, suggesting they regulate insulin/IGF signaling by binding insulin-like peptides as decoy receptors (DLAKIC 2002), analogous to the recently reported function of the *daf-2b* truncated isoform of *daf-2/InsR* (MARTINEZ *et al.* 2020). However, the *irld* genes also have homology to the EGF receptor, and family members *hpa-1* and *hpa-2* affect healthspan by acting through EGF signaling (IWASA *et al.* 2010). The *C. elegans* reference genome encodes 40 insulin-like peptides (PIERCE *et al.* 2001), and parallel expansion of these two families has been proposed (DLAKIC 2002).

**IRLD-39 and IRLD-52 regulate insulin/IGF signaling**

Because *irld-39(duk1)* and *irld-52(duk17)* alleles correspond to the exact variants present in starvation-resistant wild strains, we chose to focus on them for further analysis. We hypothesized that phenotypic effects may have been detected only in MY2147 because it is relatively sensitive, and multiple *irld* variants may reveal an effect in N2. Indeed, an *irld-39(duk1); irld-52(duk17)* double mutant in the N2 background exhibited improved starvation recovery (Fig 4A).

Given homology between IRLD proteins and DAF-2/InsR, we asked whether *irld* genes have genetic interactions with insulin/IGF signaling. We found no difference in starvation recovery between the null mutant *daf-16(mu86)* and *daf-16(mu86); irld-39(duk1); irld-52(duk17)*, indicating that increased starvation resistance of *irld-39(duk1); irld-52(duk17)* is dependent on DAF-16 (Fig 4B). We further hypothesized that IRLDs affect DAF-16 activity. In support of this hypothesis, intestinal nuclear localization of endogenous DAF-16 (AGHAYEVA *et al.* 2020) was significantly increased at 36 hours of starvation in *irld-39(duk1); irld-52(duk17)* mutants (Fig 4C). Altered DAF-16 localization was not apparent at six hours of starvation, suggesting that IRLDs may affect the rate at which DAF-16 exits the nucleus after mounting the starvation response (MATA-CABANA *et al.* 2020).

We performed two mRNA-seq experiments to further address interactions between *irld* gene function and insulin/IGF signaling (Data S3). In a transcriptome-wide epistasis analysis with all four relevant genotypes

(Fig 4D) (ANGELES-ALBORES *et al.* 2018), *daf-2* was epistatic to *irld-39* and *irld-52* early in starvation (Fig 4E). That is, gene expression in the triple mutant differed significantly from an additive model in which *daf-2* function is independent of *irld-39* and *irld-52*, instead suggesting *daf-2* functions downstream of *irld-39* and *irld-52*. In our second mRNA-seq experiment designed to interrogate dynamics of DAF-16 function (Fig 4F), DAF-16 targets were significantly differentially expressed as a group in the *irld-39(duk1); irld-52(duk17)* mutant compared to N2 throughout starvation (Fig 4G), reflecting altered DAF-16 activity. Interestingly, DAF-16 positive targets are expressed at higher levels in *irld-39(duk1); irld-52(duk17)* early in starvation compared to N2, but this effect reverses later in starvation. Like nuclear localization, differential expression of DAF-16 targets suggests that *irld* genes affect DAF-16 functional dynamics, which are presumably complex over space and time. Due to differences in these assays (*e.g.*, DAF-16 localization vs. target expression, intestinal cells vs. whole animals), it is difficult to directly compare them, but both support the conclusion that *irld-39* and *irld-52* affect starvation resistance by modifying insulin/IGF signaling.

Single-cell analysis (TAYLOR *et al.* 2020) suggests that *irld* gene expression is typically restricted to sensory neurons (Fig S6). Furthermore, *irld-39* and *irld-52* are expressed nearly exclusively in ASJ and ADL sensory neurons, respectively (Fig 4H). *C. elegans* sensory neurons are polymodal and influence life-history traits regulated by insulin/IGF signaling, including dauer formation, aging, and L1-stage developmental arrest (BARGMANN AND HORVITZ 1991; VOWELS AND THOMAS 1992; APFELD AND KENYON 1999). Furthermore, ASJ is known to express the relatively potent insulin-like peptide DAF-28 in nutrient-dependent fashion (LI *et al.* 2003; KAPLAN *et al.* 2018), and *daf-28* affects starvation survival (CHEN AND BAUGH 2014). Together these observations suggest that IRLD-39 and IRLD-52 function in specific sensory neurons to modify insulin/IGF signaling and starvation resistance (Fig 4I). Insulin-like peptides have complex spatiotemporal expression patterns and distinct loss-of-function phenotypes (RITTER *et al.* 2013; FERNANDES DE ABREU *et al.* 2014). Insulin-like peptides can function as agonists or antagonists of DAF-2/InsR (PIERCE *et al.* 2001), and relative expression of these functional classes changes over time during starvation (CHEN AND BAUGH 2014). Whether IRLDs have Insulin-like peptide-binding specificity is unclear, but specificity would likely alter the function of individual *irld* genes. Insulin-like peptides also participate in extensive feedback regulation, affecting signaling dynamics (KAPLAN *et al.* 2019). We imagine that IRLD proteins function relatively locally such that the

anatomical expression patterns of individual *irld* genes impact how they affect the insulin/IGF signaling network and organismal phenotype.

## CONCLUSIONS

Our results illustrate the power of MIP-seq as a population selection and sequencing approach for analysis of complex traits in whole animals. Increasing population complexity and sequencing depth will allow detection of meaningful phenotypic differences too small or variable to be detected by manual assays, leading to improved understanding of gene-by-environment interactions and the genotype-to-phenotype map. With polygenic and omnigenic architectures increasingly proposed as the basis of complex traits (BOYLE *et al.* 2017), leveraging the power of sequencing will be critical to elucidating their genetic basis. MIP-seq can be used in any organism with known sequence variants and that can be cultured in sufficiently large numbers with the ability to select on the trait of interest.

We identified the first natural genetic variants affecting starvation resistance in *C. elegans*. Multiple lines of evidence suggest that *irld* genes modify insulin/IGF signaling, although they could affect other signaling pathways as well (*e.g.*, EGF). We implicated multiple members of the *irld* gene family, providing an example of an expanded gene family that modifies a conserved pathway to affect a fitness-proximal and biomedically important trait. We propose that IRLD proteins alter insulin/IGF signaling by binding insulin-like peptides that otherwise bind DAF-2/InsR. *irld* genes are expressed in chemosensory neurons, and we believe they function near the top of the insulin/IGF signaling network to tune organismal physiology, analogous to the proposed function of IGF-binding proteins (ALLARD AND DUAN 2018). The proposed decoy-receptor mechanism of IRLD function could be generalized to other receptor homologs or isoforms to affect a variety of signaling pathways, including deeply conserved pathways affecting human health.

## MATERIALS AND METHODS

### *Strains used in this study*

In addition to N2, wild isolates CB4854, CB4856, CX11254, CX11264, CX11271, CX11276, CX11285, CX11307, DL200, DL226, DL238, ED3049, ECA189, ECA191, ECA36, ECA363, ECA369, ECA372, ECA396, ED3017, ED3052, ED3077, EG4724, EG4725, GXW1, JU1212, JU1400, JU1581, JU1652, JU1793, JU1896, JU2001, JU2007, JU2017, JU2106, JU2234, JU2316, JU2464, JU2519, JU2526, JU2576, JU258, JU2592,

JU2619, JU2811, JU2829, JU2593, JU2838, JU2841, JU2878, JU2879, JU3137, JU561, JU774, JU775, JU782, KR314, LKC34, MY10, MY16, MY18, MY2147, MY23, MY2453, MY2741, NIC195, NIC199, NIC251, NIC252, NIC256, NIC527, NIC258, NIC261, NIC262, NIC265, NIC266, NIC268, NIC271, NIC3, NIC501, NIC523, NIC526, NIC528, PB306, PS2025, QG2075, QG556, QW947, QX1211, QX1212, QX1791, QX1792, QX1793, QX1794, WN2001, XZ1513, XZ1514, XZ1515, and XZ1516 were used for MIP-seq. QX1430 was used for validation assays.

CB1370 *daf-2(e1370) III,* CF1038 *daf-16(mu86) I,* and OH16024 *daf-16(ot971[daf-16::GFP]) I* were used to assess the interaction of *irld* genes with insulin signaling.

*Strains generated in this study*

Near-isogenic lines include:

LRB392 – *dukIR7(V, EG4725>MY2147)*

LRB393 – *dukIR8(V, EG4725>MY2147)*

LRB395 – *dukIR10(V, EG4725>MY2147)*

LRB396 – *dukIR11(V, MY2147>EG4725)*

LRB397 – *dukIR12(V, MY2147>EG4725)*

LRB398 – *dukIR13(V, MY2147>EG4725)*

LRB399 – *dukIR14(V, MY2147>EG4725)*

LRB400 – *dukIR15(V, MY2147>EG4725)*

LRB401 – *dukIR16(V, MY2147>EG4725)*

LRB402 – *dukIR17(V, EG4725>MY2147)*

LRB403 – *dukIR18(V, EG4725>MY2147)*

LRB407 – *dukIR19(V, EG4725>MY2147)*

LRB408 – *dukIR20(V, EG4725>MY2147)*

LRB409 – *dukIR21(V, EG4725>MY2147)*

LRB410 – *dukIR22(IV, DL238>N2)*

LRB411 – *dukIR23(IV, DL238>N2)*

See Figure S3 for wild isolate composition.

CRISPR-edited strains and new crosses include:

LRB412 *irld-39* in N2 background – *irld-39(duk1) IV*

LRB413: *irld-39* in MY2147 background – *irld-39(duk2*[MY2147]) *IV*

LRB414: *irld-39* in MY2147 background – *irld-39(duk3*[MY2147]) *IV*

LRB415: *irld-39* in MY2147 background – *irld-39(duk4*[MY2147]) *IV*

LRB420: *irld-11* in MY2147 background – *irld-11(duk9*[MY2147]) *V*

LRB421: *irld-52* in N2 background – *irld-52(duk10) V*

LRB422: *irld-52* in MY2147 background - *irld-52(duk11*[MY2147]) *V*

LRB423: *irld-11* in N2 background – *irld-11(duk12) V*

LRB425: *irld-57* in N2 background – *irld-57(duk13) V*

LRB426: *irld-57* in N2 background – *irld-57(duk14) V*

LRB427: *irld-57* in MY2147 background – *irld-57(duk15*[MY2147]) *V*

LRB428: *irld-57* in MY2147 background – *irld-57(duk16*[MY2147]) *V*

LRB431: *irld-52* in N2 background – *irld-52(duk17) V*

LRB444: *irld-39; irld-52* (generated from crossing LRB412 and LRB431) - *irld-39(duk1) IV; irld-52(duk17) V*

LRB456: *daf-16(mu86) I; irld-39(duk1) IV; irld-52(duk17) V*

LRB457, LRB458: *daf-2(e1370) III; irld-39(duk1 IV); irld-52(duk17) V*

LRB463: *daf-16(ot971) I; irld-39(duk1) IV; irld-52(duk17) V*

Multiple strain names for the same genotype indicate independent lines.

*MIP-seq experimental set-up*

Wild strains were independently passaged on 10 cm NGM plates with OP50 *E. coli* every two to three days to ensure they did not starve for at least three generations prior to the experiment. For each biological replicate, a single non-starved plate with gravid adults was selected per strain to ensure initial representation of all strains. Strains were pooled for hypochlorite treatment to obtain pure populations of embryos (HIBSHMAN *et al.* 2021). Embryo concentration was calculated by repeated sampling, and 500,000 embryos were resuspended at 10/µL in S-basal (50 mL total culture) and placed in a 20ºC shaker at 180 rpm to hatch without food and enter L1 arrest. On day 1 (24 hours after hypochlorite treatment), 5 mL of culture (50,000 L1s) was taken as a baseline sample, spun down at 3000 rpm, aspirated down to approximately 100 µL in an Eppendorf tube, flash frozen in liquid nitrogen, and stored at -80ºC until DNA isolation. At days 1, 9, 13, and 17, aliquots

from the L1 arrest culture were set up in recovery cultures at 5 L1s/µL, 1x HB101 (25 mg/mL), and S-complete. Recovery cultures were 10 mL for days 1 and 9, 20 mL for day 13, and 50 mL for day 17 to account for lethality late in starvation by ensuring adequate population sizes. Four days after recovery culture set-up, samples were collected for DNA isolation. For days 1 and 9, the recovery culture was freshly starved with adults and next-generation L1 larvae. At day 13, the culture was typically near starved, with adults and some L1 larvae. At day 17, the culture was typically not starved. If HB101 was still present at collection, samples were washed 3-4 times with S-basal. Samples were flash frozen in liquid nitrogen and stored at -80ºC until DNA isolation.

*DNA isolation*

Frozen samples were rapidly freeze-thawed three times, cycling between liquid nitrogen and a 45°C water bath. Genomic DNA was isolated using the Quick-DNA Miniprep Kit (Zymo Research# D3024) following the manufacturer's protocol. The DNA concentration was determined for each sample using the Qubit dsDNA HS Assay kit (Invitrogen# Q32854).

*MIP design*

MIPgen (BOYLE *et al.* 2014) was used to design four MIPs for each of 103 strains. Unique homozygous SNVs were parsed from the VCF file WI.20170531.vcf.gz (available at https://storage.googleapis.com/elegansvariation.org/releases/20170531/WI.20170531.vcf.gz). Target regions in BED format were generated using the makeBedForMipgen.pl script. MIPgen was used against *C. elegans* genome version WS245 with the following parameters: -min_capture_size 100 -max_capture_size 100 -tag_sizes 0, 10. MIPs are 80 base-pairs (bp) long and include 20 bp ligation and extension arms that are complementary to DNA surrounding the unique SNV of interest for each strain. In addition, P5 and P7 Illumina sequences are included as part of the MIP to facilitate Illumina sequencing. Each MIP molecule includes a 10 bp unique molecular identifier (UMI) adjacent to the ligation arm. Only MIPs that capture the SNV within a 50 bp sequencing read were used, meaning the SNV was no more than 40 bp away from the UMI. SNPs located within 40 bases of the sequencing start site were parsed with the parseMipsPerSNPposition.pl script. These scripts can be found at https://github.com/amykwebster/MIPseq_2021.

*MIP-seq library preparation and sequencing*

For pilots and the starvation-resistance experiment, 500 ng genomic DNA from each sample was used for MIP-seq libraries. Libraries were generated as described previously (HIATT *et al.* 2013) with the following modifications. We included 1000 copies of each MIP for every individual copy of the worm genome in the 500 ng input DNA, which corresponded to 0.0083 picomoles of each individual MIP. All 412 MIPs (sequences available in Data S1) were first pooled in an equimolar ratio at a concentration of 100 μM. The MIP pool was diluted to 5 μM in 1 mM Tris buffer, and 50 μL of this pool was used in the 100 μL phosphorylation reaction. Next, the probe hybridization reaction for each sample was set with 500 ng DNA and 3.42 picomoles (0.0083 picomoles x 412) of the phosphorylated probe mixture. Following hybridization, gap filling, ligation, and exonuclease steps were performed as described previously. PCR amplification of the captured DNA (primer sequences available in Data S1) was performed in a 50 μL reaction with 18 cycles. The PCR libraries were purified using the SPRIselect beads (Beckman# B23318), and library concentrations were assessed with the Qubit dsDNA HS Assay kit (Invitrogen# Q32854). Sequencing was performed on the Illumina HiSeq 4000 to obtain 50-bp single-end reads.

*MIP-seq analysis*

FASTQ files from sequencing reads were processed using the script parseMIPGenotypeUMI.pl, also available at github.com/amykwebster/MIPseq_2021. This script accepts as input the list of MIPs produced from MIPgen, the UMI length, and FASTQ files in order to count the number of reads corresponding to each MIP and whether they have the reference allele, alternative allele, or one of two other alleles. While we included UMIs in our MIP design, use of the UMI to filter duplicate reads in pilot standard curves did not improve data quality (likely due to the relatively large mass of DNA used to prepare libraries), and so the UMI was not used in the published analysis. For each MIP, the frequency of the strain for which the MIP captures its unique SNV was calculated as the alternative read count divided by the total of alternative and reference read counts. For the MIP pilot with all strains in an equimolar ratio, there were 246,986,236 total mapped reads to all MIPs. Individual MIPs were filtered out if they did not meet the following criteria: 1) They were within 3.5-fold of expected frequency (that is, alt / (alt + ref) was within 3.5-fold of 1/103), 2) 'other' reads (those that are not alternative or reference alleles) were < 20,000 total, and 3) alternative and reference allele totals were between

20,000 and 2,000,000 total reads. 321 of 412 MIPs met these criteria, and reads from these 321 MIPs were included in subsequent analysis. N2 has very few unique SNVs making it difficult to design optimal MIPs, and N2 MIPs did not meet these criteria but were included nonetheless (see Data S1). For the standard curve experiment (Fig 1F), DNA from seven strains (CB4856, DL200, ED3077, JU258, JU561, JU1652, and N2) was pooled in defined concentrations ('expected'), and MIPs that met the criteria defined above were used to calculate strain frequencies ('observed', see Data S1).

For the starvation-resistance experiment, an average of 51.7 million reads (standard deviation 7.2 million reads) were sequenced per library (one library per time point, replicate, and condition – 25 libraries total). An average of 94% of reads (standard deviation 0.5%) matched the ligation probe, and 71.8% (standard deviation 3.4%) matched the ligation probe and scan sequence. Strain frequencies were determined by averaging the frequencies calculated across MIPs included in the 321 MIPs for each strain. A dataframe of all strains and their frequencies at day 1 baseline, as well as days 1, 9, 13, and 17 after recovery for all replicates was used to obtain trait values for subsequent analysis. PCA was performed on the dataframe following normalization of day 1, 9, 13, and 17 time points by the baseline day 1 sample and log2 transformation. PC1 loadings were extracted for each strain. For 'Slope', day 1, 9, and 13 recovery samples were normalized by day 1 frequencies and log2 transformed. For each strain, a line was fit to day 1, 9, and 13 normalized data with intercept at 0, and the slope of the line was taken as the trait value.

*Comparison of MIP-seq and RAD-seq*

To determine how well MIP-seq trait values correlated with RAD-seq trait values from previous work (WEBSTER *et al.* 2019), RAD-seq data were normalized the same way that we normalized the MIP-seq data. Specifically, data from one biological replicate from RAD-seq that had data at time points over the course of starvation, including days 1, 7, 14, 21, and 24, was used. The frequency of each strain at each time point was divided by its frequency on day 1. These values were log2 transformed, so positive values indicate an increase in frequency over time and negative values indicate a decrease in frequency over time. A linear regression was then fit to each with a y-intercept of 0 through the data points over time. The slope of the line was calculated as the trait value for each strain. RAD-seq and MIP-seq data were filtered to include only the 34 strains that were present in both analyses. The values were plotted against each other and a linear regression was fit through these points to determine their correlation ($R^2 = 0.24$, $p = 0.002$).

*GWA analysis*

Slope and PC1 trait values for each strain were used for GWA using the R package cegwas2 (https://github.com/AndersenLab/cegwas2-nf). Genotype data were acquired from the latest VCF release (release 20200815) from CeNDR. BCFtools (LI 2011) was used to filter variants below a 5% minor allele frequency and variants with missing genotypes and used PINKv1.9 (PURCELL *et al.* 2007; CHANG *et al.* 2015) to prune genotypes using linkage disequilibrium. The additive kinship matrix was generated from 45,733 markers using the A.mat function in the rrBLUP package. Because these markers have high LD, eigen decomposition of the correlation matrix of the genotype matrix was performed to identify 570 independent tests. GWA was performed using the GWAS function of the rrBLUP package (ENDELMAN 2011). Significance was determined by an eigenvalue threshold by the number of independent tests in the genotype matrix. Confidence intervals were defined as +/- 150 SNVs from the rightmost and leftmost markers passing the significance threshold.

To identify enriched gene groups, fine mapping data from Slope and PC1 results were merged with WS273 gene names. Unique sequence names were extracted (see Data S2), and the 867 sequence names with variants in Slope or PC1 QTL were used in WormCat (HOLDORF *et al.* 2020) to identify functional category enrichments (Fig 3C). The most specific enrichments (those in Category 3) are shown. For genes with unique symbol names (555 genes), those with the same symbol prefix were considered part of the same gene family, and enrichments were calculated for families with at least five members with variants in the QTL. *hpa* and *irld* gene prefixes were grouped under the *irld* prefix (IWASA *et al.* 2010). P-values were calculated from a hypergeometric test.

ALT and REF information for *irld-39* and *irld-52* high-impact variants was obtained from fine mapping and is available as part of Data S2. To determine whether *irld-57* and *irld-11* overlapped with hyper-divergent regions in each strain, coordinates of hyper-divergent regions for each strain were obtained from (LEE *et al.* 2021), and coordinates of *irld-11* and *irld-57* were obtained from WormBase. If the hyper-divergent region and gene overlapped for a strain, then the strain was considered hyper-divergent at the locus. Hyper-divergent status of each strain is available in Data S2.

*NIL generation*

To validate chromosome IV and V QTL, pairs of strains that differ for starvation resistance and the alternative vs reference allele for the associated SNV marker were chosen. Compatibility at the *peel-1/zeel-1* and *pha-1/sup-35* loci was considered (SEIDEL *et al.* 2008; BEN-DAVID *et al.* 2017). For chromosome V QTL, EG4725 and MY2147 were compatible at both loci, and we generated reciprocal NILs for the left and right arms of chromosome V. EG4725 did not have the alternative allele associated with starvation resistance for chromosome IV, so we used DL238 and N2 as the parental strains. DL238 and N2 are incompatible for reciprocal crosses, but we introgressed the DL238 chromosome IV QTL into the N2 background. To generate NILs, the two parental strains were first crossed, then F2 progeny were genotyped on each end of the desired QTL for introgression to identify homozygotes from one parental background (*e.g.*, MY2147). Then these homozygotes were repeatedly backcrossed to the opposite background (*e.g.*, EG4725) and repeatedly genotyped to maintain homozygotes at the introgressed region. Genotyping was performed using PCR to amplify a genomic region whose sensitivity to a particular restriction enzyme depends on parental genetic background. Primers were designed using VCF-kit (COOK AND ANDERSEN 2017). Primers and enzymes used can be found in Data S2. NILs were backcrossed a minimum of six times. Final NILs were sequenced at ~1x coverage to determine the parental contributions over the entire genome (Fig S5 and Data S2).

*CRISPR design and implementation to edit irld genes*

For genes of interest, CRISPR guide design was done in Benchling using genome version WBcel235 and importing sequence for genes of interest. To generate *irld-39* and *irld-52* variants (5 bp deletions), a single guide RNA (sgRNA) and repair template was generated. For *irld-11* and *irld-57,* two sgRNAs were generated per gene to delete the entire gene and a single repair template was used. sgRNAs (2 nmol) and 100 bp repair templates (highest purity at 4 nmol) were ordered from IDT. The *dpy-10* co-CRISPR method was used to generate and screen for edits (PAIX *et al.* 2015). The injection mix used was: sgRNA for *dpy-10* (0.2 µL of 100 µM stock), sgRNA of gene of interest (0.5 µL of 100 µM stock), *dpy-10* repair template (0.5 µL of 10 µM stock), repair template for gene of interest (0.6 µL of 100 µM stock), Cas9 (0.8 µL of 61 µM stock), and water up to 10 µL total. Injection mix components were stored at -20ºC, and injection mix was incubated at room temperature for one hour before injections. N2 and MY2147 L4s were picked the day before injecting, and young adults were injected in the gonad and singled to new plates. After 3-4 days, next-generation adults were screened for

rollers, which are heterozygous for the *dpy-10* edit and have increased likelihood of also having the desired edit. Non-roller F2 progeny of F1 roller worms were then genotyped to identify worms homozygous for the desired edit, and edits were confirmed by Sanger sequencing. Sequences of sgRNAs, repair templates, and PCR primers for genotyping are available in Data S2.

*Starvation recovery (worm length measurements)*

Strains were maintained well-fed for at least three generations prior to beginning experiments. Gravid adults were hypochlorite treated to obtain embryos, which were resuspended at 1 embryo/µL in 5 mL of S-basal in a glass test tube and placed on a roller drum at 20ºC so they hatch and enter L1 arrest. After 1 and 8 days of L1 arrest, an aliquot of 500-1000 µL (a consistent volume was used between conditions within the same experiment) per strain was plated on a 10 cm plate with OP50 and allowed to recover for 48 hr. After 48 hr, worms were washed onto an unseeded NGM plate. Images were then taken of worms using a ZeissDiscovery V20 stereomicroscope. To determine lengths of worms, the WormSizer plugin for Fiji was used and worms were manually passed or failed (MOORE *et al.* 2013). To determine differences in starvation recovery between strains, a linear mixed-effect model was fit to the length data for all individual worms with duration of starvation (1 or 8 days) and strain as fixed effects and biological replicate as a random effect using the package nlme in R. The summary function was used to calculate a p-value from the t-value.

*Starvation survival*

L1 arrest cultures were set up as described for starvation recovery. Starting on the first day of arrest and proceeding every other day, a 100 µL aliquot of culture was pipetted onto a 5 cm NGM plate with a spot of OP50 at the center. The aliquot was placed around the periphery of the lawn, and the number of worms plated was counted. Two days later, the number of worms that had made it to the bacterial lawn and were alive was counted. Live worms that have developed and were outside the lawn were also counted. The total number of live worms after two days was divided by the total plated to determine the proportion alive. For each replicate, a logistic curve was fit to the data, and the half-life (time at 50% survival) was calculated, and a t-test was performed on half-lives between strains of interest.

*Nuclear localization*

L1 arrest cultures were set up as described for starvation recovery. At 6 and 36 hours of L1 arrest, an aliquot of 700 µL was spun down in a 1.7 mL Eppendorf tube at 3000 rpm for 30 seconds to pellet L1 larvae.

1.5 µL of worm pellet was pipetted into the center of a slide with a 4% Noble agar pad, and a glass cover slip was immediately placed on top. A timer was set for 3 minutes, and the slide was systematically scanned with each individual worm scored for nuclear localization at 40x or 100x with a Zeiss compound microscope. Nuclear localization of DAF-16 was scored in intestinal cells and assigned as one of four categories: nuclear, more nuclear, more cytoplasmic, and cytoplasmic. 'More nuclear' and 'more cytoplasmic' are intermediate categories between nuclear and cytoplasmic, with localization closer to being nuclear or cytoplasmic, respectively. See Fig 4C for representative images. Scoring for each slide stopped after 3 minutes.

*mRNA-seq sample collection*

N2, LRB444, CB1370, and LR457 and LRB458 were grown on large plates with OP50, and gravid adults were hypochlorite treated to obtain embryos. Embryos were resuspended at 1 embryo/µL in 10 mL of S-basal in a 50 mL Erlenmeyer flask and put in a shaker at 20ºC and 180 rpm. Technical replicates were set up if > 10,000 embryos were obtained for a condition. Indicated time points were collected for each mRNA-seq experiment (12 hr of L1 arrest for *daf-2* epistasis experiment, and 6, 12, 24, 48, and 72 hr of L1 arrest for time-series experiment), samples were spun down at 3000 rpm for 1 min in conical tubes, supernatant was aspirated down to ~100 µL, and worms were transferred to an Eppendorf tube, spun down again, and aspirated down to approximately 10 µL. Samples were flash frozen in liquid nitrogen and stored at -80ºC until RNA isolation.

*RNA isolation and mRNA-seq library preparation*

RNA was isolated using TRIzol Reagent (Invitrogen# 15596026) following the manufacturer's instructions with some exceptions noted below. The procedure was scaled down linearly, using 100 µL Trizol. 5 µg linear polyacrylamide (Sigma# 56575) was included as a neutral carrier for RNA precipitation. RNA was eluted in nuclease-free water and incubated at 55ºC for approximately four minutes to resuspend the pellet. 150 ng of total RNA was used for each library preparation. NEBNext Poly(A) mRNA Isolation Module (New England Biolabs# E7490) and NEBNext Ultra II RNA Library Prep kit (New England Biolabs# E7770) were used to perform poly-A selection and prepare libraries for sequencing, respectively. The final libraries were enriched with 11 cycles of PCR. Libraries were then sequenced using Illumina NovaSeq 6000 to obtain 50 bp paired-end reads.

*mRNA-seq analysis*

Bowtie (LANGMEAD 2010) was used to map FASTQ reads to version WS273 of the genome using the following command: bowtie -I 0 -X 500 chunkmbs 256 -k 1 -m 2 -S -p 2. Reads were counted using HTSeq. Count tables were analyzed for differential expression in R using edgeR (ROBINSON *et al.* 2010). First, counts were filtered to only protein-coding genes, then further filtered to genes expressed above one count-per-million (CPM) in at least the number of libraries corresponding to the number of biological replicates (four libraries for *daf-2* epistasis and three for the time series). The exact test was used to determine $\log_2$ fold change differences for pairwise comparisons. DAF-16 positive targets are the overlap between Class I targets and significantly down-regulated genes in *daf-16* mutants during L1 arrest (TEPPER *et al.* 2013; KAPLAN *et al.* 2015). DAF-16 negative targets are the overlap between Class II targets and significantly up-regulated genes in *daf-16* mutants during L1 arrest. Significance of the differences between $\log_2$ fold-change distributions of DAF-16 positive and negative targets was assessed using the Kolmogorov-Smirnov test.

For epistasis analysis, edgeR was used to perform an ANOVA-like test using the GLM functionality, resulting in 3427 significant genes (FDR < 0.05). DESeq was used to perform differential expression analysis with all the background genes that had passed the low-expression filtering conducted in edgeR. The result was restricted to the 3427 significant genes and then passed to the epistasis analysis described in Angeles-Albores *et al.* to calculate an epistasis coefficient, calculate probabilities for pre-defined models, and perform model selection (ANGELES-ALBORES *et al.* 2018), in order to determine the relationship between *irld* genes and *daf-2* in this context. Pre-defined models were rejected when the odds ratio was greater than $10^3$.

For single-cell data from (TAYLOR *et al.* 2020), expression values for *irld* genes were obtained from Supplementary Table 11, which includes genes considered expressed at the most stringent threshold (threshold 4).

*Data and materials availability:*

Raw MIP-seq data for the starvation-resistance experiment and the pilot experiments to test individual MIPs is available as part of NCBI BioProject PRJNA730178. Raw and processed mRNA-seq data is available as part of the NCBI Gene Expression Omnibus at accession GSE174327. Code for processing MIP-seq data is available at github.com/amykwebster/MIPseq_2021.

*Acknowledgments*

We thank Oliver Hobert for providing OH16024 *daf-16(ot971[daf-16::GFP])*, Jon Hibshman for sharing a starvation survival curve-fitting script, Chelsea Shoben for help passaging wild isolates, Clay Dilks for CRISPR advice, and Jim Jordan for helpful discussions. Funding was provided by the NIH (R01GM117408 to LRB and R01ES02993 to ECA and LRB). AKW was supported by an NSF Graduate Research Fellowship. Some strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). We would also like to thank WormBase.

*Competing interests*

Authors declare that they have no competing interests.

*Author contributions*

Conceptualization: AKW, LRB

Investigation: AKW, RC, MP, JC, KF, RT

Formal analysis: AKW, LS, KE, IA

Visualization: AKW

Funding acquisition: LRB, ECA

Supervision: LRB, ECA

Writing – original draft: AKW, LRB

Writing – review & editing: AKW, LRB, ECA, RC, KE, IA

## REFERENCES


Aghayeva, U., A. Bhattacharya and O. Hobert, 2020 A panel of fluorophore-tagged daf-16 alleles. MicroPubl Biol 2020.

Allard, J. B., and C. Duan, 2018 IGF-Binding Proteins: Why Do They Exist and Why Are There So Many? Front Endocrinol (Lausanne) 9: 117.

Angeles-Albores, D., C. Puckett Robinson, B. A. Williams, B. J. Wold and P. W. Sternberg, 2018 Reconstructing a metazoan genetic pathway with transcriptome-wide epistasis measurements. Proc Natl Acad Sci U S A 115: E2930-E2939.

Apfeld, J., and C. Kenyon, 1999 Regulation of lifespan by sensory perception in Caenorhabditis elegans. Nature 402: 804-809.

Bargmann, C. I., and H. R. Horvitz, 1991 Control of larval development by chemosensory neurons in Caenorhabditis elegans. Science 251: 1243-1246.

Baugh, L. R., 2013 To grow or not to grow: nutritional control of development during Caenorhabditis elegans L1 arrest. Genetics 194: 539-555.

Baugh, L. R., and P. J. Hu, 2020 Starvation Responses Throughout the Caenorhabditis elegans Life Cycle. Genetics 216: 837-878.

Baugh, L. R., and P. W. Sternberg, 2006 DAF-16/FOXO regulates transcription of cki-1/Cip/Kip and repression of lin-4 during C. elegans L1 arrest. Curr Biol 16: 780-785.

Ben-David, E., A. Burga and L. Kruglyak, 2017 A maternal-effect selfish genetic element in Caenorhabditis elegans. Science 356: 1051-1055.

Boyle, E. A., Y. I. Li and J. K. Pritchard, 2017 An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell 169: 1177-1186.

Boyle, E. A., B. J. O'Roak, B. K. Martin, A. Kumar and J. Shendure, 2014 MIPgen: optimized modeling and design of molecular inversion probes for targeted resequencing. Bioinformatics 30: 2670-2672.

Cantsilieris, S., H. A. Stessman, J. Shendure and E. E. Eichler, 2017 Targeted Capture and High-Throughput Sequencing Using Molecular Inversion Probes (MIPs). Methods Mol Biol 1492: 95-106.

Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell *et al.*, 2015 Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4: 7.

Chen, Y., and L. R. Baugh, 2014 Ins-4 and daf-28 function redundantly to regulate C. elegans L1 arrest. Dev Biol 394: 314-326.

Cook, D. E., and E. C. Andersen, 2017 VCF-kit: assorted utilities for the variant call format. Bioinformatics 33: 1581-1582.

Cook, D. E., S. Zdraljevic, J. P. Roberts and E. C. Andersen, 2017 CeNDR, the Caenorhabditis elegans natural diversity resource. Nucleic Acids Res 45: D650-D657.

Dlakic, M., 2002 A new family of putative insulin receptor-like proteins in C. elegans. Curr Biol 12: R155-157.

Endelman, J. B., 2011 Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. The Plant Genome 4.

Fernandes de Abreu, D. A., A. Caballero, P. Fardel, N. Stroustrup, Z. Chen *et al.*, 2014 An insulin-to-insulin regulatory network orchestrates phenotypic specificity in development and physiology. PLoS Genet 10: e1004225.

Flier, J. S., 2019 Starvation in the Midst of Plenty: Reflections on the History and Biology of Insulin and Leptin. Endocr Rev 40: 1-16.

Hiatt, J. B., C. C. Pritchard, S. J. Salipante, B. J. O'Roak and J. Shendure, 2013 Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. Genome Res 23: 843-854.

Hibshman, J., A. Webster and L. Baugh, 2021 Liquid-culture protocols for synchronous starvation, growth, dauer formation, and dietary restriction of Caenorhabditis elegans. STAR Protocols 2: 100276.

Hibshman, J. D., A. E. Doan, B. T. Moore, R. E. Kaplan, A. Hung *et al.*, 2017 daf-16/FoxO promotes gluconeogenesis and trehalose synthesis during starvation to support survival. Elife 6.

Holdorf, A. D., D. P. Higgins, A. C. Hart, P. R. Boag, G. J. Pazour *et al.*, 2020 WormCat: An Online Tool for Annotation and Visualization of Caenorhabditis elegans Genome-Scale Data. Genetics 214: 279-294.

Iwasa, H., S. Yu, J. Xue and M. Driscoll, 2010 Novel EGF pathway regulators modulate C. elegans healthspan and lifespan via EGF receptor, PLC-gamma, and IP3R activation. Aging Cell 9: 490-505.

Jobson, M. A., J. M. Jordan, M. A. Sandrof, J. D. Hibshman, A. L. Lennox *et al.*, 2015 Transgenerational Effects of Early Life Starvation on Growth, Reproduction, and Stress Resistance in Caenorhabditis elegans. Genetics 201**:** 201-212.

Jordan, J. M., J. D. Hibshman, A. K. Webster, R. E. W. Kaplan, A. Leinroth *et al.*, 2019 Insulin/IGF Signaling and Vitellogenin Provisioning Mediate Intergenerational Adaptation to Nutrient Stress. Curr Biol 29**:** 2380-2388 e2385.

Kaplan, R. E., Y. Chen, B. T. Moore, J. M. Jordan, C. S. Maxwell *et al.*, 2015 dbl-1/TGF-beta and daf-12/NHR Signaling Mediate Cell-Nonautonomous Effects of daf-16/FOXO on Starvation-Induced Developmental Arrest. PLoS Genet 11**:** e1005731.

Kaplan, R. E. W., C. S. Maxwell, N. K. Codd and L. R. Baugh, 2019 Pervasive Positive and Negative Feedback Regulation of Insulin-Like Signaling in Caenorhabditis elegans. Genetics 211**:** 349-361.

Kaplan, R. E. W., A. K. Webster, R. Chitrakar, J. A. Dent and L. R. Baugh, 2018 Food perception without ingestion leads to metabolic changes and irreversible developmental arrest in C. elegans. BMC Biol 16**:** 112.

Langmead, B., 2010 Aligning short sequencing reads with Bowtie. Curr Protoc Bioinformatics Chapter 11**:** Unit 11 17.

Lee, D., S. Zdraljevic, L. Stevens, Y. Wang, R. E. Tanny *et al.*, 2021 Balancing selection maintains hyper-divergent haplotypes in Caenorhabditis elegans. Nature Ecology & Evolution.

Li, H., 2011 A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27**:** 2987-2993.

Li, W., S. G. Kennedy and G. Ruvkun, 2003 daf-28 encodes a C. elegans insulin superfamily member that is regulated by environmental cues and acts in the DAF-2 signaling pathway. Genes Dev 17**:** 844-858.

Lin, K., J. B. Dorman, A. Rodan and C. Kenyon, 1997 daf-16: An HNF-3/forkhead family member that can function to double the life-span of Caenorhabditis elegans. Science 278**:** 1319-1322.

Martinez, B. A., P. Reis Rodrigues, R. M. Nunez Medina, P. Mondal, N. J. Harrison *et al.*, 2020 An alternatively spliced, non-signaling insulin receptor modulates insulin sensitivity via insulin peptide sequestration in C. elegans. Elife 9.

Mata-Cabana, A., L. Gomez-Delgado, F. J. Romero-Exposito, M. J. Rodriguez-Palero, M. Artal-Sanz *et al.*, 2020 Social Chemical Communication Determines Recovery From L1 Arrest via DAF-16 Activation. Front Cell Dev Biol 8**:** 588686.

Mok, C. A., V. Au, O. A. Thompson, M. L. Edgley, L. Gevirtzman *et al.*, 2017 MIP-MAP: High-Throughput Mapping of Caenorhabditis elegans Temperature-Sensitive Mutants via Molecular Inversion Probes. Genetics 207**:** 447-463.

Moore, B. T., J. M. Jordan and L. R. Baugh, 2013 WormSizer: high-throughput analysis of nematode size and shape. PLoS One 8**:** e57142.

Munoz, M. J., and D. L. Riddle, 2003 Positive selection of Caenorhabditis elegans mutants with increased stress resistance and longevity. Genetics 163**:** 171-180.

Ogg, S., S. Paradis, S. Gottlieb, G. I. Patterson, L. Lee *et al.*, 1997 The Fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in C. elegans. Nature 389**:** 994-999.

Paix, A., A. Folkmann, D. Rasoloson and G. Seydoux, 2015 High Efficiency, Homology-Directed Genome Editing in Caenorhabditis elegans Using CRISPR-Cas9 Ribonucleoprotein Complexes. Genetics 201**:** 47-54.

Pierce, S. B., M. Costa, R. Wisotzkey, S. Devadhar, S. A. Homburger *et al.*, 2001 Regulation of DAF-2 receptor signaling by human insulin and ins-1, a member of the unusually large and diverse C. elegans insulin gene family. Genes Dev 15**:** 672-686.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira *et al.*, 2007 PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81**:** 559-575.

Ritter, A. D., Y. Shen, J. Fuxman Bass, S. Jeyaraj, B. Deplancke *et al.*, 2013 Complex expression dynamics and robustness in C. elegans insulin networks. Genome Res 23**:** 954-965.

Robinson, M. D., D. J. McCarthy and G. K. Smyth, 2010 edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26**:** 139-140.

Roux, A. E., K. Langhans, W. Huynh and C. Kenyon, 2016 Reversible Age-Related Phenotypes Induced during Larval Quiescence in C. elegans. Cell Metab 23**:** 1113-1126.

Schulenburg, H., and M. A. Felix, 2017 The Natural Biotic Environment of Caenorhabditis elegans. Genetics 206**:** 55-86.

Seidel, H. S., M. V. Rockman and L. Kruglyak, 2008 Widespread genetic incompatibility in C. elegans maintained by balancing selection. Science 319**:** 589-594.
Soeters, M. R., and P. B. Soeters, 2012 The evolutionary benefit of insulin resistance. Clin Nutr 31**:** 1002-1007.
Taylor, S. R., G. Santpere, A. Weinreb, A. Barrett, M. B. Reilly *et al.*, 2020 Molecular topography of an entire nervous system. bioRxiv**:** 2020.2012.2015.422897.
Tepper, R. G., J. Ashraf, R. Kaletsky, G. Kleemann, C. T. Murphy *et al.*, 2013 PQM-1 complements DAF-16 as a key transcriptional regulator of DAF-2-mediated development and longevity. Cell 154**:** 676-690.
Vowels, J. J., and J. H. Thomas, 1992 Genetic analysis of chemosensory control of dauer formation in Caenorhabditis elegans. Genetics 130**:** 105-123.
Webster, A. K., A. Hung, B. T. Moore, R. Guzman, J. M. Jordan *et al.*, 2019 Population Selection and Sequencing of Caenorhabditis elegans Wild Isolates Identifies a Region on Chromosome III Affecting Starvation Resistance. G3 (Bethesda) 9**:** 3477-3488.

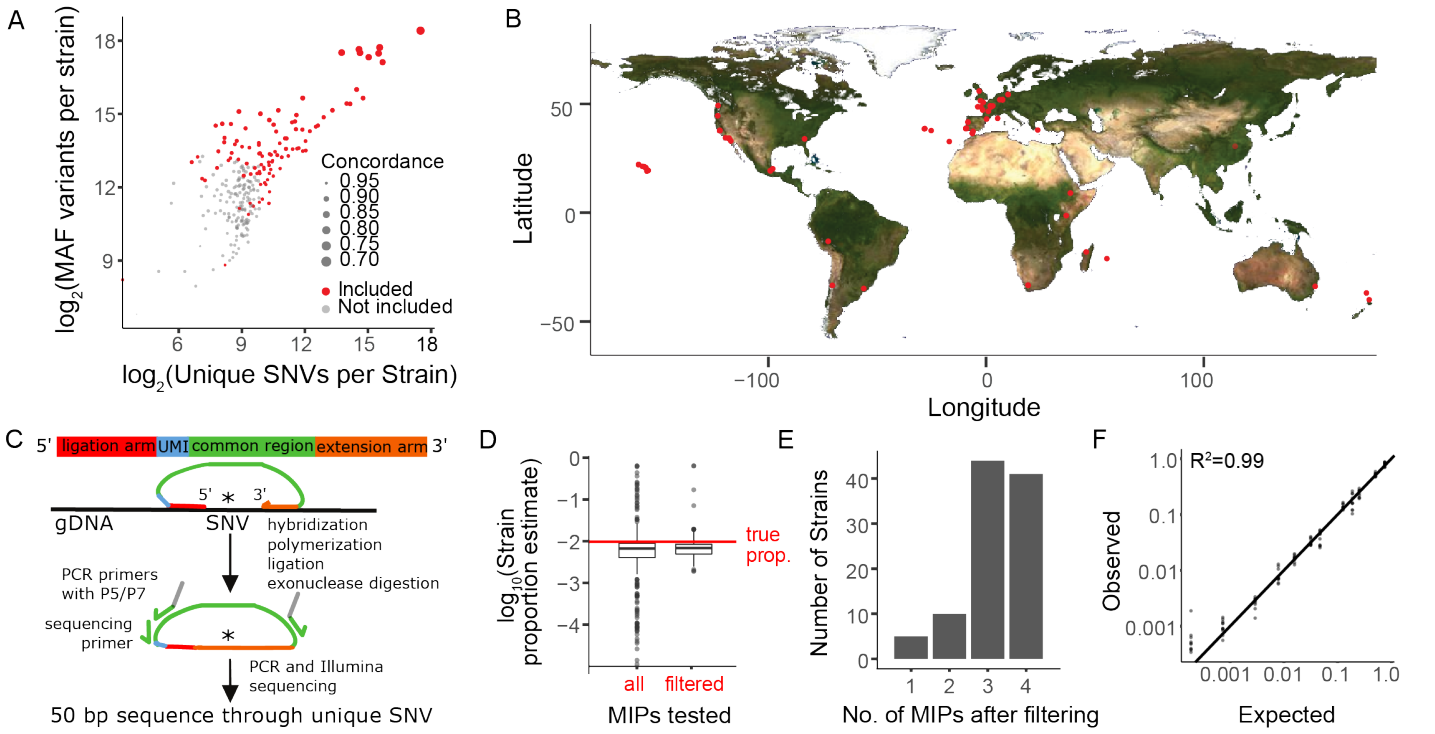


**Fig. 1. Sensitive and precise measurement of strain frequency in pooled culture using MIP-seq.** (**A**) The three metrics used to identify the most diverse *C. elegans* strains are plotted. "MAF" stands for minor allele frequency. Concordance refers to the average pairwise concordance for the focal strain compared to all other strains, which is calculated as the number of shared variant sites divided by the total number of variants for each pair. The strains included in the MIP-seq experiment are in red. (**B**) Geographic locations of the strains used. (**C**) Schematic of MIP-seq. MIPs are designed for loci with SNVs unique to each strain. Four MIPs were designed per strain. MIPs are 80 nt long and include ligation and extension arms to match DNA sequence surrounding the SNV, a unique molecular identifier (UMI), and P5 and P7 sequences for Illumina sequencing. MIPs are hybridized to genomic DNA, polymerized, ligated, and used as PCR template to generate an Illumina sequencing library. The alternative-to-total read frequency for each MIP/SNV locus indicates strain frequency. (**D**) Empirical testing of 412 MIPs with an equimolar mix of genomic DNA from 103 strains to identify reliable MIPs. 321 MIPs passed filtering and were analyzed in the starvation experiment. Outliers for filtered MIPs are for N2, which has hardly any unique SNVs because it is the reference genome, and N2 MIPs were included despite poor performance. (**E**) Number of MIPs per strain of the 321 filtered MIPs. (**F**). Genomic DNA was also combined at different known concentrations, and MIP-seq was used to generate a standard curve. $R^2$ = 0.99.

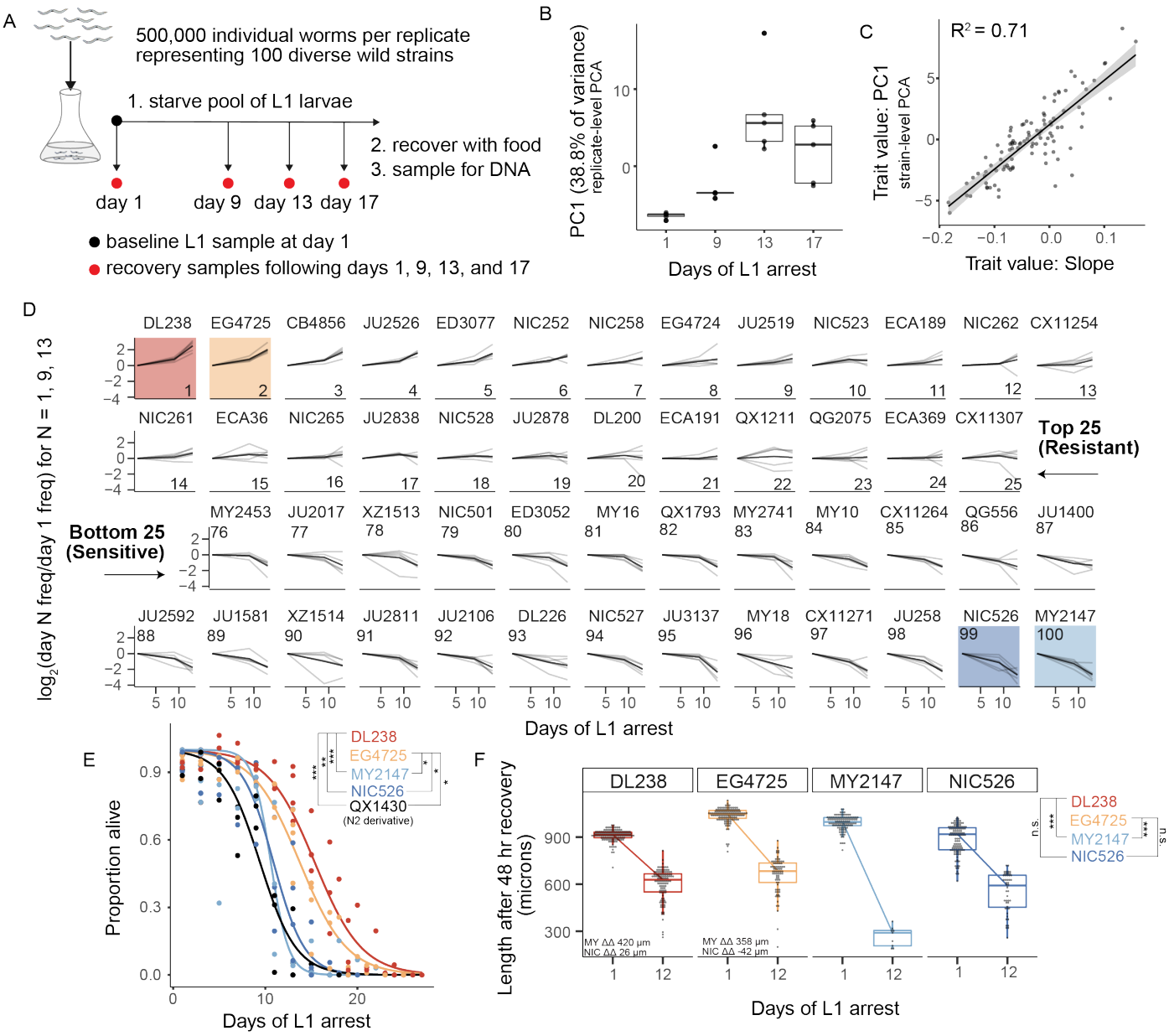


**Fig. 2. MIP-seq determines relative starvation resistance of 100 strains. (A)** Experimental design. Worms were starved at the L1 stage ('L1 arrest'). ~5,000 L1 larvae per strain were starved (~500,000 total), the population was sampled initially ('baseline'), and then sampled on the days indicated. Samples (except baseline) were allowed to recover with food in liquid culture and then frozen for DNA isolation. Five biological replicates were performed. (**B**) Principal component 1 of normalized and processed data from all replicates (replicate-level) and strains is plotted, revealing association with duration of starvation. Each point is an individual sample (MIP-seq library). (**C**) The relationship between two starvation-resistance metrics (Slope and PC1) produced from strain-level data (replicates averaged) is plotted. Each point is a different strain. (**D**) $Log_2$-normalized strain frequency is plotted over time for the 25 most resistant and 25 most sensitive strains in rank

order (based on Slope). Only days 1, 9, and 13 are plotted. See Supplementary Figure 2 for full data. Grey lines are biological replicates and black line is the mean. DL238 and EG4725 are most starvation-resistant, and NIC526 and MY2147 are most sensitive, and they are color-coded accordingly. (**E**) L1 starvation survival curves are plotted for starvation-resistant and sensitive strains. Individual replicate measurements are included as points to which curves were fit with logistic regression. T-tests on 50% survival time of four biological replicates. (**F**) Worm length following 48 hr recovery with food after 1 or 8 days of L1 starvation. ΔΔ indicates effect size of interaction between duration of starvation and strain data plotted in that panel compared to the strain listed (the difference in differences between strains' mean length at days 1 and 8). 'MY' is an abbreviation for MY2147 and 'NIC' is an abbreviation for NIC526. Linear mixed-effects model; one-way p-value of interaction between duration of starvation and strain. E, F. ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$

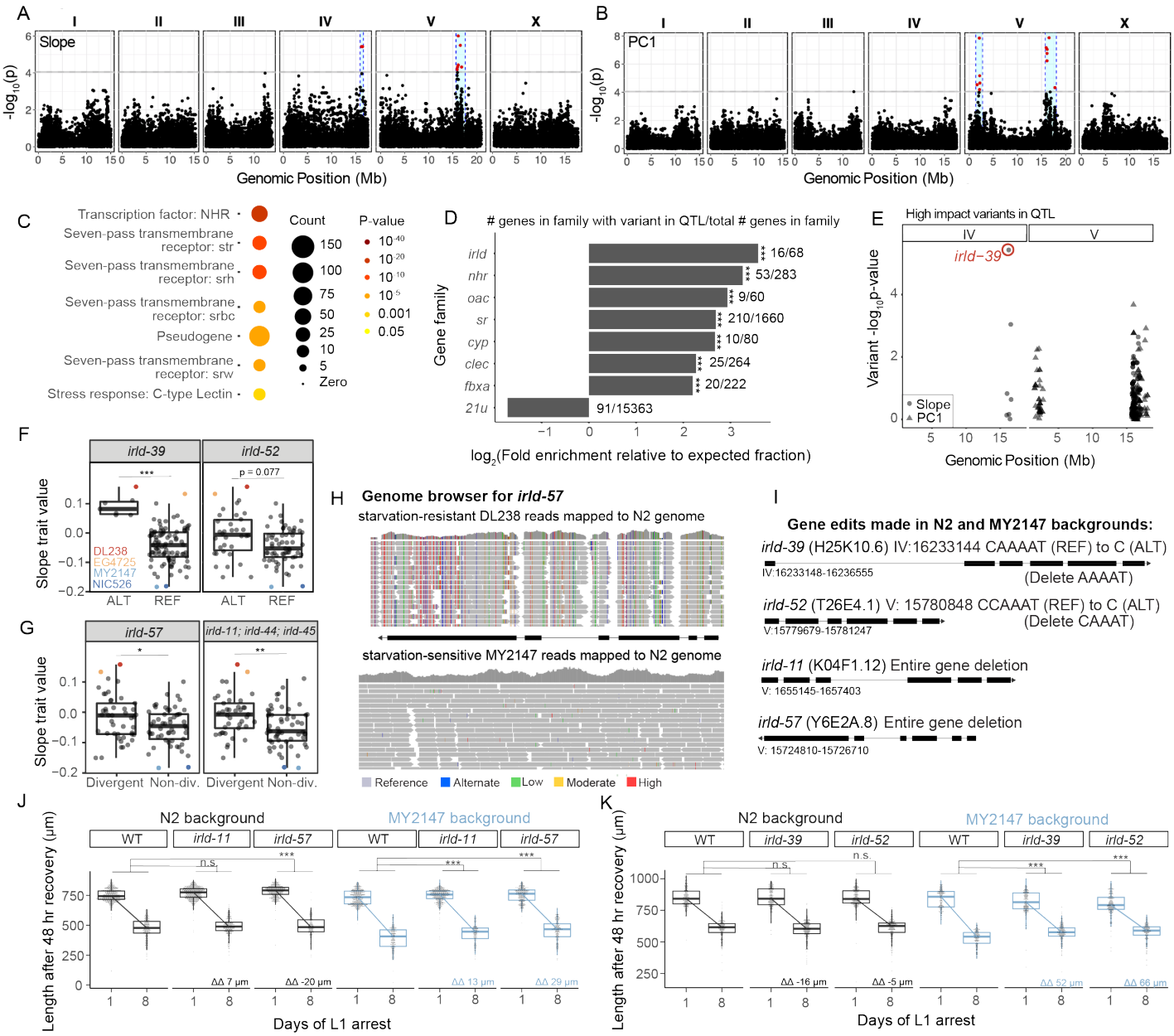


**Fig. 3. Genetic variation in the *irld* gene family underlies differences in starvation resistance. (A)** GWA output using Slope as a trait value. Significant QTL intervals are IV: 15939340-16613710 and V: 15660911-17615557. (**B**) GWA output using PC1 as a trait value. Significant QTL intervals are V: 1345848-2764788 and V: 15775895-18065050. (**C**) WormCat Category 3 enrichments for all genes with variants in the QTL. (**D**) Enrichment of gene families with at least 5 members with variants within QTL. (**E**) $Log_{10}$ p-value for all high-impact variants within QTL for Slope and PC1 GWA. The high-impact variant in *irld-39* is indicated in red. (**F**) Slope trait values for strains based on whether they have ALT and REF alleles for *irld-39* and *irld-52* high-impact variants. The *irld-52* high-impact variant p-value is p = 0.007 for PCA trait values (not shown). Significance determined from GWA fine mapping. (**G**) Slope trait values for strains based on whether they are

hyper-divergent or not at *irld-11* and *irld-57* loci. T-test on trait values between hyper-divergent and non-divergent strains. (**F-G**) DL238, EG4725, NIC526, and MY2147 are color-coded as indicated. (**H**) Genome browser screenshot for *irld-57* in a strain that is hyper-divergent, DL238, and a strain that is not hyper-divergent, MY2147. Gray rectangles are sequencing reads that map to the N2 reference genome. (**I**) Gene models for the four *irld* genes for which new strains were made to test *irld* function. For *irld-39 and irld-52,* N2 and MY2147 have the REF allele and were edited to have the ALT allele. *irld-11* and *irld-57* are hyper-divergent in DL238 and EG2745 backgrounds, so full gene deletions were generated in N2 and MY2147 backgrounds. (**J-K**). Worm length following 48 hr recovery with food after 1 or 8 d of L1 starvation for indicated genotypes. Linear mixed-effects model; one-way p-value for interaction between strain and duration of starvation; 4-5 biological replicates per condition. ΔΔ indicates effect size of interaction between duration of starvation and strain compared to control (the difference in differences between strains' mean length at days 1 and 8). ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, n.s. not significant

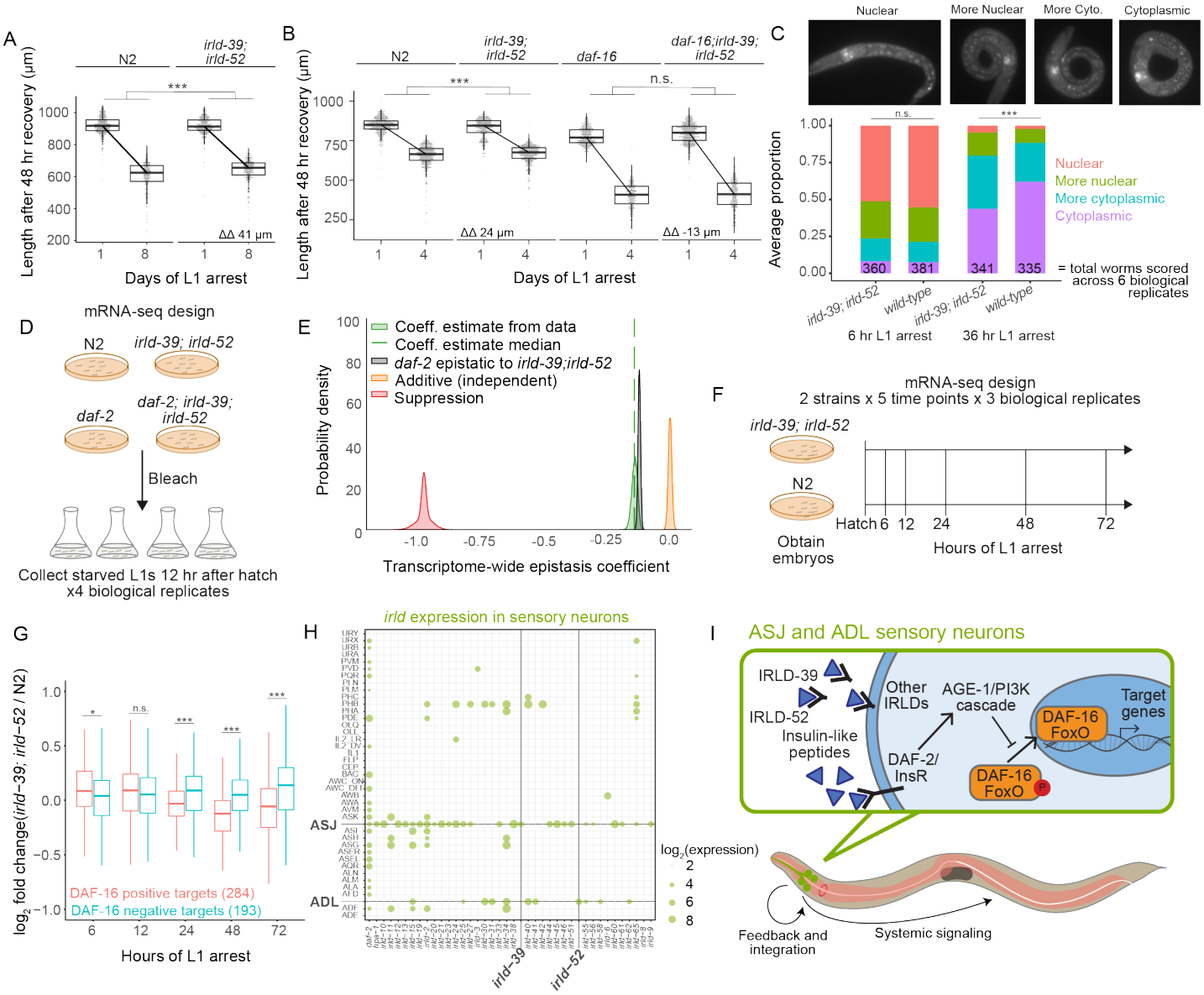


**Fig. 4. IRLD-39 and IRLD-52 regulate insulin/IGF signaling to affect starvation resistance.** (**A-B**) Worm length following 48 hr recovery with food after 1 or 8 d of L1 starvation for indicated genotypes. Linear mixed-effects model; two-way p-value for interaction between strain and duration of starvation; 4-5 biological replicates per condition. ΔΔ indicates effect size of interaction between duration of starvation and strain compared to control (the difference in differences between strains' mean length at days 1 and 8). (**C**) Nuclear localization of DAF-16 in intestinal cells of starved L1s at indicated time points following hatching. Representative images of the classification of nuclear localization are shown above the graph. The Cochran-Mantel-Haenszel test was used to determine differences in the distribution of the four categories between *daf-16(ot971)* and *daf-16(ot971); irld-39(duk1); irld-52(duk17)* at each time point. (**D**) mRNA-seq experimental design. N2, *daf-2(e1370), irld-39(duk1); irld-52(duk17), and daf-2(e1370), irld-39(duk1); irld-52(duk17)* were

bleached and collected as starved L1s 12 hours after hatch. (**E**) Probability densities are plotted for transcriptome-wide epistasis coefficients based on formal models of different types of genetic interactions (additive, epistasis, suppression) as well as the observed epistasis coefficient computed from mRNA-seq results with all four genotypes. (**F**) mRNA-seq experimental design. N2 and *irld-39(duk1); irld-52(duk17)* were bleached and collected during L1 arrest at the indicated time points. (**G**) Boxplots of $\log_2$ fold-changes between *irld-39(duk1); irld-52(duk17)* and wild type (N2) for DAF-16 positive and negative targets (pink and teal lines, respectively) (TEPPER *et al.* 2013; KAPLAN *et al.* 2015). Kolmogorov-Smirnov test used to assess differences in distribution of $\log_2$ fold changes at each time point. (**H**) $\text{Log}_2$-normalized expression levels of *irld* genes in sensory neurons. Bolded horizontal lines indicate ASJ and ADL neurons, and bolded vertical lines indicate *irld-39* and *irld-52.* (**I**) Model for function of IRLD-39 and IRLD-52. ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, n.s. not significant.

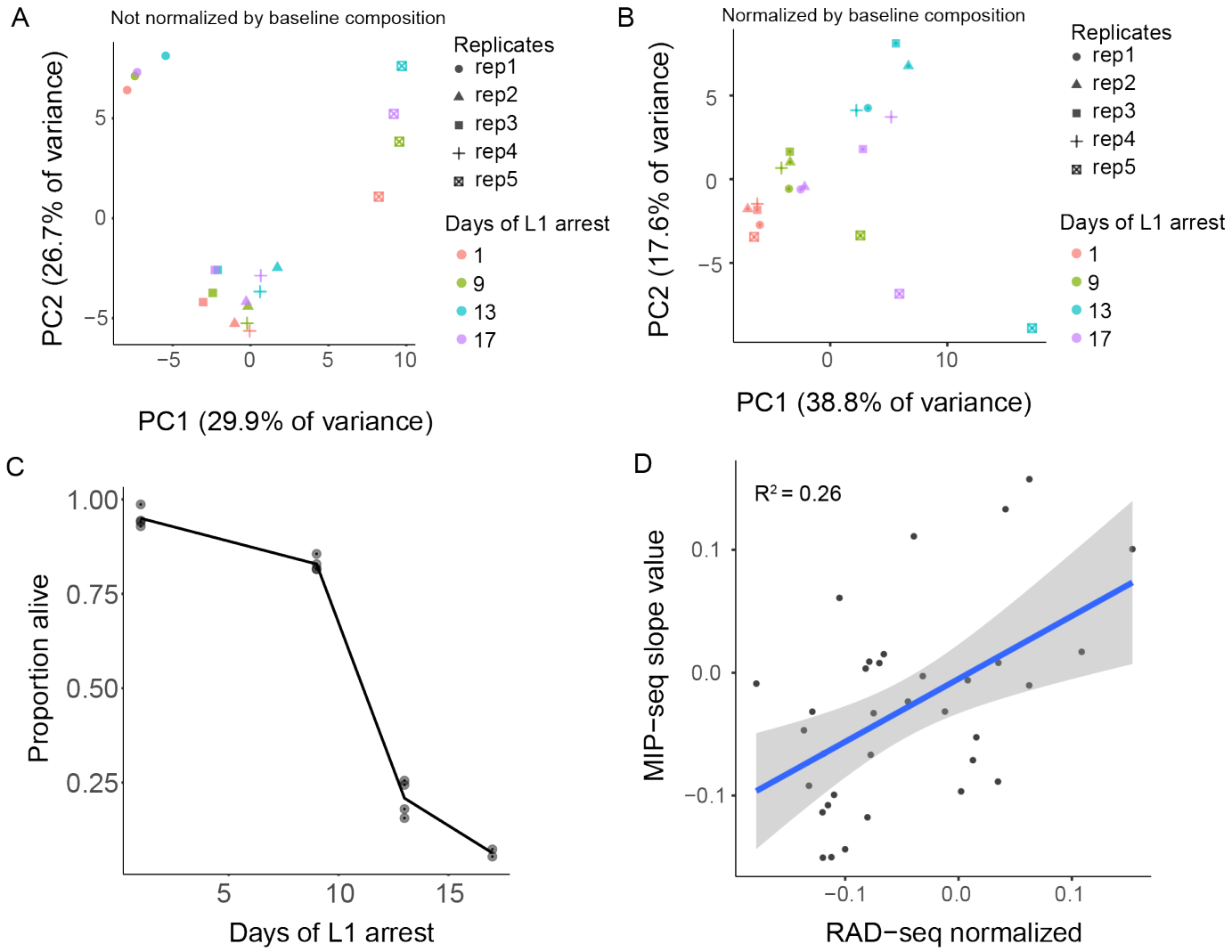


**Fig. S1. MIP-seq data analysis and comparison to RAD-seq analysis of starvation resistance. (A**) PCA of all replicates and recovery time points assayed not normalized by baseline composition. Samples cluster by replicate. (**B**) PCA of all replicates and recovery time points normalized by baseline composition of each replicate pool on day 1 of starvation without recovery (See Fig. 2A). PC1 is plotted in Fig. 2B. (**C**). Survival of cultures used to collect samples for MIP-seq experiment (Fig. 2) at days 1, 9, 13, and 17. (**D**) Comparison of MIP-seq trait value Slope to similar trait value derived from published results using RAD-seq for population sequencing (*17*). 34 strains included in both experiments are plotted. RAD-seq results are based on only a single replicate, and the vast majority of reads from RAD-seq do not cover unique SNV greatly limiting sequencing depth at informative loci. $R^2 = 0.26$ ($p = 0.002$).

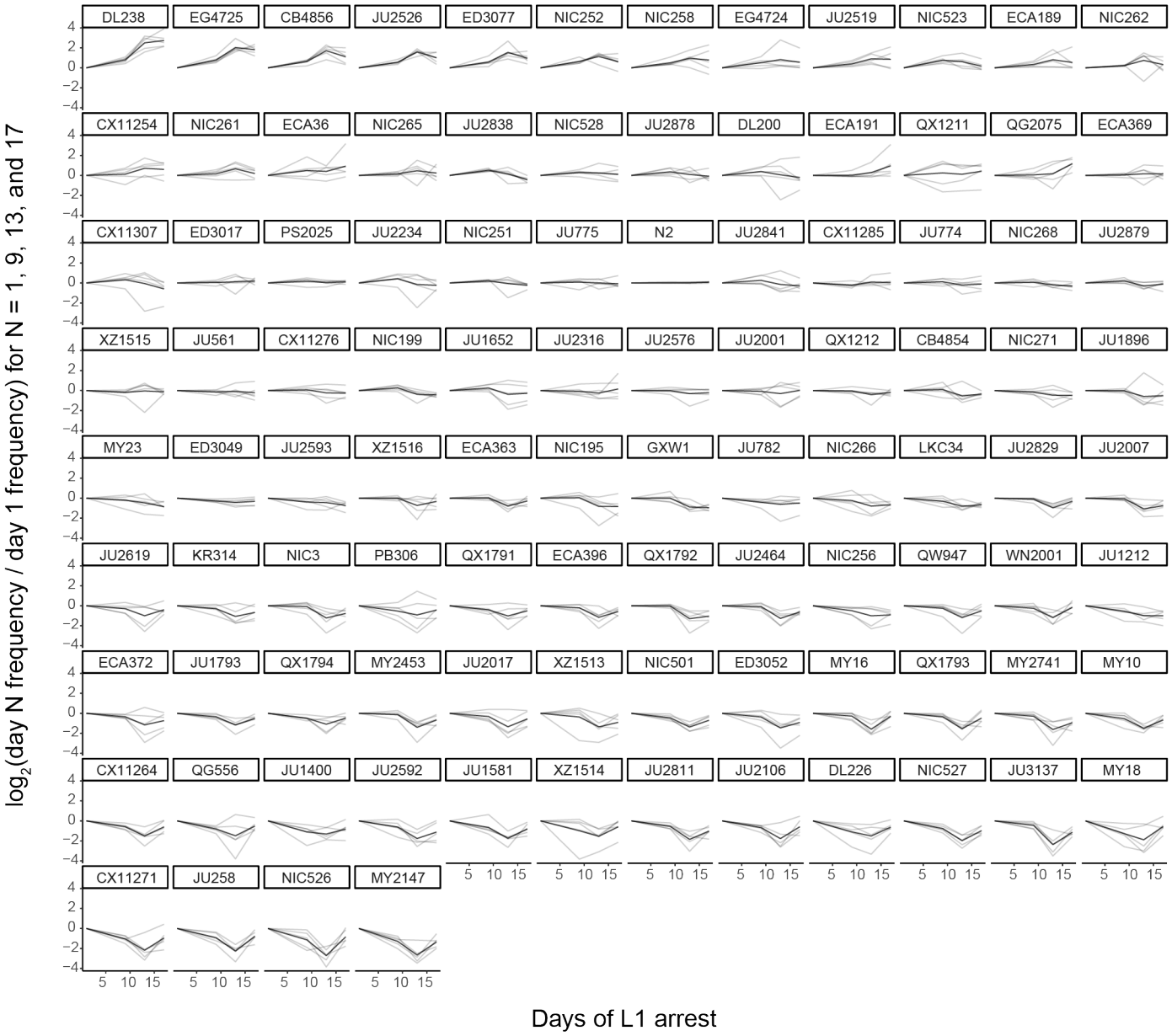


**Fig. S2. MIP-seq strain frequencies throughout starvation for all strains assayed.** MIP-seq replicates were first normalized by baseline and then by day 1. Strains are rank-ordered from most starvation-resistant to most sensitive by Slope trait values. Gray lines are individual replicates and black lines are the mean for each strain.

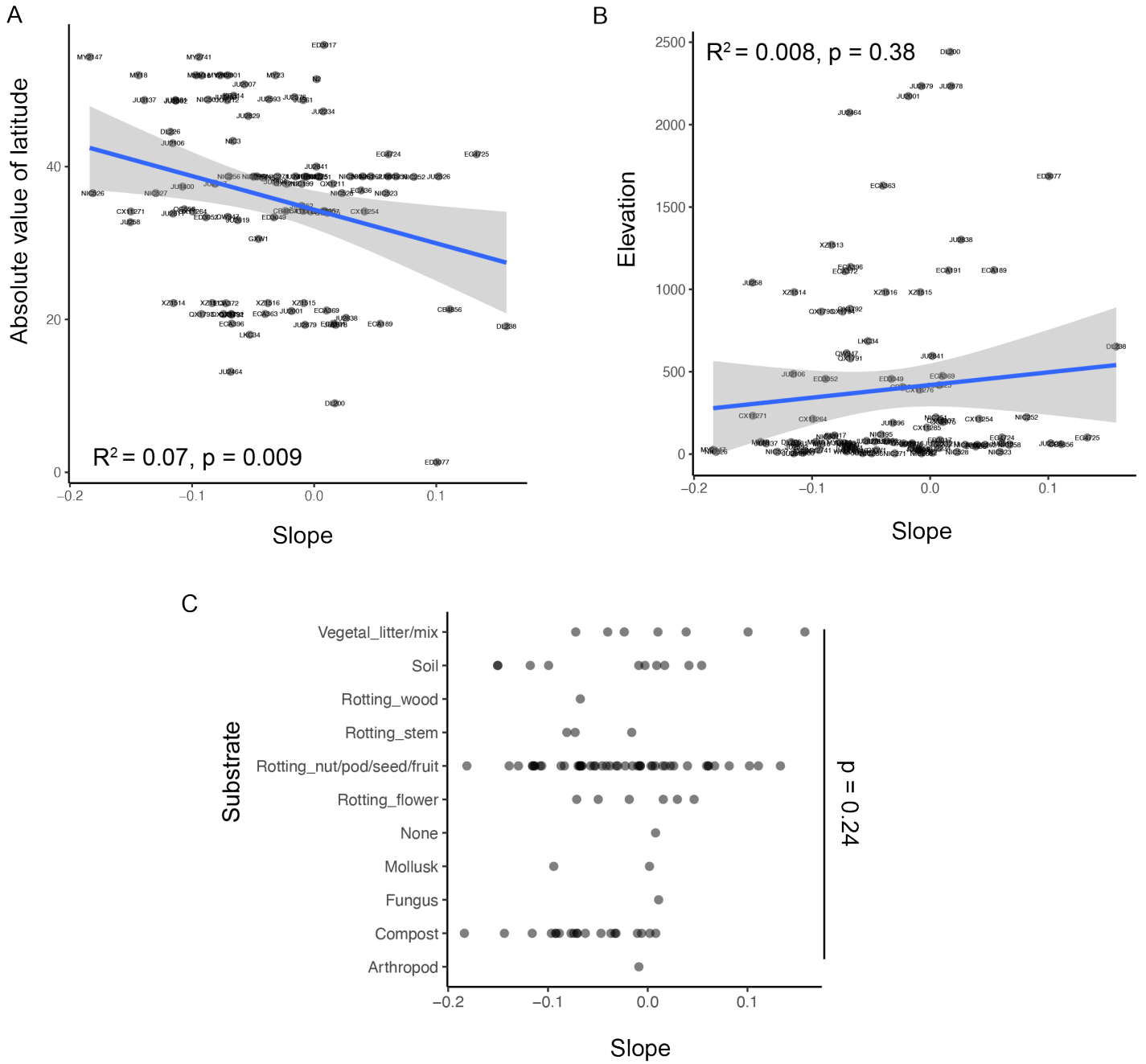


**Fig. S3. Starvation resistance of wild strains is associated with latitude at collection site, but not elevation or substrate.** (**A**) 'Slope' trait value for each strain plotted against the absolute value of its latitude of collection. $R^2 = 0.07$, p = 0.009. (**B**) 'Slope' trait value for each strain plotted against the elevation from which strains were collected. $R^2 = 0.008$, p = 0.38. (**A-B**) Significance determined by t-test on t-statistic of slope coefficient from linear model. (**C**) Slope trait values plotted for strains collected from each type of substrate. One-way ANOVA across all substrates, p = 0.24.

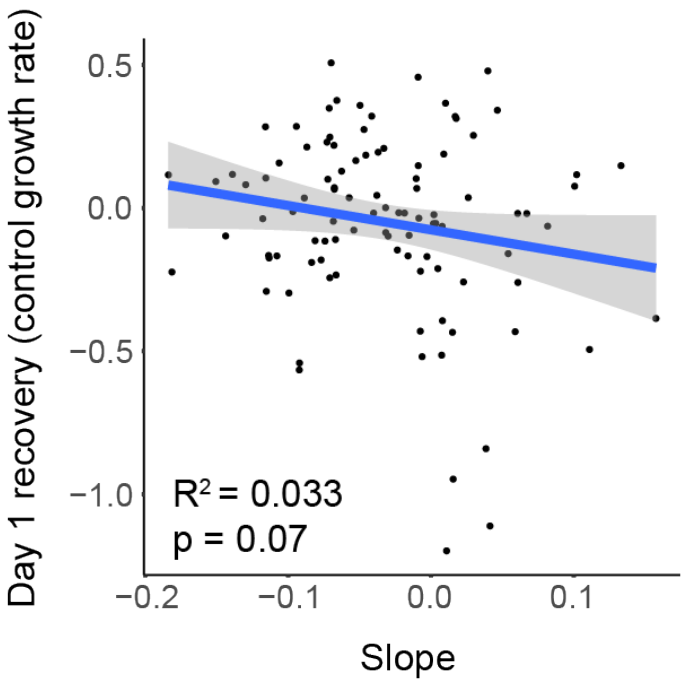


**Fig. S4. Starvation resistance is negatively correlated with growth rate.** Slope trait value is plotted against 'Early Growth' trait value determined by determining which strains became over- or under-represented after recovery from day 1 of L1 starvation relative to baseline composition at day 1 of L1 starvation. $R^2 = 0.033$, significance determined by t-test on t-statistic of slope coefficient from linear model.

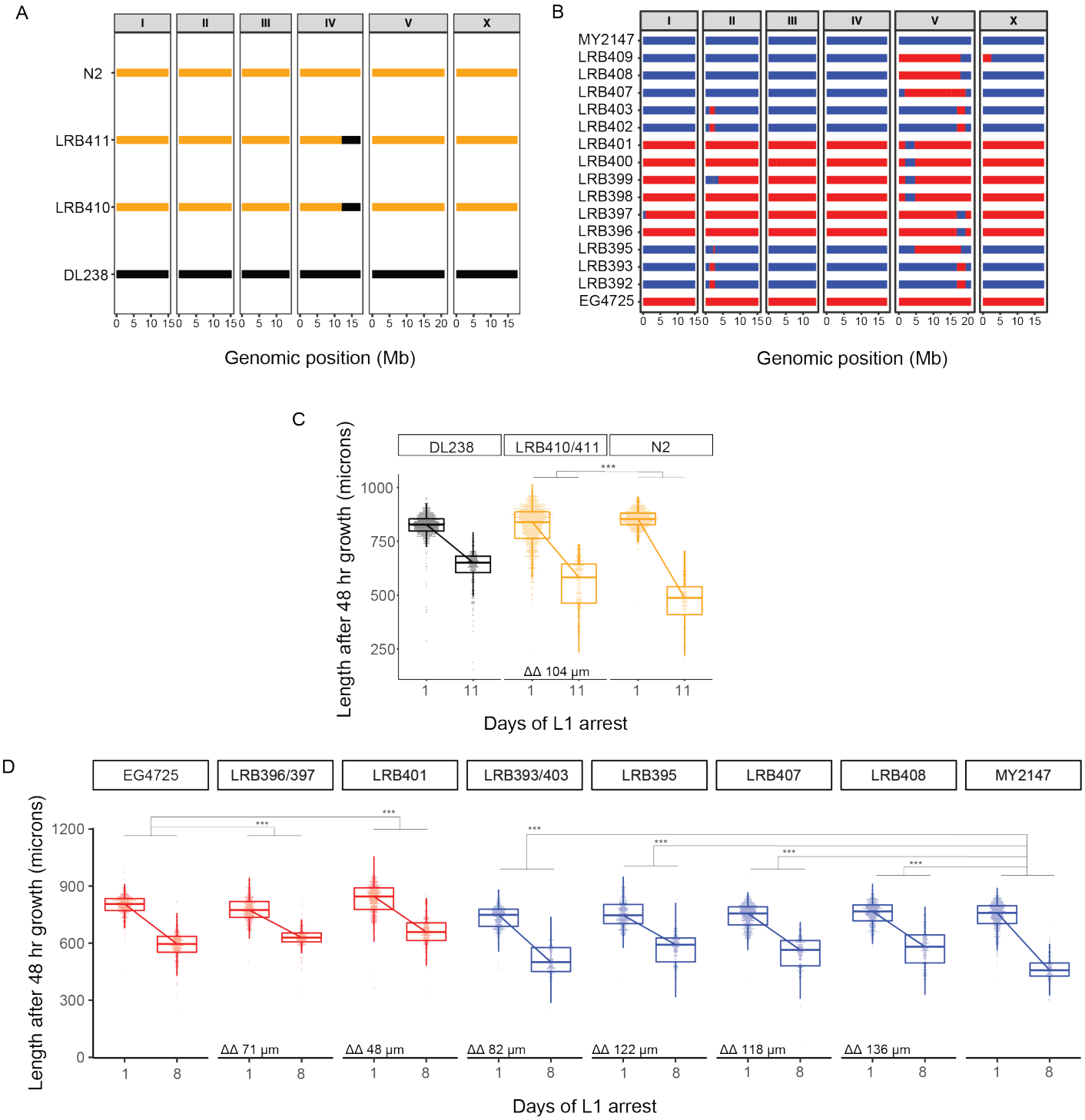


**Fig. S5. Validation of QTL with sequenced near-isogenic lines (NILs).** (**A**) Schematic of low-depth (~1x) genome-sequencing results of NILs used to validate chromosome IV QTL. Parental strains are N2 and DL238. DL238 can be introgressed into N2 but not *vice versa* due to genetic incompatibility (*46*). (**B**) Schematic of low-depth (~1x) genome-sequencing results of NILs used to validate chromosome V QTL. Parental strains are MY2147 and EG4725, which are compatible and were introgressed reciprocally. (**C**) Worm length following 48

hr recovery with food after 1 or 11 d of L1 starvation for indicated strains. Data were merged for essentially duplicate strains LRB410 and LRB411. (**D**) Worm length following 48 hr recovery with food after 1 or 8 d of L1 starvation for indicated strains. Data were merged for essentially duplicate strain pairs LRB396 and LRB397, and LRB393 and LRB403. (**C-D**) ***$p < 0.001$; linear mixed-effects model; p-value of interaction between strain and duration of starvation; eight biological replicates. ΔΔ indicates effect size of interaction between duration of starvation and strain data plotted in that panel compared to the predominant parental strain background (the difference in differences between strains' mean length at days 1 and 8).

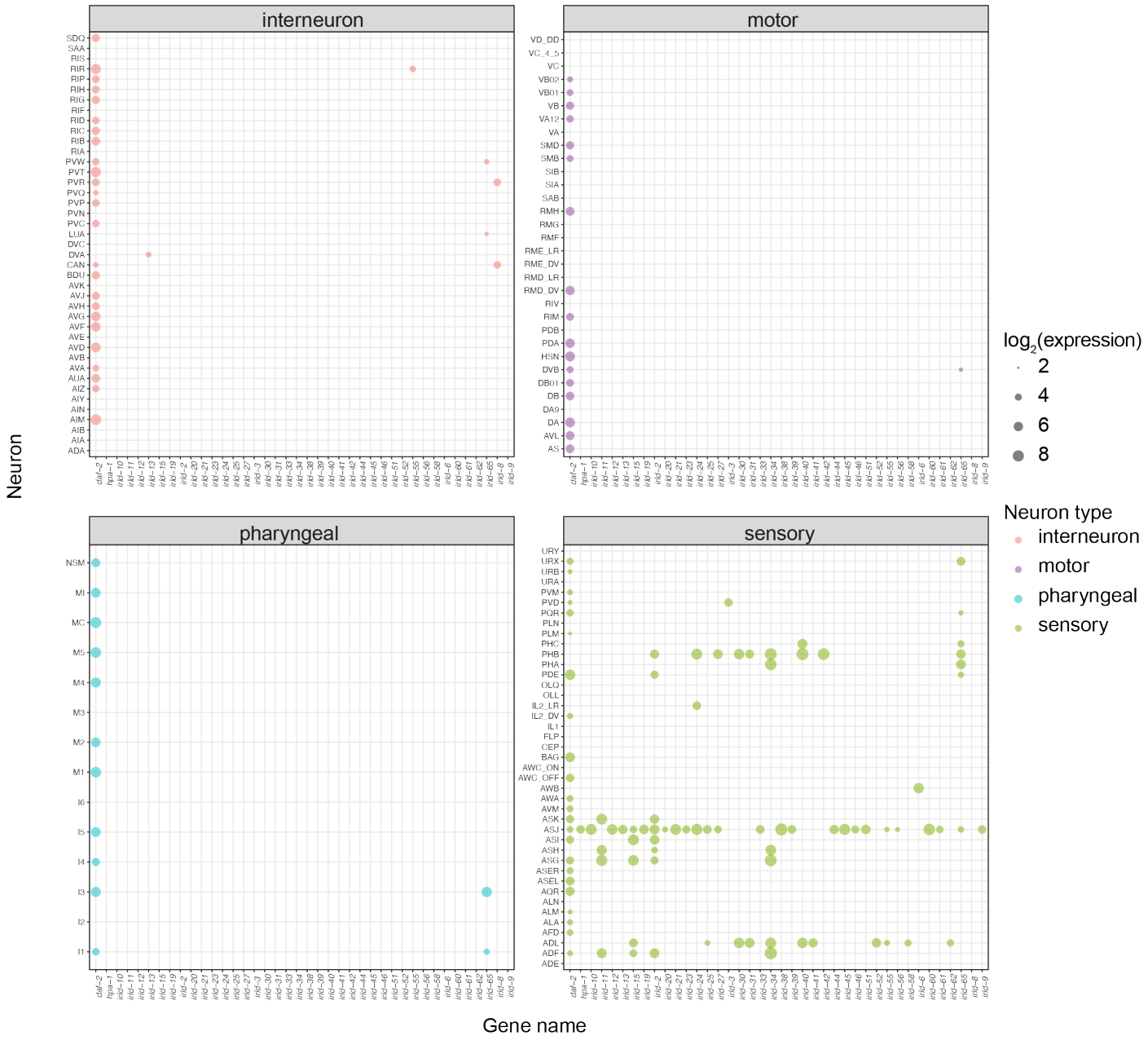


**Fig. S6. *irld* genes are expressed in specific sensory neurons but generally not other neuron types.** Single-cell RNA-seq expression levels from L2-stage larvae for all neurons are plotted for all *irld* genes reported with *daf-2* included for comparison (*26*). Neurons are sorted into interneuron, motor, pharyngeal, and sensory classifications. Sensory neurons plotted in Fig 4H shown again here for comparison. Bubble size indicates $\log_2$ expression level.

**Data S1. (Data_S1.xlsx)**

This file includes all MIP-seq processed data: Slope and PC1 trait values used in GWAS, output from MIPgen, MIP sequences, MIPs included in the final experiment, MIP primer sequences, count data for MIP-seq starvation resistance experiment and two pilot experiments.

**Data S2. (Data_S2.xlsx)**

This file includes GWAS output and follow-up on *irld* candidates: GWAS output for both Slope and PC1, genes within QTL, output from WormCat and enrichment analyses, hyper-divergence status for each strain, CRISPR sequences, genotyping primers, and NIL sequencing results.

**Data S3. (Data_S3.xlsb)**

This file includes output from RNA-seq analysis: non-normalized count data, counts-per-million-normalized count data, and output from differential expression analysis.