Natural variation in the *irld* gene family affects starvation resistance in *C. elegans*

Amy K. Webster[1,5], Rojin Chitrakar[1], Maya Powell[1,6], Jingxian Chen[1], Kinsey Fisher[1], Robyn Tanny[2], Lewis Stevens[2,7], Kathryn Evans[2], Angela Wei[1], 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 60208, 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.

[5]Current address: Institute of Ecology and Evolution, University of Oregon, Eugene, OR 97403, USA.

[6]Current address: Environment, Ecology, and Energy Program, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514, USA

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

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

ABSTRACT (239 words)

Starvation resistance is important to disease and fitness, but the genetic basis of its natural variation is unknown. We developed a synthetic-population (re)sequencing approach using molecular inversion probes (MIP-seq) to measure relative fitness during and after larval starvation in *C. elegans*. We applied this competitive assay to 100 genetically diverse, sequenced, wild strains, revealing natural variation in starvation resistance. We confirmed that the most starvation-resistant strains survive and recover from starvation better than the most starvation-sensitive strain, MY2147, using standard assays. We performed genome-wide association with the MIP-seq trait data and identified three quantitative trait loci (QTL) for starvation resistance. These QTL contain several members of the Insulin/EGF Receptor-L Domain (*irld*) family with sequence variation associated with variation in starvation resistance. We used genome editing to show that individual modification of four *irld* genes increases starvation resistance of MY2147. Modification of *irld-39* and *irld-52* together increases starvation resistance of the laboratory-reference strain N2. Increased starvation resistance of the *irld-39; irld-52* double mutant depends on *daf-16/FoxO*, and these worms also show increased nuclear localization of DAF-16 during starvation. DAF-16/FoxO is a widely conserved transcriptional effector of insulin/IGF signaling (IIS), and these results suggest that IRLD proteins modify IIS. This work demonstrates efficacy of using MIP-seq to dissect a complex trait, identifies *irld* genes as natural modifiers of starvation resistance in *C. elegans*, and suggests that an expanded gene family affects a deeply conserved signaling pathway to alter a fitness-proximal trait.

INTRODUCTION

Enduring periods of starvation is a near-ubiquitous feature of animal life that affects survival, growth, and reproduction, making starvation resistance a fitness-proximal trait. Starvation resistance is also important to human health and disease, with direct relevance to diabetes, obesity, aging, and cancer. However, the genetic basis of natural variation in starvation resistance is unclear, despite its importance to understanding animal evolution and informing therapeutic strategies.

The nematode *C. elegans* is frequently starved in the wild and has robust starvation responses (SCHULENBURG AND FÉLIX 2017; BAUGH AND HU 2020). Larvae that hatch in the absence of food arrest development in the first larval stage (L1 arrest or L1 diapause) and can survive starvation for weeks (BAUGH 2013). In addition to causing mortality, extended starvation reduces growth and reproductive success upon feeding (JOBSON *et al.* 2015; JORDAN *et al.* 2019), and these effects can be uncoupled (BAUGH AND HU 2020). For example, mutations affecting the endoplasmic reticulum unfolded-protein response do not affect starvation survival during L1 arrest, but they render starved worms incapable of growth after feeding (ROUX *et al.* 2016). Likewise, sensory perception of food during L1 starvation does not affect survival but causes irreversible developmental arrest (KAPLAN *et al.* 2018). In addition, *daf-18/PTEN* mutant worms display reduced L1 starvation survival and growth upon recovery, but their reproduction is particularly sensitive to starvation, with only 12 hours of L1 starvation causing dramatic decline in brood size, including sterility in some individuals (CHEN *et al.* 2022). Starvation resistance therefore integrates survival, growth rate, and reproductive success, and different genes and conditions can affect these three phenotypes independently.

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). Insulin/IGF signaling (IIS) is a critical regulator of *C. elegans* L1 arrest (BAUGH AND STERNBERG 2006). There is a single known insulin/IGF-like receptor in *C. elegans*, DAF-2/InsR, which signals through a conserved phosphatidylinositol 3-kinase (PI3K) signaling pathway to

antagonize the forkhead boxO transcription factor DAF-16/FoxO (LIN *et al.* 1997; OGG *et al.* 1997). When IIS is active, DAF-16 is phosphorylated by AKT-1/AKT, sequestering it in the cytoplasm, and when IIS is reduced, such as during starvation, DAF-16 moves to the nucleus and regulates transcription (HENDERSON AND JOHNSON 2001; LEE *et al.* 2001; LIN *et al.* 2001). *daf-16/FoxO* promotes starvation resistance, with *daf-16* and *daf-2* mutants displaying starvation sensitivity and resistance, respectively (MUNOZ AND RIDDLE 2003; BAUGH AND STERNBERG 2006; HIBSHMAN *et al.* 2017)

IIS is subject to extensive regulation, presumably reflecting its critical role in maintaining metabolic and developmental homeostasis during fluctuations in nutrient availability. There are 40 insulin-like peptides (ILPs) encoded by the *C. elegans* genome, and they function as either agonists or antagonists of DAF-2/InsR (PIERCE *et al.* 2001). There is evidence of functional redundancy, but genetic analysis also suggests functional specificity among ILPs beyond their broad classification as either agonists or antagonists (FERNANDES DE ABREU *et al.* 2014). ILPs are expressed in sensory and other neurons, the intestine, and other cell types (RITTER *et al.* 2013), and they contribute to systemic regulation (CHEN AND BAUGH 2014; HUNG *et al.* 2014). IIS and *daf-16/FoxO* regulate transcription of most ILPs, producing extensive feedback regulation (KAPLAN *et al.* 2019), and comprising a putative insulin-to-insulin signaling network (FERNANDES DE ABREU *et al.* 2014). The *daf-2/InsR* locus encodes a truncated transcript isoform, *daf-2b*, that produces a product lacking a cytoplasmic domain such that it is incapable of signaling. Instead, it presumably binds ILPs, competing with full-length DAF-2 isoforms (MARTINEZ *et al.* 2020). In a search for other potential insulin/IGF receptors, *C. elegans* proteins with homology to various insulin receptors were identified by PSI-BLAST, leading to the discovery of a large family of proteins with weak homology to DAF-2/InsR (DLAKIC 2002). These extracellular proteins lack a tyrosine kinase domain, like DAF-2B, but they have an atypical ligand-binding domain that also shares homology with epidermal growth factor (EGF) receptors. 68 members of this Insulin/EGF-Receptor L Domain (IRLD) protein family are now recognized in *C. elegans* (HOBERT 2013). It was originally hypothesized that the IRLD proteins function as decoy insulin/IGF receptors (DLAKIC 2002), similar to how DAF-2B is thought to function (MARTINEZ *et al.* 2020),

87 but genetic analysis of two of them (*hpa-1* and *hpa-2*) suggests that they modify EGF signaling to affect
88 healthspan (IWASA *et al.* 2010). The IRLD family is not conserved outside of the genus (Oliver Hobert,
89 personal communication), but there is a family of IGF-binding proteins in vertebrates that is thought to fine-
90 tune IGF signaling by affecting circulation and receptor binding of IGF proteins (ALLARD AND DUAN 2018).

91 Hundreds of genome-sequenced, wild *C. elegans* strains are available for genetic analysis of
92 natural variation on the *Caenorhabditis elegans* Natural Diversity Resource (CeNDR) (COOK *et al.* 2017).
93 Short-read sequencing was mapped to the N2 reference genome, and millions of single-nucleotide variants
94 (SNVs) were called. Analysis of these SNVs in conjunction with read mapping efficiency for each strain led
95 to the identification of hyper-divergent regions of the genome, which exhibit substantially more variation
96 than other regions of the genome for some strains (LEE *et al.* 2021). These regions are typically found on
97 the arms of chromosomes and are enriched for environmental-response genes.

98 We were interested in taking advantage of the genomic resources on CeNDR to investigate natural
99 variation of starvation resistance. Assaying starvation resistance for so many strains is prohibitively
100 laborious using traditional methods. We previously developed an approach based on pooling wild strains,
101 selecting for starvation resistance, and sequencing to identify relatively starvation-resistant and sensitive
102 strains (WEBSTER *et al.* 2019). This approach identified a quantitative trait locus (QTL), but it lacked power
103 since only very rare sequencing reads that include SNVs unique to a strain in the synthetic population
104 informed inference of relative strain frequency. In contrast, by capturing targeted sequences, molecular
105 inversion probe sequencing (MIP-seq) enables extremely deep sequencing of polymorphic loci
106 (CANTSILIERIS *et al.* 2017; MOK *et al.* 2017), providing sensitive and precise estimation of strain frequencies
107 in a complex pool as a means of high-throughput phenotyping. Here, we used MIP-seq to phenotype
108 starvation resistance in a pool of genetically diverse wild *C. elegans* strains. We identified three QTL and
109 four members of the *irld* gene family that affect starvation resistance, and our results suggest that they do
110 so at least in part by modifying IIS.
111

RESULTS

Sensitive and precise measurement of strain frequency in pooled culture using MIP-seq

We selected 103 genetically diverse, wild *C. elegans* strains from around the world including the laboratory reference N2 to test MIP-seq, ultimately phenotyping 100 for starvation resistance (Fig 1A-B). Molecular inversion probes (MIPs) are designed to capture a specific region of the genome for targeted multiplex sequencing (Fig 1C). We designed MIPs to target a region containing a SNV unique to each of 103 strains. Thus, the relative frequency of each strain in a pool can be determined by the SNV frequency. We designed four such MIPs per strain to provide redundancy and increase precision. To pilot MIP-seq, we prepared sequencing libraries from an equimolar mix of genomic DNA from each of 103 strains. We determined the frequency of strain-specific reads for each MIP, and we censored probes that produced frequencies substantially different than the expected value of approximately 0.01 (Fig 1D; criteria in Materials and Methods), leaving three or four reliable probes for 85% of strains and at least one MIP for 100 strains (Fig 1E, Data S1), which were used in the starvation resistance experiment. Three strains with no MIPs passing filtering were excluded from subsequent analysis. As an additional pilot, we mixed genomic DNA from a subset of strains 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 could theoretically expand the dynamic range.

Using MIP-seq to characterize natural variation of starvation resistance

We used MIP-seq to phenotype 100 diverse strains for starvation resistance. We cultured the strains in standard laboratory conditions, pooled them, and subjected them to starvation during L1 arrest. We aimed for approximately 5,000 L1 larvae per strain in the pooled starvation culture in order to ensure representative sampling. However, we expected actual representation to vary to some extent across strains and replicates, so we collected DNA from an aliquot of L1 larvae on the first day of starvation as a "baseline" sample to capture initial population composition. In addition, aliquots were taken from the pool at

137 days 1, 9, 13, and 17 of starvation, and sampled larvae were allowed to recover with food for four or five
138 days (depending on the duration of starvation), enabling reproduction for one day, and then the entire
139 population was collected for DNA preparation (Fig 2A). DNA from baseline samples, as well as samples
140 allowed to recover and reproduce following starvation, were sequenced with MIP-seq for each of five
141 biological replicates. It is critical to point out that by incorporating recovery and early fecundity, this
142 sampling scheme integrates effects of starvation on mortality as well as growth rate and reproductive
143 success, each of which are important for starvation resistance (*i.e.*, fitness) and can be uncoupled by
144 certain genotypes and conditions.

145 DNA from 'baseline' samples allowed us to effectively normalize differences in pool composition in
146 each replicate, revealing effects of starvation on strain frequency. Differences in pool composition
147 explained the first principal component in principal component analysis (PCA) when strain frequencies over
148 time were analyzed without consideration of baseline frequencies (Fig S1A). However, once the data were
149 normalized for initial strain composition using the baseline sample for each replicate, the first principal
150 component correlated with duration of starvation, especially across the first three time points (Fig 2B, Fig
151 S1B). Substantial mortality occurred by day 17 (Fig S1C), and day 17 recovery cultures thus produced
152 relatively few progeny. Consequently, differences in strain frequencies were actually smaller at day 17 than
153 13, but relative differences were conserved (Fig S2). Therefore, after normalization, duration of starvation
154 is the major factor accounting for differences in strain frequency across all samples, and this is robust to
155 differences in the initial composition of the pool across replicates.

156 We developed two metrics to quantify relative starvation resistance for each strain. "Slope" is a
157 measure of how much a strain increases or decreases in frequency over time across days 1, 9, and 13,
158 calculated as the slope of a linear model (Data S1). "PC1" is the value of the first principal component for
159 each strain from strain-level PCA (Fig 2B, Fig S1B). These two metrics are correlated but also show some
160 differences (Fig 2C), suggesting they capture related but also distinct features of the data. While Slope is
161 intuitive, it is limited by the use of a linear model. Nonetheless, Slope values are correlated with starvation-

162 resistance values produced from a previously published population-sequencing approach with less power
163 that included some of the same strains (Fig S1D) (WEBSTER *et al.* 2019). In addition, Slope is modestly
164 correlated with the latitude from which strains were collected, suggesting possible adaptation to starvation
165 or other correlated traits based on location (Fig S3). There is also a modest negative correlation between
166 Slope and growth rate after only one day of starvation (control condition) (Fig S4), suggesting a possible
167 trade-off between starvation resistance and population growth rate in the absence of stress. We used the
168 Slope metric to order strains from most resistant to sensitive, revealing differences in starvation resistance
169 between wild strains (Fig 2D, Fig S2). In contrast to Slope, PC1 does not assume linearity, it includes the
170 results from day 17 of starvation, and it may be less affected by noise. PCA is also a common way to
171 obtain trait values for genome-wide association (GWA) studies (RIED *et al.* 2016; YANO *et al.* 2019).

172 Our recovery-based sampling approach integrates starvation survival, recovery, and early fecundity
173 into a single fitness assay. It is therefore unclear whether a given strain is more or less resistant because of
174 differences in mortality, growth rate, or progeny production. For example, a particular strain could survive
175 no better than others, but it could grow faster upon recovery or retain greater reproductive success in order
176 to display relative resistance to starvation. Conversely, a strain may not survive as long as others despite
177 comparable growth or fecundity upon recovery in order to display relative sensitivity to starvation. It is also
178 unclear what the absolute effect sizes are between the most resistant and sensitive strains in this
179 competition assay. Nonetheless, our approach is intended to model the impact of larval starvation on
180 fitness broadly, while traditional assays can be used to isolate specific effects of starvation on survival,
181 growth, and reproduction in follow-up experiments.

182 We performed manual assays for starvation survival, growth rate, and early fecundity for the most
183 resistant and sensitive strains. We found starvation-resistant strains DL238 and EG4725 survived
184 starvation significantly longer during L1 arrest than sensitive strains MY2147 and NIC526 (Fig 2E).
185 Differences in starvation survival among wild strains are relatively small compared to many published
186 mutants in the N2 reference background (BAUGH AND HU 2020), suggesting that large-effect mutations are

187 purged by selection. After extended L1 arrest, DL238 and EG4725 recovered from starvation better than
188 MY2147 but not NIC526, as assessed by their size following 48 hours of recovery (Fig 2F). Finally, DL238
189 and EG4725 exhibited a larger early brood size following extended starvation compared to both MY2147
190 and NIC526. Overall, this demonstrates that differences in starvation resistance among wild strains are
191 driven by differences in survival, recovery, and early fecundity, but that sensitivity of NIC526 is apparently
192 driven by differences in survival and early fecundity without an appreciable effect on growth. These results
193 validate the MIP-seq approach and reveal the extent of natural variation in starvation resistance.

194

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

196 We used Slope and PC1 as trait values to perform GWA on CeNDR (COOK *et al.* 2017). GWA
197 identified QTL on the right arm of chromosome IV and on the left and right arms of chromosome V (Fig 3A-
198 B, Data S2). We confirmed that each QTL affected starvation resistance by generating near-isogenic lines
199 and measuring growth rate upon recovery from starvation (Fig S5). We chose this assay, as opposed to
200 starvation survival or early fecundity, because it revealed relatively robust differences between
201 DL238/EG4725 (resistant) and MY2147 (sensitive) (Fig 2F).

202 These QTL include enrichments for several large gene families. WormCat analysis identified
203 significant enrichments of serpentine receptors, nuclear hormone receptors, and C-type lectins (Fig 3C)
204 (HOLDORF *et al.* 2020). Likewise, protein-domain enrichment analysis (FINN *et al.* 2011) identified seven-
205 pass transmembrane domains and hormone receptor domains (Fig 3D). In addition, the receptor L domain
206 was significantly enriched, which is found in proteins comprising Insulin/EGF-Receptor L Domain (IRLD)
207 family (DLAKIC 2002). Given weak homology to DAF-2/InsR, and the critical role of IIS in regulation of
208 starvation resistance, we were intrigued at the possibility that natural variation in *irld* family genes may
209 impact starvation resistance. Across all three QTL identified, there are genetic variants in 16 *irld* genes,
210 and 68 genes have been identified as part of this family in *C. elegans* (HOBERT 2013). Multiple variants are

211 present for each *irld* gene, and variants differed in the degree to which they were associated with variation
212 in starvation resistance (Fig 3E; each gene name is indicated for its most significantly associated variant).

213 We selected at least one *irld* gene from each QTL for functional analysis. We used multiple criteria
214 to select *irld* genes, including each variant's association with starvation resistance, the predicted functional
215 consequence of each variant (favoring variants that clearly disrupt protein function; for example, by
216 introducing an early stop codon, frameshift, etc.), and its context among other variants. On the left arm of
217 chromosome IV, a variant in *irld-39* was the strongest individual candidate both because of strong
218 association with starvation resistance and because the variant is predicted to disrupt the start codon of the
219 gene (Fig 3E, F, Data S2), likely rendering *irld-39* a functional null in the starvation-resistant strain DL238.
220 On the right arm of chromosome V, *irld-52* was identified through both Slope and PC1 phenotype metrics
221 and contains a variant associated with starvation resistance predicted to disrupt its fifth exon (Fig 3E, F). In
222 this case, it is likely that protein function is disrupted but unclear if the variant causes a null mutation. We
223 were thus interested in testing the function of both specific variants for *irld-39* and *irld-52* (Fig 3F). While
224 analyzing variants on the left arm of chromosome V, we noticed that many *irld* genes are adjacent to each
225 other and that each contain many variants, precluding identification of a single best candidate variant. In
226 particular, *irld-11, irld-44,* and *irld-45* are directly adjacent to each other, and each gene contains over 50
227 genetic variants. This pattern of some loci containing many variants relative to N2 has been observed on a
228 broader scale, leading to identification of so-called "hyper-divergent" regions of the genome that contain
229 exceptional amounts of variation (LEE *et al.* 2021). A genomic region for which at least some strains are
230 classified as hyper-divergent is considered a hyper-divergent region, but a given strain may or may not be
231 hyper-divergent in a hyper-divergent region depending on its genomic sequence. *irld-11, irld-44,* and *irld-45*
232 are part of a hyper-divergent region, and because they are in such close proximity, they are hyper-
233 divergent in the same strains. We considered hyper-divergence status across these three genes as a
234 summary of the variation in them, and found that hyper-divergence at these loci was associated with
235 starvation resistance (Fig 3G). Finally, *irld-57* provides another example of a gene overlapping with a

236 hyper-divergent region on the right arm of chromosome V, and hyper-divergence at this locus is also
237 associated with starvation resistance (Fig 3G). We thus aimed to follow-up on the *irld* gene family by
238 testing the functional consequences of variation in *irld-39, irld-52, irld-11,* and *irld-57.* Notably, the
239 associations between variants or hyper-divergence and Slope (Fig 3F, G) together with their inferred
240 negative impacts on gene function (Fig 3H) suggests that disruption of these *irld* genes in backgrounds
241 where they are functional will increase starvation resistance.

242 We used CRISPR-Cas9 genome editing to determine functional consequences of genetic
243 modification of our candidate *irld* genes. Because *irld-39* and *irld-52* contain singular variants associated
244 with starvation resistance and predicted to disrupt protein function, we generated these specific variants in
245 the starvation-sensitive MY2147 and the laboratory-reference N2 backgrounds (Fig 3H). Since *irld-11* and
246 *irld-57* contain so many candidate variants, we deleted these genes in MY2147 and N2, rendering them
247 null mutants at each locus (Fig 3H). We note that although the strains we generated are nulls for *irld-11*
248 and *irld-57*, it is possible that the starvation-resistant wild strains may retain some function in these genes.
249 In this sense, edits of *irld-39* and *irld-52* are more likely to approximate the effect of specific variants in the
250 wild, because they are the exact variants present in starvation-resistant wild strains. We tested all eight
251 strains and compared each to the appropriate control parental background, N2 or MY2147, to determine if
252 these alleles affect starvation resistance. None of the alleles in either background significantly affected
253 survival (Fig 3I, J). A power analysis suggests there is sufficient statistical power to detect differences of
254 approximately two days or greater, suggesting there is not a difference of at least this magnitude. However,
255 alleles for all four *irld* genes mitigated the effect of starvation on growth rate in the MY2147 background but
256 not N2 (Fig 3K, L). This suggests that MY2147, as a more starvation-sensitive background than N2,
257 facilitates detection of alleles that increase starvation resistance. In addition, these results suggest that
258 these *irld* genes may uncouple survival and recovery from starvation to impact starvation resistance. It is
259 also possible that we were able to detect an effect on starvation recovery simply because the assay is
260 more sensitive. In summary, these results show that multiple types of variants in different *irld* family

261 members reduce the effect of extended L1 starvation on recovery, identifying four individual genes from
262 one family that affect starvation resistance in wild strains.

263

264 **IRLD-39 and IRLD-52 act through DAF-16/FoxO to affect starvation resistance**

265 Because *irld-39(duk1)* and *irld-52(duk17)* alleles correspond to the exact variants present in
266 starvation-resistant wild strains, we chose to focus on them for further analysis. We hypothesized that
267 phenotypic effects may have been detected only in MY2147 because it is relatively sensitive to starvation,
268 and multiple *irld* variants may reveal an effect in N2. An *irld-39(duk1); irld-52(duk17)* double mutant in the
269 N2 background increased starvation survival but without statistical significance (Fig 4A). In this case there
270 was sufficient statistical power to detect differences of approximately 1.5 days or greater, suggesting there
271 is not a difference of at least this magnitude. However, the double mutant displayed a modest but
272 significant increase in growth following 8 days of starvation (Fig 4B). Furthermore, the double mutant
273 significantly increased early fecundity following starvation (Fig 4C). These results further support the
274 conclusion that natural variation in *irld-39* and *irld-52* affects starvation resistance. Notably, the two variants
275 are both present in the most starvation-resistant strain identified, DL238 (Fig 3F).

276 Given weak homology between IRLD proteins and the extracellular domain of DAF-2/InsR, we
277 wondered if IRLD-39 and IRLD-52 modify IIS, as originally proposed (DLAKIC 2002). DAF-2/InsR activity
278 inhibits the transcription factor DAF-16/FoxO, and *daf-2* mutants exhibit increased starvation resistance
279 that is dependent on *daf-16* (BAUGH AND STERNBERG 2006). We therefore hypothesized that increased
280 starvation resistance with disruption of *irld-39* and *irld-52* depends on *daf-16/FoxO.* To test this, we
281 generated a triple mutant including the null reference allele for *daf-16* (*daf-16(mu86); irld-39(duk1); irld-*
282 *52(duk17)*). Notably, the *irld-39; irld-52* double mutant displayed significant mitigation of the effect of
283 starvation on growth (Fig 4D). This result corroborates the effect of the double mutant after 8 days of
284 starvation (Fig 4B), except after only 4 days of starvation in this case (4 days of starvation was used since
285 this experiment included a *daf-16* null mutant which is sensitive to starvation). Consistent with our

286 hypothesis, we found no significant difference in the effect of starvation on growth between the null mutant
287 *daf-16(mu86)* and *daf-16(mu86); irld-39(duk1); irld-52(duk17)*, indicating that increased starvation
288 resistance of *irld-39(duk1); irld-52(duk17)* is dependent on *daf-16* (Fig 4D). This example of genetic
289 epistasis suggests that DAF-16 activity is increased in the *irld-39(duk1); irld-52(duk17)* double mutant. As a
290 transcription factor regulated by phosphorylation-dependent subcellular localization, the degree of DAF-16
291 nuclear localization is a proxy of its activity. In support of our hypothesis, nuclear localization of
292 endogenous DAF-16 (AGHAYEVA *et al.* 2020) in intestinal cells was significantly increased at 36 hours of L1
293 starvation in *irld-39(duk1); irld-52(duk17)* mutants (Fig 4E), suggesting that IRLD-39 and IRLD-52 inhibit
294 DAF-16 activity during starvation in wild-type N2 larvae. Together, genetic epistasis and nuclear
295 localization assays suggest that IRLD-39 and IRLD-52 act through DAF-16/FoxO to affect starvation
296 resistance during L1 arrest.

297

298 **DISCUSSION**

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

307 We characterized natural variation in starvation resistance in a set of genetically diverse, wild
308 strains of *C. elegans* using MIP-seq and traditional assays. Our results suggest relatively little phenotypic
309 variation of this presumably fitness-proximal trait. Nonetheless, we identified the first natural genetic
310 variants affecting starvation resistance in this species. Moreover, the four genes we implicated are all from

311 the same expanded *irld* gene family, suggesting that expansion (or contraction) of gene families influences
312 natural variation and possibly evolutionary adaptation in this context. In addition, two of the *irld* genes
313 identified are in hyper-divergent regions of the genome, consistent with genes in these regions contributing
314 to environmental responses (LEE *et al.* 2021).

315 Genetic epistasis analysis showed that the effect of *irld-39* and *irld-52* on starvation resistance
316 depends on *daf-16/FoxO*, and we found that nuclear localization of DAF-16/FoxO is affected by *irld-39* and
317 *irld-52.* These observations suggest that these *irld* genes modify IIS, as originally proposed for the gene
318 family (DLAKIC 2002). However, *irld-39* and *irld-52* could affect DAF-16 activity independent of IIS. In
319 addition, these *irld* genes could affect other signaling pathways as well. For example, they also bear weak
320 homology to EGF receptors, and *irld* family members *hpa-1* and *hpa-2* affect healthspan by modifying EGF
321 signaling (IWASA *et al.* 2010). It is not known whether EGF signaling affects starvation resistance or other
322 aspects of L1 arrest, and future work is needed to address the possible role of *irld* genes affecting EGF
323 signaling in this context.

324 Given the proposal that IRLD proteins modify IIS, it is intriguing to speculate that they do so by
325 binding ILPs that would otherwise bind DAF-2/InsR, as suggested previously (DLAKIC 2002). However, it is
326 imperative to emphasize that this is speculation. DAF-2B is an alternative isoform of DAF-2/InsR that
327 includes the extracellular domain but lacks the tyrosine kinase domain, like the IRLD proteins, and it is also
328 thought to act this way (MARTINEZ *et al.* 2020). This hypothetical mechanism is also analogous to the
329 proposed function of IGF-binding proteins, which affect circulation and receptor binding of IGF proteins
330 (ALLARD AND DUAN 2018). These parallels suggest the possibility that natural variation in the IGF-binding
331 protein family (ROTWEIN 2017) contributes to phenotypic variation in humans. The proposed decoy-receptor
332 mechanism could be generalized to other receptor homologs or isoforms to affect a variety of signaling
333 pathways, including deeply conserved pathways. However, we have not shown that IRLD proteins actually
334 bind ILPs, and a variety of uncertainties remain regarding their function in the context of a multicellular
335 animal whose developmental physiology is governed by a complex organismal signaling network.

336 ILPs have complex spatiotemporal expression patterns and distinct loss-of-function phenotypes
337 (RITTER *et al.* 2013; FERNANDES DE ABREU *et al.* 2014). ILPs can function as agonists or antagonists of
338 DAF-2/InsR (PIERCE *et al.* 2001), and relative expression of these functional classes changes over time
339 during starvation (CHEN AND BAUGH 2014). Assuming that IRLD proteins do bind ILPs, specificity of binding
340 is unclear, but specificity would likely alter the function of individual *irld* genes. In particular, IRLD proteins
341 could differ in binding to DAF-2/InsR agonists or antagonists, or they could bind ILPs indiscriminately, with
342 distinct consequences on IIS in each case. Many but not all IRLD proteins have signal sequences
343 suggesting they are secreted. The extent to which they act locally as opposed to moving freely throughout
344 the body cavity to directly exert systemic effects is unclear. To the extent that IRLD proteins do act locally,
345 their anatomical expression patterns should impact how they affect the IIS network, as well as other
346 signaling pathways they may modify, and possibly organismal phenotype.

347 Expression analysis provides possible clues to how *irld* genes influence starvation resistance. We
348 used genome editing to knock mScarlet into the *irld-39* and *irld-52* loci to generate endogenous reporter
349 genes, but we were unable to detect their expression during L1 arrest. In an effort to increase sensitivity,
350 we used putative promoter sequences for *irld-39* and *irld-52* to generate transcriptional reporter constructs,
351 and we introduced these as high-copy transgenes, but again we were unable to detect their expression.
352 Published whole-animal mRNA-seq analysis of fed and starved L1 larvae (WEBSTER *et al.* 2018) revealed
353 relatively low expression levels of the entire *irld* family (detected at low levels or not at all) (Fig S6).
354 However, about half of the *irld* genes were differentially expressed, and all of them were expressed at
355 higher levels in starved than fed larvae, suggesting a relatively specific role in starvation. We interrogated
356 two existing single-cell RNA-seq datasets, both of which analyzed expression in fed larvae (either L2-stage
357 (CAO *et al.* 2017) or L4-stage (TAYLOR *et al.* 2021) larvae). One included each of the major tissue types
358 (CAO *et al.* 2017), and it suggests that *irld* genes are most prominently expressed in ciliated sensory
359 neurons, though there is expression in other types of neurons and other tissues (Fig S7). The other study
360 focused on neurons (TAYLOR *et al.* 2021), and it also suggests that *irld* gene expression is more prominent

361 in sensory neurons than other neuron types (Fig S7). The authors noted that *irld* family expression is more
362 likely than other gene families to be restricted to a specific neuronal cell type (TAYLOR *et al.* 2021). *irld-39* is
363 expressed in the ASJ sensory neurons, along with distal tip cells and vulval precursors (Fig S7). *irld-52* is
364 expressed in the ADL sensory neurons and also intestinal rectal muscle cells. *C. elegans* sensory neurons
365 are polymodal and influence life-history traits regulated by IIS, including dauer formation, aging, and L1
366 arrest (BARGMANN AND HORVITZ 1991; VOWELS AND THOMAS 1992; APFELD AND KENYON 1999).
367 Furthermore, ASJ is known to express the relatively potent ILP DAF-28 in nutrient and sensory-dependent
368 fashion (LI *et al.* 2003; KAPLAN *et al.* 2018), and *daf-28* affects starvation survival (CHEN AND BAUGH 2014).
369 *ins-4* is also expressed in ASJ, and it too affects starvation survival (CHEN AND BAUGH 2014). Together
370 these data suggest that *irld* genes, including *irld-39* and *irld-52*, are expressed in sensory neurons and
371 other, seemingly unrelated cell types. If IRLD proteins are translated and function in the vicinity of these
372 sensory neurons, then that would allow them to exert their influence at the interface of the animal and its
373 environment.
374

375 **METHODS**

376 *Strains used in this study*

377 In addition to N2, wild isolates CB4854, CB4856, CX11254, CX11264, CX11271, CX11276,
378 CX11285, CX11307, DL200, DL226, DL238, ED3049, ECA189, ECA191, ECA36, ECA363, ECA369,
379 ECA372, ECA396, ED3017, ED3052, ED3077, EG4724, EG4725, GXW1, JU1212, JU1400, JU1581,
380 JU1652, JU1793, JU1896, JU2001, JU2007, JU2017, JU2106, JU2234, JU2316, JU2464, JU2519,
381 JU2526, JU2576, JU258, JU2592, JU2619, JU2811, JU2829, JU2593, JU2838, JU2841, JU2878, JU2879,
382 JU3137, JU561, JU774, JU775, JU782, KR314, LKC34, MY10, MY16, MY18, MY2147, MY23, MY2453,
383 MY2741, NIC195, NIC199, NIC251, NIC252, NIC256, NIC527, NIC258, NIC261, NIC262, NIC265, NIC266,
384 NIC268, NIC271, NIC3, NIC501, NIC523, NIC526, NIC528, PB306, PS2025, QG2075, QG556, QW947,
385 QX1211, QX1212, QX1791, QX1792, QX1793, QX1794, WN2001, XZ1513, XZ1514, XZ1515, and XZ1516
386 were phenotyped for starvation resistance using MIP-seq. In addition to these 100 strains, CX11262,
387 ECA348, and NIC260 were included in the MIP-seq pilot but excluded from subsequent analysis based on
388 quality-control metrics described in the 'MIP-seq analysis' section. QX1430 was used for validation assays.
389 CB1370 *daf-2(e1370) III,* CF1038 *daf-16(mu86) I,* and OH16024 *daf-16(ot971[daf-16::GFP]) I* were used to
390 assess the interaction of *irld* genes with insulin signaling.

391

392 *Strains generated in this study*

393 Near-isogenic lines include:

394 LRB392 – *dukIR7(V, EG4725>MY2147)*

395 LRB393 – *dukIR8(V, EG4725>MY2147)*

396 LRB395 – *dukIR10(V, EG4725>MY2147)*

397 LRB396 – *dukIR11(V, MY2147>EG4725)*

398 LRB397 – *dukIR12(V, MY2147>EG4725)*

399 LRB398 – *dukIR13(V, MY2147>EG4725)*

400 LRB399 – *dukIR14(V, MY2147>EG4725)*

401 LRB400 – *dukIR15(V, MY2147>EG4725)*

402 LRB401 – *dukIR16(V, MY2147>EG4725)*

403 LRB402 – *dukIR17(V, EG4725>MY2147)*

404 LRB403 – *dukIR18(V, EG4725>MY2147)*

405 LRB407 – *dukIR19(V, EG4725>MY2147)*

406 LRB408 – *dukIR20(V, EG4725>MY2147)*

407 LRB409 – *dukIR21(V, EG4725>MY2147)*

408 LRB410 – *dukIR22(IV, DL238>N2)*

409 LRB411 – *dukIR23(IV, DL238>N2)*

410 See Figure S3 for wild isolate composition.

411 CRISPR-edited strains and new crosses include:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

429 Multiple strain names for the same genotype indicates independent lines.

430

431 *MIP-seq experimental set-up*

432 Wild strains were independently passaged on 10 cm NGM plates with OP50 *E. coli* every two to
433 three days to ensure they did not starve for at least three generations prior to the experiment. For each
434 biological replicate, a single non-starved plate with gravid adults was selected per strain to ensure initial
435 representation of all strains. Strains were pooled for hypochlorite treatment to obtain pure populations of
436 embryos (HIBSHMAN *et al.* 2021). Embryo concentration was calculated by repeated sampling, and 500,000
437 embryos were resuspended at 10/µL in S-basal (50 mL total culture) and placed in a 20ºC shaker at 180
438 rpm to hatch without food and enter L1 arrest. On day 1 (24 hours after hypochlorite treatment), 5 mL of
439 culture (50,000 L1s) was taken as a baseline sample, spun down at 3000 rpm, aspirated down to
440 approximately 100 µL in an Eppendorf tube, flash frozen in liquid nitrogen, and stored at -80ºC until DNA
441 isolation. At days 1, 9, 13, and 17, aliquots from the L1 arrest culture were set up in recovery cultures at 5
442 L1s/µL, 1x HB101 (25 mg/mL), and S-complete. Recovery cultures were 10 mL for days 1 and 9, 20 mL for
443 day 13, and 50 mL for day 17 to account for lethality late in starvation by ensuring adequate population
444 sizes. Four days after recovery culture set-up, samples were collected for DNA isolation. For days 1 and 9,
445 the recovery culture was freshly starved with adults and next-generation L1 larvae. At day 13, the culture
446 was typically near starved, with adults and some L1 larvae. At day 17, the culture was typically not starved.
447 If HB101 was still present at collection, samples were washed 3-4 times with S-basal. Samples were flash
448 frozen in liquid nitrogen and stored at -80ºC until DNA isolation.

449

450 *DNA isolation*

451 Frozen samples were rapidly freeze-thawed three times, cycling between liquid nitrogen and a 45°C
452 water bath. Genomic DNA was isolated using the Quick-DNA Miniprep Kit (Zymo Research# D3024)

453 following the manufacturer's protocol. The DNA concentration was determined for each sample using the
454 Qubit dsDNA HS Assay kit (Invitrogen# Q32854).
455
456 *MIP design*
457 MIPgen (BOYLE *et al.* 2014) was used to design four MIPs for each of 103 strains. Unique
458 homozygous SNVs were parsed from the VCF file WI.20170531.vcf.gz (available at
459 https://storage.googleapis.com/elegansvariation.org/releases/20170531/WI.20170531.vcf.gz). Target
460 regions in BED format were generated using the makeBedForMipgen.pl script. MIPgen was used against
461 *C. elegans* genome version WS245 with the following parameters: -min_capture_size 100 -
462 max_capture_size 100 -tag_sizes 0, 10. MIPs are 80 base-pairs (bp) long and include 20 bp ligation and
463 extension arms that are complementary to DNA surrounding the unique SNV of interest for each strain. In
464 addition, P5 and P7 Illumina sequences are included as part of the MIP to facilitate Illumina sequencing.
465 Each MIP molecule includes a 10 bp unique molecular identifier (UMI) adjacent to the ligation arm. Only
466 MIPs that capture the SNV within a 50 bp sequencing read were used, meaning the SNV was no more
467 than 40 bp away from the UMI. SNPs located within 40 bases of the sequencing start site were parsed with
468 the parseMipsPerSNPposition.pl script. These scripts can be found at
469 https://github.com/amykwebster/MIPseq_2021.
470
471 *MIP-seq library preparation and sequencing*
472 For pilots and the starvation-resistance experiment, 500 ng genomic DNA from each sample was
473 used for MIP-seq libraries. Libraries were generated as described previously (HIATT *et al.* 2013) with the
474 following modifications. We included 1000 copies of each MIP for every individual copy of the worm
475 genome in the 500 ng input DNA, which corresponded to 0.0083 picomoles of each individual MIP. All 412
476 MIPs (sequences available in Data S1) were first pooled in an equimolar ratio at a concentration of 100 μM.
477 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

478 phosphorylation reaction. Next, the probe hybridization reaction for each sample was set with 500 ng DNA
479 and 3.42 picomoles (0.0083 picomoles x 412) of the phosphorylated probe mixture. Following hybridization,
480 gap filling, ligation, and exonuclease steps were performed as described previously. PCR amplification of
481 the captured DNA (primer sequences available in Data S1) was performed in a 50 μL reaction with 18
482 cycles. The PCR libraries were purified using the SPRIselect beads (Beckman# B23318), and library
483 concentrations were assessed with the Qubit dsDNA HS Assay kit (Invitrogen# Q32854). Sequencing was
484 performed on the Illumina HiSeq 4000 to obtain 50-bp single-end reads.
485
486 *MIP-seq analysis*
487 FASTQ files from sequencing reads were processed using the script parseMIPGenotypeUMI.pl,
488 also available at github.com/amykwebster/MIPseq_2021. This script accepts as input the list of MIPs
489 produced from MIPgen, the UMI length, and FASTQ files in order to count the number of reads
490 corresponding to each MIP and whether they have the reference allele, alternative allele, or one of two
491 other alleles. While we included UMIs in our MIP design, use of the UMI to filter duplicate reads in pilot
492 standard curves did not improve data quality (likely due to the relatively large mass of DNA used to prepare
493 libraries), and so the UMI was not used in the published analysis. For each MIP, the frequency of the strain
494 for which the MIP captures its unique SNV was calculated as the alternative read count divided by the total
495 of alternative and reference read counts. For the MIP pilot with all strains in an equimolar ratio, there were
496 246,986,236 total mapped reads to all MIPs. Individual MIPs were filtered out if they did not meet the
497 following criteria: 1) They were within 3.5-fold of expected frequency (that is, alt / (alt + ref) was within 3.5-
498 fold of 1/103), 2) 'other' reads (those that are not alternative or reference alleles) were < 20,000 total, and
499 3) alternative and reference allele totals were between 20,000 and 2,000,000 total reads. 321 of 412 MIPs
500 met these criteria, and reads from these 321 MIPs were included in subsequent analysis. N2 has very few
501 unique SNVs making it difficult to design optimal MIPs, and N2 MIPs did not meet these criteria but were
502 included nonetheless (see Data S1). For the standard curve experiment (Fig 1F), DNA from seven strains
503 (CB4856, DL200, ED3077, JU258, JU561, JU1652, and N2) was pooled in defined concentrations

504 ('expected'), and MIPs that met the criteria defined above were used to calculate strain frequencies
505 ('observed', see Data S1).
506 For the starvation-resistance experiment, an average of 51.7 million reads (standard deviation 7.2
507 million reads) were sequenced per library (one library per time point, replicate, and condition – 25 libraries
508 total). An average of 94% of reads (standard deviation 0.5%) matched the ligation probe, and 71.8%
509 (standard deviation 3.4%) matched the ligation probe and scan sequence. Strain frequencies were
510 determined by averaging the frequencies calculated across MIPs included in the 321 MIPs for each strain.
511 A dataframe of all strains and their frequencies at day 1 baseline, as well as days 1, 9, 13, and 17 after
512 recovery for all replicates was used to obtain trait values for subsequent analysis. PCA was performed on
513 the dataframe following normalization of day 1, 9, 13, and 17 time points by the baseline day 1 sample and
514 log2 transformation. PC1 loadings were extracted for each strain. For 'Slope', day 1, 9, and 13 recovery
515 samples were normalized by day 1 frequencies and log2 transformed. For each strain, a line was fit to day
516 1, 9, and 13 normalized data with intercept at 0, and the slope of the line was taken as the trait value.
517
518 *Comparison of MIP-seq and RAD-seq*
519 To determine how well MIP-seq trait values correlated with RAD-seq trait values from previous work
520 (WEBSTER *et al.* 2019), RAD-seq data were normalized the same way that we normalized the MIP-seq
521 data. Specifically, data from one biological replicate from RAD-seq that had data at time points over the
522 course of starvation, including days 1, 7, 14, 21, and 24, was used. The frequency of each strain at each
523 time point was divided by its frequency on day 1. These values were log2 transformed, so positive values
524 indicate an increase in frequency over time and negative values indicate a decrease in frequency over
525 time. A linear regression was then fit to each with a y-intercept of 0 through the data points over time. The
526 slope of the line was calculated as the trait value for each strain. RAD-seq and MIP-seq data were filtered
527 to include only the 34 strains that were present in both analyses. The values were plotted against each
528 other and a linear regression was fit through these points to determine their correlation ($R^2 = 0.24$, $p =$
529 0.002).

530

531 *GWA analysis*

532 Slope and PC1 trait values for each strain were used for GWA using the R package cegwas2
533 (https://github.com/AndersenLab/cegwas2-nf). Genotype data were acquired from the latest VCF release
534 (release 20200815) from CeNDR. BCFtools (LI 2011) was used to filter variants below a 5% minor allele
535 frequency and variants with missing genotypes and used PINKv1.9 (PURCELL *et al.* 2007; CHANG *et al.*
536 2015) to prune genotypes using linkage disequilibrium. The additive kinship matrix was generated from
537 45,733 markers using the A.mat function in the rrBLUP package. Because these markers have high LD,
538 eigen decomposition of the correlation matrix of the genotype matrix was performed to identify 570
539 independent tests. GWA was performed using the GWAS function of the rrBLUP package (ENDELMAN
540 2011). Significance was determined by an eigenvalue threshold by the number of independent tests in the
541 genotype matrix. Confidence intervals were defined as +/- 150 SNVs from the rightmost and leftmost
542 markers passing the significance threshold.

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

549

550 *Enrichment analyses*

551 To identify enriched gene groups, fine mapping data from Slope and PC1 results were merged with
552 WS273 gene names. Unique sequence names were extracted (see Data S2), and the 867 sequence
553 names with variants in Slope or PC1 QTL were used in WormCat (HOLDORF *et al.* 2020) to identify
554 functional category enrichments (Fig 3C). The most specific enrichments (those in Category 3) are shown.
555 For genes with unique symbol names (555 genes), those with the same symbol prefix were considered part

556 of the same gene family, and enrichments were calculated for families with at least five members with
557 variants in the QTL. *hpa* and *irld* gene prefixes were grouped under the *irld* prefix (IWASA *et al.* 2010). P-
558 values were calculated from a hypergeometric test.

559 For protein domain enrichment analysis, a protein fasta file was downloaded from Wormbase
560 (c_elegans.PRJNA13758.WS281.protein.fa). To determine enriched protein domains, the 867 sequence
561 names present among genes with variants in significant QTL were first used to subset this fasta file. In
562 cases in which a gene had multiple versions within the fasta file, the 'a' isoform of the gene was used. 644
563 of the 867 sequence names had protein sequences in the fasta file; most others are annotated as
564 pseudogenes and presumably do not have protein sequences. The protein sequences were used as input
565 using the hmmscan program (https://www.ebi.ac.uk/Tools/hmmer/search/hmmscan) and searching the
566 Pfam database (FINN *et al.* 2011). To obtain a background set of protein domains, the genome fasta file
567 was also used to search the Pfam database. Fasta files were split into groups of 500 sequences that are
568 between 10 and 5000 peptides to comply with the hmmscan search algorithm. To calculate enrichment of
569 protein domains, hypergeometric p-values were calculated for each protein domain present among genes
570 with variants in significant QTL. 102 protein domains were present, so a Bonferroni-corrected p-value of
571 0.00049 was used as a significance threshold. Protein domains were excluded if the domain was not
572 present at least five times among genes with variants in significant QTL.

573

574 *NIL generation*

575 To validate chromosome IV and V QTL, pairs of strains that differ for starvation resistance and the
576 alternative vs reference allele for the associated SNV marker were chosen. Compatibility at the *peel-1/zeel-*
577 *1* and *pha-1/sup-35* loci was considered (SEIDEL *et al.* 2008; BEN-DAVID *et al.* 2017). For chromosome V
578 QTL, EG4725 and MY2147 were compatible at both loci, and we generated reciprocal NILs for the left and
579 right arms of chromosome V. EG4725 did not have the alternative allele associated with starvation
580 resistance for chromosome IV, so we used DL238 and N2 as the parental strains. DL238 and N2 are
581 incompatible for reciprocal crosses, but we introgressed the DL238 chromosome IV QTL into the N2

582 background. To generate NILs, the two parental strains were first crossed, then F2 progeny were
583 genotyped on each end of the desired QTL for introgression to identify homozygotes from one parental
584 background (*e.g.*, MY2147). Then these homozygotes were repeatedly backcrossed to the opposite
585 background (*e.g.*, EG4725) and repeatedly genotyped to maintain homozygotes at the introgressed region.
586 Genotyping was performed using PCR to amplify a genomic region whose sensitivity to a particular
587 restriction enzyme depends on parental genetic background. Primers were designed using VCF-kit (COOK
588 AND ANDERSEN 2017). Primers and enzymes used can be found in Data S2. NILs were backcrossed a
589 minimum of six times. Final NILs were sequenced at ~1x coverage to determine the parental contributions
590 over the entire genome (Fig S5 and Data S2).
591
592 *CRISPR design and implementation to edit* irld *genes*
593 For genes of interest, CRISPR guide design was done in Benchling using genome version
594 WBcel235 and importing sequence for genes of interest. To generate *irld-39* and *irld-52* variants (5 bp
595 deletions), a single guide RNA (sgRNA) and repair template was generated. For *irld-11* and *irld-57,* two
596 sgRNAs were generated per gene to delete the entire gene and a single repair template was used.
597 sgRNAs (2 nmol) and 100 bp repair templates (highest purity at 4 nmol) were ordered from IDT. The *dpy-*
598 *10* co-CRISPR method was used to generate and screen for edits (PAIX *et al.* 2015). The injection mix used
599 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-*
600 *10* repair template (0.5 µL of 10 µM stock), repair template for gene of interest (0.6 µL of 100 µM stock),
601 Cas9 (0.8 µL of 61 µM stock), and water up to 10 µL total. Injection mix components were stored at -20ºC,
602 and injection mix was incubated at room temperature for one hour before injections. N2 and MY2147 L4s
603 were picked the day before injecting, and young adults were injected in the gonad and singled to new
604 plates. After 3-4 days, next-generation adults were screened for rollers, which are heterozygous for the
605 *dpy-10* edit and have increased likelihood of also having the desired edit. Non-roller F2 progeny of F1 roller
606 worms were then genotyped to identify worms homozygous for the desired edit, and edits were confirmed

607 by Sanger sequencing. Sequences of sgRNAs, repair templates, and PCR primers for genotyping are
608 available in Data S2.
609
610 *Starvation recovery (worm length measurements)*
611 Strains were maintained well-fed for at least three generations prior to beginning experiments.
612 Gravid adults were hypochlorite treated to obtain embryos, which were resuspended at 1 embryo/μL in 5
613 mL of S-basal in a glass test tube and placed on a roller drum at 20ºC so they hatch and enter L1 arrest.
614 After the number of days of L1 arrest indicated on each graph, an aliquot of 500-1000 μL (a consistent
615 volume was used between conditions within the same experiment) per strain was plated on a 10 cm plate
616 with OP50 and allowed to recover for 48 hr. After 48 hr, worms were washed onto an unseeded NGM
617 plate. Images were then taken of worms using a ZeissDiscovery V20 stereomicroscope. To determine
618 lengths of worms, the WormSizer plugin for Fiji was used and worms were manually passed or failed
619 (MOORE *et al.* 2013). To determine differences in starvation recovery between strains, a linear mixed-effect
620 model was fit to the length data for all individual worms with duration of starvation and strain as fixed
621 effects and biological replicate as a random effect using the package nlme in R. The summary function was
622 used to calculate a p-value from the t-value.
623
624 *Starvation survival*
625 L1 arrest cultures were set up as described for starvation recovery. Starting on the first day of arrest
626 and proceeding every other day, a 100 μL aliquot of culture was pipetted onto a 5 cm NGM plate with a
627 spot of OP50 at the center. The aliquot was placed around the periphery of the lawn, and the number of
628 worms plated was counted. Two days later, the number of worms that had made it to the bacterial lawn and
629 were alive was counted. Live worms that have developed and were outside the lawn were also counted.
630 The total number of live worms after two days was divided by the total plated to determine the proportion
631 alive. For each replicate, a logistic curve was fit to the data, and the half-life (time at 50% survival) was
632 calculated, and a t-test was performed on half-lives between strains of interest. Power analysis was

633 performed in R using the pwr.t.test function in the “pwr” package, with parameters n = 5, sig.level = 0.05,
634 power = 0.5, and type = two.sample. The value of d was calculated, and this was multiplied by the standard
635 deviation of control median survival for that experiment to determine the detectable effect size.
636
637 *Early fecundity following starvation*
638 For the assay in Figure 2G, L1 arrest cultures were set up as described for starvation recovery. For
639 the assay in Figure 4C, the experiment was done in a different lab and worms were arrested in a 15 mL
640 conical tube instead of a glass test tube. Conical tubes were rotated continuously at 20ºC. For both figure
641 panels, ~500 L1 larvae were plated on 10 cm OP50 plates at the indicated time point, then allowed to
642 recover for 48 hours. Worms were singled to new 5 cm OP50 plates at approximately 48 hours, then
643 allowed lay progeny until 72 hours, at which time the singled worm was removed. Progeny were counted
644 on these plates 2-3 days later to determine the early fecundity of the individual worms.
645
646 *Nuclear localization*
647 L1 arrest cultures were set up as described for starvation recovery. At 36 hours of L1 arrest, an
648 aliquot of 700 µL was spun down in a 1.7 mL Eppendorf tube at 3000 rpm for 30 seconds to pellet L1
649 larvae. 1.5 µL of worm pellet was pipetted into the center of a slide with a 4% Noble agar pad, and a glass
650 cover slip was immediately placed on top. A timer was set for 3 minutes, and the slide was systematically
651 scanned with each individual worm scored for nuclear localization at 40x or 100x with a Zeiss compound
652 microscope. Nuclear localization of DAF-16::GFP was scored in intestinal cells and assigned as one of four
653 categories: nuclear, more nuclear, more cytoplasmic, and cytoplasmic. ‘More nuclear’ and ‘more
654 cytoplasmic’ are intermediate categories between nuclear and cytoplasmic, with localization closer to being
655 nuclear or cytoplasmic, respectively. Scoring for each slide stopped after 3 minutes. For statistical analysis,
656 nuclear and more nuclear categories were pooled as ‘nuclear’, while cytoplasmic and more cytoplasmic
657 were pooled as ‘cytoplasmic’. The Cochran-Mantel-Haenszel test was used to determine differences in the
658 distribution of the two categories while controlling for biological replicate. See Fig 4E for representative

659 images. DAF-16 is initially very nuclear during L1 starvation, and it moves back to the cytoplasm over time
660 during starvation (MATA-CABANA *et al.* 2020). The 36-hour time point was chosen since it is intermediate in
661 this dynamic process.
662
663 *Analysis of published RNA-seq data*
664 We analyzed data from three existing publications (Fig S6, S7) (CAO *et al.* 2017; WEBSTER *et al.*
665 2018; TAYLOR *et al.* 2021). First, we re-analyzed whole worm bulk mRNA-seq data from fed and starved N2
666 L1 larvae (four replicates of each condition from a single batch) (WEBSTER *et al.* 2018). Count data was
667 analyzed using edgeR. 60 *irld* genes were part of the protein-coding gene dataset for genome version
668 WS273, and no minimum expression filter was used to restrict the gene set. The calcNormFactors,
669 estimateCommonDisp, and estimateTagwiseDisp functions were used prior to running the exactTest. An
670 FDR cutoff of 0.05 was used to determine significance. For single-cell data across all major worm tissues
671 (CAO *et al.* 2017),Table S4 from the paper was subset to include only *irld* genes, 63 of which were present
672 in the table. Gene expression is represented in Fig S7 when expression levels of transcripts-per-million are
673 at least 1 for that tissue type. For single-cell neuronal data (TAYLOR *et al.* 2021), expression values for *irld*
674 genes were obtained from Supplementary Table 11 of the published paper, which includes genes
675 considered expressed at a variety of thresholds and neuronal cell types. Data plotted in Figure S7 uses
676 threshold 3 and data from sensory neurons.
677
678 *Data availability*
679 Raw MIP-seq data for the starvation-resistance experiment and the pilot experiments to test
680 individual MIPs is available as part of NCBI BioProject PRJNA730178. Code for processing MIP-seq data
681 is available at github.com/amykwebster/MIPseq_2021.
682
683 *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, Sophia Gomez for genotyping assistance, Seth Taylor for strain organization and maintenance, and Jim Jordan for helpful discussions. Funding was provided by the NIH (R01GM117408 and R01GM143159 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.

Author contributions

Conceptualization: AKW, LRB

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

Formal analysis: AKW, LS, KE, IA, AW

Visualization: AKW

Funding acquisition: LRB, ECA

Supervision: LRB, ECA

Writing – original draft: AKW, LRB

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

Competing interests

Authors declare that they have no competing interests.

Materials and correspondence

Correspondence and material requests should be addressed to ryan.baugh@duke.edu.

707 **REFERENCES**

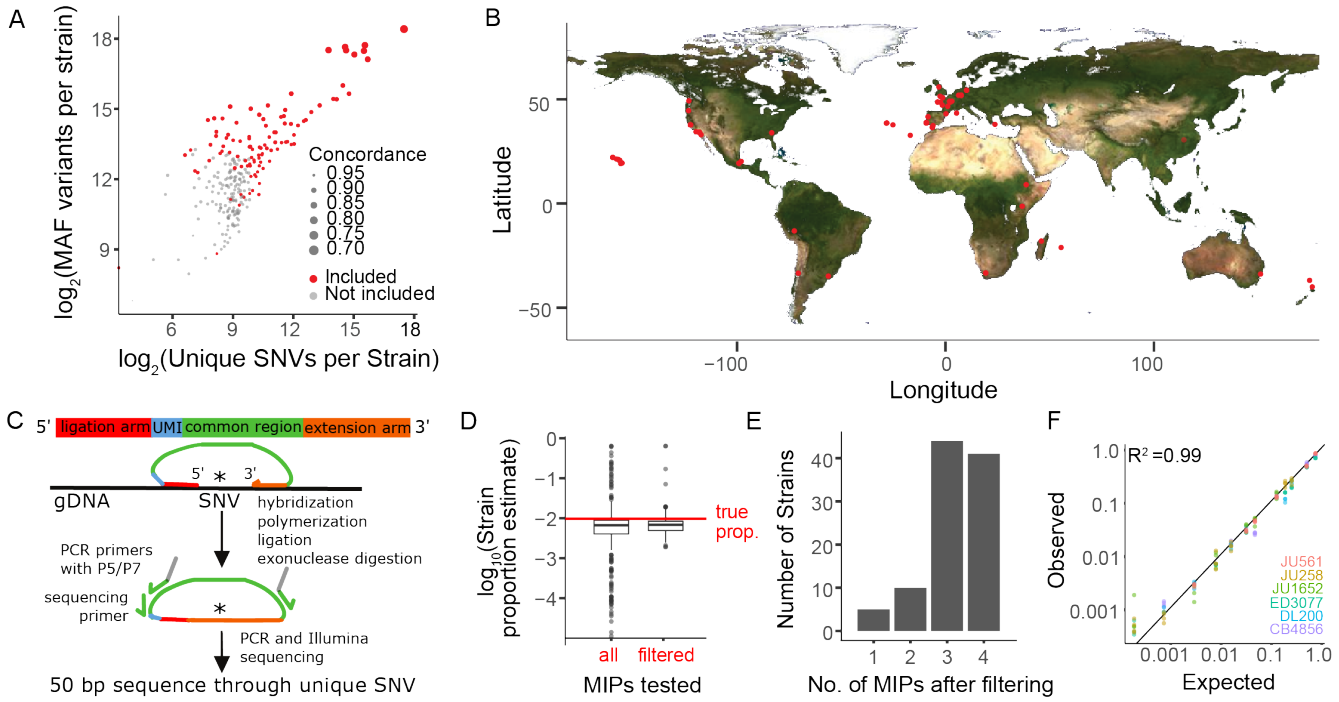
708 Aghayeva, U., A. Bhattacharya and O. Hobert, 2020 A panel of fluorophore-tagged daf-16 alleles.
709 MicroPubl Biol 2020.
710 Allard, J. B., and C. Duan, 2018 IGF-Binding Proteins: Why Do They Exist and Why Are There So Many?
711 Front Endocrinol (Lausanne) 9**:** 117.
712 Apfeld, J., and C. Kenyon, 1999 Regulation of lifespan by sensory perception in Caenorhabditis elegans.
713 Nature 402**:** 804-809.
714 Bargmann, C. I., and H. R. Horvitz, 1991 Control of larval development by chemosensory neurons in
715 Caenorhabditis elegans. Science 251**:** 1243-1246.
716 Baugh, L. R., 2013 To Grow or Not to Grow: Nutritional Control of Development During Caenorhabditis
717 elegans L1 Arrest. Genetics 194**:** 539-555.
718 Baugh, L. R., and P. J. Hu, 2020 Starvation Responses Throughout the Caenorhabditis elegans Life Cycle.
719 Genetics 216**:** 837-878.
720 Baugh, L. R., and P. W. Sternberg, 2006 DAF-16/FOXO regulates transcription of cki-1/Cip/Kip and
721 repression of lin-4 during C. elegans L1 arrest. Curr Biol 16**:** 780-785.
722 Ben-David, E., A. Burga and L. Kruglyak, 2017 A maternal-effect selfish genetic element in Caenorhabditis
723 elegans. Science 356**:** 1051-1055.
724 Boyle, E. A., Y. I. Li and J. K. Pritchard, 2017 An Expanded View of Complex Traits: From Polygenic to
725 Omnigenic. Cell 169**:** 1177-1186.
726 Boyle, E. A., B. J. O'Roak, B. K. Martin, A. Kumar and J. Shendure, 2014 MIPgen: optimized modeling and
727 design of molecular inversion probes for targeted resequencing. Bioinformatics 30**:** 2670-2672.
728 Cantsilieris, S., H. A. Stessman, J. Shendure and E. E. Eichler, 2017 Targeted Capture and High-Throughput
729 Sequencing Using Molecular Inversion Probes (MIPs). Methods Mol Biol 1492**:** 95-106.
730 Cao, J., J. S. Packer, V. Ramani, D. A. Cusanovich, C. Huynh *et al.*, 2017 Comprehensive single-cell
731 transcriptional profiling of a multicellular organism. Science 357**:** 661-667.
732 Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell *et al.*, 2015 Second-generation PLINK: rising
733 to the challenge of larger and richer datasets. Gigascience 4**:** 7.
734 Chen, J., L. Y. Tang, M. E. Powell, J. M. Jordan and L. R. Baugh, 2022 Genetic analysis of daf-18/PTEN
735 missense mutants for starvation resistance and developmental regulation during Caenorhabditis
736 elegans L1 arrest. G3 (Bethesda).
737 Chen, Y., and L. R. Baugh, 2014 Ins-4 and daf-28 function redundantly to regulate C. elegans L1 arrest. Dev
738 Biol 394**:** 314-326.
739 Cook, D. E., and E. C. Andersen, 2017 VCF-kit: assorted utilities for the variant call format. Bioinformatics
740 33**:** 1581-1582.
741 Cook, D. E., S. Zdraljevic, J. P. Roberts and E. C. Andersen, 2017 CeNDR, the Caenorhabditis elegans natural
742 diversity resource. Nucleic Acids Res 45**:** D650-D657.
743 Dlakic, M., 2002 A new family of putative insulin receptor-like proteins in C. elegans. Curr Biol 12**:** R155-
744 157.
745 Endelman, J. B., 2011 Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP.
746 The Plant Genome 4.
747 Fernandes de Abreu, D. A., A. Caballero, P. Fardel, N. Stroustrup, Z. Chen *et al.*, 2014 An insulin-to-insulin
748 regulatory network orchestrates phenotypic specificity in development and physiology. PLoS Genet
749 10**:** e1004225.

750 Finn, R. D., J. Clements and S. R. Eddy, 2011 HMMER web server: interactive sequence similarity searching.
751 Nucleic Acids Res 39: W29-37.
752 Flier, J. S., 2019 Starvation in the Midst of Plenty: Reflections on the History and Biology of Insulin and
753 Leptin. Endocr Rev 40: 1-16.
754 Henderson, S. T., and T. E. Johnson, 2001 daf-16 integrates developmental and environmental inputs to
755 mediate aging in the nematode Caenorhabditis elegans. Curr Biol 11: 1975-1980.
756 Hiatt, J. B., C. C. Pritchard, S. J. Salipante, B. J. O'Roak and J. Shendure, 2013 Single molecule molecular
757 inversion probes for targeted, high-accuracy detection of low-frequency variation. Genome Res 23:
758 843-854.
759 Hibshman, J., A. Webster and L. Baugh, 2021 Liquid-culture protocols for synchronous starvation, growth,
760 dauer formation, and dietary restriction of Caenorhabditis elegans. STAR Protocols 2: 100276.
761 Hibshman, J. D., A. E. Doan, B. T. Moore, R. E. Kaplan, A. Hung *et al.*, 2017 daf-16/FoxO promotes
762 gluconeogenesis and trehalose synthesis during starvation to support survival. Elife 6.
763 Hobert, O., 2013 The neuronal genome of Caenorhabditis elegans. WormBook: 1-106.
764 Holdorf, A. D., D. P. Higgins, A. C. Hart, P. R. Boag, G. J. Pazour *et al.*, 2020 WormCat: An Online Tool for
765 Annotation and Visualization of Caenorhabditis elegans Genome-Scale Data. Genetics 214: 279-
766 294.
767 Hung, W. L., Y. Wang, J. Chitturi and M. Zhen, 2014 A Caenorhabditis elegans developmental decision
768 requires insulin signaling-mediated neuron-intestine communication. Development 141: 1767-
769 1779.
770 Iwasa, H., S. Yu, J. Xue and M. Driscoll, 2010 Novel EGF pathway regulators modulate C. elegans healthspan
771 and lifespan via EGF receptor, PLC-gamma, and IP3R activation. Aging Cell 9: 490-505.
772 Jobson, M. A., J. M. Jordan, M. A. Sandrof, J. D. Hibshman, A. L. Lennox *et al.*, 2015 Transgenerational
773 Effects of Early Life Starvation on Growth, Reproduction, and Stress Resistance in Caenorhabditis
774 elegans. Genetics 201: 201-212.
775 Jordan, J. M., J. D. Hibshman, A. K. Webster, R. E. W. Kaplan, A. Leinroth *et al.*, 2019 Insulin/IGF Signaling
776 and Vitellogenin Provisioning Mediate Intergenerational Adaptation to Nutrient Stress. Curr Biol 29:
777 2380-2388 e2385.
778 Kaplan, R. E. W., C. S. Maxwell, N. K. Codd and L. R. Baugh, 2019 Pervasive Positive and Negative Feedback
779 Regulation of Insulin-Like Signaling in Caenorhabditis elegans. Genetics 211: 349-361.
780 Kaplan, R. E. W., A. K. Webster, R. Chitrakar, J. A. Dent and L. R. Baugh, 2018 Food perception without
781 ingestion leads to metabolic changes and irreversible developmental arrest in C. elegans. BMC Biol
782 16: 112.
783 Lee, D., S. Zdraljevic, L. Stevens, Y. Wang, R. E. Tanny *et al.*, 2021 Balancing selection maintains hyper-
784 divergent haplotypes in Caenorhabditis elegans. Nature Ecology & Evolution.
785 Lee, R. Y., J. Hench and G. Ruvkun, 2001 Regulation of C. elegans DAF-16 and its human ortholog FKHRL1
786 by the daf-2 insulin-like signaling pathway. Curr Biol 11: 1950-1957.
787 Li, H., 2011 A statistical framework for SNP calling, mutation discovery, association mapping and
788 population genetical parameter estimation from sequencing data. Bioinformatics 27: 2987-2993.
789 Li, W., S. G. Kennedy and G. Ruvkun, 2003 daf-28 encodes a C. elegans insulin superfamily member that is
790 regulated by environmental cues and acts in the DAF-2 signaling pathway. Genes Dev 17: 844-858.
791 Lin, K., J. B. Dorman, A. Rodan and C. Kenyon, 1997 daf-16: An HNF-3/forkhead family member that can
792 function to double the life-span of Caenorhabditis elegans. Science 278: 1319-1322.

793 Lin, K., H. Hsin, N. Libina and C. Kenyon, 2001 Regulation of the Caenorhabditis elegans longevity protein
794 DAF-16 by insulin/IGF-1 and germline signaling. Nat Genet 28: 139-145.
795 Martinez, B. A., P. Reis Rodrigues, R. M. Nunez Medina, P. Mondal, N. J. Harrison *et al.*, 2020 An
796 alternatively spliced, non-signaling insulin receptor modulates insulin sensitivity via insulin peptide
797 sequestration in C. elegans. Elife 9.
798 Mata-Cabana, A., L. Gomez-Delgado, F. J. Romero-Exposito, M. J. Rodriguez-Palero, M. Artal-Sanz *et al.*,
799 2020 Social Chemical Communication Determines Recovery From L1 Arrest via DAF-16 Activation.
800 Front Cell Dev Biol 8: 588686.
801 Mok, C. A., V. Au, O. A. Thompson, M. L. Edgley, L. Gevirtzman *et al.*, 2017 MIP-MAP: High-Throughput
802 Mapping of Caenorhabditis elegans Temperature-Sensitive Mutants via Molecular Inversion
803 Probes. Genetics 207: 447-463.
804 Moore, B. T., J. M. Jordan and L. R. Baugh, 2013 WormSizer: high-throughput analysis of nematode size
805 and shape. PLoS One 8: e57142.
806 Munoz, M. J., and D. L. Riddle, 2003 Positive selection of Caenorhabditis elegans mutants with increased
807 stress resistance and longevity. Genetics 163: 171-180.
808 Ogg, S., S. Paradis, S. Gottlieb, G. I. Patterson, L. Lee *et al.*, 1997 The Fork head transcription factor DAF-16
809 transduces insulin-like metabolic and longevity signals in C. elegans. Nature 389: 994-999.
810 Paix, A., A. Folkmann, D. Rasoloson and G. Seydoux, 2015 High Efficiency, Homology-Directed Genome
811 Editing in Caenorhabditis elegans Using CRISPR-Cas9 Ribonucleoprotein Complexes. Genetics 201:
812 47-54.
813 Pierce, S. B., M. Costa, R. Wisotzkey, S. Devadhar, S. A. Homburger *et al.*, 2001 Regulation of DAF-2
814 receptor signaling by human insulin and ins-1, a member of the unusually large and diverse C.
815 elegans insulin gene family. Genes Dev 15: 672-686.
816 Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira *et al.*, 2007 PLINK: a tool set for whole-
817 genome association and population-based linkage analyses. Am J Hum Genet 81: 559-575.
818 Ried, J. S., M. J. Jeff, A. Y. Chu, J. L. Bragg-Gresham, J. van Dongen *et al.*, 2016 A principal component meta-
819 analysis on multiple anthropometric traits identifies novel loci for body shape. Nat Commun 7:
820 13357.
821 Ritter, A. D., Y. Shen, J. Fuxman Bass, S. Jeyaraj, B. Deplancke *et al.*, 2013 Complex expression dynamics
822 and robustness in C. elegans insulin networks. Genome Res 23: 954-965.
823 Rotwein, P., 2017 Large-scale analysis of variation in the insulin-like growth factor family in humans reveals
824 rare disease links and common polymorphisms. J Biol Chem 292: 9252-9261.
825 Roux, A. E., K. Langhans, W. Huynh and C. Kenyon, 2016 Reversible Age-Related Phenotypes Induced
826 during Larval Quiescence in C. elegans. Cell Metab 23: 1113-1126.
827 Schulenburg, H., and M. A. Félix, 2017 The Natural Biotic Environment of Caenorhabditis elegans. Genetics
828 206: 55-86.
829 Seidel, H. S., M. V. Rockman and L. Kruglyak, 2008 Widespread genetic incompatibility in C. elegans
830 maintained by balancing selection. Science 319: 589-594.
831 Soeters, M. R., and P. B. Soeters, 2012 The evolutionary benefit of insulin resistance. Clin Nutr 31: 1002-
832 1007.
833 Taylor, S. R., G. Santpere, A. Weinreb, A. Barrett, M. B. Reilly *et al.*, 2021 Molecular topography of an entire
834 nervous system. Cell 184: 4329-4347 e4323.
835 Vowels, J. J., and J. H. Thomas, 1992 Genetic analysis of chemosensory control of dauer formation in
836 Caenorhabditis elegans. Genetics 130: 105-123.

837 Webster, A. K., A. Hung, B. T. Moore, R. Guzman, J. M. Jordan *et al.*, 2019 Population Selection and
838 Sequencing of Caenorhabditis elegans Wild Isolates Identifies a Region on Chromosome III Affecting
839 Starvation Resistance. G3 (Bethesda) 9**:** 3477-3488.
840 Webster, A. K., J. M. Jordan, J. D. Hibshman, R. Chitrakar and L. R. Baugh, 2018 Transgenerational Effects of
841 Extended Dauer Diapause on Starvation Survival and Gene Expression Plasticity in Caenorhabditis
842 elegans. Genetics 210**:** 263-274.
843 Yano, K., Y. Morinaka, F. Wang, P. Huang, S. Takehara *et al.*, 2019 GWAS with principal component analysis
844 identifies a gene comprehensively controlling rice architecture. Proc Natl Acad Sci U S A 116**:**
845 21262-21267.
846

847 **FIGURES**



848

849 **Fig. 1. Sensitive and precise measurement of strain frequency in pooled culture using MIP-seq.** (**A**)
850 The three metrics used to identify the most diverse *C. elegans* strains are plotted. "MAF" stands for minor
851 allele frequency. Concordance refers to the average pairwise concordance for the focal strain compared to
852 all other strains, which is calculated as the number of shared variant sites divided by the total number of
853 variants for each pair. The strains included in the MIP-seq experiments are in red. (**B**) Geographic locations
854 of the strains assayed for starvation resistance. (**C**) Schematic of MIP-seq. MIPs are designed for loci with
855 SNVs unique to each strain. Four MIPs were designed per strain. MIPs are 80 nt long and include ligation
856 and extension arms to match DNA sequence surrounding the SNV, a unique molecular identifier (UMI),
857 and P5 and P7 sequences for Illumina sequencing. MIPs are hybridized to genomic DNA, polymerized,
858 ligated, and used as PCR template to generate an Illumina sequencing library. The alternative-to-total read
859 frequency for each MIP/SNV locus indicates strain frequency. (**D**) Empirical testing of 412 MIPs with an
860 equimolar mix of genomic DNA from 103 strains to identify reliable MIPs. 321 MIPs passed filtering and
861 were analyzed in the starvation experiment. Outliers for filtered MIPs are for N2, which has hardly any

862 unique SNVs because it is derived from the reference genome. N2 MIPs were included despite poor
863 performance. (**E**) Number of MIPs per strain of the 321 filtered MIPs that passed filtering. (**F**) Genomic
864 DNA from 7 strains was combined at different known concentrations, and MIP-seq was used to generate a
865 standard curve. Included MIPs all passed filtering. (R^2 = 0.99).
866

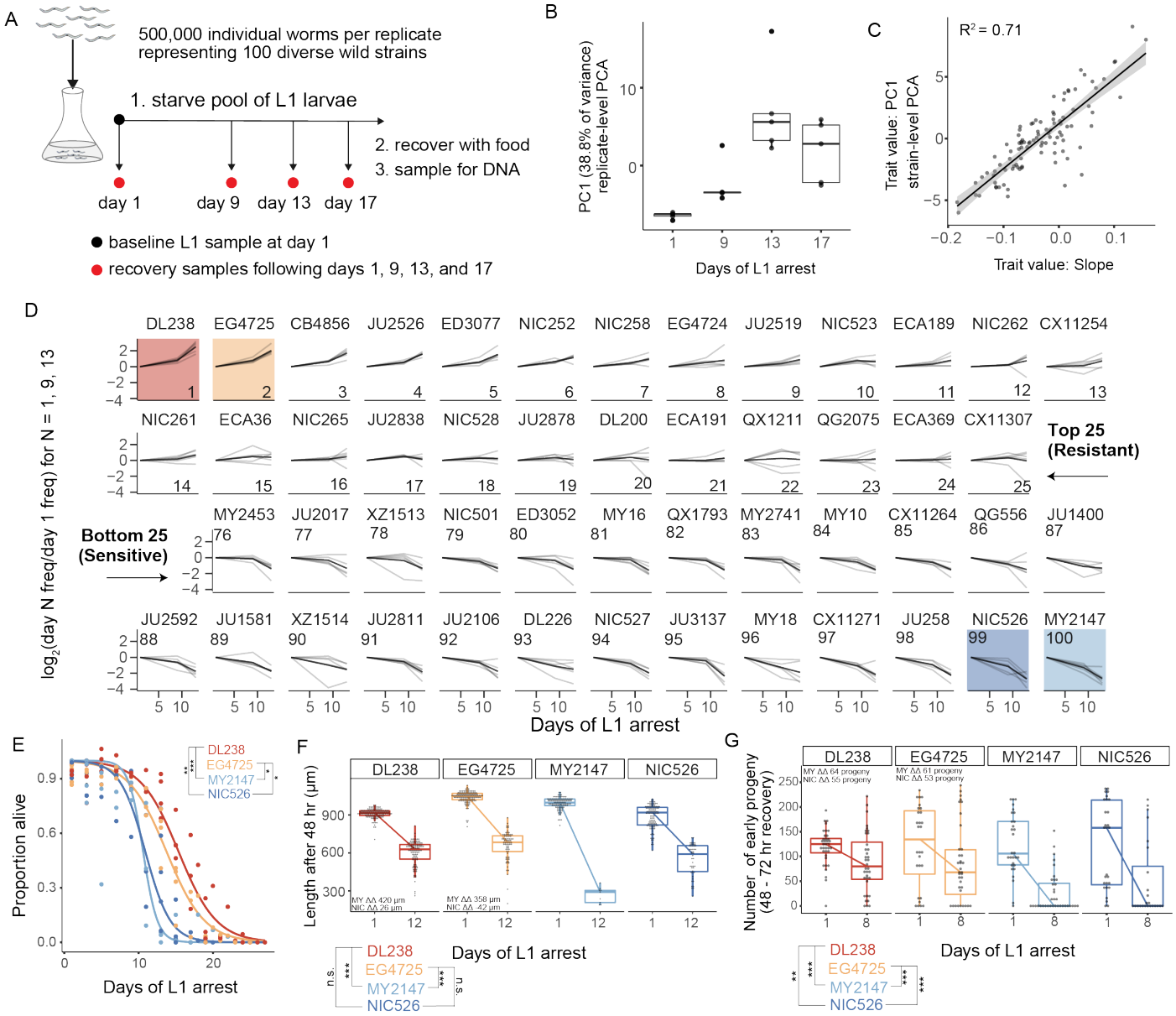
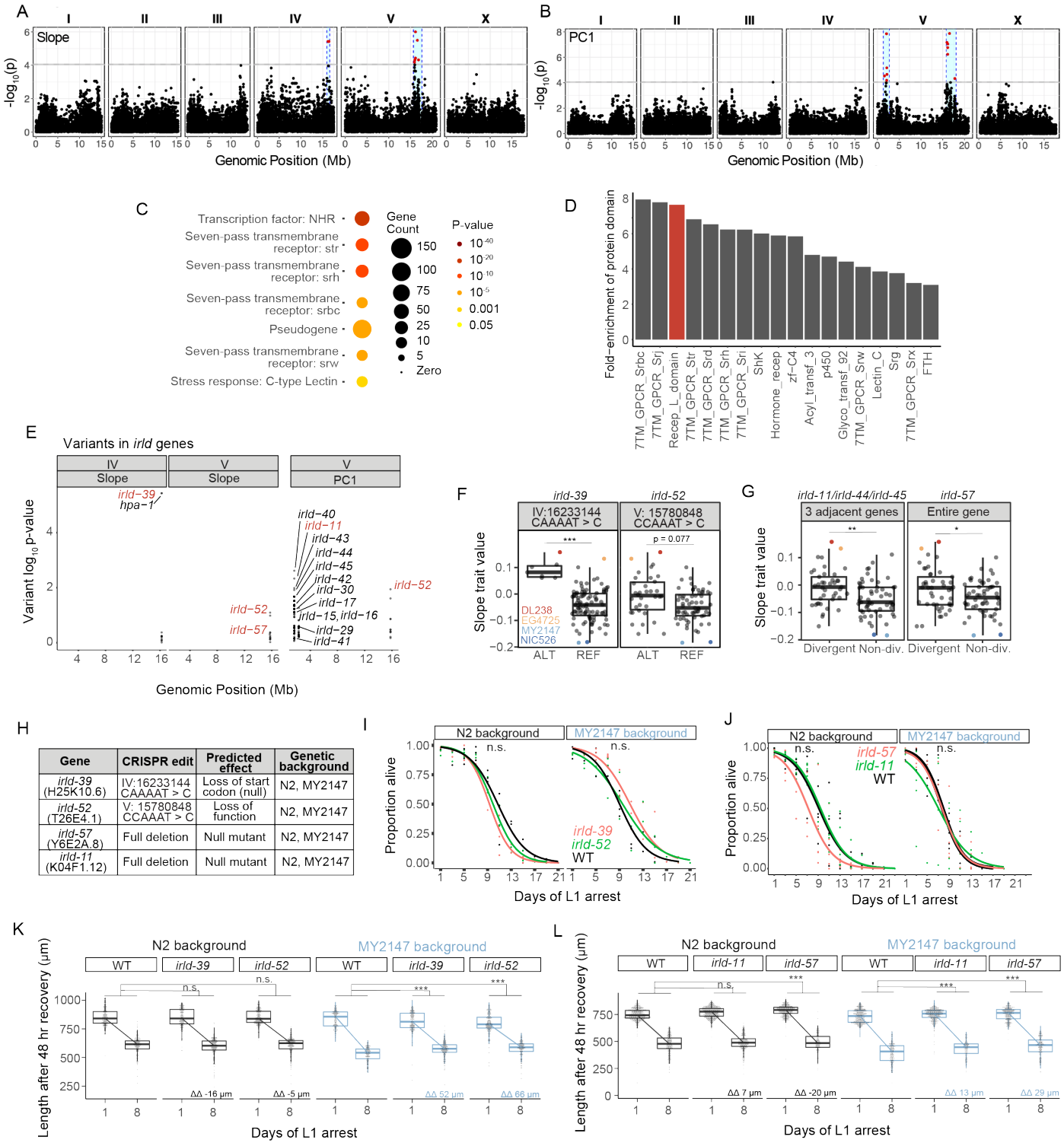


Fig. 2. MIP-seq determines relative starvation resistance of 100 strains. (A) Experimental design. Worms were starved at the L1 stage ('L1 arrest'). ~5,000 L1 larvae per strain were starved (~500,000 total). The population of starved L1 larvae was sampled initially ("baseline" on day 1), and then sampled on the days indicated. Samples (except baseline) were recovered with food in liquid culture, reaching adulthood and producing progeny for one day, and the entire population was 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

875 individual sample (MIP-seq library). (**C**) The relationship between two starvation-resistance metrics (Slope
876 and PC1) produced from strain-level data (replicates averaged) is plotted. Each point is a different strain.
877 (**D**) Log_2-normalized strain frequency is plotted over time for the 25 most resistant and 25 most sensitive
878 strains in rank order (based on Slope). Only days 1, 9, and 13 are plotted. See Supplementary Figure 2 for
879 full data. Grey lines are biological replicates and black line is the mean. DL238 and EG4725 are most
880 starvation-resistant, and NIC526 and MY2147 are most sensitive, and they are color-coded accordingly.
881 (**E**) L1 starvation survival curves are plotted for starvation-resistant and sensitive strains. Individual
882 replicate measurements are included as points to which curves were fit with logistic regression. T-tests on
883 50% survival time of four biological replicates. (**F**) Worm length following 48 hours of recovery with food
884 after 1 or 12 days of L1 starvation. (**G**) Number of progeny produced between 48 and 72 hours of recovery
885 on food following 1 or 8 days of starvation. (**F,G**) ΔΔ indicates effect size of interaction between duration of
886 starvation and strain data plotted in that panel compared to the strain listed (the difference in differences
887 between strains' mean length at days 1 and 12 or between mean number of progeny at days 1 and 8). 'MY'
888 is an abbreviation for MY2147 and 'NIC' is an abbreviation for NIC526. Linear mixed-effects model; one-
889 way p-value of interaction between duration of starvation and strain. (**E-G**) ***$p < 0.001$, **$p < 0.01$, *$p <$
890 0.05
891



Gene	CRISPR edit	Predicted effect	Genetic background
irld-39 (H25K10.6)	IV:16233144 CAAAAT > C	Loss of start codon (null)	N2, MY2147
irld-52 (T26E4.1)	V: 15780848 CCAAAT > C	Loss of function	N2, MY2147
irld-57 (Y6E2A.8)	Full deletion	Null mutant	N2, MY2147
irld-11 (K04F1.12)	Full deletion	Null mutant	N2, MY2147

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-

896 2764788 and V: 15775895-18065050. (**C**) WormCat Category 3 enrichments for all genes with variants in
897 the QTL. (**D**) Fold-enrichment of protein domains significantly enriched among genes with variants in QTL.
898 A hypergeometric p-value was calculated for each of 102 protein domains present, and a Bonferroni-
899 corrected p-value of 0.00049 was used as a threshold to determine significance. Red indicates the receptor
900 L domain, which is found in *irld* genes. (**E**) All variants in *irld* genes that are within significant QTL and their
901 association with the starvation-resistance traits, Slope and PC. Each gene name is shown next to the most
902 significant variant for that gene, but multiple variants are plotted for each gene when present. Red indicates
903 genes selected for functional validation. (**F**) Slope trait values for strains based on whether they have ALT
904 and REF alleles for specific *irld-39* and *irld-52* variants predicated to disrupt protein function. The *irld-52*
905 variant p-value is $p = 0.007$ for the PC1 trait value (only the Slope trait value is shown here). Significance
906 determined from GWA fine mapping. (**G**) Slope trait values for strains based on whether they are hyper-
907 divergent or not at *irld-11* and *irld-57* loci. T-test on trait values between hyper-divergent and non-divergent
908 strains. (**F-G**) DL238, EG4725, NIC526, and MY2147 are color-coded as indicated. (**H**) The four *irld* genes
909 selected for genome editing and the edits generated for each. For *irld-39 and irld-52,* N2 and MY2147 have
910 the REF allele and were edited to have the ALT allele. *irld-11* and *irld-57* are hyper-divergent in DL238 and
911 EG2745 backgrounds, so full gene deletions were generated in N2 and MY2147 backgrounds. (**I**) L1
912 starvation survival assays on *irld-39* and *irld-52* ALT alleles in N2 and MY2147 backgrounds. There were
913 no significant differences between strains within a background. (**J**) L1 starvation survival assays on *irld-11*
914 and *irld-57* deletions in N2 and MY2147 backgrounds. There were no significant differences between
915 strains within a background. (**K-L**). Worm length following 48 hr recovery with food after 1 or 8 d of L1
916 starvation for indicated genotypes. Linear mixed-effects model; one-way p-value for interaction between
917 strain and duration of starvation; 4-5 biological replicates per condition. ΔΔ indicates effect size of
918 interaction between duration of starvation and strain compared to control (the difference in differences
919 between strains' mean length at days 1 and 8). (**F,G,K,L**) ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, n.s. not
920 significant.
921

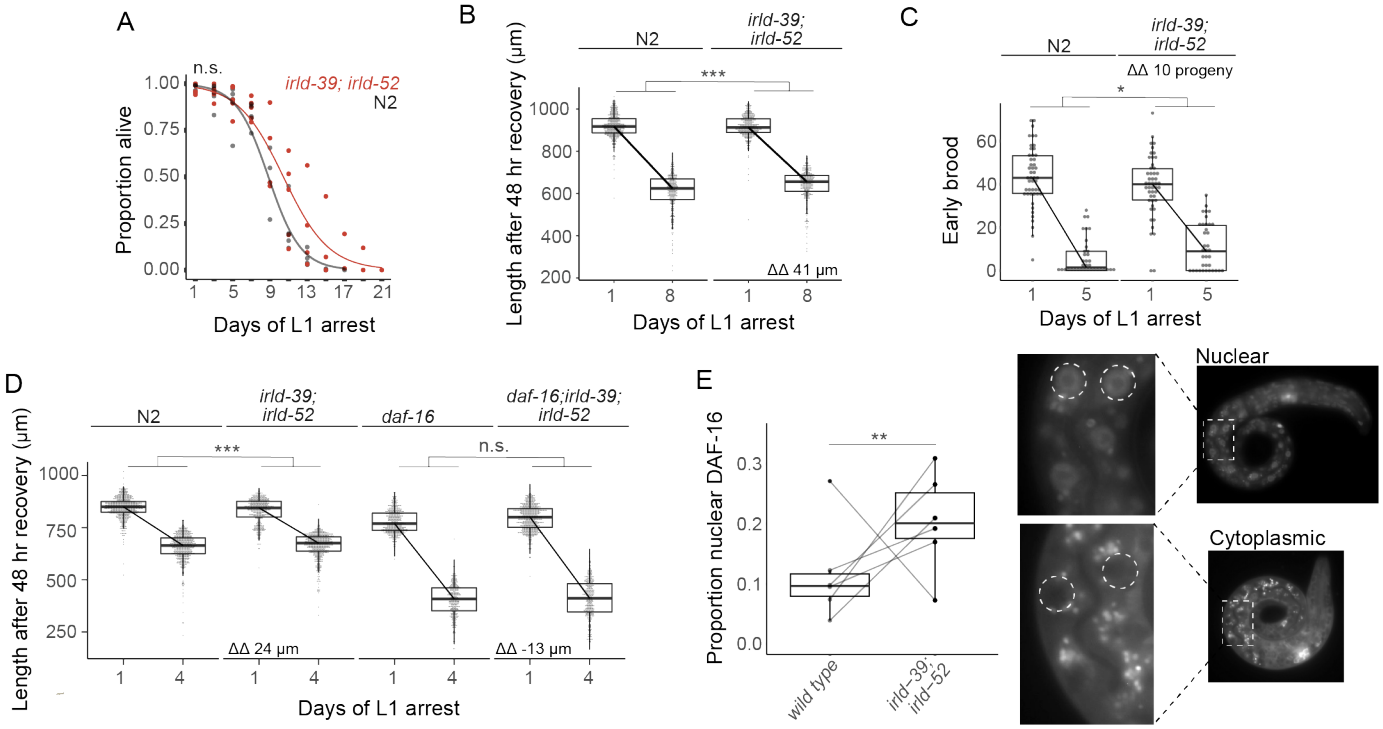


Fig. 4. IRLD-39 and IRLD-52 together impact starvation resistance and depend on DAF-16. (**A**) Survival curves of *irld-39(duk1); irld-52(duk17)* and N2 throughout L1 starvation. The apparent increase in starvation survival in the double mutant is not statistically significant (p = 0.14). (**B**) Worm length of *irld-39(duk1); irld-52(duk17)* and N2 following 48 hours of recovery with food after 1 or 8 days of L1 starvation. (**C**) Number of progeny produced between 48 and 72 hours of recovery with food after 1 or 5 days of L1 starvation. (**D**) Worm length of N2, *irld-39(duk1); irld-52(duk17), daf-16(mu86),* and *daf-16(mu86); irld-39(duk1); irld-52(duk17)* following 48 hours of recovery with food after 1 or 4 days of L1 starvation. (**B-D**) Linear mixed-effects model with duration of L1 starvation and genotype as fixed effects and the number of replicates as a random effect; p-value calculated for interaction between fixed effects. ΔΔ indicates effect size of interaction between duration of starvation and strain compared to control. (**E**) Nuclear localization of DAF-16::GFP in intestinal cells of starved L1s ~36 hours after hatching. Each point represents the result of a single independent biological replicate with 51-64 worms scored for each condition and replicate, with a line connecting the two genotypes in each replicate. The Cochran-Mantel-Haenszel test was used to determine differences in the distribution of the two categories (nuclear and cytoplasmic) between *daf-*

937 *16(ot971)* (wild type) and *daf-16(ot971); irld-39(duk1); irld-52(duk17)* (*irld-39; irld-52*). Images of intestinal

938 nuclear and cytoplasmic localization are shown. (**A-E**) 4-6 biological replicates were performed per

939 experiment. ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, n.s. not significant.

940

SUPPLEMENTARY FIGURES

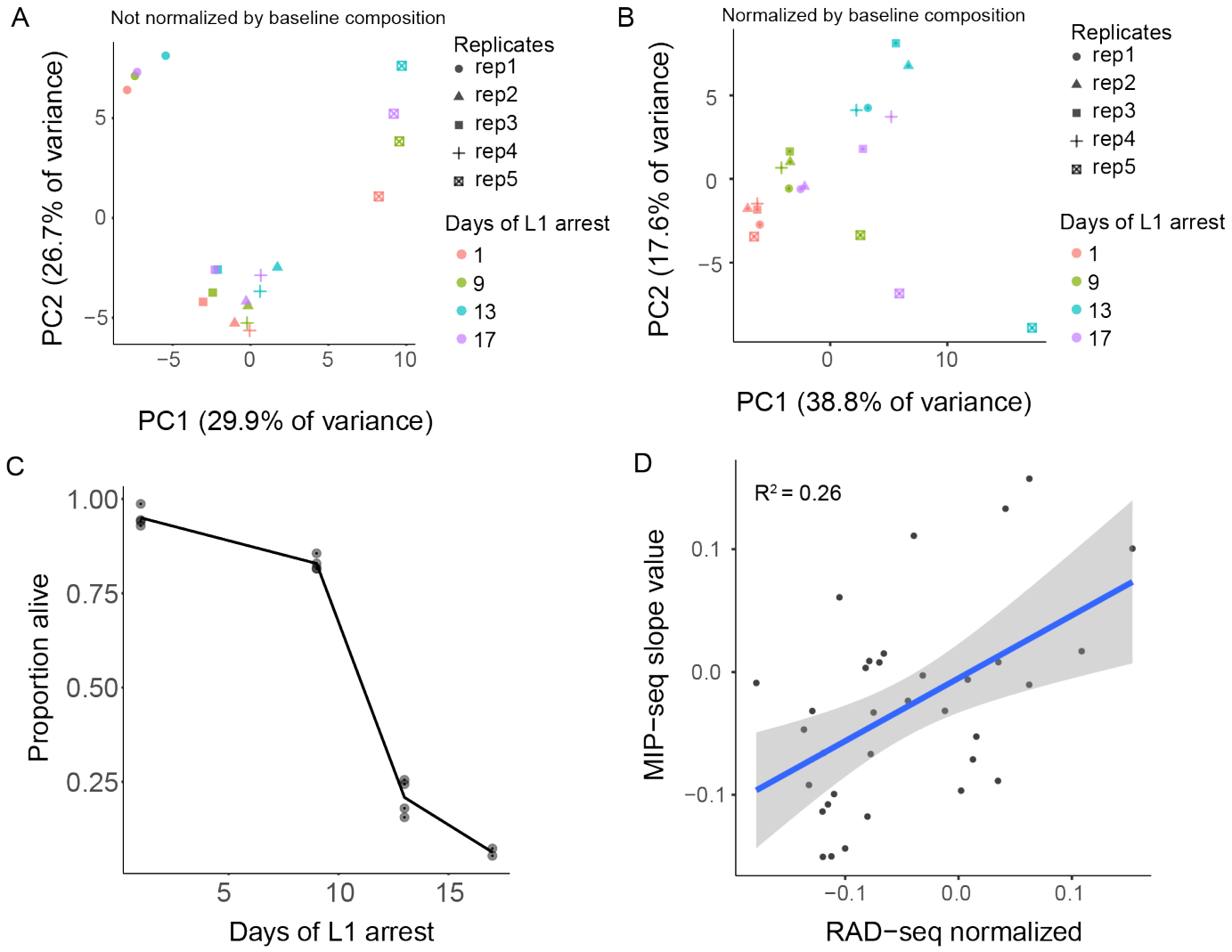


Fig. S1. MIP-seq data analysis and comparison to RAD-seq analysis of starvation resistance. (A) PCA of all replicates and recovery time points assayed not normalized by baseline composition. Samples cluster by replicate. (**B**) PCA of all replicates and recovery time points normalized by baseline composition 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 (WEBSTER *et al.* 2019). 34 strains included in both experiments are plotted.

950 RAD-seq results are based on only a single replicate, and the vast majority of reads from RAD-seq do not

951 cover unique SNV greatly limiting sequencing depth at informative loci. $R^2 = 0.26$ ($p = 0.002$).

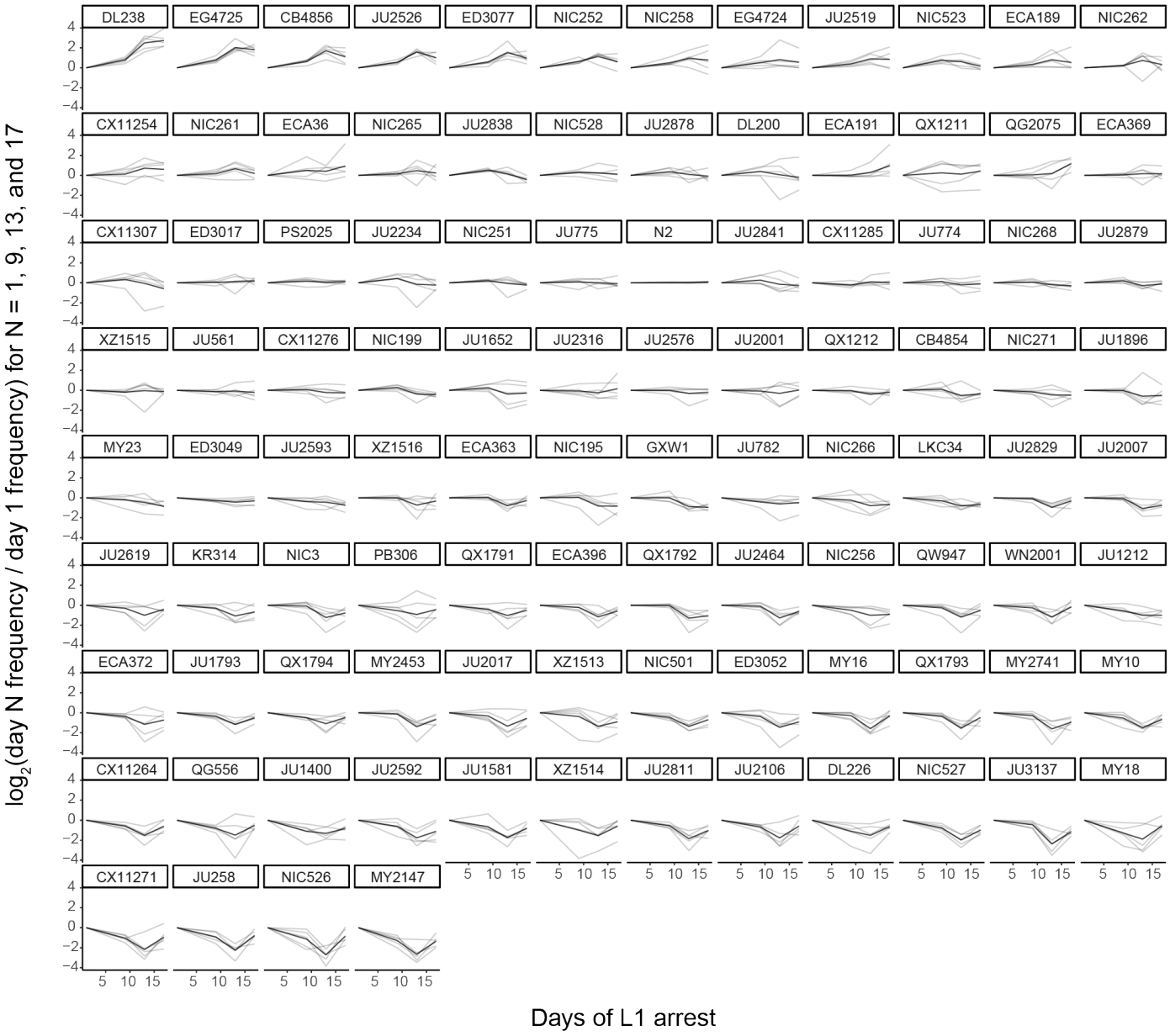


Fig. S2. MIP-seq strain frequencies throughout starvation for all strains assayed. MIP-seq replicates were first normalized by baseline and then by day 1. Strains are rank-ordered from most starvation-resistant to most sensitive by Slope trait values. Gray lines are individual replicates and black lines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

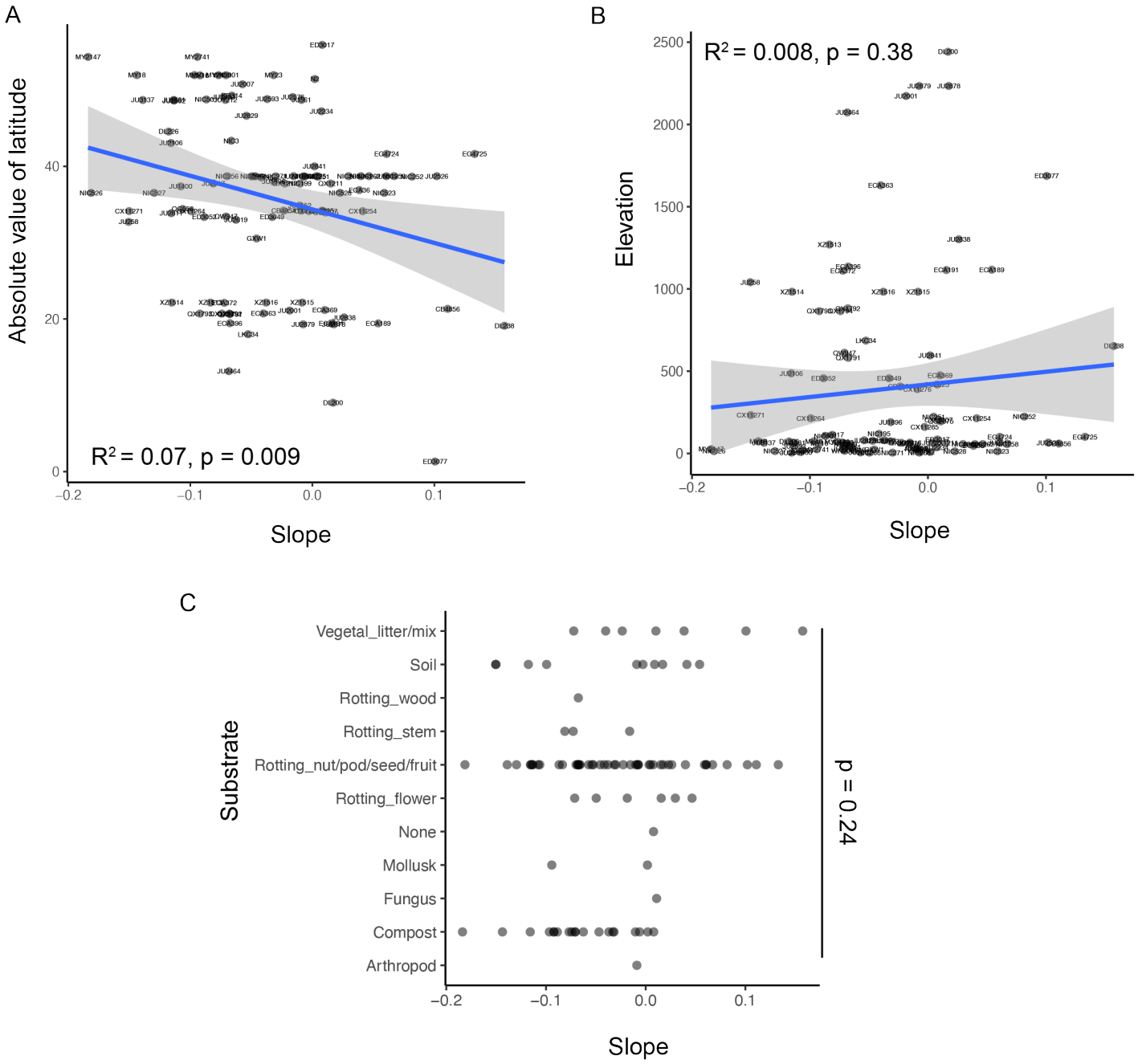
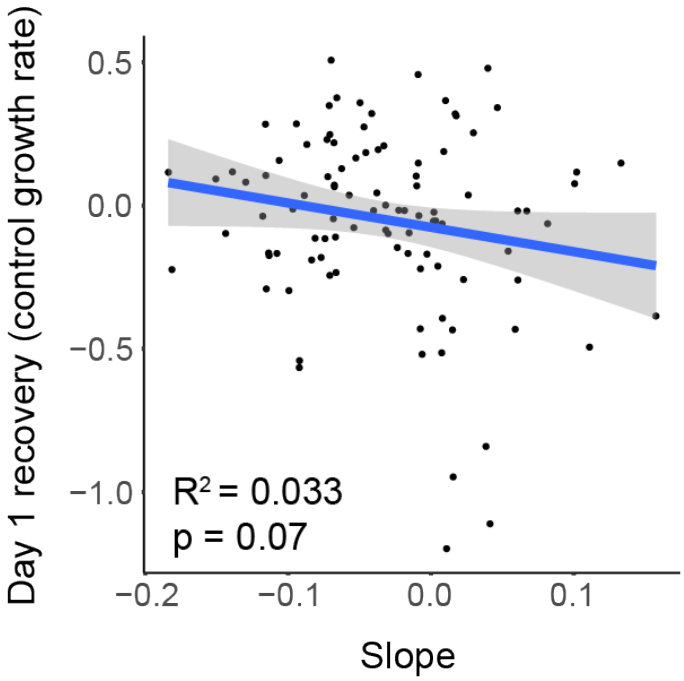


Fig. S3. Starvation resistance of wild strains is associated with latitude at collection site, but not elevation or substrate. (**A**) 'Slope' trait value for each strain plotted against the absolute value of its latitude of collection. $R^2 = 0.07$, $p = 0.009$. (**B**) 'Slope' trait value for each 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-

962 statistic of slope coefficient from linear model. (**C**) Slope trait values plotted for strains collected from each

963 type of substrate. One-way ANOVA across all substrates, $p = 0.24$.



964

965 **Fig. S4. Starvation resistance is negatively correlated with growth rate.** Slope trait value is plotted

966 against 'Early Growth' trait value calculated by determining which strains became over- or under-

967 represented after recovery from day 1 of L1 starvation relative to baseline composition at day 1 of L1

968 starvation. $R^2 = 0.033$, significance determined by t-test on t-statistic of slope coefficient from linear model.

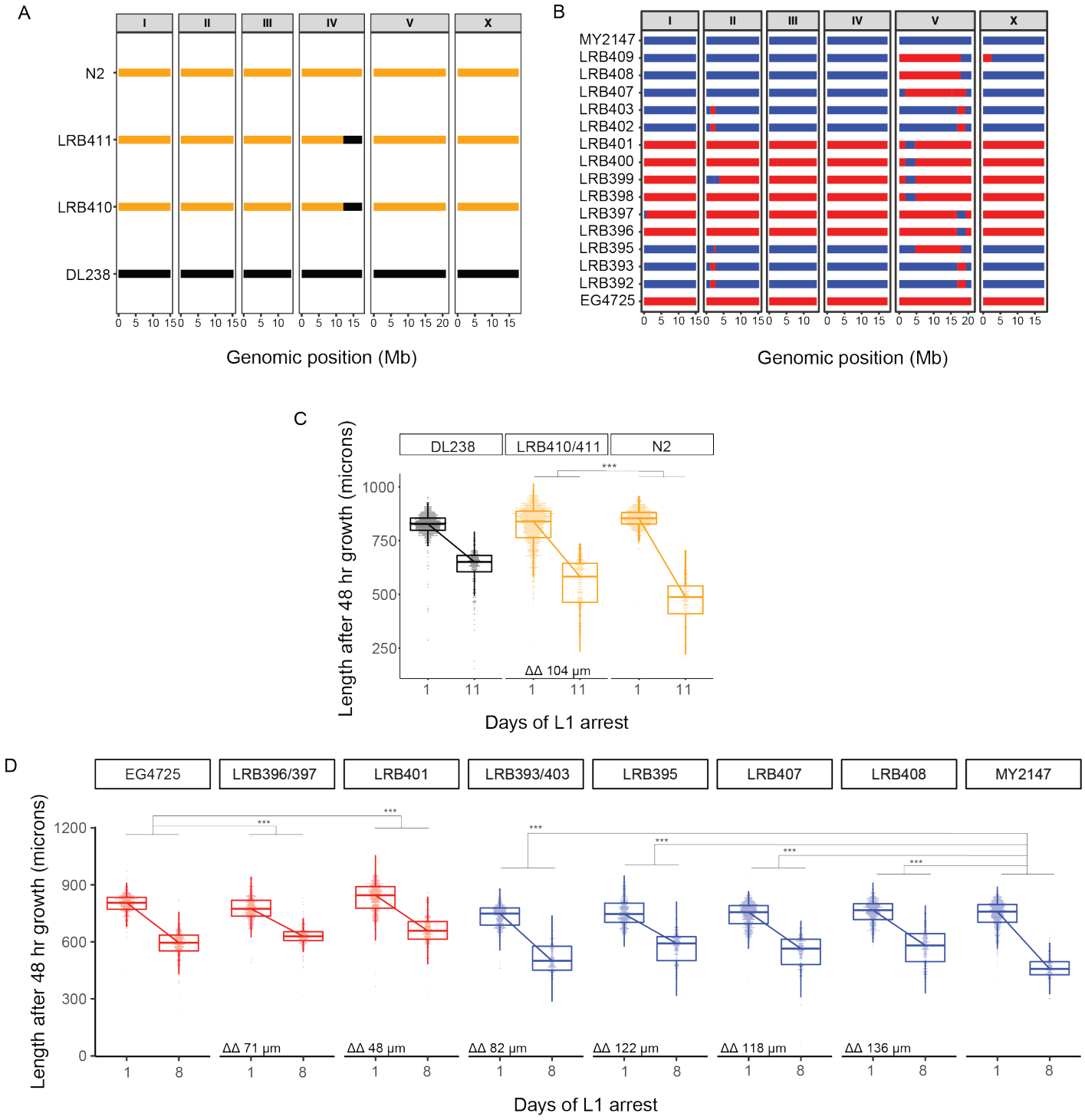


Fig. S5. Validation of QTL with sequenced near-isogenic lines (NILs). (**A**) Schematic of low-depth (~1x) genome-sequencing results of NILs used to validate chromosome IV QTL. Parental strains are N2 and DL238. DL238 can be introgressed into N2 but not *vice versa* due to genetic incompatibility (BEN-DAVID *et al.* 2017). N2 sequence is in gold, while DL238 sequence is in black. (**B**) Schematic of low-depth

974 (~1x) genome-sequencing results of NILs used to validate chromosome V QTL. Parental strains are
975 MY2147 and EG4725, which are compatible and were introgressed reciprocally. MY2147 sequence is in
976 blue, while EG4725 sequence is in red. (**C**) Worm length following 48 hr recovery with food after 1 or 11 d
977 of L1 starvation for indicated strains. Data were merged for essentially duplicate strains LRB410 and
978 LRB411. Color-coding indicates the primary parental background (*e.g.,* LRB410 is primarily the N2
979 background so is plotted in yellow). (**D**) Worm length following 48 hr recovery with food after 1 or 8 d of L1
980 starvation for indicated strains. Data were merged for essentially duplicate strain pairs LRB396 and
981 LRB397, and LRB393 and LRB403. Color-coding indicates the primary parental background (*e.g.,* LRB401
982 is red because its sequence is primarily EG4725). (**C-D**) ***$p < 0.001$; linear mixed-effects model; p-value
983 of interaction between strain and duration of starvation; eight biological replicates. ΔΔ indicates effect size
984 of interaction between duration of starvation and strain data plotted in that panel compared to the
985 predominant parental strain background (the difference in differences between strains' mean length at days
986 1 and 8).

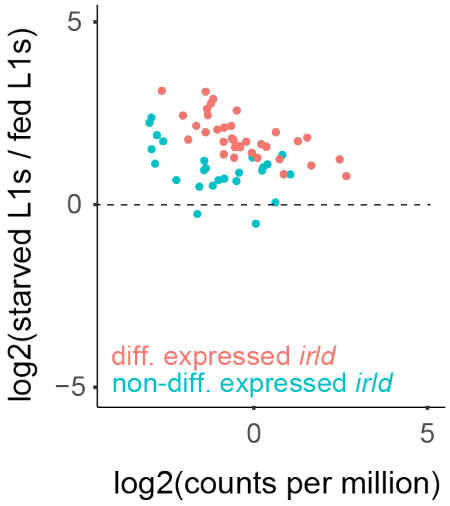
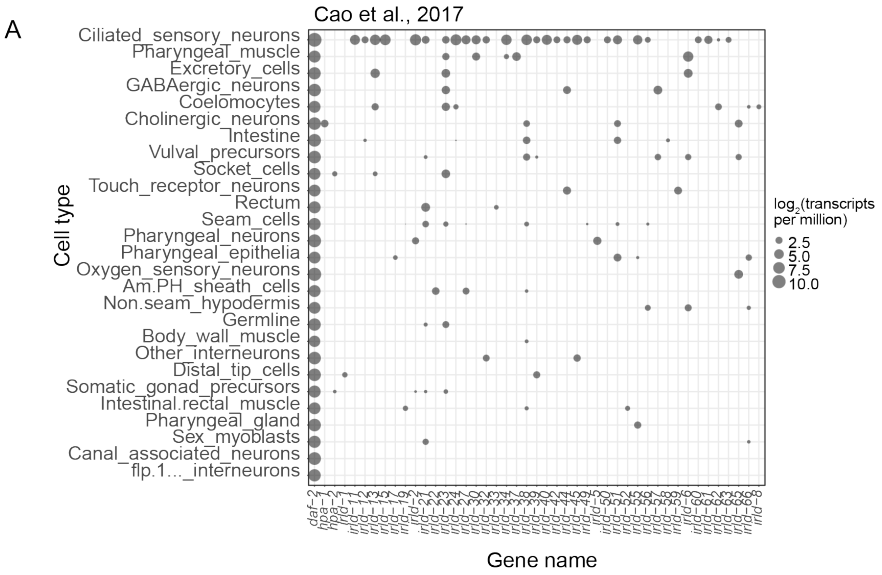
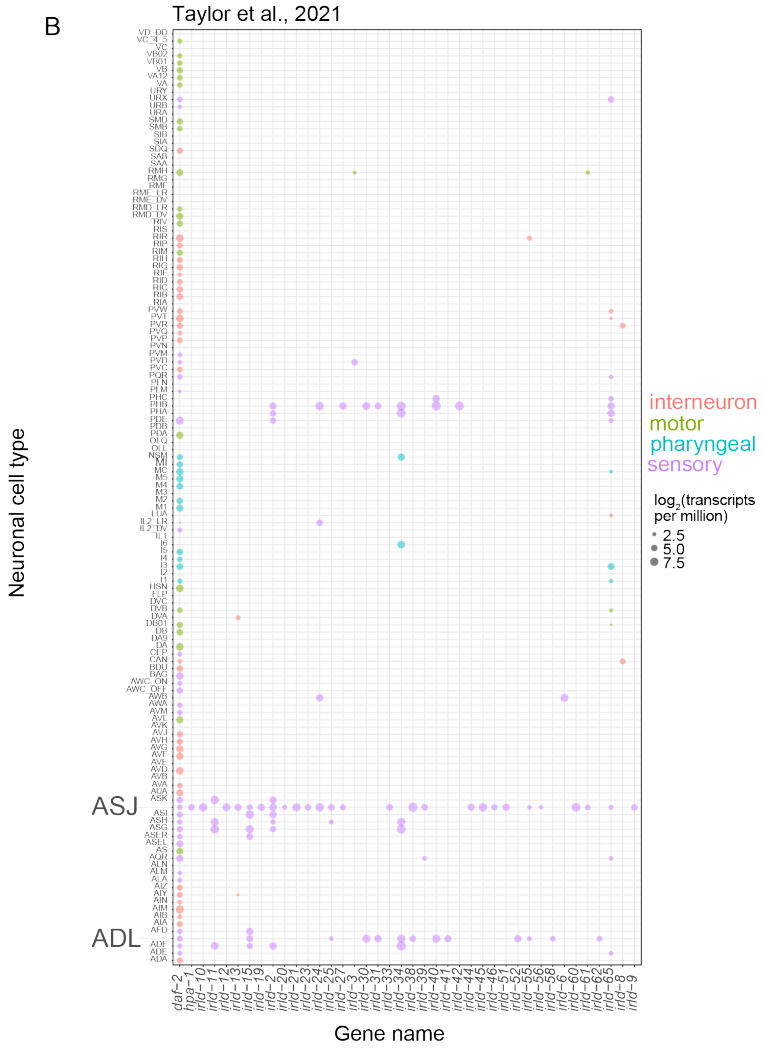


Fig. S6. *irld* genes are up-regulated in starved L1s compared to fed L1s.

Starved and fed L1 mRNA-seq data from whole, wild-type (N2) worms were re-analyzed from Webster et al. (2018) without a minimum expression filter. *irld* genes are expressed at low levels (typically below one count-per-million reads) but exhibit a consistent pattern of up-regulation in starved compared to fed L1s. Color-coding indicates significance at FDR < 0.05. "diff." differentially; "non-diff." non-differentially.


A
Cao et al., 2017
Cell type
Ciliated_sensory_neurons
Pharyngeal_muscle
Excretory_cells
GABAergic_neurons
Coelomocytes
Cholinergic_neurons
Intestine
Vulval_precursors
Socket_cells
Touch_receptor_neurons
Rectum
Seam_cells
Pharyngeal_neurons
Pharyngeal_epithelia
Oxygen_sensory_neurons
Am.PH_sheath_cells
Non.seam_hypodermis
Germline
Body_wall_muscle
Other_interneurons
Distal_tip_cells
Somatic_gonad_precursors
Intestinal.rectal_muscle
Pharyngeal_gland
Sex_myoblasts
Canal_associated_neurons
flp.1..._interneurons
Gene name
log2(transcripts per million)
2.5
5.0
7.5
10.0


B
Taylor et al., 2021
Neuronal cell type
ASJ
ADL
Gene name
interneuron
motor
pharyngeal
sensory
log2(transcripts per million)
2.5
5.0
7.5

1007

1008 **Fig. S7. *irld* genes exhibit a bias toward expression in sensory neurons with some expression in**
1009 **other cell types.** Single-cell RNA-seq expression levels for all detectable *irld* genes and the insulin/IGF
1010 receptor *daf-2* included for comparison. (**A**) Expression levels (transcripts-per-million) from L2-stage larvae
1011 for all available cell types are plotted (CAO *et al.* 2017). (**B**) Expression levels from L4 larvae for all neurons
1012 (TAYLOR *et al.* 2021) are plotted. Neurons are sorted into interneuron, motor, pharyngeal, and sensory
1013 classifications. Bubble size indicates $\log_2$ expression level in transcripts-per-million.

1014

1015 **Data S1. (Data_S1.xlsx)**

1016 This file includes all MIP-seq processed data: Slope and PC1 trait values used in GWA, output from
1017 MIPgen, MIP sequences, MIPs included in the final experiment, MIP primer sequences, count data for MIP-
1018 seq starvation resistance experiment and two pilot experiments.

1019 **Data S2. (Data_S2.xlsx)**

1020 This file includes GWA output and follow-up on *irld* candidates: GWA output for both Slope and PC1, genes
1021 within QTL, output from WormCat and protein domain enrichment analyses, hyper-divergence status for
1022 each strain, CRISPR sequences, genotyping primers, and NIL sequencing results.

1023 **Data S3. (Data_S3.xlsx)**

1024 This file includes input and output for RNA-seq analysis, including count tables and differential expression
1025 output from Webster et al. (2018) used in Fig S6, transcripts-per-million from Cao et al. (2017) plotted in
1026 Fig S7, and transcripts-per-million for threshold 3 from Taylor et al. (2021) plotted in Fig S7.