**Modifiers of Propionate Sensitivity in *C. elegans*: a Model for Human Inborn Errors in Metabolism [BROAD BUT VAGUE]**

**Or**

**Natural Variation in Glucuronyltransferase GLCT-3 Modulates Propionate Sensitivity in *C. elegans* [BORING]**

Huimin Na1#, Stefan Zdraljevic2#, Albertha J.M. Walhout1\* and Erik C. Andersen2\*

1. Program in Systems Biology and Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, MA 01605, USA

2. Department of Molecular Biosciences, Northwestern University, Evanston, IL 60208, USA

# These authors contribute equally to this work

\* Corresponding authors

Email: [erik.andersen@northwestern.edu](mailto:erik.andersen@northwestern.edu); [marian.walhout@umassmed.edu](mailto:marian.walhout@umassmed.edu);

**ABSTRACT**

**Recessive mutations in human metabolic genes can lead to rare diseases known as inborn errors of human metabolism. One class of these are acidurias, which are characterized by a toxic buildup of small organic acids. For instance, patients with recessive mutations in either subunit of propionyl-CoA carboxylase suffer from propionic acidemia because they cannot catabolize propionate, leading to its harmful accumulation. Interestingly, both the penetrance and expressivity of inborn errors of metabolism can be modulated by genetic background. However, modifiers of these diseases have not been identified, in part because of the lack of statistical power for rare diseases in human population genetics. Here, we use the nematode *Caenorhabditis elegans* to identify genetic modifiers of propionate sensitivity. We use a panel of 132 natural strains in genome-wide association mapping to identify two quantitative trait loci (QTL) associated with reduced propionate sensitivity. We find that natural variation in the putative glucuronosyltransferase GLCT-3, a homolog of human B3GAT, partly explains the difference in propionate sensitivity between a sensitive strain (BRC20067) and a resistant strain (DL238). By CRISPR/Cas9, we demonstrate that loss-of-function alleles in *glct-3* indeed render the animals less sensitive to propionate. [something about the family here?]. Our findings demonstrate that natural variation in metabolic enzymes can contribute to propionate sensitivity and provide a framework for characterizing genetic background contribution to a wide variety of inborn errors in human metabolism using *C. elegans* as a model.**

**INTRODUCTION**

Inborn errors of human metabolism are rare genetic diseases in which dietary nutrients or cellular metabolites cannot be broken down to generate energy, biomass or remove toxic compounds. Most of these disorders are caused by recessive mutations in genes encoding metabolic enzymes or metabolite transporters. Inborn errors of metabolism are often considered monogenic disorders, however, the penetrance and expressivity of these diseases can vary {Argmann, 2016 #3436}. Therefore, it has been proposed that such diseases should be viewed as more complex traits in which not only environmental factors such as diet, but also genetic background, affect the age of onset and severity of the disease {Argmann, 2016 #3436}. In different genetic backgrounds, modifier genes can harbor genetic variation, and affect inborn error of metabolism penetrance and expressivity. However, because such diseases are rare, often with incidences of 1:50,000 or fewer, identifying modifier genes in human populations has been exceedingly difficult {Argmann, 2016 #3436}{Saudubray, 2018 #3598}.

Propionic and methylmalonic acidemia are inborn errors of metabolism in which the short chain fatty acid propionate cannot be broken down {Deodato, 2006 #3135}. Patients with propionic acidemia carry recessive mutations in either one of two genes, PCCA or PCCB, which encode the two proteins comprising propionyl-CoA carboxylase that converts propionyl-CoA to D-methylmalonyl-CoA. Methylmalonic acidemia is a bit more complicated because it can be caused by mutations in either methylmalonyl-CoA racemase, methylmalonyl-CoA mutase, or in enzymes involved in the processing of vitamin B12, which is an essential cofactor for the latter protein {Banerjee, 2003 #2684}{Deodato, 2006 #3135}. Propionyl-CoA is generated in the natural breakdown of the branched chain amino acids isoleucine and valine, as well as the catabolism of methionine, threonine and odd-chain fatty acids. It can be inter-converted with propionate, which is generated by our gut microbiota in the process of the digestion of plant fibers. While propionate has been found to have beneficial functions {Kasubuchi, 2015 #3142}{Hosseini, 2011 #2770}, it is toxic when it accumulates, as exemplified by patients with propionic acidemia {Deodato, 2006 #3135}. Propionic acidemia is a rare disorder with a worldwide live birth incidence of 1:50,000 to 1:100,000. It is diagnosed by newborn screening by the detection of elevated levels of propionylcarnitine, 3-hydroxypropionate, and other aberrant metabolites {Matsumoto, 1996 #3068}.

The nematode *Caenorhabditis elegans* is a bacterivore found in temperate climates around the world. In the laboratory, *C. elegans* can be fed different bacterial species and strains and the vast majority of studies have used *Escherichia coli* OP50, a uracil auxotroph, as a diet. *E. coli* OP50 forms thin lawns, enabling facile visualization of the transparent animal. However, *E. coli* OP50 cannot synthesize vitamin B12 and therefore cannot support the efficient breakdown of propionate by the canonical pathway {Watson, 2013 #2432;Watson, 2014 #2714}. Previously, we found that *C. elegans* transcriptionally activates an alternative propionate breakdown pathway, or shunt, when flux through the canonical pathway is low due to genetic perturbations or low dietary vitamin B12 {Watson, 2016 #3248}{Bulcha, 2019 #3502}. This -oxidation pathway comprises five genes, and results in the generation of acetyl-CoA {Watson, 2016 #3248}(**Figure 1A**). *C. elegans* may have evolved a dedicated pathway for alternate propionate breakdown to be able to thrive on bacteria that do not synthesize vitamin B12. It only activates the expression of propionate shunt genes when propionate accumulation is persistent, via a specific regulatory circuit known as a type 1 feed-forward loop with AND-logic gate comprising the nuclear hormone receptors *nhr-10* and *nhr-68* {Bulcha, 2019 #3502}. In humans, there is evidence for the presence of the propionate shunt, although its activity is not sufficient to mitigate propionate toxicity. This is likely because in humans the enzymes functioning in other metabolic pathways are repurposed {Watson, 2016 #3248}.

The vast majority of studies using *C. elegans* have relied on a strain named N2, from Bristol, England. Over the last years, hundreds of other wild *C. elegans* isolates have been collected from all over the world. *C. elegans* is a self-fertilizing hermaphrodite and, therefore, different wild strains can be easily maintained as fully isogenic strains. These different strains have been used to identify quantitative trait loci that contribute to a variety of phenotypes, including anthelmintic and cancer chemotherapeutic resistance, and in several cases the precise genotypic variation that is causal to phenotypic variation has been determined {Reddy, 2009 #3369}{Ghosh, 2012 #3368}{Zdraljevic, 2017 #3594}{Brady, 2019 #3595}{Greene, 2016 #3377}{Burga, 2019 #3611}. Genomic information about the different strains is tabulated in the *C. elegans* Natural Diversity Resource (CeNDR), together with different tools for genome wide association (GWA) mapping {Cook, 2017 #3365}.

Here, we use wild *C. elegans* isolates to identify natural variation in loci that modifies the resistance to exogenous propionate supplementation. To mimic propionic acidemia metabolic conditions, we used animals fed a diet of *Escherichia coli* OP50, which is low in vitamin B12, resulting in low flux through the canonical propionate breakdown pathway {Watson, 2014 #2714;Watson, 2016 #3248}. GWA mapping using 132 strains identified two independent quantitative trait loci (QTL) associated with propionate resistance. For one of these loci, we identify the causal variant in *glct-3,* which encodes a predicted beta-1,3-glucuronyltransferase, and is an ortholog of human B3GAT1, 2, and 3. Glucuronyltransferases catalyze reactions between metabolites, specifically the addition of glucuronic acid to toxic metabolites such as drugs {Rowland, 2013 #3612}. [when I read a bit more, it seems that they can modify glycosylated proteins? We need more info here and for the Discussion from Nana/Stefan]. Interestingly, loss of function mutations in *glct-3* confer resistance to propionate. Our data show that quantitative toxicity phenotyping can be used to identify candidate modifier genes of traits associated with inborn errors in human metabolism.

**RESULTS**

**Different Sensitivities to Propionate in Wild *C. elegans* Strains**

We have previously used propionate dose-response curves (DRCs) using larval stage 1 (L1) arrest as a readout with the N2 *C. elegans* strain, and found an LD50 of ~80 mM in wild type animals {Watson, 2014 #2714}{Watson, 2016 #3248}{Bulcha, 2019 #3502}. When supplemented with vitamin B12, the LD50 was increased to 120 mM, while in animals carrying a deletion in the propionyl-CoA carboxylase ortholog *pcca-1,* or the first gene of the propionate shunt, *acdh-1*, the LD50 was reduced to 50 mM {Watson, 2014 #2714}{Watson, 2016 #3248}. It is technically not feasible to perform DRCs for large numbers of wild strains. Therefore, we first asked whether wild *C. elegans* would exhibit differences in propionate sensitivity using 12 strains that represent high genetic diversity {Andersen, 2012 #3363}. To mimic metabolic conditions of human propionic acidemia, we fed the animals vitamin B12-deplete *E. coli* OP50 bacteria which ensures that flux through the canonical propionate breakdown pathway was low {Watson, 2014 #2714}{Watson, 2016 #3248}. We performed three biological replicate experiments, each consisting of three technical replicates, and found that the 12 strains exhibit varying degrees of propionate sensitivity (**Figures 1B, 1C**). Nine of the strains had similar propionate sensitivities as the N2 strain with an LD50 of ~85 mM. The other three strains were more resistant to propionate with an LD50 of ~100-110 mM. Next, we carefully titrated propionate at concentrations between 80 and 120 mM with 10 mM increments. We confirmed that most strains exhibited sensitivity similar to the N2 strain, while two, DL238 and EG4725, were significantly more resistant (**Figure 1D**).

In order to perform GWA, we needed to test propionate sensitivity in a broader panel of wild *C. elegans* strains. We reasoned that a single dose of propionate that could discriminate between more sensitive and more resistant strains would be a sensible approach. Propionate sensitivity assays can be noisy, in part because of slight variations in experimental and environmental factors such as incubator and room temperature, propionate concentrations (which can change slightly due to evaporation and dilution), *etc*. We needed to determine which dose of propionate to use in the broader panel of strains, and found that a concentration of 100 mM had the highest broad-sense heritability (*H2* = 0.79)(**Figures 1E, 1F**). Therefore, we selected this dose for further experiments.

As mentioned above, propionate sensitivity assays are noisy, and indeed we observed substantial variability across the dose-response experiments. Therefore, we performed power analysis to determine the number of replicate experiments that needed to be performed prior to testing a large number of wild strains. We found that five independent experiments, each with four technical replicates, would give us 80% power to detect a 20% difference in propionate sensitivity (**Figure S1**).

**Genome-Wide Association Mapping Identifies Loci that Modify Propionate Sensitivity**

To identify the genetic basis of propionate response variation in *C. elegans*, we exposed 132 wild strains to 100 mM propionate and measured L1 survival (Table S1 – GWA strains and phenotypes)[this table needs to be made asap]. We tested the strains in three batches, and included six strains in every batch to control for potential batch effects (**Figure S2**). We observed a broad range of propionate sensitivities from strains exhibiting high survival, indicating propionate resistance, to sensitive strains with low survival (**Figure 2A**). Next, we performed GWA mapping with this dataset and identified five QTL that were above the Bonferroni-corrected significance threshold (**Figure 2B**). We identified two QTL on chromosome II (II:1880662-1993488, II:13859466-13979658; All chromosome coordinates are based on N2 WS245), two QTL on chromosome V (V:3213649-4284434, V:19229887-19390858), and one QTL on chromosome X (X:9987812-10370303). To test the independence of these five QTL, we calculated the pairwise linkage disequilibrium (LD) between each of the peak QTL markers (**Figure S3 - LD**). We observed low levels of LD for the majority of QTL pairs, with the exception of the two QTL on chromosome V (*r2* = 0.87, peak markers - V:3929669 and V:19356375). Because multiple QTL were associated with propionate sensitivity, it was difficult to decide which QTL to characterize in more detail. Therefore, we next used the sequence kernel association test (SKAT), which looks for an association between the phenotype of interest and the cumulative variation on a gene-by-gene basis {Wu, 2011 #3599}. This approach identified two QTL, one that overlaps with the QTL on chromosome V (V:3213649-4284434) identified using the single-marker mapping approach (**Figure 2B**), and a novel QTL on chromosome I (I:12374204-12388791) (**Figure 2C**)[is this really novel compared to figure 1B?]. This additional support for the QTL on the left arm of chromosome V motivated us to further investigate this genomic region.

**Chromosome V Near-Isogenic Lines do not Recapitulate Propionate Resistance**

To test the contribution of the QTL on chromosome V, we constructed near-isogenic lines (NILs) in which the region associated with propionate resistance (V:3213649-4284434) was crossed from a resistant into the genome of a sensitive strain. To identify candidate parental strains for NIL construction, we focused on the 12 strains that were phenotyped in the dose-response experiment (**Figure 1C**). Of these 12 strains, two were significantly resistant to propionate and ten were sensitive. Next, we verified that the propionate-resistant strains had the alternative genotype at the peak QTL marker identified using the single-marker mapping method and were compatible with propionate sensitive strains at the *peel-1-zeel-1* {Seidel, 2008 #3607} and *sup-35-pha-1* {Ben-David, 2017 #3608} incompatibility loci. Using these criteria, we identified DL238 (propionate-resistant) and BRC20067 (propionate-sensitive) as suitable parental strains for NIL construction. We constructed nine NILs that contained the DL238 genomic region surrounding the chromosome V QTL introgressed into the BRC20067 genetic background (**Figure 3A**). When we exposed these NILs to propionate, we observed that the DL238 introgressed regions that correspond to the chromosome V QTL confidence interval did not confer propionate resistance (**Figure 3B**). Because the genomic region spanned by these NILs is larger than the QTL confidence interval, these results suggest that the chromosome V QTL we identified might have been the result of a spurious association with genetic variation in this region. Alternatively, the other QTL identified by the single-marker mapping approach might contain genetic variation that contributes to propionate sensitivity, there may be complex epistatic interactions among these QTL, or the unique QTL identified by SKAT mapping contains variation that contributes to propionate sensitivity. Therefore, we next focused on the QTL on chromosome I (**Figure 2C**).

**Variation in the *glct-3* Gene Confers Propionate Resistance**

Closer inspection of the QTL on chromosome I showed that it encompasses only four genes [I suggested a cartoon of the four genes as Figure 4A – also panels A and B are crazy large and don’t need to be]. Burden mapping revealed that variation in *glct-3* is most correlated with propionate sensitivity in the *C. elegans* population (**Figure S4**). Within the *glct-3* gene, we observed eight distinct combinations of alleles (haplotypes) among the phenotyped population (**Figure S4**). Five of these distinct haplotypes included a stop-gained variant at amino acid position 16 (Gly16\*). These five haplotypes include Gly16\* as well as other variants: Gly17Arg, Ser50Ala, Ser111Thr, Tyr231Cys in the QX1793 strain; Gly17Arg Ser50Ala in the CX11276, DL238, ED3046, ED3049, and NIC252 strains; Leu184Phe in the ECA36 strain; Gly16\* Ile46Thr in the QX1792 strain, and only Gly16\* in the MY23 and QX1791 strains. Strains with the Gly16\* variant were on average 20% more resistant to propionate treatment than strains with no variation in *glct-3* (**Figure 4A**). [note: we should discuss that this is not black and white and that there are thus more modifiers in *C. elegans* that remain to be discovered]. These results suggest that putative loss-of-function variation in the *glct-3* gene is associated with resistance to propionate.

To test whether variation in the *glct-3* is (partially) causative of the difference in propionate sensitivity, we generated two independent *glct-3* alleles in the propionate sensitive BRC20067 strain using CRISPR-Cas9 genome editing {Kim, 2014 #3034}. The *ww62* allele has a one base pair deletion at position 57 in the first exon that causes a frameshift in the reading frame leading to an early stop codon, and the *ww63* allele contains the Gly16\* variant that is present in DL238. We found that both the *ww62* and the *ww63* allele conferred resistance to propionate (**Figure 4B**)[the alleles are not indicated in the figure], demonstrating that the loss of *glct-3* function confers resistance to propionate. [for discussion: how much of the overall heritability is explained by *glct-3* variation?].

There are ~330 wild *C. elegans* isolates currently available in CeNDR {Cook, 2017 #3365}, 42 of which contain the Gly16\* variant in *glct-3*. A majority of strains that contain the Gly16\* variant (33/42) were isolated on the Hawaiian islands (**Figure 4C**), which are known to harbor the most genetically divergent *C. elegans* individuals (REFs). An additional three strains also have variants that are predicted to cause a loss of *glct-3* function (ECA733, JU1395, and ECA723). In agreement with the geographic distribution of the Gly16\* allele, this allele [which allele?] is predominantly found in genetically divergent *C. elegans* strains (**Figure 4D**). However, a subset of genetically divergent strains contain a functional copy of *glct-3* and two that have undergone the global selective sweep harbor the Gly16\* allele (REF)[I still have no clue what we’r talking about here – EA please clarify/expand], indicating that this locus has a complex evolutionary history in the *C. elegans* lineage (**Figure SX**).

**The *glct* Gene Family is Expanded in *C. elegans***

To further explore the evolutionary history of the *glct-3* gene, we examined the conservation of *glct-3* paralogs and their orthologs. The *glct-3* gene encodes a glucuronosyl transferase-like protein and has six paralogs in *C. elegans*, including five closely related genes *glct-1*, *glct-2*, *glct-4*, *glct-5*, and *glct-6*, and one distantly related paralog *sqv-8*, which we will not discuss further (**Figure 5A**)[not sure what Figure 5B is supposed to add]. Five *glct* genes (1-5) are located on an 80 kb region on chromosome I, while *glct-6* is located on chromosome IV. The close proximity of five of the six paralogs suggests that these genes are the products of gene duplication events (REF). We observed elevated levels of variation in the genomic region that contains *glct-1* through *glct-5* (**Figure 5B-C**), which supports the hypothesis that these sequences duplicated at some point in the *C. elegans* lineage and then diverged.

Next, we explored the conservation of the *glct* gene family across 20 species of *Caenorhabditis* nematodes, including ten for which the genome assembly was recently released (REF). We found that nine species contained only one *glct-3* ortholog, five contain two *glct-3* orthologs, three contain three *glct-3* orthologs, and one species each contain four, five, or six *glct-3* orthologs (**Figure SX**). The prevalence of low-copy numbers of *glct* genes among a majority of *Caenorhabditis* species suggests that the ancestral copy number state is fewer than the six copies found in the *C. elegans* genome. This idea is supported by the presence of one and two *glct-3* orthologs in the outgroup species *Heterorhabditis bacteriophora* and *Oscheius tipulae*, respectively.

Phylogenetic analysis of the DNA (**Figure SX**) and protein (**Figure SX**) sequences of the *C. elegans glct* genes suggests that *glct-6* is likely most homologous to the ancestral state of this gene family [I did not have these at this time]. This is because the GLCT-6 protein sequence more closely resembles orthologous proteins than its paralogs in *C. elegans*. Furthermore, the pattern of polymorphism in the six *C. elegans glct* paralogs suggests that after the initial duplication event, the function of the *glct-6* gene was retained, which is indicated by the absence of deleterious variants in this gene among wild isolates (Figure?). Similarly, *glct-4* has no variation that is predicted to cause a loss of gene function. By contrast, *glct-1*, *glct-2*, *glct-3*, and *glct-5* contain variants predicted to have a large effect on gene function. Among the 330 *C. elegans* strains, 24 have variation that is predicted to remove the function of two or more of these four genes, with two strains that have predicted loss-of-function alleles in all four genes. Taken together, these results suggest that the copy number of *glct* genes might affect fitness in the wild.

**DISCUSSION**

Points for Discussion

* Our findings
* The strongest QTL was not the right one – statistics does not always rule
* GLCTs and their function and anything related to human disease
* How much does glct-3 explain heritability of propionate sensitivity
* End with how awesome *C. elegans* is as model to study IEM and their modifiers

**MATERIALS AND METHODS**

**Strains**

All the wild strains were obtained from the CeNDR (**Table S1**){Cook, 2017 #3365 and maintained at 20°C on nematode growth medium (NGM) Plates on a diet of *E. coli* OP50. Near isogenic lines (NILs) were generated using a procedure described previously {Evans, 2018 #3548} by crossing BRC20067 and DL238. Each NIL strain harbors double recombinations occurred at different locations of ChrⅤ[not sure we use proper Roman symbols elsewhere, check consistency] generated by crossing two single recombinant strains, followed by six times backcrossing with BRC20067 to change the other five chromosomes into the BRC20067 background.

**Propionate Sensitivity Assay**

A 2 M propionic acid stock solution was prepared in a chemical hood. For 40 ml solution, 6 ml propionic acid (sigma, #402907), 13.5 ml 5 M NaOH, and 20.5 ml H2O were mixed together, and the pH was adjusted to 6.0 with NaOH. The solution was filter sterilized and stored at 4°C. On day 0, arrested L1 animals were placed on seeded plates with propionate and after incubation for two days, animals that proceeded through development we evaluated as survivors. Propionic acid survival rate was calculated as the proportion of survivors over the total number of L1 animals at day 0. Biological triplicate experiments with three technical replicates were performed. For the large panel of wild isolates, 100 mM propionate was used in five biological replicates, each with four technical replicates.

**Genome-Wide Association Mapping**

**Heritability Calculations**

For dose-response experiments, broad-sense heritability () estimates were calculated using the *lmer* function in the lme4 package with the following linear mixed-model (phenotype ~ 1 + (1|strain)). was then calculated as the fraction of the total variance that can be explained by the random component (strain) of the mixed model. For the complete dose-response experiment, we calculated per dose. For the fine-scale dose response experiment, we subsampled three replicates twelve independent times for calculations.

**Power Analysis**

To determine the number of replicate measures we needed to collect for association mapping, we measured L1 survival of the DL238 strain after exposure to 100 mM propionic acid in 40 replicates. The 40 replicates consisted of eight technical replicates across five independent preparations of agar plates with propionate. For a range of mean differences (0.01 to 1, in increments of 0.01), we subsampled two to eight replicates for each of the five plate preparations 100 times and calculated the standard deviation of L1 survival for the subsamples. To calculate the power to detect a difference across a range of replicates and mean differences, we used the *power.t.test* function in the pwr R package with the following parameters - *n* = number of subsampled replicates, *delta* = (0.01 to 1, in increments of 0.01), *sd* = mean of the standard deviation subsamples, *sig*.*level* = 0.00001, alternative = “two.sided”, type = “two.sample”. With four technical replicates across five independent plate preparations we were able to detect a 20% difference in means 80% of the time.

**Wild Isolate Propionate Sensitivity Phenotyping for GWA mapping**

We quantified propionate sensitivity of 132 wild isolates in five independent experiments with four technical replicates per experiment. The scale of this experiment required us to split the strains into three sets. To process the data, we first took the mean of the four technical replicates and removed biological replicate outliers, which were defined by mean + 1.5 x standard deviation. Next, we corrected the strain phenotype data for biological replicate and strain set using a linear model with the formula (phenotype ~ biological replicate + strain set). After this regression analysis, we removed outlier replicates as above and took the mean of the remaining replicate residuals per strain (**Supp Data XX - input mapping phenotype**). The resulting phenotype data was used for association mapping.

**Single-Marker Association Mapping**

GWA mapping was performed using phenotype data from 133 *C. elegans* isotypes. We performed the same mapping procedure as described previously {Zdraljevic, 2019 #3600}. Briefly, genotype data were acquired from the latest variant call format (VCF) release (Release 20180527) from CeNDR that was imputed using IBDseq, with the following parameters: minalleles = 5%, r2window = 1500, ibdtrim = 0, r2max = 0.8 {Browning, 2013 #3601}. We used BCFtools to filter variants that had any missing genotype calls and variants that were below 5% minor allele frequency {Li, 2011 #3602}. We used PLINK v1.9 to LD-prune the genotypes at a threshold of 0.8, using *--indep-pairwise 50 10 0.8* {Chang, 2015 #3603}{Purcell, 2007 #3604})*.* This resulting genotype set consisted of 59,241 markers that were used to generate the realized additive kinship matrix using the *A.mat* function in the *rrBLUP* R package (doi:10.3835/plantgenome2011.08.0024). These markers were also used for genome-wide mapping. However, because these markers still have substantial LD within this genotype set, we performed eigen decomposition of the correlation matrix of the genotype matrix using *eigs\_sym* function in Rspectra package {Li, 2005 #3606}( https://github.com/yixuan/RSpectra). The correlation matrix was generated using the *cor* function in the correlateR R package (https://github.com/AEBilgrau/correlateR). We set any eigenvalue greater than one from this analysis to one and summed all of the resulting eigenvalues. This number was 772, which corresponds to the number of independent tests within the genotype matrix. We used the *GWAS* function in the rrBLUP package to perform genome-wide mapping with the following command: *rrBLUP::GWAS(pheno = PC1, geno = Pruned\_Markers, K = KINSHIP, min.MAF = 0.05, n.core = 1, P3D = FALSE, plot = FALSE)*. To perform fine-mapping, we defined confidence intervals from the genome-wide mapping as +/- 100 SNVs from the rightmost and leftmost markers above the Bonferroni significance threshold. We then generated a QTL region of interest genotype matrix that was filtered as described above, with the one exception that we did not perform LD pruning. We used PLINK v1.9 to extract the LD between the markers used for fine mapping and the peak QTL marker identified from the genome-wide scan. We used the same command as above to perform fine mapping, but with the reduced variant set. The workflow for performing GWA mapping can be found on <https://github.com/AndersenLab/cegwas2-nf>.

**Sequence Kernel Association Test (SKAT) Mapping**

In parallel to single-marker mappings, *cegwas2-nf* performs burden-based mapping using SKAT was performed, which is implemented in the RVtests software {Wu, 2011 #3599}{Zhan, 2016 #3605}). We set the maximum allele frequency for SKAT to 5% using the --freqUpper from flag *cegwas2-nf*, and the minimum number of strains to share a variant for it to be considered to two using the --minburden flag.

**Linkage Disequilibrium**

We used the *LD* function from the genetics package in R to calculate linkage disequilibrium and report the *r2* correlation coefficient between the markers (https://CRAN.R-project.org/package=genetics).

**REFERENCES**

**FIGURES**

**Figure 1. Natural Variation in Propionate Sensitivity in 12 Genetically Diverse *C. elegans* Strains.**

A. Cartoon of propionate breakdown pathways in *C. elegans*. MM – methylmalonyl; TCA – tricarboxylic acid; MSA – malonic semialdehyde; HP – hydroxypropionate.

B. Dose response curves of 12 genetically distinct wild *C. elegans* strains. The Loess-smoothed fits of three biological replicates, each comprising three technical replicates, is shown by solid lines and the standard error of the fit is shown in gray. Propionate concentrations were tested from 0 to 140 mM in 20 mM increments. For reference, the N2 (orange, sensitive) and DL238 (blue, resistant) strains are colored. The dashed red line indicates 100 mM propionate.

C. LD50 values for L1 arrest in response to propionate supplementation for the 12 wild *C. elegans* strains. Three biological replicates each with three technical replicates were performed. (\*indicates P<0.05).

D. Dose response curves of 12 genetically distinct wild *C. elegans* strains. Propionate concentrations were tested from 80 to 120 mM in 10 mM increments. The dashed red line indicates 100 mM propionate.

E. Broad-sense heritability (*H2*) estimates for different concentrations of propionate.

F. Tukey boxplots of broad-sense heritability (*H2*) estimates. Each dot corresponds to an *H2* estimate after subsampling three replicate measures.

**Figure 2. Multiple QTL are Associated with Variable Propionate Sensitivity Among *C. elegans* Strains**

(A) Normalized L1 survival in the presence of 100 mM propionate for 132 wild *C. elegans* strains. L1 survival percentages were normalized by dividing each strain measurement by the maximum L1 survival percentage of all strains. Error bars show the standard deviation of replicate strain measurements.

(B) Manhattan plot from single-marker association mapping for the L1 survival percentage after propionate exposure. Each point represents an SNV [has this been defined?] that is present in at least 5% of the assayed wild population. The genomic position in Mb, separated by chromosome, is plotted on the x-axis and the *-log10(p)* for each SNV is plotted on the y-axis. SNVs are colored red if they pass the genome-wide Bonferroni-corrected significance (BF) threshold, which is denoted by the gray horizontal line. SNVs are colored pink if they pass the genome-wide Eigen-decomposition significance (ED) [do we use this abbreviation anywhere?] threshold, which is denoted by the dotted gray horizontal line. The genomic regions of interest surrounding the QTL that pass the BF indicated in cyan.

(C) Manhattan plot from SKAT burden mapping for L1 survival after propionate exposure. Each point represents a gene and is colored red if it passes the genome-wide BF threshold. The genomic position in Mb, separated by chromosome, is plotted on the x-axis and the *-log10(p)* for each gene is plotted on the y-axis.

**Figure 3. Chromosome V Near-Isogenic Lines do not Recapitulate Propionate Resistance**

(A) Chromosome V genotypes of near-isogenic lines (NILs) generated between BRC20067 and DL238. Blue corresponds to the DL238 genotype and pink corresponds to the BRC20067 genotype.

(B) Tukey box plots of the L1 survival phenotypes of each NIL and parental strains is shown on the x axis. Each dot represents a replicate L1 survival measurement. The red dotted line represents the mean L1 survival of the parental BRC20067 strain.

**Figure 4.** **Variation in *glct-3* Underlies Differential Propionate Sensitivity in *C. elegans***

(A) Box plot showing normalized L1 survival in response to 100 mM propionate supplementation for strains carrying the two different *glct-3* alleles [should we indicate gene rather than protein in figure? I know Gly indicates the protein but I think we use the gene throughout, definitely in Figure 4B]. Each dot represents strain.

(B) Tukey box plots of the L1 survival of each CRISPR-edited and parental strain upon supplementation of 100 mM propionate. Each dot represents a replicate L1 survival measurement.

(C) Map indicating where propionate resistant and sensitive strains were found.

(D) Phylogenetic tree indicating evolutionary history of different *C. elegans* strains.

**Figure 5.** ***glct* Family Expansion [EA: please provide legend]**

(A) An unrooted maximum-likelihood phylogeny of the six glucuronosyl transferase-like gene coding sequences present in the *C. elegans* genome. Branch colors correspond to the bootstrap support for the split, with pink indicating higher support.

(B)

(C)

**Figure S1. Power Calculation**

**Figure S2. Experimental Procedure and Individual Data Sets**

(A) L1 survival rate in the presence of 100 mM propionate for six *C. elegans* strains with four technical plus five biological replicates.

(B) Experimental diagram for large panel assays. 132 wild *C. elegans* strains were divided into three batches to test 100 mM propionate survival rate of L1 animals, and each batch contains 48 strains, including six control strains to control for batch effects.

(C) L1 survival rate in the presence of 100 mM propionate for 132 *C. elegans* strains in three batches. Red bars indicate averages for the six control strains.

**Figure S3. Linkage Disequilibrium of QTL Peaks Identified by Genome-Wide Association mapping [I think it is weird to use abbreviation for QTL but not GWA or LD]**

Linkage disequilibrium (r2) of peak QTL markers identified by genome-wide association mapping.

**Figure S4. Variation in chromosome I genes associated with propionate sensitivity [I don’t have this figure so I stopped here]**

The normalized L1 survival in the presence of propionate for each phenotyped C. elegans strain is shown on the x axis. The y axis represents unique haplotypes constructed from variants with moderate-to-severe predicted effects on gene function for each gene found to be significantly associated with propionate sensitivity. The haplotype notation ALT\_CT:x Hap:y refers to the alternate genotype count (ALT\_CT) and the unique haplotype (Hap) for each gene, (A) bgnt-1.7, (B) gly-17, (C) nhr-77, and (D) glct-3. If a variant with a high predicted effect on gene function was identified, we plotted it separately. The red diamonds represent the median phenotype value for each unique haplotype. The blue and pink diamonds represent the DL238 and BRC20067 strains, respectively.