# Article

# *C. elegans* as a model for inter-individual variation in metabolism

Check for updates

Bennett W. Fox[1], Olga Ponomarova[2,4], Yong-Uk Lee[2,4], Gaotian Zhang[3,4], Gabrielle E. Giese[2], Melissa Walker[2], Nicole M. Roberto[3], Huimin Na[2], Pedro R. Rodrigues[1], Brian J. Curtis[1], Aiden R. Kolodziej[1], Timothy A. Crombie[3], Stefan Zdraljevic[3], L. Safak Yilmaz[2], Erik C. Andersen[3✉], Frank C. Schroeder[1✉] & Albertha J. M. Walhout[2✉]

Individuals can exhibit differences in metabolism that are caused by the interplay of genetic background, nutritional input, microbiota and other environmental factors[1–4]. It is difficult to connect differences in metabolism to genomic variation and derive underlying molecular mechanisms in humans, owing to differences in diet and lifestyle, among others. Here we use the nematode *Caenorhabditis elegans* as a model to study inter-individual variation in metabolism. By comparing three wild strains and the commonly used N2 laboratory strain, we find differences in the abundances of both known metabolites and those that have not to our knowledge been previously described. The latter metabolites include conjugates between 3-hydroxypropionate (3HP) and several amino acids (3HP-AAs), which are much higher in abundance in one of the wild strains. 3HP is an intermediate in the propionate shunt pathway, which is activated when flux through the canonical, vitamin-$B_{12}$-dependent propionate breakdown pathway is perturbed[5]. We show that increased accumulation of 3HP-AAs is caused by genetic variation in HPHD-1, for which 3HP is a substrate. Our results suggest that the production of 3HP-AAs represents a 'shunt-within-a-shunt' pathway to accommodate a reduction-of-function allele in *hphd-1*. This study provides a step towards the development of metabolic network models that capture individual-specific differences of metabolism and more closely represent the diversity that is found in entire species.

Inborn errors in metabolism are characterized by the accumulation or depletion of metabolites as a result of mutations in genes that encode metabolic enzymes[6]. For example, people suffering from propionic acidaemia have mutations in either subunit of propionyl-CoA carboxylase and are diagnosed in newborn screening by the accumulation of 3HP (ref. [7]). Altered metabolism is also associated with a wide variety of complex diseases, including obesity, type 2 diabetes and cancer, which often have a genetic basis as well[8,9]. Connecting genomic and gene expression variation to differences in metabolism is a daunting task in humans because, in addition to obtaining an individual's genome sequence, it requires replicate measurements of the metabolome and transcriptome in age- and gender-matched individuals who, ideally, would consume exactly the same diet and experience identical environments.

Here we introduce the nematode *C. elegans* as a model to study individual differences in metabolism. A key advantage of using *C. elegans* is that it is a self-fertilizing hermaphrodite that can be fed single bacterial diets[10]. Therefore, isogenic *C. elegans* strains can be grown to large numbers of animals under standardized conditions, almost as easily as bacteria, which allows dual use of the same samples for both metabolomics and transcriptomics. These measurements can be integrated with genome-scale metabolic network models that cover annotated biochemical reactions in *C. elegans*, such as iCEL1314, to obtain functional predictions about the effects of genetic variation[11,12]. Numerous wild strains of *C. elegans* that represent genetically unique individuals have been collected and used to connect genomic variation to a plethora of traits[13]. Finally, CRISPR–Cas9 genome editing facilitates the validation of mechanistic predictions in intact animals[14].

## Variation in metabolite abundance

To connect differences in metabolite abundance to genomic variation, we selected four *C. elegans* strains on the basis of high-quality genome sequencing data, including the N2 reference strain and three wild strains: BRC20067, CB4856 and DL238 (Fig. 1a). The N2 (Bristol, England) and BRC20067 (Taiwan) strains are closely related despite the physical distance between the locations where they were found[15], whereas the CB4856 and DL238 strains, which were isolated from different islands of Hawaii in the USA, are divergent from the first two strains and from each other (Extended Data Fig. 1). In brief, synchronized animals at the first larval stage (L1) were grown in chemically defined K medium[16] with lysed and lyophilized *Escherichia coli* HB101 (referred to

---

[1]Boyce Thompson Institute and Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY, USA. [2]Department of Systems Biology, University of Massachusetts Chan Medical School, Worcester, MA, USA. [3]Department of Molecular Biosciences, Northwestern University, Evanston, IL, USA. [4]These authors contributed equally: Olga Ponomarova, Yong-Uk Lee, Gaotian Zhang. ✉e-mail: erik.andersen@northwestern.edu; schroeder@cornell.edu; marian.walhout@umassmed.edu

# Article
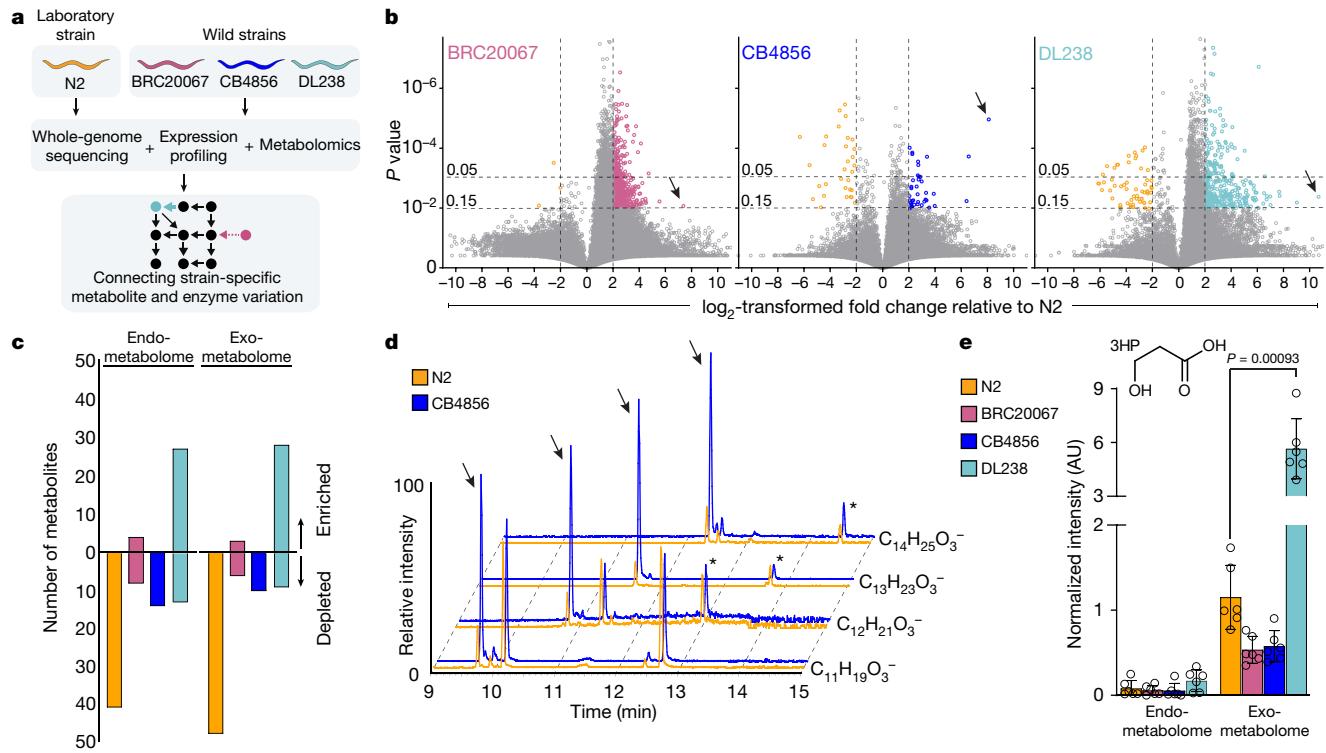


Fig. 1 | Inter-individual variation in metabolism. a, Study outline. b, Volcano plots for the subset of features detected by high-performance liquid chromatography–mass spectrometry (HPLC–MS; negative ion) in the exo-metabolomes for each wild strain versus the N2 reference strain (orange). See Methods for more details. Arrows indicate iglu#93, a metabolite that is observed in the three wild strains but not in N2 (see also Extended Data Fig. 2). Dashed lines represent Benjamini–Hochberg-adjusted significance thresholds at false discovery rates of 15% and 5%. c, Number of manually curated strain-specific metabolites, conservatively defined as any metabolite more than fivefold enriched or depleted in a given strain relative to the average of the other three strains at a significance threshold of $P < 0.01$ (unpaired, two-sided $t$-test). d, HPLC–MS (negative ion) extracted ion chromatograms (EICs) representing a homologous series of singly unsaturated hydroxy-fatty acids that are enriched in the wild strains (marked by arrows) relative to N2. A later-eluting isobaric series does not significantly vary between the four strains (marked by asterisks; see also Extended Data Fig. 3). e, Quantification of 3HP in the endo- and exo-metabolome extracts of the four strains. AU, arbitrary units. Data represent six biologically independent experiments and error bars indicate mean ± s.d. $P$ values were calculated by unpaired, two-sided $t$-test with Welch correction.

as *E. coli* powder) as a diet to avoid any potentially confounding effects of active bacterial metabolism (Extended Data Fig. 1). We collected *C. elegans* individuals as young adults and used 90% and 10% of animals for metabolomic and gene expression analysis, respectively (Extended Data Fig. 1). For the metabolomic analyses, we separately extracted animals and supernatants for analysis of the endo- and exo-metabolomes; that is, cellular and secreted metabolites, respectively. To maximize the coverage of metabolites, we analysed metabolome samples with multiple liquid chromatography– and gas chromatography–mass spectrometry (LC–MS and GC–MS, respectively) approaches (Extended Data Fig. 1).

Detected iCEL1314 metabolites include amino acids, a variety of fatty acids and other lipids, as well as carbohydrates and nucleosides (Supplementary Table 1). iCEL1314 metabolites that could not be measured include reactive molecules, such as acyl-CoA derivatives and nucleoside triphosphates, as well as biomass components such as bulk carbohydrates and triacyl glycerides. Most features detected by LC–MS (more than 500,000) could not be linked to known compounds using existing databases and are likely to represent more than 20,000 metabolites (Supplementary Tables 2–8).

To characterize metabolome differences among the four strains, we used comparative metabolomics with the XCMS and Metaboseek analysis platforms[17,18]. Pairwise comparisons with the N2 reference strain revealed several hundred metabolites that had significantly different abundances in the three natural strains (Fig. 1b). In addition, approximately 200 metabolites (representing around 1% of the detected metabolites) were highly strain-specific, defined as greater than fivefold enrichment or depletion in one strain relative to the average of the other three strains (unadjusted $P < 0.01$) (Fig. 1c and Supplementary Table 9). For example, the three natural strains produced iglu#93, an indole glycoside that was absent from the metabolome of the N2 strain and especially abundant in DL238 (Fig. 1b; see also Extended Data Fig. 2). Similarly, several oxidized fatty acid derivatives were more abundant in the three wild strains than in N2 (Fig. 1d), although the levels of the corresponding free fatty acids were only slightly or not increased in the wild strains (Extended Data Fig. 3). Among detected iCEL1314 metabolites, the most significant strain-specific compound was 3HP, which was more than sevenfold enriched in the exo-metabolome of DL238 compared to the average of the other three strains (Fig. 1e and Supplementary Table 10). 3HP is produced by the propionate shunt, which is transcriptionally activated when flux through the canonical, vitamin-$B_{12}$-dependent propionate breakdown pathway is perturbed; for example, on bacterial diets that are low in vitamin $B_{12}$ (such as *E. coli*)[5]. The increased excretion of 3HP suggests that flux through the propionate shunt is impaired in the DL238 strain.

Most of the compounds that differ greatly in abundance between strains represent metabolites that have not to our knowledge been previously described. These were analysed by tandem mass spectrometry (MS/MS) molecular networking, in which the MS/MS spectra of different compounds are compared and ranked by similarity of fragmentation patterns[19,20]. The metabolites (represented by nodes) are connected by edges that indicate structural similarity, and therefore metabolites that are connected in the network may share common biogenic pathways. A set of structurally related compounds, the abundances of which
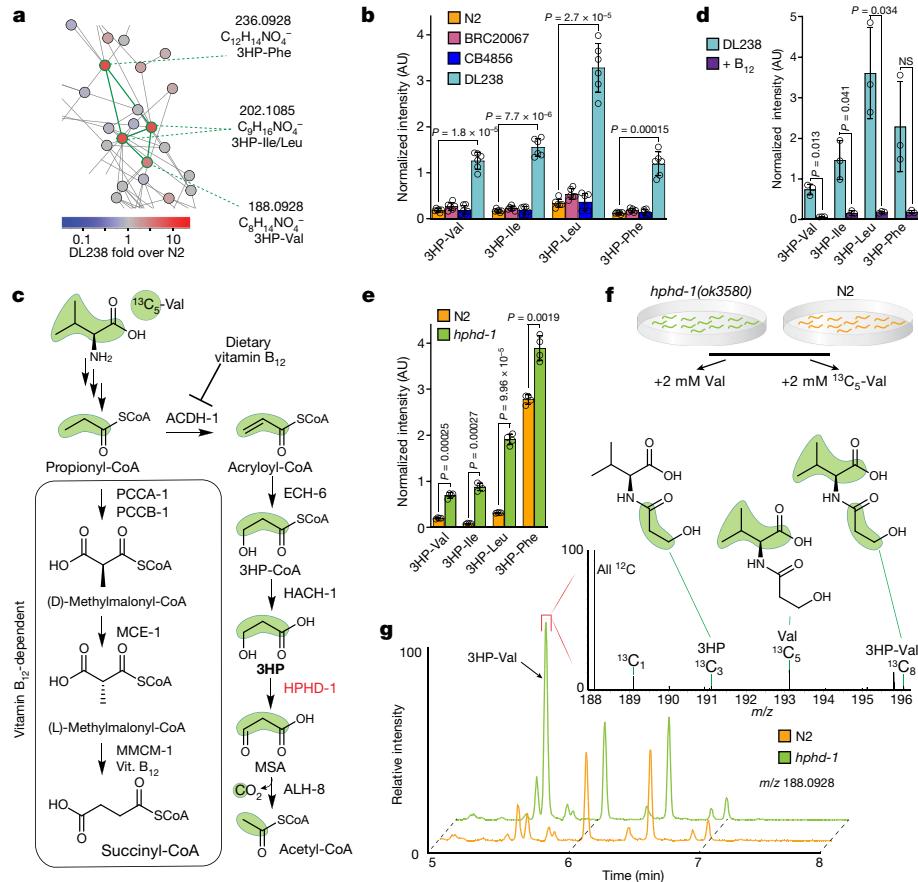
**Fig. 2 | A shunt-within-a-shunt for propionate degradation. a**, MS/MS networking revealed four structurally related metabolites (highlighted with green edges) that are enriched in DL238. MS/MS fragmentation and molecular formulae suggested that these compounds are 3HP-AA conjugates. **b**, Quantification of 3HP-AAs from exo-metabolome extracts of the four strains. Data represent six biologically independent experiments and error bars indicate mean ± s.d. **c**, Canonical (boxed) and shunt metabolic pathways for propionyl-CoA degradation. Under $B_{12}$-replete conditions, propionate shunt genes are transcriptionally repressed, and propionyl-CoA is metabolized to succinyl-CoA. Under $B_{12}$-deficient conditions, propionyl-CoA is degraded by the propionate shunt pathway. 3HP is an intermediate in the propionate shunt and accumulates in *hphd-1(ok3580)* mutant animals. **d**, Quantification of 3HP-AAs from exo-metabolome extracts of animals of the DL238 strain that were fed the standard *E. coli* OP50 diet without (light blue) or with (purple)

supplementation of 64 nM vitamin $B_{12}$. Data represent three biologically independent experiments and error bars indicate mean ± s.d. **e**, Quantification of 3HP-AAs from exo-metabolome extracts of N2 and *hphd-1(ok3580)* animals. Data represent four biologically independent experiments and error bars indicate mean ± s.d. **f**, $^{13}C_5$-Val isotope-tracing experiment. N2 and *hphd-1(ok3580)* animals were supplemented with Val or $^{13}C_5$-Val and analysed for isotopic enrichment. **g**, EICs (negative ion) for *m/z* 188.0928 in N2 and *hphd-1(ok3580)* mutants supplemented with $^{13}C_5$-Val. 3HP-Val (arrow) is more abundant in the *hphd-1(ok3580)* mutant. Inset: averaged mass spectra from the region of interest (denoted by red bracket) reveal $^{13}C_3$, $^{13}C_5$ and $^{13}C_8$ enrichment in 3HP-Val ($^{13}C_1$ natural abundance shown for scale). The distributions of $^{13}C$ enrichment in 3HP-Val resulting from $^{13}C_5$-Val supplement are highlighted with green shading. For **b**,**d**,**e**, *P* values were calculated by unpaired, two-sided *t*-test with Welch correction. NS, not significant.

were five- to eightfold increased in DL238 relative to the N2 strain and the other wild strains, were among the most abundant significantly differential metabolites detected (Fig. 2a,b; full network in Extended Data Fig. 4). These metabolites were 10- to 100-fold more abundant in the exo- relative to the endo-metabolome extracts, which indicates that they are primarily excreted (Extended Data Fig. 5). Analysis of their molecular formulae and MS/MS fragmentation patterns suggested that these compounds consist of 2-hydroxypropionate (lactate) or 3HP conjugated to different amino acids (Extended Data Fig. 5). Given that the abundance of 3HP was strongly increased in DL238 (Fig. 1e), we hypothesized that these metabolites represent 3HP-amino acid (3HP-AA) conjugates, analogous to the formation of *N*-lactoyl amino acids in response to increased concentrations of lactic acid in humans[21]. Given that propionate metabolism is largely conserved from *C. elegans* to humans[5], the hypothesis that 3HP is metabolized differently in DL238 relative to the other three strains seemed particularly noteworthy. Therefore, we selected the putative 3HP-derived metabolites for detailed follow-up studies.

## A shunt-within-a-shunt

We reasoned that production of the putative 3HP-AA conjugates should be abolished when propionate is degraded through the canonical, vitamin-$B_{12}$-dependent pathway (Fig. 2c). Consistent with this reasoning, supplementation with vitamin $B_{12}$ strongly reduced the production of the putative 3HP-AA conjugates in DL238 animals (Fig. 2d). 3HP is oxidized to malonic semialdehyde by HPHD-1 (Fig. 2c), a conserved hydroxyacid-oxoacid transhydrogenase[5], and disruption of *hphd-1* in the N2 background causes markedly increased levels of 3HP[5]. Correspondingly, we found that the abundance of the putative 3HP-AA conjugates was also greatly increased in *hphd-1(ok3580)* mutant animals (Fig. 2e). To confirm the conjugation between 3HP and amino acids, we performed isotope-tracing experiments with $^{13}C_5$-Val in *hphd-1(ok3580)* mutants and the N2 strain (Fig. 2f). Valine catabolism produces propionyl-CoA, which feeds into the propionate shunt[22,23]. Supplementation with $^{13}C_5$-Val resulted in $^{13}C_3$-, $^{13}C_5$- and a small amount of $^{13}C_8$-labelling in 3HP-valine in the *hphd-1(ok3580)* mutant (Fig. 2g),
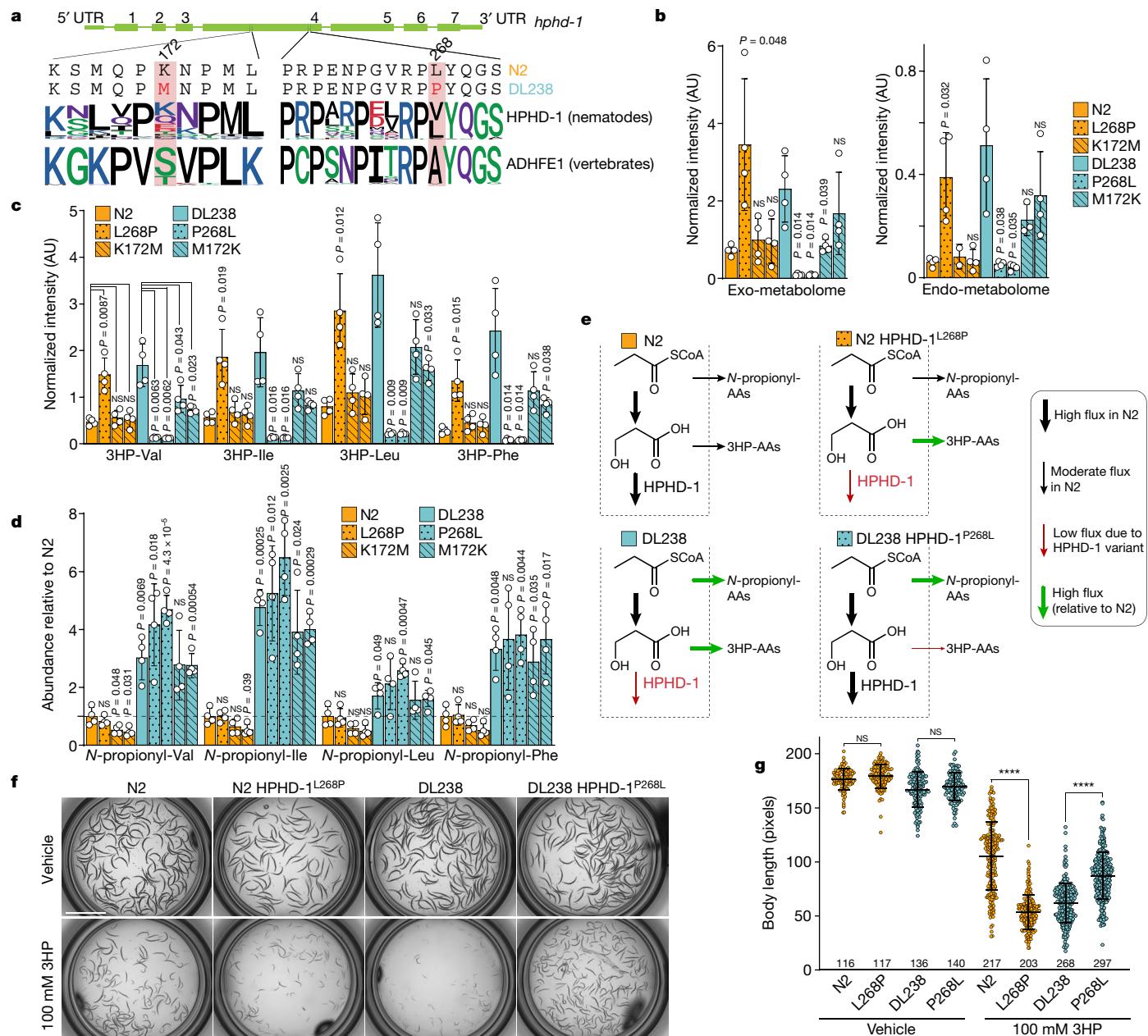
**Fig. 3 | Genomic variation in *hphd-1* causes the accumulation of 3HP and 3HP-AA in DL238. a**, DL238 contains two SNVs in *hphd-1*, which result in the amino acid substitutions K172M and L268P. Sequence logos of amino acid conservation proximal to the missense mutations are shown for nematode HPHD-1 and its vertebrate orthologue, ADHFE1. Logo colours: acidic, red; basic, blue; hydrophobic, black; neutral, purple; and polar, green. UTR, untranslated region. **b**, Quantification of 3HP from exo- and endo-metabolome extracts of N2, DL238 and CRISPR–Cas9-edited HPHD-1 strains. Data represent four biologically independent experiments and error bars indicate mean ± s.d. **c**, Quantification of 3HP-AAs from exo-metabolome extracts of the indicated strains. Data represent four biologically independent experiments and error bars indicate mean ± s.d. **d**, Relative abundance of *N*-propionyl-AAs in endo-metabolomes of the indicated strains. Data represent four biologically independent experiments and error bars indicate mean ± s.d. **e**, Conceptual

model of flux through the propionate shunt in N2, DL238 and CRISPR–Cas9-edited HPHD-1 strains. Thickness of arrows represents relative flux through enzymatic reactions, with green and red arrows representing higher and lower flux relative to N2, respectively. **f,g**, Representative images (**f**) and quantification of the body length (**g**) of N2, DL238 and CRISPR–Cas9-edited HPHD-1 strains after 72 h on normal medium (vehicle) or on medium containing 100 mM ethyl-3HP (see Methods). Scale bar, 2 mm. Data represent two biologically independent experiments of 50–150 animals per experiment; the total number of animals measured is indicated above the *x*-axis. Horizontal line indicates mean ± s.d. \*\*\*\**P* < .0001, as calculated by Welch's ANOVA with post-hoc comparison using Games-Howell's multiple comparisons test. For **b,c,d**, *P* values were calculated by unpaired, two-sided *t*-test with Welch correction, in which edited strains were compared against the corresponding parental strains (**b,c**) or against the N2 strain (**d**).

consistent with the generation of propionyl-CoA from valine catabolism (Fig. 2c). In line with this, $^{13}C_3$ enrichment was also observed for the other putative 3HP-AA conjugates (Extended Data Fig. 5).

Parallel $^{13}C_6$-Leu tracing experiments showed $^{13}C_6$-labelling of the 3HP-Leu conjugate, as well as two additional isobaric metabolites that

were not enriched in the DL238 strain relative to the other strains and that are likely to be diastereomers of *N*-lactoyl-Leu (Extended Data Fig. 6). These $^{13}C_6$-labelled metabolites eluted as the later peak of an isobaric pair of features with identical MS/MS spectra, which suggests that the earlier eluting peaks represented the corresponding
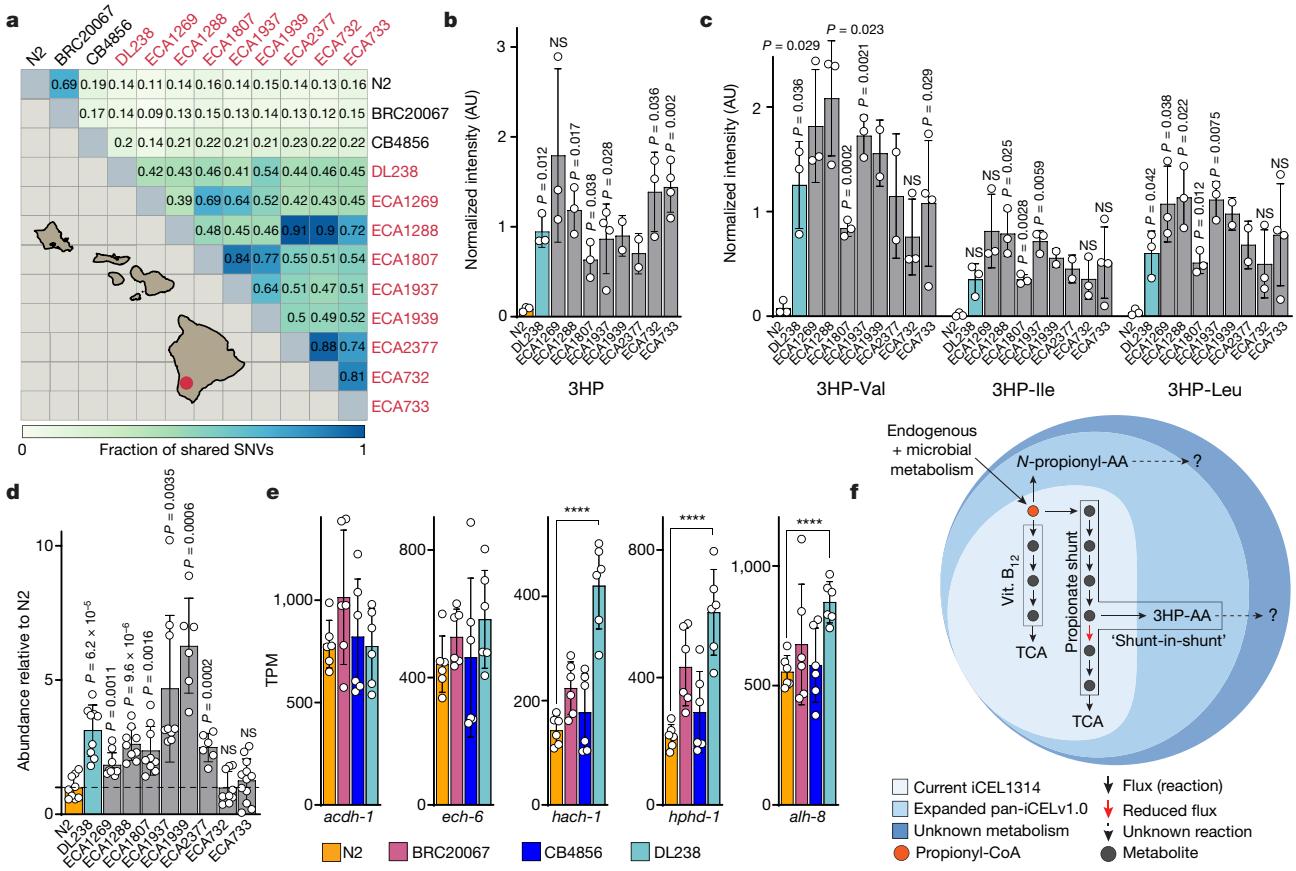
**Fig. 4 | Natural variation in *hphd-1* and propionate shunt gene expression.** **a**, Similarity matrix depicting the genetic relatedness of *C. elegans* strains in this study. Pairwise similarities were calculated on the basis of 2,604,844 bi-allelic SNVs; see Methods for details. All strains that contain the HPHD-1 variant were isolated near Manuka State Wayside Park, Hawaii, USA (red dot). **b**,**c**, Quantification of 3HP (**b**) and 3HP-Val, 3HP-Ile and 3HP-Leu (**c**) from endo-metabolome extracts of N2, DL238 and eight additional wild strains that contain the DL238 HPHD-1 variant. Error bars indicate mean ± s.d. **d**, Relative abundance of *N*-propionyl-Val, *N*-propionyl-Ile and *N*-propionyl-Leu in endo-metabolome extracts; error bars indicate mean ± s.d. **e**, Propionate shunt gene expression (TPM, transcripts per million). Data represent six biologically independent experiments and error bars indicate mean ± s.d. ****$P_{adj}$ < 0.0001, as calculated using the Wald test and corrected for multiple comparisons using the Benjamini–Hochberg method. **f**, The iCEL1314 metabolic network model is based on enzymes and metabolites identified in the N2 strain. Expansion of the model to accommodate strain-specific differences will better represent the species (pan-iCELv1.0). For **b**,**c**,**d**, *P* values were calculated by unpaired, two-sided *t*-test with Welch correction, in which all strains were compared to the N2 strain. The exact *n* value for each strain is presented in Supplementary Table 12.

isoleucine conjugates (Extended Data Fig. 6). Isotope-tracing experiments also revealed a series of predominantly excreted *N*-propionyl-AA conjugates that were two- to fivefold enriched in DL238 relative to N2 (Extended Data Fig. 7). Notably, *N*-propionyl-AAs were neither enriched in *hphd-1(ok3580)* mutants nor suppressed by supplementation with vitamin B_{12} (Extended Data Fig. 7).

Finally, we synthesized authentic samples of representative members of the identified series of amino acid conjugates, including 3HP-Leu, 3HP-Phe and *N*-propionyl-Leu, which showed chromatographic retention times and MS/MS spectra that matched the natural compounds (Extended Data Fig. 8). Collectively, our analyses revealed the accumulation of 3HP as well as 3HP- and *N*-propionyl-AA conjugates in the DL238 strain. These observations indicate that DL238 rewires propionate metabolism by conjugating both 3HP and propionate to amino acids.

## Coding variation in *hphd-1*

As disruption of *hphd-1* in the N2 strain results in the accumulation of both 3HP and 3HP-AAs, we asked whether the high abundance of these metabolites in the DL238 strain may be due to variation in this gene. Compared to N2, DL238 has two variants in HPHD-1: K172M and L268P (Fig. 3a). The K172M variant is found in a region of the enzyme

that is not highly conserved, but the L268P variant is embedded in a highly conserved region and abuts a Pro-rich motif, PXPXXPXXRP(L/P) YQGS (ref. [24]). The L to P variant probably causes a conformationally restricted Pro-Pro dipeptide in the DL238 version of HPHD-1 (refs. [25,26]) (Fig. 3a). To test whether the DL238 HPHD-1 variants contribute to the observed differences in abundance of 3HP and 3HP-AA conjugates, we used CRISPR–Cas9 genome editing to individually introduce the N2 *hphd-1* variants into the DL238 background and vice versa.

Targeted analysis of the metabolomes of N2, DL238 and two independent strains for each edit (except for N2^{L268P}) revealed that 3HP and 3HP-AAs were sharply reduced in both the exo- and the endo-metabolome of DL238^{P268L} animals relative to DL238 (Fig. 3b,c and Extended Data Fig. 9). Furthermore, 3HP and 3HP-AAs were strongly increased in N2^{L268P} animals, similar to the levels observed in DL238 (Fig. 3b,c and Extended Data Fig. 9). 3HP and 3HP-AAs in N2^{K172M} strains were unchanged relative to N2 but were approximately twofold reduced in DL238^{M172K} as compared to DL238, suggesting that this variant may contribute to the observed phenotype in DL238 but is latent in the N2 genetic context (Fig. 3b,c and Extended Data Fig. 9). These results indicate that the HPHD-1 L268P variant in DL238 is a reduction-of-function allele that is primarily responsible for the increased abundance of 3HP and 3HP-AAs in this strain. By contrast, the abundances of *N*-propionyl-AAs in DL238 and edited DL238 animals

# Article

were above the levels that were observed in N2 or edited N2 animals (Fig. 3d and Extended Data Fig. 10). The finding that propionate is directed to *N*-propionyl-AAs in DL238 animals independent of the HPHD-1 edits indicates that propionate metabolism in DL238 exhibits additional differences compared to N2 (Fig. 3e). In fact, diversion of propionate through the increased formation of *N*-propionyl-AAs could explain how 3HP and 3HP-AAs are reduced to even lower levels in the DL238[P268L] exo-metabolome than in N2 animals (Fig. 3b,c and Extended Data Fig. 9).

The conjugation of 3HP to amino acids and the excretion of both 3HP and 3HP-AAs suggested that 3HP may be toxic at high levels. Indeed, we found that exposure to exogenous 3HP strongly delayed the development of DL238 animals, whereas N2 animals were less affected (Fig. 3f,g). Notably, introducing the L268P variant into HPHD-1 in N2 animals delayed development to an extent similar to that in DL238, whereas reverting this amino acid from P to L in DL238 mitigated the developmental delay caused by 3HP exposure (Fig. 3g). These results show that a single amino acid substitution underlies the physiological response to excess 3HP.

Finally, we asked whether other wild strains contain the L268P variant in HPHD-1 and whether these strains also exhibit increased levels of 3HP and 3HP-AAs. In addition to DL238, 9 strains contain both the K172M and L268P variants, representing around 2% of the 540 curated wild strains that are at present in the *C. elegans* Natural Diversity Resource (CeNDR) database[27]. All of these strains, including DL238, were collected from the Big Island of Hawaii, USA, near the Manuka State Wayside Park (Fig. 4a). An analysis of genetic similarity among the Manuka cohort revealed that these strains are more genetically related to DL238 than DL238 is to N2, but that they do not share as many common single-nucleotide variants (SNVs) as N2 and BRC20067 (Fig. 4a). Despite their geographical proximity and shared *hphd-1* alleles, these strains contain many independent SNVs, including a missense variant in ECH-6 (A12S) that is unique to DL238. There are no other known missense variants in any genes in the propionate shunt or canonical propionate metabolic pathway in the Manuka cohort (Supplementary Table 11).

We grew additional cultures of eight wild strains containing the HPHD-1 variant in parallel with N2 and DL238, and found that both 3HP and 3HP-AAs were enriched in the exo- and endo-metabolomes of all tested natural strains that contain the DL238 HPHD-1 variant (Fig. 4b,c and Extended Data Fig. 11). These findings reinforce the notion that DL238 and other natural strains that have the HPHD-1 variant are adapted to low flux through the reaction catalysed by HPHD-1 by increasing the biosynthesis of 3HP-AAs. In addition, *N*-propionyl-AAs were enriched in six of the eight tested wild strains relative to N2 (Fig. 4d and Extended Data Fig. 12), thus supporting a model in which amino acid conjugation serves to divert propionate, and illustrating different mechanisms of propionate metabolism in wild strains.

## Towards a pan-species metabolic model

Using the exact same samples that were used for the initial metabolomic analyses, we found that three propionate shunt genes—*hach-1*, *hphd-1* and *alh-8*—were significantly upregulated in DL238 (Fig. 4e). This result suggests that this strain responds to reduced flux in the reaction catalysed by HPHD-1 in two ways: first, it converts 3HP into 3HP-AAs, and second, it increases the expression of *hphd-1* and the two adjacent enzymes (*hach-1* and *alh-8*) in the pathway. The latter has important implications for the interpretation of gene expression data from different strains. Increased expression of *hach-1*, *hphd-1* and *alh-8* would normally indicate a higher degradation flux capacity for 3HP; for example, when different tissues are compared within the same strain[12]. However, when comparing different strains, higher expression levels may actually reflect compensation for reduction-of-function variants, thus challenging conventional assumptions about the relationship between flux modelling and gene expression[28].

The iCEL1314 model is based on the N2 reference strain[11,12], and the work presented here affirms that large numbers of metabolites can be discovered in wild strains[29–31]. We postulate that 3HP- and *N*-propionyl-AAs result from conjugation between 3HP or propionate (or their corresponding coenzyme A (CoA) derivatives) and the relevant amino acids, similar to the conjugation between lactate and amino acids[21]. We added reactions to the iCEL1314 model that combine 3HP or propionate (or their CoA derivatives) with amino acids to form conjugates, including transport and exchange reactions (Fig. 4f). These additions represent a step towards a pan-iCEL model that captures metabolism of the entire species; we expect that the current pan-iCELv1.0 will grow as metabolomes for more strains become available and metabolites are identified and characterized.

## Discussion

We introduce the nematode *C. elegans* as a tractable model to connect inter-individual variation in metabolism to genomic variation. Our analysis—despite initially being limited to four 'individuals'—revealed a large number of compounds that were specifically enriched or depleted in one strain relative to the others. By combining comparative metabolomics with genomics, we connected the increased production of a new class of compounds to a single amino acid change in a central metabolism enzyme, HPHD-1. Moreover, the HPHD-1 L268P variant in the N2 strain confers sensitivity to exogenous 3HP, whereas the converse P268L edit confers increased resistance in the DL238 strain. These findings expand our understanding of the metabolism of propionate, a ubiquitous short chain fatty acid that is abundantly produced in both eukaryotes and bacteria, including species that are found in the human gut[32]. Because Val and Ile produce propionyl-CoA when metabolized, production of 3HP-Val and 3HP-Ile may represent a mechanism to simultaneously prevent the accumulation of shunt intermediates and reduce the catabolic production of propionate. The biosynthesis of *N*-propionyl-AAs, although not linked to variation in *hphd-1*, likewise reduces flux through the propionate shunt. It seems plausible that the latter contributes to the detoxification of excess propionate, in addition to genetic variation in *glct-3* (ref. [33]).

3HP-AAs are derived from 3HP (or 3HP-CoA), which accumulates under conditions of low flux through the vitamin-B$_{12}$-dependent propionate degradation pathway. Given that vitamin B$_{12}$ is produced by only a small subset of bacteria and archaea[34], our findings suggest that 3HP-AA production may represent a metabolic adaptation to changing nutrient conditions. Notably, a substantial portion of abnormal metabolic profiles observed in newborn screening can be attributed to maternal vitamin B$_{12}$ deficiency, rather than inborn genetic mutations[35]. Further studies with different diets and an increased cohort of *C. elegans* strains will help to contextualize strain-specific metabolites, expand the pan-iCEL metabolic network and develop individualized metabolic network models.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-04951-3.

1. Montgomery, M. K. et al. Mouse strain-dependent variation in obesity and glucose homeostasis in response to high-fat feeding. *Diabetologia* **56**, 1129–1139 (2013).
2. Braun, J. M. Early-life exposure to EDCs: role in childhood obesity and neurodevelopment. *Nat. Rev. Endocrinol.* **13**, 161–173 (2017).
3. Visconti, A. et al. Interplay between the human gut microbiome and host metabolism. *Nat. Commun.* **10**, 4505 (2019).
4. Loos, R. J. F. & Yeo, G. S. H. The genetics of obesity: from discovery to biology. *Nat. Rev. Genet.* **23**, 120–133 (2022).

5. Watson, E. et al. Metabolic network rewiring of propionate flux compensates vitamin B12 deficiency in *C. elegans*. *eLife* **5**, e17670 (2016).

6. Wilcken, B., Wiley, V., Hammond, J. & Carpenter, K. Screening newborns for inborn errors of metabolism by tandem mass spectrometry. *N. Engl. J. Med.* **348**, 2304–2312 (2003).

7. Deodato, F., Boenzi, S., Santorelli, F. M. & Dionisi-Vici, C. Methylmalonic and propionic aciduria. *Am. J. Med. Genet. C* **142C**, 104–112 (2006).

8. Pillon, N. J., Loos, R. J. F., Marshall, S. M. & Zierath, J. R. Metabolic consequences of obesity and type 2 diabetes: balancing genes and environment for personalized care. *Cell* **184**, 1530–1544 (2021).

9. Min, H. Y. & Lee, H. Y. Oncogene-driven metabolic alterations in cancer. *Biomol. Ther.* **26**, 45–56 (2018).

10. Zhang, J., Holdorf, A. D. & Walhout, A. J. *C. elegans* and its bacterial diet as a model for systems-level understanding of host-microbiota interactions. *Curr. Opin. Biotechnol.* **46**, 74–80 (2017).

11. Yilmaz, L. S. & Walhout, A. J. A *Caenorhabditis elegans* genome-scale metabolic network model. *Cell Syst.* **2**, 297–311 (2016).

12. Yilmaz, L. S. et al. Modeling tissue-relevant *Caenorhabditis elegans* metabolism at network, pathway, reaction, and metabolite levels. *Mol. Syst. Biol.* **16**, e9649 (2020).

13. Evans, K. S., van Wijk, M. H., McGrath, P. T., Andersen, E. C. & Sterken, M. G. From QTL to gene: *C. elegans* facilitates discoveries of the genetic mechanisms underlying natural variation. *Trends Genet.* **37**, 933–947 (2021).

14. Nance, J. & Frokjaer-Jensen, C. The *Caenorhabditis elegans* transgenic toolbox. *Genetics* **212**, 959–990 (2019).

15. Lee, D. et al. Balancing selection maintains hyper-divergent haplotypes in *Caenorhabditis elegans*. *Nat. Ecol. Evol.* **5**, 794–807 (2021).

16. Boyd, W. A., McBride, S. J., Rice, J. R., Snyder, D. W. & Freedman, J. H. A high-throughput method for assessing chemical toxicity using a *Caenorhabditis elegans* reproduction assay. *Toxicol. Appl. Pharmacol.* **245**, 153–159 (2010).

17. Domingo-Almenara, X. & Siuzdak, G. Metabolomics data processing using XCMS. *Methods Mol. Biol.* **2104**, 11–24 (2020).

18. Helf, M. J., Fox, B. W., Artyukhin, A. B., Zhang, Y. K. & Schroeder, F. C. Comparative metabolomics with Metaboseek reveals functions of a conserved fat metabolism pathway in *C. elegans*. *Nat. Commun.* **13**, 782 (2022).

19. Nguyen, D. D. et al. MS/MS networking guided analysis of molecule and gene cluster families. *Proc. Natl Acad. Sci. USA* **110**, E2611–E2620 (2013).

20. Nothias, L. F. et al. Feature-based molecular networking in the GNPS analysis environment. *Nat. Methods* **17**, 905–908 (2020).

21. Jansen, R. S. et al. *N*-lactoyl-amino acids are ubiquitous metabolites that originate from CNDP2-mediated reverse proteolysis of lactate and amino acids. *Proc. Natl Acad. Sci. USA* **112**, 6601–6606 (2015).

22. Ando, T., Rasmussen, K., Nyhan, W. L. & Hull, D. 3-hydroxypropionate: significance of -oxidation of propionate in patients with propionic acidemia and methylmalonic acidemia. *Proc. Natl Acad. Sci. USA* **69**, 2807–2811 (1972).

23. Walker, M. D. et al. WormPaths: *Caenorhabditis elegans* metabolic pathway annotation and visualization. *Genetics* **219**, iyab089 (2021).

24. Kay, B. K., Williamson, M. P. & Sudol, M. The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J.* **14**, 231–241 (2000).

25. Hinderaker, M. P. & Raines, R. T. An electronic effect on protein structure. *Protein Sci.* **12**, 1188–1194 (2003).

26. Vanhoof, G., Goossens, F., De Meester, I., Hendriks, D. & Scharpe, S. Proline motifs in peptides and their biological processing. *FASEB J.* **9**, 736–744 (1995).

27. Cook, D. E., Zdraljevic, S., Roberts, J. P. & Andersen, E. C. CeNDR, the *Caenorhabditis elegans* natural diversity resource. *Nucleic Acids Res.* **45**, D650–D657 (2017).

28. Machado, D. & Herrgard, M. Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. *PLoS Comput. Biol.* **10**, e1003580 (2014).

29. Falcke, J. M. et al. Linking genomic and metabolomic natural variation uncovers nematode pheromone biosynthesis. *Cell Chem. Biol.* **25**, 787–796 (2018).

30. Keurentjes, J. J. et al. The genetics of plant metabolism. *Nat. Genet.* **38**, 842–849 (2006).

31. Wen, W. et al. Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nat. Commun.* **5**, 3438 (2014).

32. Reichardt, N. et al. Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. *ISME J.* **8**, 1323–1335 (2014).

33. Na, H., Zdraljevic, S., Tanny, R. E., Walhout, A. J. M. & Andersen, E. C. Natural variation in a glucuronosyltransferase modulates propionate sensitivity in a *C. elegans* propionic acidemia model. *PLoS Genet.* **16**, e1008984 (2020).

34. Fang, H., Kang, J. & Zhang, D. Microbial production of vitamin B$_{12}$: a review and future perspectives. *Microb. Cell. Fact.* **16**, 15 (2017).

35. Scolamiero, E. et al. Targeted metabolomics in the expanded newborn screening for inborn errors of metabolism. *Mol. Biosyst.* **11**, 1525–1535 (2015).

36. Andersen, E. C., Bloom, J. S., Gerke, J. P. & Kruglyak, L. A variant in the neuropeptide receptor *npr-1* is a major determinant of *Caenorhabditis elegans* growth and physiology. *PLoS Genet.* **10**, e1004156 (2014).

37. Hahnel, S. R. et al. Extreme allelic heterogeneity at a *Caenorhabditis elegans* beta-tubulin locus explains natural resistance to benzimidazoles. *PLoS Pathog.* **14**, e1007226 (2018).

38. Zheng, X. et al. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012).

39. Boyd, W. A., Smith, M. V. & Freedman, J. H. *Caenorhabditis elegans* as a model in developmental toxicology. *Methods Mol. Biol.* **889**, 15–24 (2012).

40. Lee, D. et al. Selection and gene flow shape niche-associated variation in pheromone response. *Nat. Ecol. Evol.* **3**, 1455–1463 (2019).

41. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).

42. Giese, G. E. et al. *Caenorhabditis elegans* methionine/S-adenosylmethionine cycle activity is sensed and adjusted by a nuclear hormone receptor. *eLife* **9**, e60259 (2020).

43. Zhang, J. et al. A delicate balance between bacterial iron and reactive oxygen species supports optimal *C. elegans* development. *Cell Host Microbe* **26**, 400–411 (2019).

44. Noble, L. M. et al. Selfing is the safest sex for *Caenorhabditis tropicalis*. *eLife* **10**, e62587 (2021).

45. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).

46. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).

47. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res.* **4**, 1521 (2015).

48. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

49. Pimentel, H., Bray, N. L., Puente, S., Melsted, P. & Pachter, L. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat. Methods* **14**, 687–690 (2017).

# Article

## Methods

### Strains of *C. elegans*

N2 (Bristol) and wild strains BRC20067, CB4856, DL238, ECA1269, ECA1288, ECA1807, ECA1937, ECA1939, ECA2377, ECA732 and ECA733 were maintained at 20 °C using *E. coli* HB101 and grown on modified nematode growth medium containing 1% agar and 0.7% agarose (NGMA) to prevent animals from burrowing[36]. Strains were obtained from the CeNDR and are available upon request. *hphd-1(ok3580)* was obtained from the *C. elegans* Gene Knockout Consortium and was backcrossed as described[5]. The CRISPR–Cas9 HPHD-1-edited strain ECA2545 *hphd-1(ean779)* N2[L268P] was generated by standard protocol[37]. The CRISPR–Cas9 HPHD-1-edited strains PHX2627 *hphd-1(syb2627)* N2[K172M], PHX2628 *hphd-1(syb2628)* N2[K172M], PHX2640 *hphd-1(syb2640)* DL238[M172K], PHX2666 *hphd-1(syb2666)* DL238[P268L], PHX2667 *hphd-1(syb2667)* DL238[P268L] and PHX2757 *hphd-1(syb2757)* DL238[M172K] were purchased from SunyBiotech.

### Genetic variants

Genotype data for the wild *C. elegans* strains along with the laboratory-adapted strain N2 were acquired from the variant call format (VCF) file (Release 20200815) from the CeNDR[27]. Pairwise dissimilarity analysis was performed among the *C. elegans* strains, using 2,604,844 bi-allelic SNVs from the VCF file and the R package SNPRelate (v1.16.0)[38].

The NCBI protein basic local alignment search tool (BLAST) was used to generate multiple sequence alignments for hydroxyacid-oxoacid transhydrogenase in 100 vertebrates (based on similarity to human ADHFE1) and in 56 available non-*C. elegans* nematodes (based on similarity to *C. elegans* HPHD-1) (Fig. 3a). Accession numbers for all sequences are provided in Supplementary Table 13. Sequence logos were created in R (v.4.1.1) using the ggseqlogo package (v.0.1).

### Growth of *C. elegans* strains

Synchronized animals at the first larval stage (L1) were prepared by two different methods: settling and bleaching. For bleaching, animals were synchronized at the L1 stage by starvation. L1 animals were transferred to twenty modified nematode growth medium (NGMA) plates[36] seeded with *E. coli* HB101. After incubation for 72 h at 20 °C, gravid adult animals were bleached using *C. elegans* bleaching solution (40 ml NaOCl (Fisher, SS290-1), 10 ml 10 N NaOH and 300 ml $H_2O$). For settling, starved L1 animals were cultured for 96 h at 20 °C until all food was depleted. Plates were rinsed with 2 ml M9 buffer (22 mM $KH_2PO_4$, 42 mM $Na_2HPO_4$, 86 mM NaCl, 1 mM $MgSO_4$ and 1 l $H_2O$) per plate and transferred to 50-ml conical tubes. Tubes were left for 10 min at room temperature, after which all non-L1 stages settled at the bottom of the tube. The top 13 ml, containing L1 animals, was transferred to a new 15-ml conical tube and settled again for 10 min at room temperature to pellet all non-L1 stages. After this second settling, L1 animals were collected by transferring 10 ml, washed twice with 12 mL M9 buffer and pelleted at 2,000 rpm in a tabletop centrifuge for 1 min.

To generate the samples for metabolomics and transcriptomics, approximately 100,000 L1 animals were placed in a 125-ml Erlenmeyer flask filled with 25 ml K medium with modified salt concentrations (51 mM NaCl, 32 mM KCl, 3 mM $CaCl_2$, 3 mM $MgSO_4$), 1.25 μl ml[−1] cholesterol, 50 μg ml[−1] kanamycin and 10 mg ml[−1] HB101 bacterial lysate (Pennsylvania State University Shared Fermentation Facility)[39,40]. Animals were cultured at 20 °C with 180 rpm orbital shaking in a shaking incubator (Multitron standard, INFORS HT) until they reached the young adult stage. Animals and culture medium were separated by 2,000 rpm centrifugation for 1 min. Images were taken of each sample. Approximately 20 ml of culture medium was immediately frozen in liquid nitrogen and stored at −80 °C for LC–MS and GC–MS analyses. Separated animals were washed three times with M9 buffer. After the third wash, animals were pelleted by centrifugation and aspiration of M9 buffer and divided into sets of around 90,000 animals for metabolomics and 10,000 animals for RNA-seq analysis.

Owing to bacterial powder supply issues, subsequent experiments were performed with live *E. coli* HB101, unless otherwise indicated. Live HB101 was independently extracted and analysed in parallel for each experiment (see Extended Data Figs. 9–12). For experiments with CRISPR-edited strains, approximately 50,000 animals were grown in a 50-ml Erlenmeyer flask filled with 10 ml modified K medium and live *E. coli* HB101 from a 100-ml overnight culture. For experiments with additional wild strains, approximately 20,000 animals were grown in a 50-ml Erlenmeyer flask filled with 10 ml modified K medium and live *E. coli* HB101 from a 100-ml overnight culture. For experiments with vitamin $B_{12}$ supplement, approximately 40,000 animals were grown in a 50-ml Erlenmeyer flask filled with 20 ml modified K medium and live *E. coli* OP50 or OP50 supplemented with 64 nM vitamin $B_{12}$ (as adenosyl cobalamin, Sigma-Aldrich, C0884). Culture medium and animals were isolated separately as described and samples were immediately frozen in liquid nitrogen and stored at −80 °C.

### Targeted quantification of metabolites using GC–MS

Animal endo- and exo-metabolome extracts were dried under vacuum. Derivatization of dried samples was performed first with 20 μl of 20 mg ml[−1] methoxyamine hydrochloride (Sigma-Aldrich) in pyridine at 37 °C for 90 min, followed by the addition of 50 μl *N*-methyl-*N*-(trimethylsilyl) trifluoroacetamide (Sigma-Aldrich) and incubation for 3 h at 37 °C. The derivatization reaction was completed by incubation for 5 h at room temperature. Measurements were performed on an Agilent 7890B single-quadrupole mass spectrometer coupled to an Agilent 5977B gas chromatograph (GC–MS) with an HP-5MS Ultra Inert capillary column (30 m × 0.25 mm × 0.25 μm). Helium was used as carrier gas at a flow rate of 1 ml per min (constant flow). The temperatures were set as follows: inlet at 230 °C, the transfer line at 280 °C, the MS source at 230 °C and quadrupole at 150 °C. One microlitre of sample was injected in a split mode with 5 ml per min split flow. Initial oven temperature was set to 80 °C and then increased to 310 °C at a rate of 5 °C per min. MS parameters were: 3 scans per s with 30–500 *m/z* range, electron impact ionization energy 70 eV. Each metabolite was identified on the basis of retention time as well as one quantifier and two qualifier ions that were manually selected using a reference compound. Peak integration and quantification of peak areas were done using MassHunter software (Agilent v.10.1); blank subtraction and normalization to total quantified metabolites were done using R software.

### Sample preparation for HPLC–MS

Animal bodies (endo-metabolome) and conditioned medium (exo-metabolome) were frozen and processed separately. For preparation of endo-metabolome extracts, synchronized adult *C. elegans* were lyophilized for 18–24 h using a VirTis BenchTop 4K Freeze Dryer. Dried pellets were transferred to 1.5-ml microfuge tubes and the dry pellet weight was recorded. Pellets were disrupted in a Spex 1600 MiniG tissue grinder after the addition of three stainless steel grinding balls to each sample. Microfuge tubes were placed in a Cryoblock (Model 1660) cooled in liquid nitrogen, and samples were disrupted at 1,100 rpm for 6 cycles of 30 s, with cooling in between. Pellets were transferred to 20-ml glass vials in 15 ml 100% ethanol and stirred overnight. Glass vials were centrifuged at 2,750*g* for 5 min in an Eppendorf 5702 Centrifuge using rotor F-35-30-17. The resulting supernatant was transferred to a clean 20-ml glass vial and concentrated to dryness in an SC250EXP Speedvac Concentrator coupled to an RVT5105 Refrigerated Vapor Trap (Thermo Fisher Scientific). The resulting powder was suspended in 100% ethanol and analysed directly by HPLC–MS, see below. For preparation of exo-metabolome extracts, frozen conditioned medium was lyophilized for around 48 h using a VirTis BenchTop 4K Freeze Dryer. Dried material was directly extracted in 15 ml 100% ethanol after transfer to 20-ml glass vials. Subsequent steps for concentration and suspension

were followed as described for preparation of the endo-metabolome samples. For CRISPR-edited samples, extractions were performed as described except that methanol was used instead of ethanol.

## HPLC–MS/MS analysis

Reversed-phase chromatography was performed using a Vanquish HPLC system controlled by Chromeleon Software (Thermo Fisher Scientific) and coupled to an Orbitrap Q-Exactive HF mass spectrometer controlled by Xcalibur software (Thermo Fisher Scientific), or by a Dionex Ultimate 3000 HPLC system coupled to an Oribtrap Q-Exactive mass spectrometer controlled by the same software. Extracts prepared as described above were separated on an Agilent Zorbax Eclipse XDB-C18 column (150 mm × 2.1 mm, particle size 1.8 μm), or on a Thermo Fisher Scientific Hypersil Gold column (150 mm × 2.1 mm, particle size 1.9 μm) maintained at 40 °C with a flow rate of 0.5 ml per min. Solvent A: 0.1% formic acid in water; solvent B: 0.1% formic acid in acetonitrile. A/B gradient started at 1% B for 3 min after injection and increased linearly to 98% B at 20 min, followed by 5 min at 98% B, then back to 1% B over 0.1 min and finally held at 1% B for an additional 2.9 min to re-equilibrate the column.

Reversed-phase post-column ion-pairing chromatography was performed using the same system as described; extracts were separated on an Agilent Zorbax Eclipse XDB-C18 column (150 mm × 2.1 mm, particle size 1.8 μm) maintained at 40 °C with a flow rate of 0.5 ml per min. Solvent A: 0.1% ammonium acetate in water; solvent B: acetonitrile. A/B gradient started at 5% B for 3 min after injection and increased linearly to 98% B at 20 min, followed by 5 min at 98% B, then back to 5% B over 0.1 min and finally held at 5% B for an additional 2.9 min to re-equilibrate the column. A second pump (Dionex 3000) controlling a solution of 800 mM ammonia in methanol was run at a constant flow rate of 0.015 ml per min for the duration of the method and mixed via micro-splitter valve (Idex #P-460S) with the eluate line from the column.

Normal-phase chromatography was performed using the same system as described; extracts were separated on an Agilent Zorbax RRHD HILIC Plus column (150 mm × 2.1 mm, particle size 1.8 μm) maintained at 40 °C with a flow rate of 0.5 ml per min. Solvent A: 0.1% formic acid in water; solvent B: 0.1% formic acid in acetonitrile. A/B gradient started at 90% B for 3 min after injection and decreased linearly to 50% B at 15 min, followed by a linear decrease to 10% B at 16 min, held at 10% B until 19 min, followed by a linear increase to 90% B at 20 min, and finally held at 90% B for an additional 5 min to re-equilibrate the column.

Mass spectrometer parameters: spray voltage (−3.0 kV, +3.5 kV), capillary temperature 380 °C, probe heater temperature 400 °C; sheath, auxiliary and sweep gas 60, 20 and 2 AU, respectively. S-Lens RF level: 50, resolution 240,000 at $m/z$ 200, AGC target 3E6. Each sample was analysed in negative and positive electrospray ionization modes with $m/z$ ranges 70–1,000 for reversed-phase, 120–800 for reversed-phase post-column ion pairing and 70–700 for normal phase. Parameters for MS/MS (dd-MS2): MS1 resolution: 60,000, AGC target: 1E6. MS/MS resolution: 30,000, AGC target: 2E5, maximum injection time: 50 ms, isolation window 1.0 $m/z$, stepped normalized collision energy (NCE) 10, 30; dynamic exclusion: 5 s, top 10 masses selected for MS/MS per scan.

LC–MS RAW data were converted to mzXML file format using MSConvert (v.3.0, ProteoWizard) and were analysed using Metaboseek software v.0.9.9.1 (see below) and normalized to the abundance of ascr#3 as an approximate measure of sample size for replicates from the same strain. To account for possible variation between strains, metabolites were normalized as a ratio to ascr#3, and the resulting quotient multiplied by the strain average of ascr#3 (performed independently for endo- and exo-metabolome), thereby removing the effect of variation between strains. Quantification was performed with Metaboseek software or via integration using Xcalibur QualBrowser v.4.1.31.9 (Thermo Fisher Scientific) using a 5 ppm window around the $m/z$ of interest. For volcano plots in Fig. 1b, exo-metabolome in negative ionization mode is depicted as a representative dataset. The list of features identified by peak picking in XCMS (including degenerate features such as adducts

and isotopes) was culled by retention time (180–1,400 s) and further filtered by using the 'Peak Shapes' functionality to calculate peak quality scores, applying a threshold of 0.96. The resulting feature list of 50,203 features was analysed by built-in ANOVA using Metaboseek software, then adjusted for significance by the Benjamini–Hochberg method using false discovery rates of 15% or 5%, as indicated[41]. Each natural strain was profiled against the N2 strain, and the $\log_2$-transformed fold change was plotted against the negative logarithm of the unadjusted $P$ value, as calculated by unpaired, two-sided $t$-test modelled as six independent experiments. Statistical analysis for metabolomics was performed with Metaboseek software (v.0.9.9.1) and with GraphPad Prism (v.9.3.0).

## Metabolomics network analysis

MS/MS networking was performed using Metaboseek MS analysis software (https://doi.org/10.5281/zenodo.3360087). Documentation and source code are available in the Metaboseek R package on GitHub: https://github.com/mjhelf/metaboseek. For additional information, installation instructions and a user guide, please visit https://metaboseek.com[18]. The XCMS package within Metaboseek was used with 'Metaboseek_default' settings, with the exception that the RT width parameter was adjusted to 3 s. MS/MS matching used the following parameters: ppm window: 3, RT window: 5 s, unique assignments on. From roughly 200,000 features detected in negative-ion ESI HPLC–MS, approximately 10,000 features were matched to MS/MS spectra; this was further culled to 4,563 features by restricting the retention time window from 180–1350 s and using the 'Peak Shapes' functionality to apply a peak quality threshold of 0.98. Networking was performed using the Compare MS/MS function in Metaboseek with the following parameters: $m/z$ tolerance: 0.002, ppm tolerance: 3, minimum number of peaks in common: 3, noise level: 2%, with parent mass matching turned on. The network was modified by removing edges below similarity score (cosine) = 0.6, and further simplified by restricting the maximum number of edges per node to the top eight ranked by similarity score.

On the basis of our comparative analysis of relative feature abundance in wild isolates and N2 (Supplementary Table 9), we identified subnetworks of potential interest. Detailed analysis of the MS/MS spectra (see 'Data availability') of these subnetworks revealed that they represent metabolites from diverse pathways, including a variety of fatty acid derivatives and other lipids, known and putative ascarosides and modular glucosides including the putative iglu#93 (Extended Data Fig. 2), as well as the 3HP-AAs. The latter were selected for detailed additional studies as described.

## Isotopic labelling experiments

Approximately 60,000 synchronized N2 (wild-type) or *hphd-1(ok3580)* L1 larvae were seeded in 125-ml Erlenmeyer flasks containing 20 ml S-Complete medium. Worms were fed with 3 mg ml⁻¹ freeze-dried OP50 powder (InVivoBiosystems, formerly NemaMetrix, OP-50-31772) and supplemented with either L-leucine (Sigma-Aldrich, L8000), $^{13}C_6$-L-leucine (Cambridge Isotope Laboratories, CLM-2262-H-PK), L-valine (Sigma-Aldrich, V0500) or $^{13}C_5$-L-valine (Cambridge Isotope Laboratories, CLM-2249-H-PK) at a final concentration of 2 mM. Worms were incubated at 20 °C with shaking at 180 rpm for approximately 70 h, at which time the population was a mixture of young and gravid adults, determined by microscopic inspection. Liquid cultures were centrifuged (500$g$, 22 °C, 1 min), and the resulting supernatant was snap-frozen. The worm pellet was washed three times with M9 before snap-freezing in liquid nitrogen. Samples were prepared for HPLC–MS analysis as described.

## 3HP toxicity assays

Animals were maintained on modified NGM plates containing soy peptone (to limit residual vitamin $B_{12}$ present in the peptone)[42] and fed *E. coli* OP50. We used the ethyl ester of 3HP (ethyl-3HP) to improve uptake (Combi-Blocks, ST-8598). Ethyl-3HP was diluted in water to achieve a final concentration of 400 mM followed by pH adjustment to around

6.5 with sodium hydroxide to yield an ethyl-3HP stock solution, which was further diluted into molten NGM to the desired concentration. Synchronized L1 larvae obtained by alkaline bleach egg prep (see above) were grown on control plates or plates containing 100 mM ethyl-3HP at 20 °C. After approximately 72 h, animals were washed off the plates and anaesthetized in a 1 mM solution of levamisole (TCI, T1215), then photographed using an Invitrogen EVOS FL imaging system. Body length was used as a proxy for developmental stage[43]; animals that were subjected to high concentrations of 3HP did eventually reach adulthood during continued surveillance. Body length was measured using a custom script (wormFinder_v2.m) written on MATLAB (MathWorks, v.R2020a)[43] and available at https://github.com/shiaway/wormFinder.

## RNA-seq and data processing

Total RNA extraction and RNA-seq library construction were performed for the 24 samples simultaneously as previously described[44]. All 24 RNA-seq libraries were pooled and sequenced on a single lane of an Illumina NovaSeq 6000 platform, yielding 150-bp paired-end (PE150) reads. On average, we obtained 34.3 million reads per sample, with a range of 25 to 45.5 million reads. Adapter sequences and low-quality reads in raw sequencing data were removed using fastp (v.0.20.0)[45]. Transcript expression of each sample was quantified using the SNV-substituted transcriptome of each strain and Kallisto (v.0.44.0)[46] with 100 bootstraps per sample. The R package tximport (v.1.10.1)[47] was used to summarize transcript-level abundances into gene-level abundances. We set a threshold of a minimal total count of 120 for each gene across the 12 samples in each pairwise comparison. We used the DESeq() function in the R package DESeq2 (v.1.22.2)[48] to normalize expression counts and perform pairwise differential expression analysis among the four strains with the default Wald test and the Benjamini–Hochberg multiple testing correction. Thresholds of differentially expressed genes were set as a fold change of greater than 1.5-fold for biological significance and a $q$-value smaller than 0.05 for statistical significance. We used the R package sleuth (v.0.30.0)[49] to normalize TPM across all samples.

## Chemical synthesis

Nuclear magnetic resonance (NMR) spectra were recorded on 400 MHz Bruker AVIII HD, 500 MHz Bruker AVIII HD and 600 MHz Varian INOVA spectrometers at Cornell University's NMR facility. For reaction schemes, see Supplementary Fig. 1.

### N-(3-O-TBDPS-propionyl)-L-leucine tert-butyl ester (2)

To a solution of 3-O-tert-butyldiphenylsilylpropionic acid (3-O-TBDPS-propionic acid, 1, 60 mg, 0.191 mmol, 1.0 equiv.) in 3 ml of dichloromethane (DCM) was added L-leucine tert-butyl ester hydrochloride (43 mg, 0.191 mmol, 1.0 equiv.), 1-ethyl-3-(3′-dimethylaminopropyl) carbodiimide (EDC) hydrochloride (88 mg, 0.461 mmol, 2.4 equiv.) and 4-dimethylaminopyridine (DMAP, 28 mg, 0.229 mmol, 1.2 equiv.). The resulting solution was stirred at room temperature overnight and concentrated in vacuo. Flash column chromatography on silica using a gradient of 0–40% methanol (MeOH) in DCM afforded 2 (85 mg, 89%) as a light yellow, viscous oil. See Supplementary Fig. 1a for structures.

**¹H NMR (400 MHz, chloroform-d).** δ 7.74–7.60 (m, 4H), 7.47–7.34 (m, 6H), 6.75 (d, $J$ = 8.5 Hz, 1H), 4.57 (td, $J$ = 8.7, 5.4 Hz, 1H), 3.93 (t, $J$ = 5.6 Hz, 2H), 2.45 (t, $J$ = 5.6 Hz, 2H), 1.76–1.56 (m, 2H), 1.54–1.47 (m, 1H), 1.47 (s, 9H), 1.06 (s, 9H), 0.94 (d, $J$ = 7.2 Hz, 3H), 0.92 (d, $J$ = 7.5 Hz, 3H).

### N-(3-hydroxypropionyl)-L-leucine tert-butyl ester (3)

To a solution of 2 (78 mg, 0.157 mmol, 1.0 equiv.) in 2.5 ml of tetrahydrofuran (THF) was added tetra-n-butylammonium fluoride (TBAF, 1M in THF, 0.21 ml, 0.21 mmol, 1.34 equiv.). The resulting solution was stirred at room temperature for 2 h and was then concentrated in vacuo. Flash column chromatography on silica using a gradient of 0–100% MeOH in DCM afforded 3 (37.5 mg, 94%) as a colourless oil. See Supplementary Fig. 1b for structures.

**¹H NMR (600 MHz, chloroform-d).** δ 6.19 (d, $J$ = 8.4 Hz, 1H), 4.51 (td, $J$ = 8.4, 5.4 Hz, 1H), 3.87 (t, $J$ = 5.4 Hz, 2H), 2.51–2.41 (m, 2H), 1.71–1.58 (m, 2H), 1.52–1.46 (m, 1H), 1.46 (s, 9H), 0.94 (d, $J$ = 6.5, 6H, overlapping signals).

### N-(3-hydroxypropionyl)-L-leucine (4)

To a solution of 3 (15 mg, 0.058 mmol, 1.0 equiv.) in 0.5 ml of DCM was added trifluoroacetic acid (TFA, 0.35 ml, 4.63 mmol, 80 equiv.). The resulting solution was stirred for 3 h and then concentrated to dryness in vacuo. The crude mixture (12 mg, containing a mixture of product alcohol and its corresponding trifluoroacetic acid ester) was dissolved in 1:1 MeOH:H$_2$O (0.4 ml), K$_2$CO$_3$ (14 mg, 0.108 mmol, 2.7 equiv.) was added, and the resulting mixture was stirred at 40 °C for 1 h. The crude mixture was then concentrated to dryness in vacuo and resuspended in 4:1 DCM:MeOH, and the supernatant was collected. After concentration, a pure sample of 4 (10 mg, 71% over two steps) was obtained as its potassium salt. See Supplementary Fig. 1c for structures and Supplementary Figs. 2 and 3 for NMR spectra.

**¹H NMR (500 MHz, methanol-$d_4$).** δ 4.30 (dd, $J$ = 9.9, 4.4 Hz, 1H), 3.86–3.75 (m, 2H), 2.51–2.38 (m, 2H), 1.77–1.51 (m, 3H), 0.94 (d, $J$ = 6.2 Hz, 3H), 0.93 (d, $J$ = 6.2 Hz, 3H).

**¹³C NMR (126 MHz, methanol-$d_4$).** δ 180.3, 173.6, 59.6, 55.0, 43.1, 40.3, 26.2, 23.8, 22.2.

HRMS (ESI) $m/z$: [M − H]⁻ calculated for C$_9$H$_{16}$NO$_4$ [M-H]⁻ 202.10848; found 202.10890.

The retention time and MS/MS spectrum of synthetic 4 were identical to those of one of the features at $m/z$ 202.10848 in the C. elegans metabolome.

### N-(3-O-TBDPS-propionyl)-L-phenylalanine tert-butyl ester (5)

To a solution of 1 (100 mg, 0.305 mmol, 1.0 equiv.) in 2 mL of DCM was added L-phenylalanine tert-butyl ester hydrochloride (78 mg, 0.304 mmol, 1.0 equiv.), EDC hydrochloride (116 mg, 0.608 mmol, 2.0 equiv.), and DMAP (126 mg, 1.04 mmol, 3.4 equiv.). The resulting solution was stirred at room temp. overnight and concentrated in vacuo. Flash column chromatography on silica using a gradient of 0–50% ethyl acetate in hexanes afforded 5 (135 mg, 82%) as a light yellow, viscous oil. See Supplementary Fig. 1d for structures.

**¹H NMR (400 MHz, chloroform-d).** δ 7.66–7.61 (m, 4H), 7.45–7.34 (m, 6H), 7.23–7.18 (m, 3H), 7.18–7.10 (m, 2H), 6.71 (d, $J$ = 7.6 Hz, 1H), 4.86–4.71 (m, 1H), 3.88 (td, $J$ = 6.0, 2.0 Hz, 2H), 3.07 (dd, $J$ = 6.0, 2.0 Hz, 2H), 2.39 (t, $J$ = 5.8 Hz, 2H), 1.37 (s, 9H), 1.01 (s, 9H).

### N-(3-hydroxypropionyl)-L-phenylalanine tert-butyl ester (6)

To a solution of 5 (80 mg, 0.150 mmol, 1.0 equiv.) in 2.5 mL of THF was added TBAF (1M in THF, 0.21 ml, 0.21 mmol, 1.40 equiv.). The resulting solution was stirred at room temperature for 0.5 h and was then concentrated in vacuo. Flash column chromatography on silica using a gradient of 0–40% MeOH in DCM afforded 6 (38.5 mg, 80%) as a colourless oil. See Supplementary Fig. 1e for structures.

**¹H NMR (400 MHz, chloroform-d).** δ 7.34–7.20 (m, 3H), 7.19–7.12 (m, 2H), 6.16 (d, $J$ = 7.9 Hz, 1H), 4.77 (dt, $J$ = 7.9, 6.0 Hz, 2H), 3.85 (q, $J$ = 5.5 Hz, 2H), 3.13 (dd, $J$ = 14.1, 6.3 Hz, 1H), 3.07 (dd, $J$ = 14.1, 6.3 Hz, 1H), 2.86 (t, $J$ = 6.1 Hz, 1H), 2.42 (td, $J$ = 5.2, 2.4 Hz, 2H), 1.42 (s, 9H).

### N-(3-hydroxypropionyl)-L-phenylalanine (7)

To a solution of 6 (15 mg, 0.046 mmol, 1.0 equiv.) in 0.5 mL of DCM was added TFA (0.28 mL, 3.70 mmol, 80 equiv.). The resulting solution was stirred for 3 h and then concentrated to dryness in vacuo. The crude mixture (12 mg, containing product alcohol and its corresponding trifluoroacetic acid ester) was dissolved in 1:1 MeOH:H$_2$O (0.4 ml), K$_2$CO$_3$

(14 mg, 0.108 mmol, 3.0 equiv.) was added and the resulting mixture was stirred at 40 °C for 1 h. The crude mixture was then concentrated to dryness in vacuo and resuspended in 4:1 DCM:MeOH, and the supernatant was collected. After concentration, a pure sample of **7** (12 mg, 95%) was obtained as its potassium salt. See Supplementary Fig. 1f for structures and Supplementary Figs. 4 and 5 for NMR spectra.

**¹H NMR (500 MHz, methanol-*d₄*).** δ 7.26–7.11 (m, 5H), 4.51 (dd, *J* = 7.8, 4.8 Hz, 1H), 3.77–3.66 (m, 2H), 3.22 (dd, *J* = 13.8, 4.8 Hz, 1H), 2.95 (dd, *J* = 13.8, 7.8 Hz, 1H), 2.42–2.30 (m, 2H).

**¹³C NMR (126 MHz, methanol-*d₄*).** δ 178.3, 173.3, 139.7, 130.6, 129.1, 127.2, 59.5, 57.5, 40.3, 39.3.

HRMS (ESI) *m/z*: [M − H]⁻ calculated for $C_{12}H_{14}O_4N$ [M-H]⁻ 236.09283; found 236.09309. The retention time and MS/MS spectrum of synthetic **8** were identical to those of a feature at *m/z* 236.09283 in the *C. elegans* metabolome.

### *N*-propionyl-ʟ-leucine *tert*-butyl ester (9)
Propionic acid (**8**, 74 mg, 1.00 mmol, 1.0 equiv.), ʟ-leucine *tert*-butyl ester hydrochloride (223 mg, 1.00 mmol, 1.0 equiv.), EDC hydrochloride (382 mg, 2.00 mmol, 2.0 equiv.) and DMAP (414 mg, 3.39 mmol, 3.39 equiv.) were combined in 5 ml DCM. The resulting solution was stirred at room temperature overnight and concentrated in vacuo. Flash column chromatography on silica using a gradient of 0–100% ethyl acetate in hexanes afforded **9** (220 mg, 90%) as a light yellow, viscous oil. See Supplementary Fig. 1g for structures.

**¹H NMR (400 MHz, chloroform-*d*).** δ 5.92 (d, *J* = 8.2 Hz, 1H), 4.51 (td, *J* = 8.5, 5.4 Hz, 1H), 2.22 (q, *J* = 7.6 Hz, 2H), 1.71–1.54 (m, 2H), 1.50–1.41 (m, 1H), 1.44 (s, 9H), 1.14 (t, *J* = 7.6 Hz, 3H), 0.92 (d, *J* = 6.3 Hz, 3H), 0.92 (d, *J* = 6.5 Hz, 3H).

### *N*-propionyl-ʟ-leucine (10)
To a solution of **9** (90 mg, 0.37 mmol, 1.0 equiv.) in 4 ml of DCM was added TFA (2.29 ml, 30.0 mmol, 81 equiv.). The resulting solution was stirred for 2 h and then concentrated to dryness in vacuo, affording **10** (69 mg, quantitative). See Supplementary Fig. 1h for structures and Supplementary Figs. 6 and 7 for NMR spectra.

**¹H NMR (500 MHz, chloroform-*d*).** δ 7.95–7.73 (br m, 1H), 6.29–6.21 (m, 1H), 4.64 (td, *J* = 8.4, 4.9 Hz, 1H), 2.36 (q, *J* = 7.6 Hz, 2H), 1.79–1.57 (m, 3H), 1.19 (t, *J* = 7.6 Hz, 3H), 0.97 (d, *J* = 6.1 Hz, 3H), 0.95 (d, *J* = 6.1 Hz, 3H).

**¹³C NMR (126 MHz, chloroform-*d*).** δ 177.0, 176.6, 51.2, 41.1, 29.5, 25.0, 22.8, 21.9, 9.84.

HRMS (ESI) *m/z*: [M − H]⁻ calculated for $C_9H_{16}O_3N$ [M-H]⁻ 186.11357; found 186.11337.

The retention time and MS/MS spectrum of synthetic **11** were identical to those of a feature at *m/z* 186.11357 in the *C. elegans* metabolome.

### Reporting summary
Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

### Data availability
The HPLC–MS/MS data generated during this study have been deposited in the MassIVE database under accession code MSV000087810 (https://doi.org/10.25345/C5C244). The raw sequencing reads for these samples are available from the NCBI Sequence Read Archive (project PRJNA669127). The metabolic network model with the added short chain fatty acid–amino acid conjugation reactions (pan-iCELv1.0) is available for download in SBML and JSON formats at the wormflux website (http://wormflux.umassmed.edu/download. php). Genotype data for *C. elegans* strains were acquired from the variant call format (VCF) file (release 20200815) available from the CeNDR (http://elegansvariation.org). Source data are provided with this paper.

50. Hoki, J. S. et al. Deep interrogation of metabolism using a pathway-targeted click-chemistry approach. *J. Am. Chem. Soc.* **142**, 18449–18459 (2020).

**Author contributions** A.J.M.W., F.C.S. and E.C.A. conceived the study. Y.-U.L., O.P., B.W.F. and H.N. prepared *C. elegans* cultures. B.W.F. performed LC–MS experiments and analysed the LC–MS data with help from P.R.R. and A.R.K. O.P. performed GC–MS experiments and analysed the GC–MS data. G.Z. performed the RNA-seq experiments. G.Z. and L.S.Y. analysed the RNA-seq data. G.E.G. and M.W. performed 3HP toxicity assays. N.M.R. created genome-edited strains. G.Z. and B.W.F. analysed genomic variants. B.J.C. synthesized the authentic standards. T.A.C. and S.Z. collected new wild strains. L.S.Y. created pan-iCELv1.0 by adding new metabolites and reactions to iCEL1314. A.J.M.W., B.W.F., L.S.Y., F.C.S. and E.C.A. wrote the paper. All authors read, edited and approved the final manuscript.

**a**



**b**



**c**
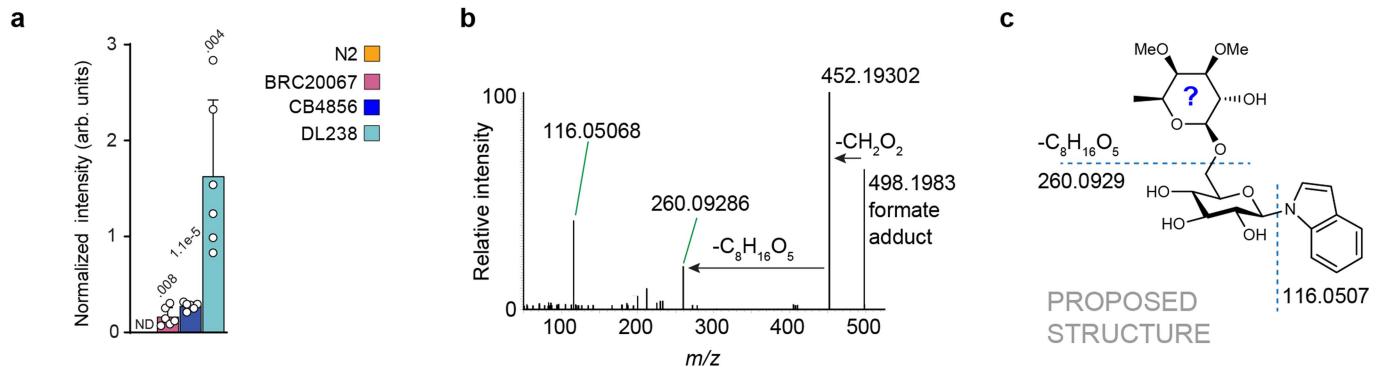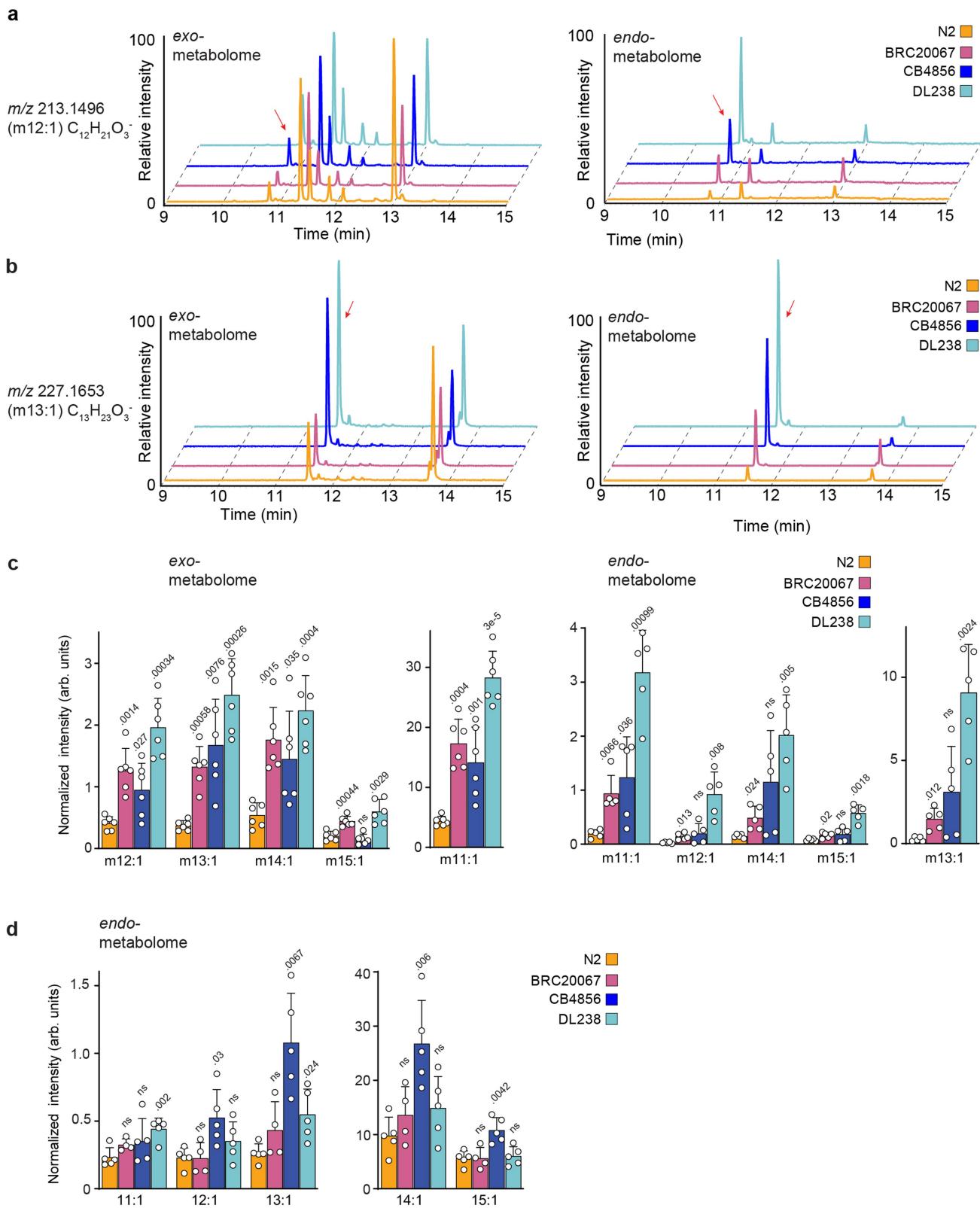


**Extended Data Fig. 1 | Genetic relatedness of the four *C. elegans* strains used in this study and experimental design. a**, Similarity matrix depicting the genetic relatedness of the four *C. elegans* strains used in this study. Pairwise similarities for strains were calculated based on 2,604,844 bi-allelic SNVs. See Methods for more details. **b**, *C. elegans* populations were synchronized either by bleaching or by settling in three biological replicates each, yielding six independent samples. Freeze-dried *E. coli* HB101 lysate powder was used as a food source to culture *C. elegans*. Animals and supernatant were collected as day 1 adults at the stage of the first egg laid and used for transcriptomic and metabolomic analyses. Created with BioRender.com. **c**, Photographs of all strains at the time of collection from each experiment. Scale bar, 250 µm.

**a**

Normalized intensity (arb. units)

N2
BRC20067
CB4856
DL238

.004

1.1e-5

.008

ND

**b**

Relative intensity

116.05068

452.19302

260.09286

498.1983
formate
adduct

-CH₂O₂

-C₈H₁₆O₅

$m/z$

**c**

MeO    OMe

?

⸱⸱⸱⸱OH

-C₈H₁₆O₅

260.0929

HO

HO

OH

N

PROPOSED
STRUCTURE

116.0507
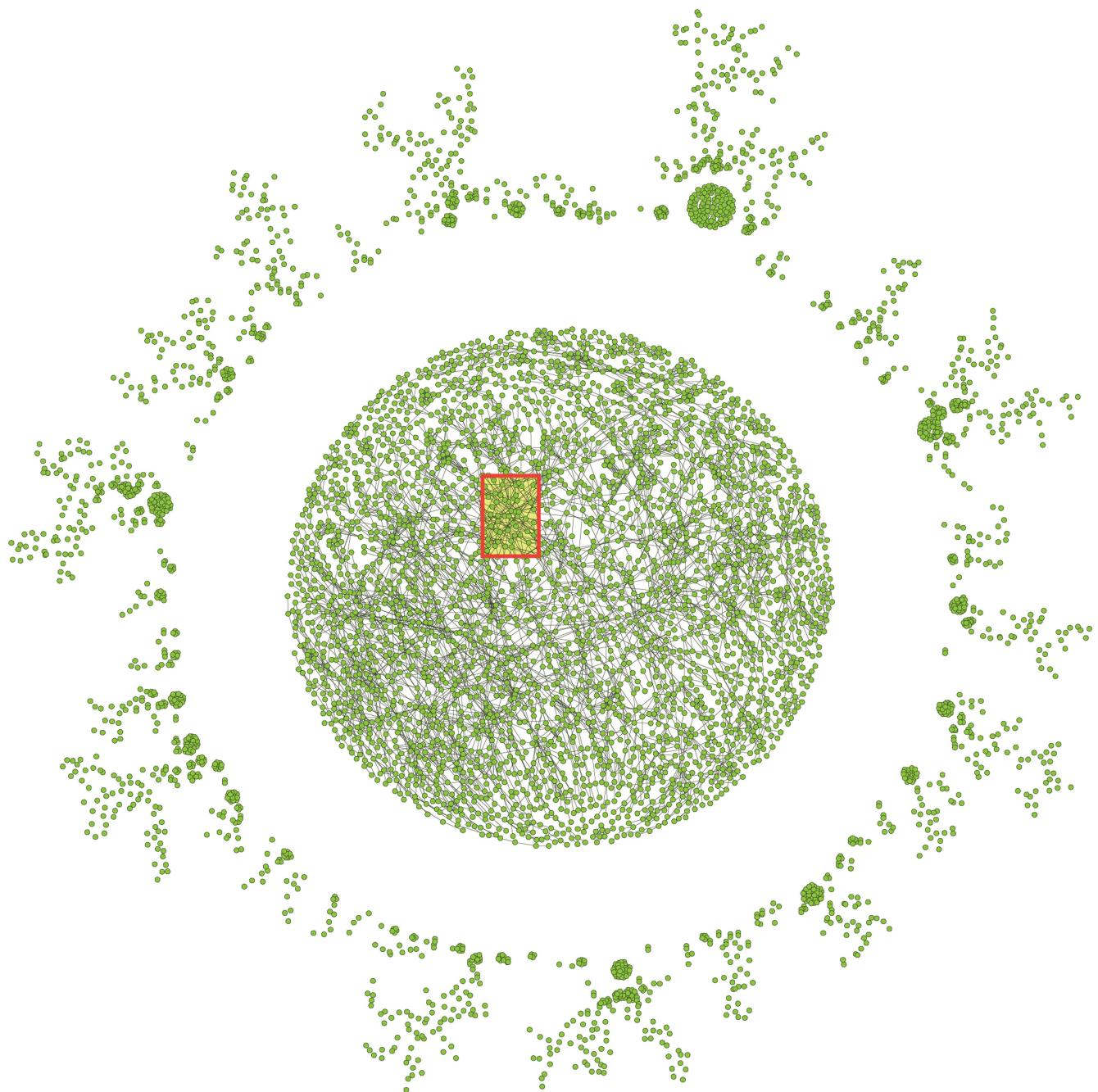
**Extended Data Fig. 2 | iglu#93 is an indole glycoside produced by wild strains but not by the N2 reference strain. a**, Quantification of indole glycoside iglu#93 in exo-metabolome extracts of the four strains. Data represent six biologically independent experiments and bars indicate mean ± s.d. $p$-values calculated by unpaired, two-sided $t$-test with Welch correction. **b**, Major fragmentation reactions of $m/z$ 498.1983 (negative ion mode) and resulting fragment ions representing free indole and indole glycoside. Fragmentation is consistent with a deoxy-dimethyl sugar moiety, such as di-$O$-methylfucose[50]. **c**, Proposed structure of iglu#93. Stereochemistry and exact substitution patterns are unknown. Source data are provided as a Source Data file.
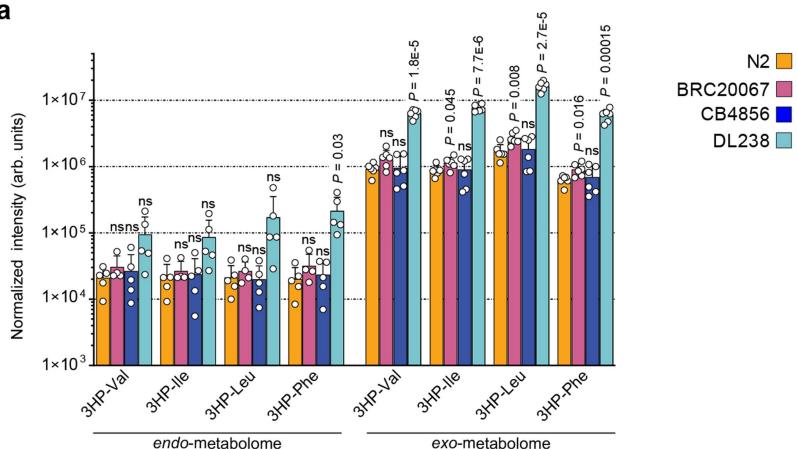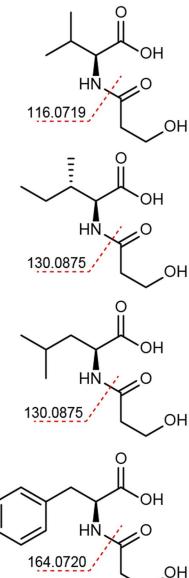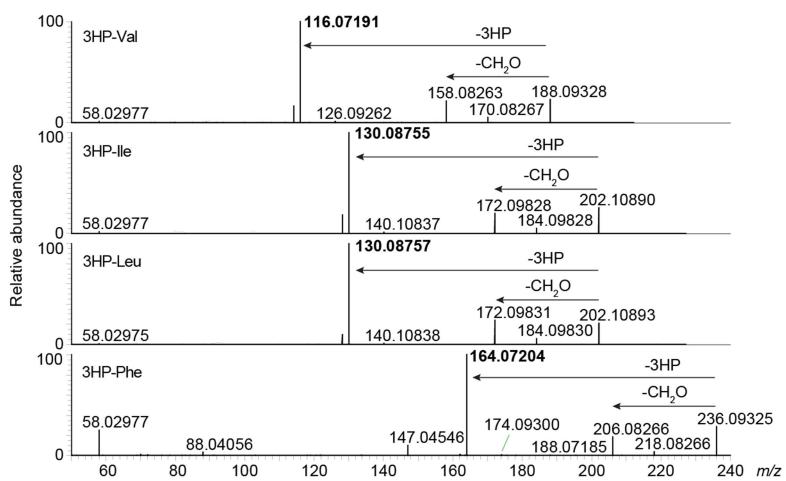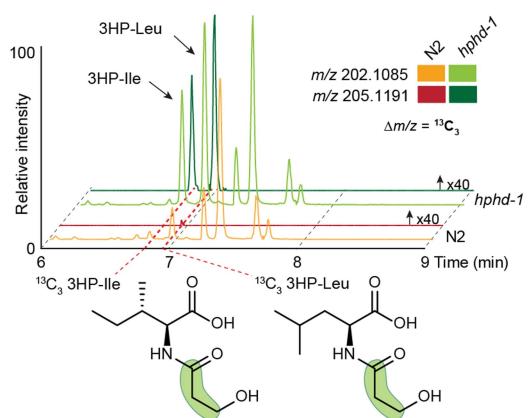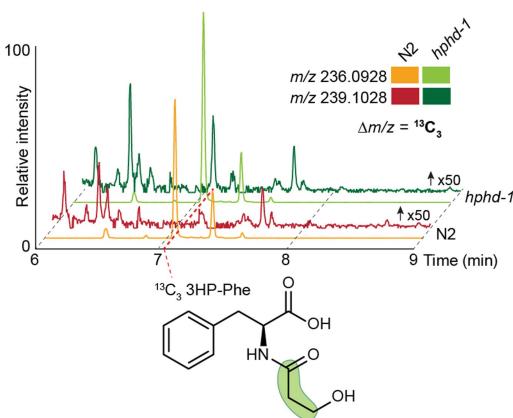
**Extended Data Fig. 3 | Homologous series of hydroxy-fatty acids reduced in the N2 strain compared to wild strains. a,b,** Representative HPLC–MS (negative ion) EICs for *m/z* 213.1496 (**a**) and *m/z* 227.1653 (**b**) in the exo- and endo-metabolomes of the four strains, as indicated. Features highly differential between strains are highlighted with red arrows. Several additional isobaric species detected primarily in the exo-metabolomes do not vary significantly among the four strains. **c,** Quantification of singly unsaturated (N:1), *m*ono-oxygenated (mN:1) fatty acids ranging from 11 to 15 carbons (m11:1 – m15:1) in exo- and endo-metabolome extracts, as indicated. Data represent

six (exo) or five (endo) biologically independent experiments and bars indicate mean ± s.d., *p*-values calculated by unpaired, two-sided *t*-test with Welch correction. ns, not significant. Y-axis scaling is maintained across graphs, which are separated for clarity. **d,** Quantification of singly unsaturated fatty acids (11:1 – 15:1) in endo-metabolome extracts, as indicated. Data represent five biologically independent experiments and bars indicate mean ± s.d., *p*-values calculated by unpaired, two-sided *t*-test with Welch correction. ns, not significant. Source data are provided as a Source Data file.
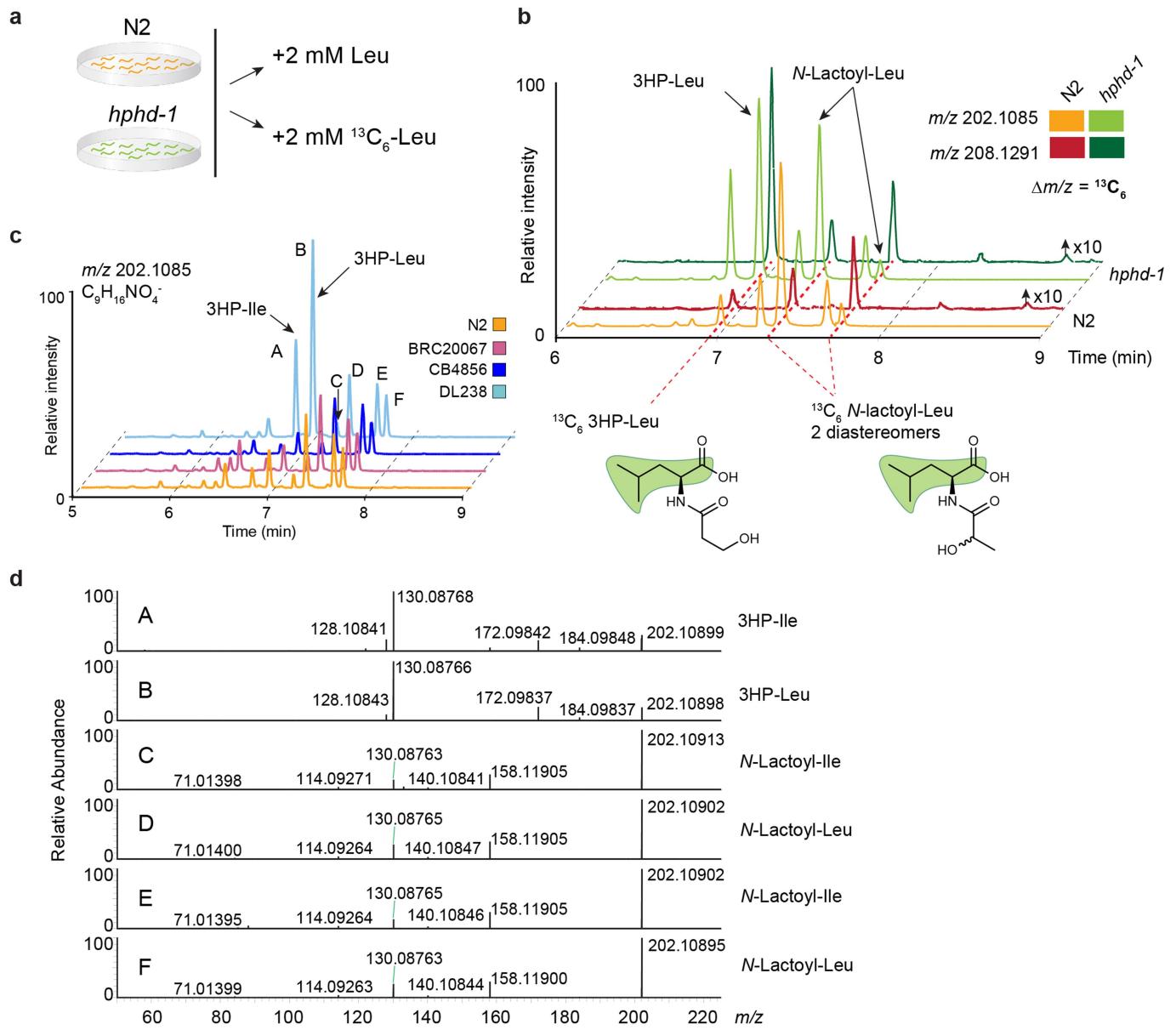
**Extended Data Fig. 4 | Full MS/MS network.** MS/MS network generated from exo-metabolome extracts in negative ionization mode. A simplified version of the boxed region containing 3HP-AAs is shown in Fig. 2a. An estimated 70% of nodes represent unique metabolites in this network, culled by standard filtering methods and some additional manual curation, see Methods for more details.

**Extended Data Fig. 5 | 3HP-AAs are more abundant in the exo-metabolome and are $^{13}C_3$-enriched from $^{13}C_5$-Val supplementation in *hphd-1(ok3580)* animals. a**, Quantification of 3HP-AAs in endo- and exo-metabolome extracts of the four strains. Data represent five or six biologically independent experiments for endo- or exo-metabolome quantification, respectively, and bars indicate mean ± s.d., *p*-values calculated by two-sided, unpaired *t*-test with Welch correction. ns, not significant. **b**, Major MS/MS fragmentation (negative ion mode) of 3HP-AAs and resulting fragment ions representing neutral loss of $CH_2O$ and free amino acids. Source data are provided as a Source Data file. **c**, EICs for 202.1085 and 205.1191, corresponding to $C_9H_{16}NO_4^-$ and $^{13}C_3$-$C_9H_{16}NO_4^-$, from exo-metabolome extracts of N2 and *hphd-1(ok3580)*. Red
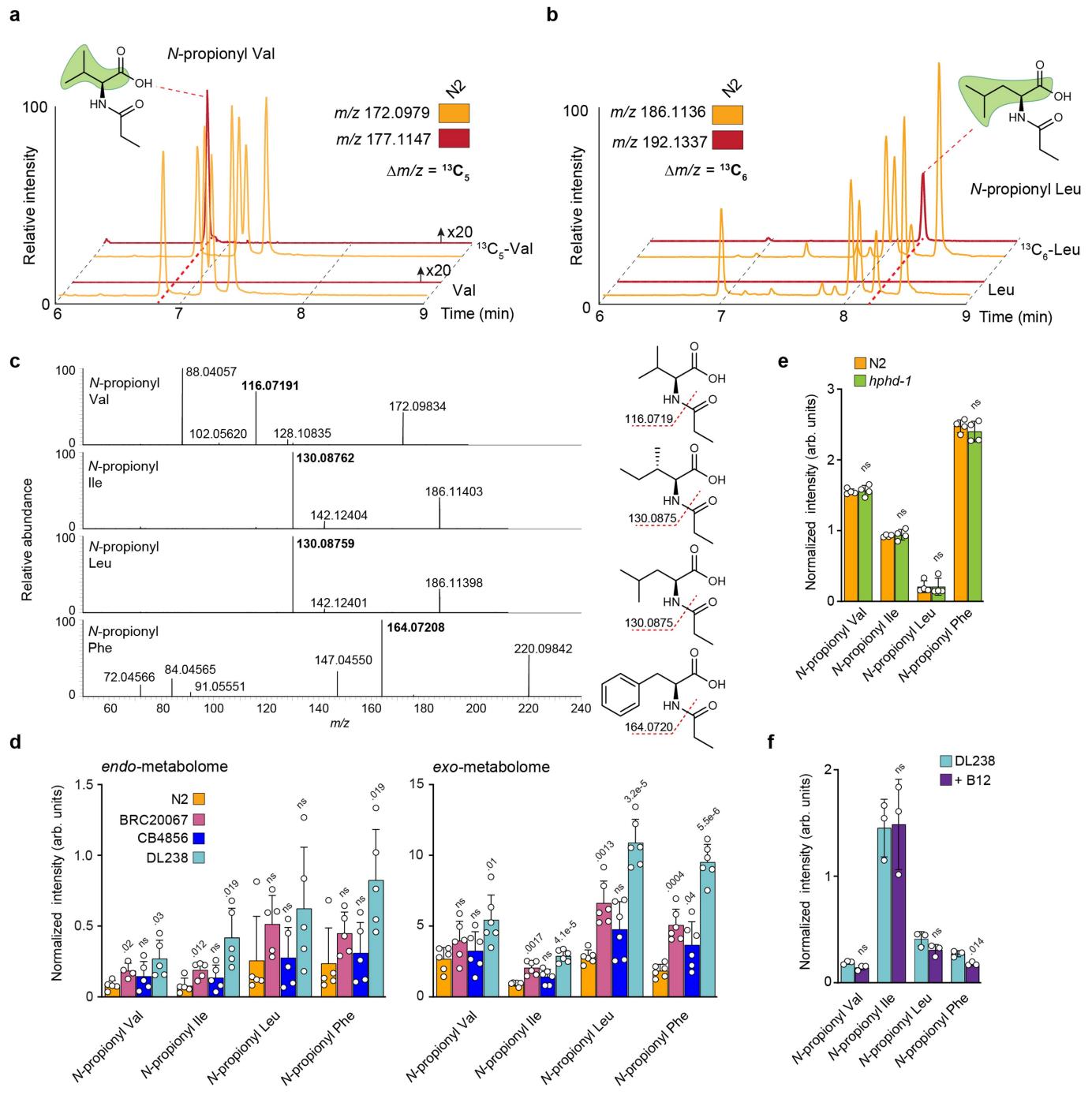
dashed lines highlight metabolites with $^{13}C_3$-enrichment that are also more abundant in *hphd-1(ok3580)*, corresponding to 3HP-Ile and 3HP-Leu. Y-axis for 205.1191 EIC is scaled 40-fold to more clearly show traces for labelled features. Isotopic labelling is highlighted with green shading in the shown compound structures. **d**, EICs for 236.0928 and 239.1028, corresponding to $C_{12}H_{14}NO_4^-$ and $^{13}C_3$-$C_{12}H_{14}NO_4^-$, from exo-metabolome extracts of N2 and *hphd-1(ok3580)*, as indicated. Red dashed line on the x-axis highlights metabolite with $^{13}C_3$-enrichment that is also more abundant in *hphd-1(ok3580)*, corresponding to 3HP-Phe. Y-axis for 239.1028 EIC is scaled 50-fold to more clearly show traces for labelled features. Isotopic labelling is highlighted with green shading in the shown compound structures.

**Extended Data Fig. 6 | ¹³C₆-Leu tracing distinguishes Leu- and Ile-conjugates that share identical MS/MS fragmentation. a**, Schematic of ¹³C₆-Leu isotope-tracing experiment. Synchronized N2 and *hphd-1(ok3580)* animals were supplemented with Leu or ¹³C₆-Leu and analysed for isotopic enrichment of metabolites of interest. **b**, EICs (negative ion mode) for *m/z* 202.1085 and 208.1291, corresponding to C₉H₁₆NO₄⁻ and ¹³C₆-C₉H₁₆NO₄⁻, from exo-metabolome extracts of N2 and *hphd-1(ok3580)*. Red dashed lines highlight metabolites with ¹³C₆-enrichment, corresponding to 3HP- and *N*-lactoyl Leu conjugates. Y-axis for *m/z* 208.1291 EIC is scaled 10-fold to more clearly show traces for labelled features. Isotopic labelling is highlighted with green shading in the shown compound structures. **c**, EIC for *m/z* 202.1085 (negative ion mode), corresponding to C₉H₁₆NO₄⁻, from exo-metabolome extracts of N2 and the three wild strains. **d**, MS/MS fragmentation for metabolites A-F in (**c**). Metabolites A and B have near-identical MS/MS spectra, that differ significantly from the near-identical MS/MS spectra of metabolites C, D, E, and F. Based on isotopic labelling and MS/MS fragmentation data, A and B were assigned as 3HP-Ile and 3HP-Leu, respectively. The remaining four metabolites represent diastereomers of *N*-lactoyl-Ile and -Leu, with the later-eluting metabolite of each pair assigned as *N*-lactoyl-Leu based on isotopic enrichment (see panel **b**).
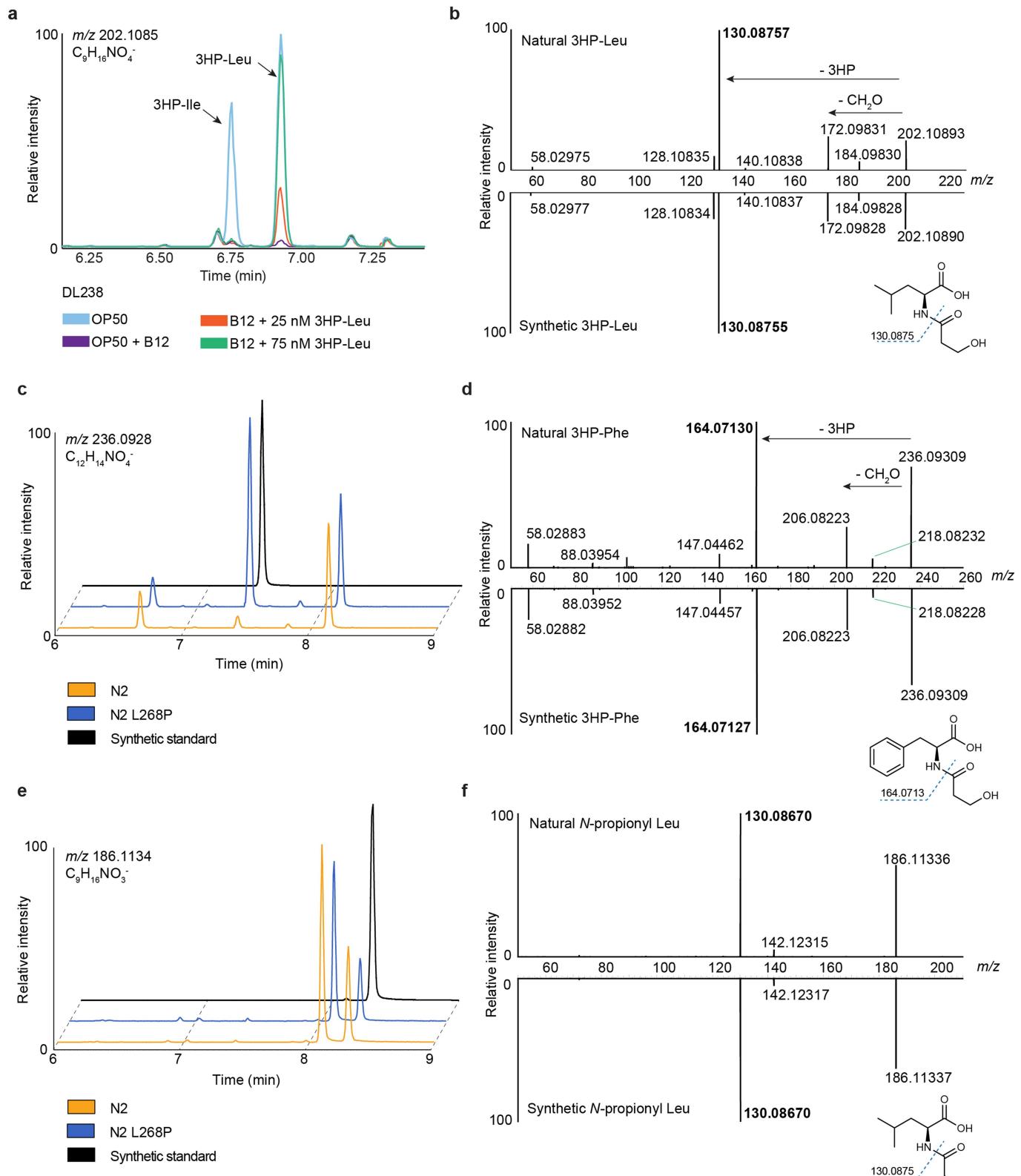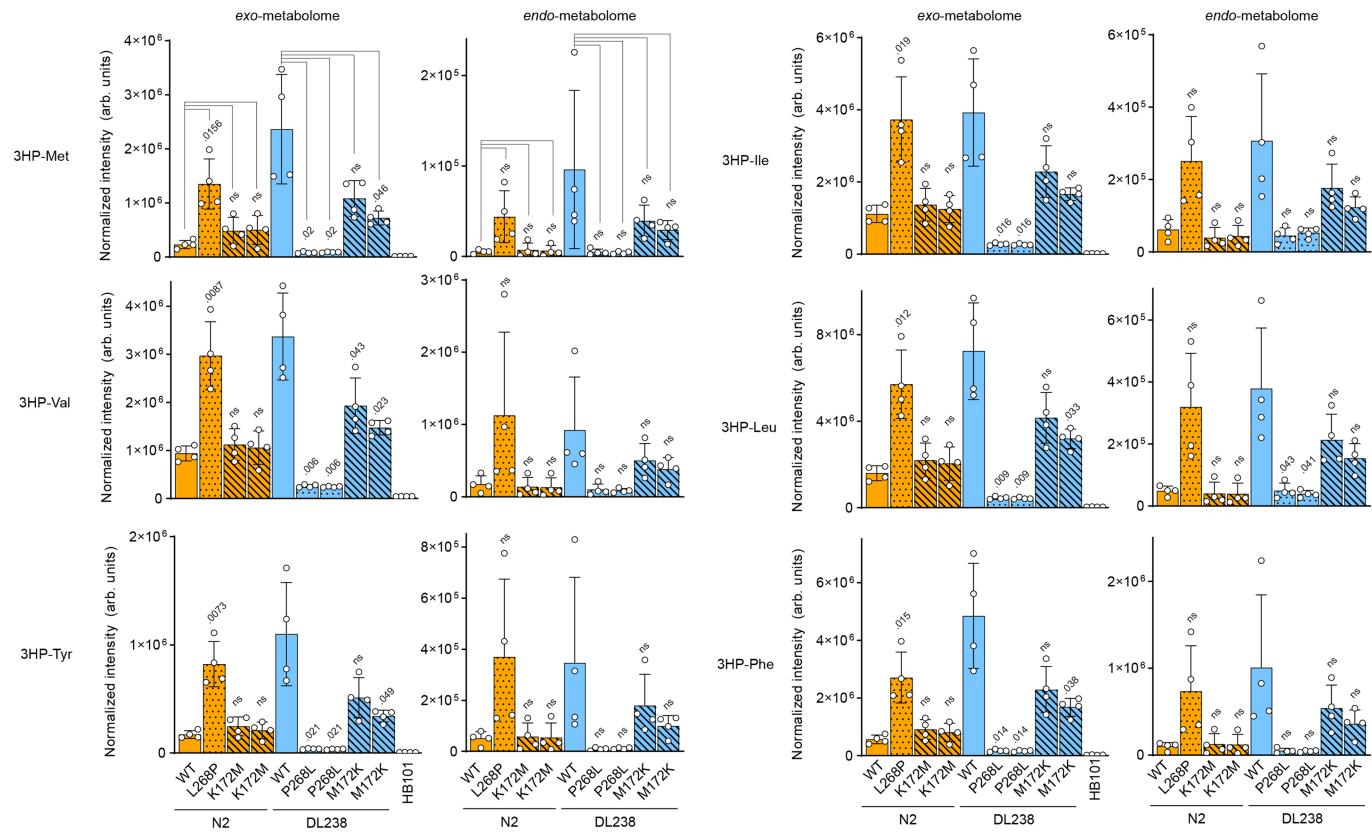
**Extended Data Fig. 7 | Isotope-tracing experiments aid the identification of N-propionyl-AAs. a**, EICs for *m/z* 172.0979 and 177.1147, corresponding to $C_8H_{14}NO_3^-$ and $^{13}C_5$-$C_8H_{14}NO_3^-$, from exo-metabolome extracts of N2 supplemented with Val or $^{13}C_5$-Val. Red dashed line highlights metabolite with $^{13}C_5$-enrichment, corresponding to *N*-propionyl-Val. EIC Y-axis for *m/z* 177.1147 is scaled 20-fold to more clearly show traces for labelled features. Isotopic labelling is highlighted by green shading in the structure of *N*-propionyl-Val. **b**, EICs for *m/z* 186.1136 and 192.1337, corresponding to $C_9H_{16}NO_3^-$ and $^{13}C_6$-$C_9H_{16}NO_3^-$, from exo-metabolome extracts of N2, supplemented with Leu or $^{13}C_6$-Leu. Red dashed line highlights metabolite with $^{13}C_6$-enrichment, corresponding to *N*-propionyl-Leu. Isotopic labelling is highlighted by green shading in the structure of *N*-propionyl-Leu. **c**, Major MS/MS fragmentation (negative ion mode) of *N*-propionyl-AAs and resulting fragment ions

representing different amino acids. **d**, Quantification of *N*-propionyl-AAs from endo- and exo-metabolome extracts of the four strains. Data represent five or six biologically independent experiments for the endo- or exo-metabolomes, respectively, and bars indicate mean ± s.d. Y-axis scaling is maintained across graphs. **e**, Quantification of *N*-propionyl-AAs in exo-metabolome extracts of N2 and *hphd-1(ok3580)* animals. Data represent four biologically independent experiments and bars indicate mean ± s.d. **f**, Quantification of *N*-propionyl-AAs in exo-metabolome extracts of the DL238 strain fed the standard *E. coli* OP50 diet without or with supplementation of 64 nM vitamin $B_{12}$, as indicated. Data represent three biologically independent experiments and bars means ± s.d. For **d**, **e**, and **f**, *p*-values calculated by two-sided, unpaired *t*-test with Welch correction. ns, not significant. Source data are provided as a Source Data file.
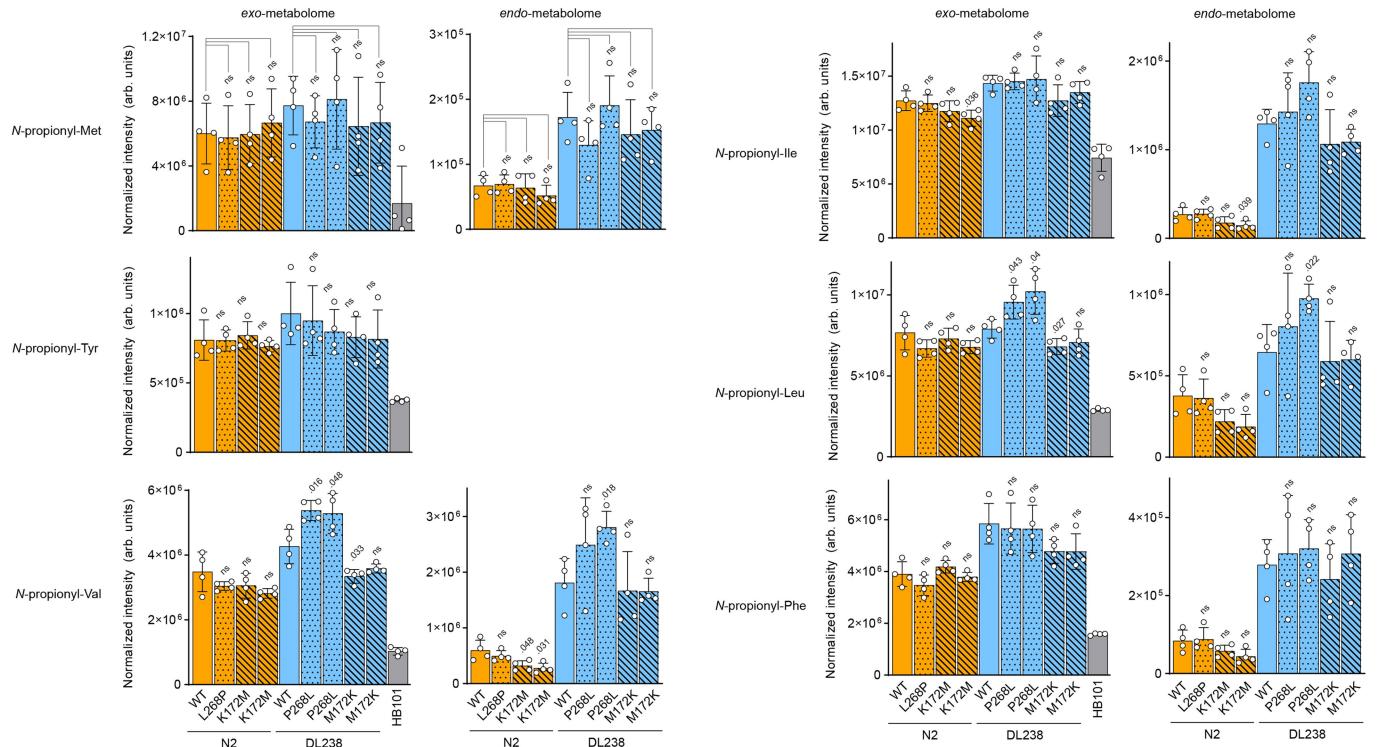
**Extended Data Fig. 8 | Validation of metabolite assignments by synthesis of 3HP-Leu, 3HP-Phe and *N*-propionyl-Leu authentic standards. a**, EIC for *m/z* 202.1085 (negative ion mode), corresponding to $C_9H_{16}NO_4^-$, from exo-metabolome extracts of DL238 fed OP50 or DL238 fed OP50 supplemented with 64 nM vitamin $B_{12}$, as indicated. To extracts of DL238 fed OP50 supplemented with vitamin $B_{12}$, synthetic 3HP-Leu was added to achieve final concentrations of 25 nM and 75 nM. **b**, MS/MS spectra (negative ion mode) of natural and synthetic 3HP-Leu with major fragmentation reactions and fragment ions. **c**, EIC for *m/z* 236.0928, corresponding to $C_{12}H_{14}NO_4^-$, from exo-metabolome extracts of N2, N2 HPHD-1$^{L268P}$, and synthetic 3HP-Phe. **d**, MS/MS spectra (negative ion mode) of natural and synthetic 3HP-Phe and major fragmentation reactions. **e**, EIC for *m/z* 186.1134, corresponding to $C_9H_{16}NO_3^-$, from exo-metabolome extracts of N2, N2 HPHD-1$^{L268P}$, or synthetic *N*-propionyl-Leu. **f**, MS/MS spectra (negative ion mode) of natural and synthetic *N*-propionyl-Leu and major fragmentation reactions.

**Extended Data Fig. 9 | Abundance of 3HP-AAs in *hphd-1* edited strains.** Quantification of 3HP-AA conjugates in exo- and endo-metabolome extracts of CRISPR–Cas9 HPHD-1-edited strains, parental strains, and HB101 bacteria only, as indicated. Data represent four biologically independent experiments and bars indicate mean ± s.d. *p*-values calculated by unpaired, two-sided *t*-test with Welch correction in which edited strains were compared to the corresponding parental strains. ns, not significant. Source data are provided as a Source Data file.
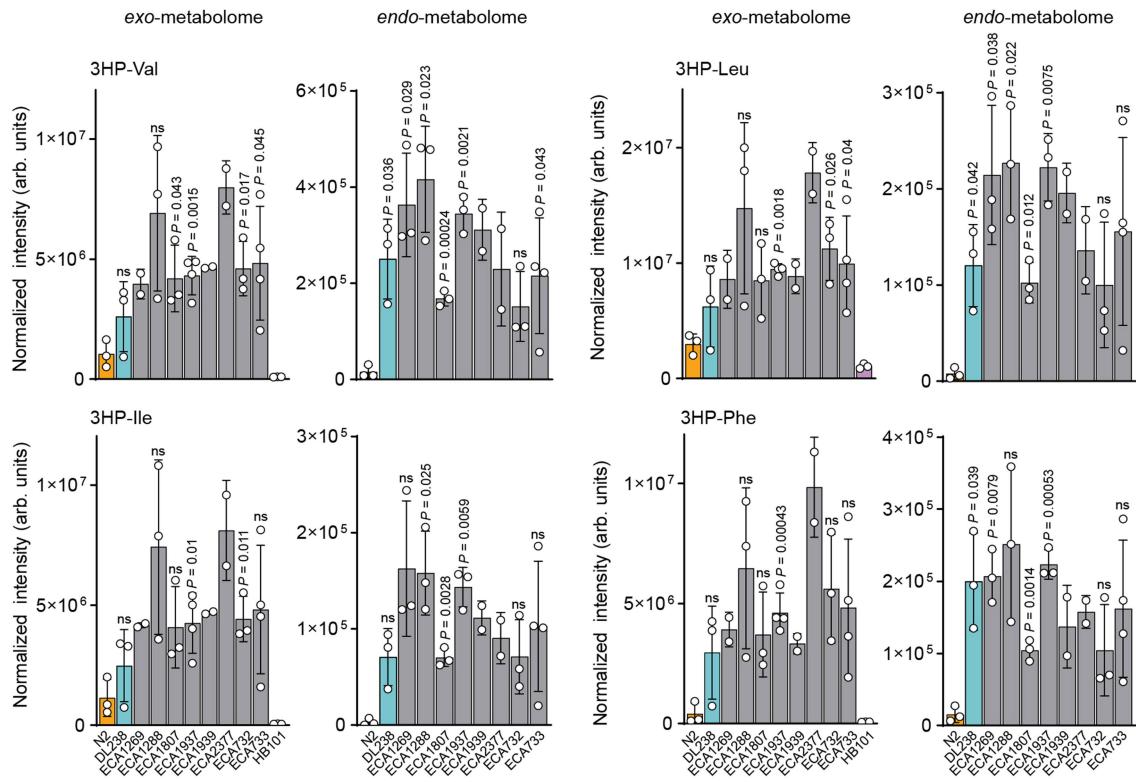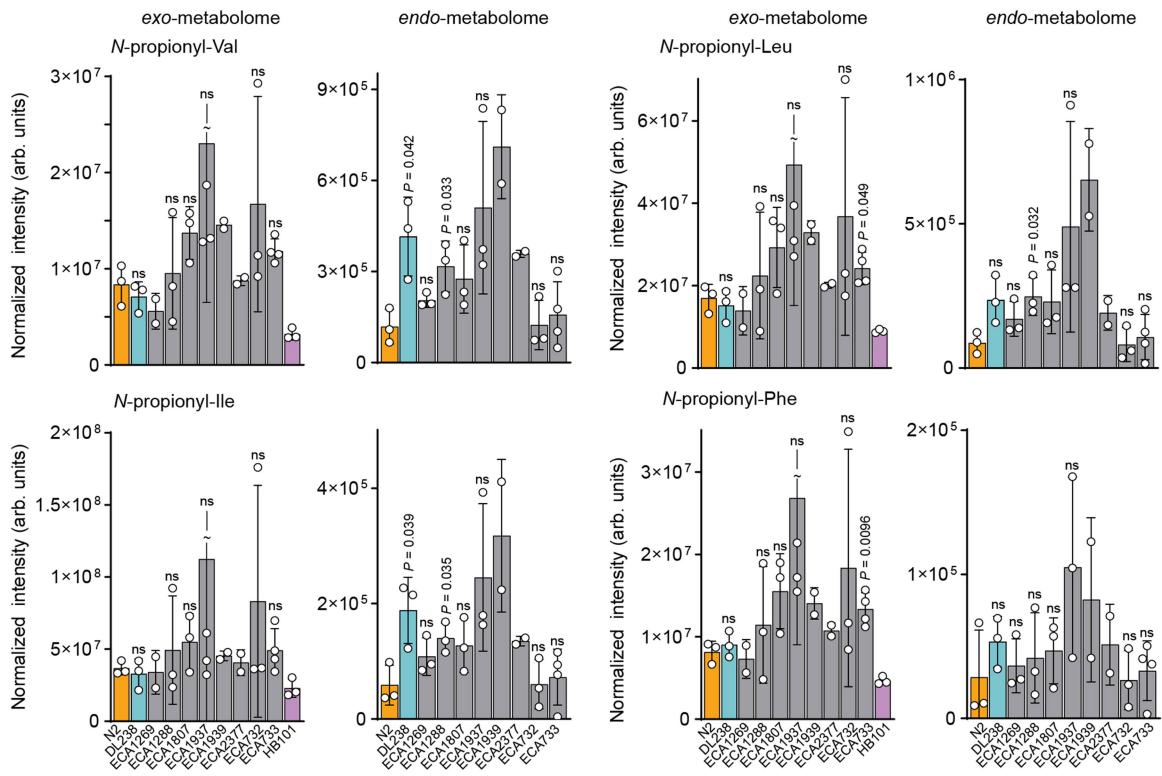
**Extended Data Fig. 10 | Abundance of *N*-propionyl-AAs in *hphd-1* edited strains.** Quantification of *N*-propionyl-AA conjugates in exo- and endo-metabolome extracts of CRISPR–Cas9 HPHD-1-edited strains, parental strains, and HB101 bacteria only, as indicated. Data represent four biologically independent experiments and bars indicate mean ± s.d. *p*-values calculated by unpaired, two-sided *t*-test with Welch correction in which edited strains were compared to the corresponding parental strains. ns, not significant. *N*-propionyl-Tyr could not be reliably quantified in the endo-metabolome due to interference from a co-eluting metabolite. Source data are provided as a Source Data file.

**Extended Data Fig. 11 | Abundance of 3HP and 3HP-AAs in wild strains.**
**a**, Quantification of 3HP and **b**, 3HP-AAs in exo- and endo-metabolome extracts of N2, DL238, additional wild strains and HB101 bacteria only, as indicated. Exact n for each strain is presented in Supplementary Table 12. Bars indicate mean ± s.d. *p*-values calculated by unpaired, two-sided *t*-test with Welch correction in which wild strains were compared to N2. ns, not significant. Source data are provided as a Source Data file.

**Extended Data Fig. 12 | Abundance of *N*-propionyl-AAs in wild strains.** Quantification of *N*-propionyl-AAs in exo- and endo-metabolome extracts of N2, DL238, additional wild strains and HB101 bacteria only, as indicated. Exact n for each strain is presented in Supplementary Table 12. Bars indicate mean ± s.d. *p*-values calculated by unpaired, two-sided *t*-test with Welch correction in which wild strains were compared to N2. ns, not significant. Source data are provided as a Source Data file.

# nature portfolio

Corresponding author(s): Andersen EC, Schroeder FC, Walhout AJM

Last updated by author(s): May 7, 2022

# Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☐ | ☒ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. $F$, $t$, $r$) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted *Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☐ | ☒ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | MS data acquisition was controlled by Chromeleon software (Thermo Scientific v7.3) and Xcalibur software (Thermo Scientific v4.1.31.9). |
| Data analysis | Gene expression data was analyzed using Kallisto (v.0.44.0), and the R packages tximport (v.1.10.1) and DESeq2 (v1.22.2). Transcript levels were normalized using the R package sleuth (v0.30.0). LC-MS RAW data were converted to mzXML file format using MSConvert (v3.0, ProteoWizard). Data was analyzed using Metaboseek software (v0.9.9.1) available here: [https://doi.org/10.5281/zenodo.3360087] and Xcalibur QualBrowser (Thermo Scientific v4.1.31.9). Statistical analysis was performed with Metaboseek software (v0.9.9.1) and GraphPad Prism (v9.3.0). GC-MS RAW data was analyzed using MassHunter software (Agilent v10.1). |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The HPLC-MS/MS data generated during this study have been deposited in the MassIVE database under accession code MSV000087810 [https://doi:10.25345/C5C244]. The raw sequencing reads for these samples are available from the NCBI Sequence Read Archive (Project PRJNA669127). Metabolic network model with

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | No sample size calculations were performed. The samples sizes in the study were chosen based on prior experience, reasonable replication, and standards in the field (see, for example, Ma et al. 2015, Chen et al. 2019, Giese et al. 2020, Helf et al. 2022). |
| Data exclusions | - In the original experiment: one replicate of endo-metabolome samples was excluded from HPLC-MS analysis due to a technical error during sample preparation.<br>- For the additional wild strains, one sample of exo-metabolome for ECA1269 was accidentally discarded, and one sample of ECA1937 endo-metabolome was not included due to technical error during sample preparation. |
| Replication | - In the original global comparison of four different C. elegans strains, 6 independent replicates were collected for the metabolomic and trancriptomic analyses.<br>- For vitamin B12 supplement in DL238, 3 independent replicates were analyzed.<br>- For isotope labeling studies, 4 independent replicates in N2 and hphd-1(ok3580) were analyzed in which each genotype was supplemented with Leu, 13C6-Leu, Val, or 13C5-Val.<br>- For the CRISPR-Cas9 HPHD-1 edit experiments, 4 independent replicates per strain were analyzed, and two independent strains for each amino acid edit were investigated (except for N2 L268P, for which there was one strain).<br>- For 3HP toxicity assays, 3 independent experiments were performed, two of which were quantified.<br>- For the additional wild strains, N2=3, DL238=3, ECA1269=3, ECA1288=3, ECA1807=3, ECA1937=4, ECA1939=2, ECA2377=2, ECA732=3, ECA733=4 independent replicates. See Supplementary Table 12 for exact n for each Figure Panel. |
| Randomization | For comparative analyses of different strains and conditions, all steps of worm culture and subsequent sample preparation were carried out in parallel. Samples were not randomized. |
| Blinding | Blinding was not possible during data collection because the acquired experimental data required manipulations of live animals. |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☒ | Antibodies |
| ☒ | Eukaryotic cell lines |
| ☒ | Palaeontology and archaeology |
| ☐ | Animals and other organisms |
| ☒ | Human research participants |
| ☒ | Clinical data |
| ☒ | Dual use research of concern |

## Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ChIP-seq |
| ☒ | Flow cytometry |
| ☒ | MRI-based neuroimaging |

# Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

| | |
|---|---|
| Laboratory animals | The C. elegans laboratory strain (N2 Bristol) and mutant hphd-1(ok3580) and CRISPR-Cas9 edited strains: ECA2545, PHX2627, PHX2628, PHX2640, PHX2666, PHX2667, and PHX2757. Wild strains: BRC20067, CB4856, DL238, ECA1269, ECA1288, ECA1807, ECA1937, ECA1939, ECA2377, ECA732, and ECA733. C. elegans are predominantly self-fertilizing hermaphrodites grown for between 1-3 days prior to harvest. |

| Wild animals | No wild animals were used in this study. |
| Field-collected samples | No field-collected samples were used in this study. |
| Ethics oversight | No ethical approval was required |

Note that full information on the approval of the study protocol must also be provided in the manuscript.