Combining metabolomics and experimental evolution reveals key mechanisms underlying longevity differences in laboratory evolved *Drosophila melanogaster* populations

Authors: Mark A. Phillips[1*], Kenneth R. Arnold[2], Zer Vue[3], Heather K. Beasley[3], Edgar Garza Lopez[4], Andrea G. Marshall[3], Derrick J. Morton[5], Melanie R. McReynolds[6], Thomas T. Barter[2], Antentor Hinton Jr.[3,4#]

*First author

#Corresponding

Affiliations:

1. Department of Integrative Biology, Oregon State University, Corvallis, OR, USA.
2. Department of Ecology and Evolutionary Biology, University of California, Irvine, Irvine, CA, 92697, USA.
3. Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN 37232, USA.
4. Hinton and Garza Lopez Family Consulting Company, Iowa City, IA
5. Department of Biological Sciences, University of Southern California, Los Angeles, CA
6. Department of Biochemistry and Molecular Biology, Huck Institute of the Life Sciences, Pennsylvania State University, University Park, PA

∗Corresponding author:

Antentor O. Hinton Jr., Ph.D.
Department of Molecular Physiology and Biophysics
Vanderbilt School of Medicine Basic Sciences
The Vanderbilt Diabetes Research and Training Center
750 Robinson Research Building 2200 Pierce Ave Nashville, TN 37232-0615, USA.
E-mail: antentor.o.hinton.jr@Vanderbilt.Edu

Abstract

Experimental evolution with *Drosophila melanogaster* has been used extensively for decades to study aging and longevity. In recent years, the addition of DNA and RNA sequencing to this framework has allowed researchers to leverage the statistical power inherent to experimental evolution study the genetic basis of longevity itself. Here we incorporated metabolomic data into to this framework to generate even deeper insights into the physiological and genetic mechanisms underlying longevity differences in three groups of experimentally evolved *D. melanogaster* populations with different aging and longevity patterns. Our metabolomic analysis found that aging alters mitochondrial metabolism through increased consumption of NAD^+ and increased usage of the TCA cycle. Combining our genomic and metabolomic data produced a list of biologically relevant candidate genes. Among these candidates, we found significant enrichment for genes and pathways associated with neurological development and function, and carbohydrate metabolism. While we do not explicitly find enrichment for aging canonical genes, neurological dysregulation and carbohydrate metabolism are both known to be associated with accelerated aging and reduced longevity. Taken together, our results provide very plausible genetic mechanisms for what might be driving longevity differences in this experimental system. More broadly, our findings demonstrate the value of combining multiple types of omic data with experimental evolution when attempting to dissect mechanisms underlying complex and highly polygenic traits like aging.

Key Words: metabolomics; aging; evolve and resequence; Drosophila melanogaster; experimental evolution; life history traits; genomics

Introduction

Understanding the factors that drive differences in life history between individuals, populations, and species is a major area of interest for evolutionary biologists [1–4]. Within this area of research, experimental evolution is a powerful tool for addressing fundamental questions about how natural selection shapes life history. *Drosophila melanogaster* is an extensively used, powerful, genetically tractable model system for studying life histories. For instance, there is a large body of work devoted to subjecting *D. melanogaster* populations to selection regimes that target variables thought to be important in life history evolution to test theoretical predictions [5–9]. Comparisons between populations selected for radically different life histories have also been successfully used to study the physiological and genetic mechanisms underlying observed differences in traits like longevity, developmental rates, and reproductive output [10]. Here we are primarily interested in longevity.

Many studies using experimental evolution to study physiological mechanisms underlying longevity differences have focused on the relationship between stress tolerance and life span. For instance, selection for decreased rates of senescence and increased longevity have repeatably been shown to associate with improved desiccation and starvation resistance and vice-versa [11–15]. Physiological assays have then shown that differences in longevity associated stress tolerances in these systems are tied to differences in lipid, glycogen, and water content [15–19]. However, deeper, more fine-scale investigations into what other physiological factors are driving differences in these types of experimental systems are currently lacking.

More recently, combining experimental evolution with next-generation sequencing technologies, termed “evolve and resequence” or “E&R”, has emerged as a powerful general tool for dissecting the underlying genetic architecture of complex traits [20,21]. Using this framework, evolutionary biologists can study the genetic basis of longevity by simply comparing patterns of genetic differentiation between

groups of evolved populations with significant differences in mean longevity. Through these comparisons, statistical associations can be made between genetic variants and longevity or related phenotypes. Here it is important to note that using evolution to drive phenotypes in opposite directions generates increased statistical power when making these associations [22]. Combined with the population-level replication featured in most studies, this approach allows for more powerful statistical inferences to be drawn than what is typically seen in conventional genome-wide association studies [23].

At present, several genomic [15,24–27] and transcriptomic [28,29] E&R studies have generated insights into the genetics of increased longevity and related stress resistances. The integration of other types of "omic" data into this framework can potentially yield similarly powerful insights. Here, we aim to do so by incorporating metabolomic data into the E&R framework. Specifically, we explore metabolomic differences between *Drosophila* populations where dozens to hundreds of generations of selection have produced large differences in longevity and life history traits. Our study also represents a more nuanced investigation of the physiological mechanisms driving observed differences than previous studies which have been largely limited to assaying broad differences in total lipid and glycogen contents.

Metabolomic studies characterize differences in the prevalence of small-molecule metabolites in biological systems along some axis of interest. These collections of metabolites are viewed as one of the intermediary layers between the genome and expressed phenotypes, and thus represent a key component to understanding how phenotypes are determined. This general approach is already being used to explore the mechanisms driving physiological declines associated with advanced age [31], and *Drosophila*-based studies are already part of this effort [32]. For instance, Hoffman et al. 2014 [33] analyzed changes in metabolomic profiles over time in different genetic backgrounds to identify aging-associated metabolic pathways. Taking a different approach, Laye et al. 2015[34] examined how dietary restriction, a proven intervention strategy for extending lifespan in fruit flies, alters the metabolome to slow aging-related physiological declines. Here, we built on these efforts using an evolution-based approach where, like in the E&R studies described above, we used selection and replicate populations to amplify signals and identify relationships between metabolomic profiles and different aging patterns.

In this study, we investigated differences in metabolomic profiles among three groups of experimentally evolved *D. melanogaster* populations: populations where selection for rapid development has resulted in reduced longevity (A), and where selection for delayed reproduction (C) and starvation resistance (S) has led to increased longevity. In the past, the C populations have been used as controls for both the A and S populations. Studies comparing the C and A populations focused on the evolution and genetics of developmental rates, reproductive rates, and longevity [9,29], while work with the C and S populations focused more on understanding physiological changes associated with starvation resistance [15]. However, with respect to longevity, we find similar mean lifespans in the C and S populations while the A populations tend to die at much younger ages (Figure 1A). Additionally, all populations are ultimately derived from a single ancestral population, differences between groups are known to be the result of selection on standing genetic variation [15,26], and populations cluster genetically based on treatment (Figure 1B). As such, comparisons across the three groups have the potential to isolate and identify mechanisms specifically associated with longevity (i.e. elements that overlap between the C and S populations but differ in A). Using metabolomic data from these populations we aimed to: 1) identify metabolites associated with longevity differences based on patterns of overlap and differentiation between groups, and 2) identify genetic mechanisms underlying based on statistical associations between patterns of genomic and metabolomic differentiation.

Materials and Methods

Experimental populations

This study involved 30 large outbred, experimentally evolved *D. melanogaster* populations evenly split between three treatments (A, C, and S). These populations are part of an experimental system founded in 1980 [7] and maintained in the Rose Lab at UC Irvine. All populations trace back to a single ancestral laboratory population created by sampling females from a wild population in Amherst, Massachusetts. The A and C populations are described in depth in Burke et al. 2016[9] and Graves et al. (2017)[26]. Briefly, selection in these treatments is defined by optimization around different reproductive windows. The A populations are maintained on a discrete 10-day generation cycle, while the C populations are maintained on a 28-day cycle. Over generations, this has resulted in accelerated development and reduced longevity in the A populations, and slower development, delayed reproduction, and increased longevity in the C populations (see Burke et al. 2016 for in-depth phenotypic characterizations[9]). Both the A and C populations consist of two sub-groups: ACO_{1-5}/AO_{1-5} and CO_{1-5}-/nCO_{1-5}, respectively. The primary difference between these groups is the number of generations under selection; AO and nCO have dozens of generations under selection, whereas ACO and CO have hundreds. Despite these differences, previous efforts have found extreme phenotypic [9], genomic [26], and transcriptomic [29] convergence between subgroups within a given treatment. As such, we now simply recognize them as the A and C populations.

Unlike the A and C populations where selection is defined by shifting reproductive windows, the S populations have been subjected to selection for increased starvation resistance. A full description of the S populations and how they were created can be found in Kezos et al. (2019)[15]. Briefly, the 10 S populations (SCO_{1-5a} and SCO_{1-5b}) were derived from the CO_{1-5} populations with two replicates derived from each CO population (e.g., SCO_{1a} and SCO_{1b} were derived from CO_1). In terms starvation selection protocols, after 2 weeks of development from eggs to adults in vials, approximately ~3000 flies are collected from each population and split between Plexiglas cages. This two-cage approach is used to maintain high census sizes through selection. At this point, flies are fed an agar media prepared with cooked bananas, corn syrup, yeast, and barley malt (see Rose et al. 2004 for exact recipe)[10]. On day 17 from egg, Petri dishes with banana agar media are removed from cages and replaced with dishes containing media made with only agar and water (i.e. a nonnutritive media). This starvation condition is maintained until 75–80% of flies have died. At this point, the two cages set up for each replicate are condensed into one cage. Flies are then provided with banana agar media for3 days at which points eggs are collected to start the next generation. From generation 50 onwards, starvation conditions typically needed to be maintained for ~10 days to reach the 75–80% mortality threshold.

Collecting flies for metabolite extraction

To collect samples for metabolomic characterization, we used the same sampling procedures described in previous genomic studies [15,26]. Briefly, cohorts were derived from each replicate population and reared on a 14-day culture cycle for two generation to reduce the impact of maternal effects. Cohorts descended from the second cycle were sampled for metabolite extraction. At creation, each cohort consisted of ~1500 flies in a Plexiglas cage and were fed the standard banana-based media with food being changed every two days. On day 21 from egg, 150 females were collected from each cohort, flash frozen in liquid nitrogen, and stored at -80°C until it was time for extraction.

The decision to collect samples on day 21 for metabolomic characterization was motivated by the findings of Barter et al. [29]. Here the authors compared gene expression in the A and C populations at two time points, day 21 and day 14, and found that differences much greater differences between the groups at the later timepoint. This was ultimately attributed to the fact that in demographic terms, the populations are not yet aging at day 14 (i.e. there is no acceleration in mortality during this period). However, by day 21, the A populations are clearly aging by this definition while the C populations are not[9]. And this is ultimately being reflected in patterns of differential gene expression at different life stages with there being greater power to identify aging related genes when one group is in its aging phase and the other is not. Given this, it stands to reason that comparing metabolomic differences at day 21across between treatments should similarly allow for more powerful identification of metabolic pathways underlying different rates of aging between the groups.

Metabolomic characterization

Gas chromatography-mass spectrometry (GC-MS)

For metabolite extraction, samples were extracted in -80°C 2:2:1 methanol/acetonitrile/water that contained a mixture of nine internal standards (d_4-Citric Acid, $^{13}C_5$-Glutamine, $^{13}C_5$-Glutamic Acid, $^{13}C_6$-Lysine, $^{13}C_5$-Methionine, $^{13}C_3$-Serine, d_4-Succinic Acid, $^{13}C_{11}$-Tryptophan, d_8-Valine; Cambridge Isotope Laboratories) at a concentration of 1 µg/ml each. The ratio of extraction solvent to sample volume was 18:1. Fly tissue samples were lyophilized overnight prior to extraction. After the addition of extraction buffer, fly tissues were homogenized using a ceramic bead mill homogenizer. The samples were then incubated at -20°C for 1 hour followed by a 10-minute centrifugation at maximum speed. Supernatants were transferred to fresh tubes. Pooled quality control (QC) samples were prepared by adding an equal volume of each sample to a fresh 1.5-ml microcentrifuge tube. Processing blanks were utilized by adding extraction solvent to microcentrifuge tubes. Samples, pooled QCs, and processing blanks were evaporated using a speed-vac, and the resulting dried extracts were derivatized using methyoxyamine hydrochloride (MOX) and N,O-Bis(trimethylsilyl)trifluoroacetamide (TMS) [both purchased from Sigma]. Briefly, dried extracts were reconstituted in 30 µl of 11.4 mg/ml MOC in anhydrous pyridine (VWR), vortexed for 10 min, and heated for 1 hour at 60°C. Next, 20 µl TMS was added to each sample, and samples were vortexed for 1 minute before heating for 30 minutes at 60°C. The derivatized samples, blanks, and pooled QCs were then immediately analyzed using GC-MS.

GC chromatographic separation was conducted on a Thermo Trace 1300 GC with a TraceGold TG-5SilMS column (0.25 um film thickness; 0.25mm ID; 30 m in length). The injection volume of 1 µl was used for all samples, blanks, and QCs. The GC was operated in split mode with the following settings: 20:1 split ratio, split flow: 24 µl/min, purge flow: 5 mL/min, Carrier mode: Constant Flow, Carrier flow rate: 1.2 mL/min. The GC inlet temperature was 250°C. The GC oven temperature gradient was as follows: 80°C for 3 minutes, ramped at 20°C/minute to a maximum temperature of 280°C, which was held for 8 minutes. The injection syringe was washed 3 times with pyridine between each sample. Metabolites were detected using a Thermo ISQ single quadrupole mass spectrometer, and data were acquired from 3.90 to 21.00 minutes in EI mode (70eV) by single-ion monitoring (SIM). Metabolite profiling data were analyzed using TraceFinder 4.1 utilizing standard verified peaks and retention times.

We used TraceFinder 4.1 to identify metabolites in extracted samples, blanks, and QCs by comparing sample metabolite peaks against an in-house library of standards prepared by processing and

analyzing authentic standards via the method described above. We created a database of retention times and three fragment ions for each metabolite standard: a target peak/ion and two confirming peaks/ions. When running biological samples, we identified metabolites that not only matched with the known retention times of the authentic standard but also with its target and confirming peaks. TraceFinder was also used for GC-MS peak integration to obtain peak areas for each metabolite. After TraceFinder analysis, we corrected for instrument drift over time using local regression analysis as described by Li et al. [35] We use the pooled QC samples, which were run in duplicate at the beginning and end of the GC-MS run for this purpose. The data are then normalized to an internal standard to control for extraction, derivatization, and/or loading effects.

Liquid chromatography-mass spectrometry (LC-MS)

Fly samples were dried to make extracts. Dried extracts were reconstituted in 40 µL acetonitrile/water (1:1 v/v), vortexed well, and transferred to LC-MS autosampler vials for analysis. LC-MS data were acquired on a Thermo Q Exactive hybrid quadrupole Orbitrap mass spectrometer with a Vanquish Flex UHPLC system or Vanquish Horizon UHPLC system. Notably, the LC column used was a Millipore SeQuant ZIC-pHILIC (2.1 ´ 150 mm, 5-µm particle size) with a ZIC-pHILIC guard column (20 x 2.1 mm). The injection volume was 2 µL. The mobile phase was composed of solvent A (20 mM ammonium carbonate [(NH4)2CO3] and 0.1% Ammonium Hydroxide [NH4OH]) and solvent B (Acetonitrile). The mobile phase gradient started at 80% solvent B, decreased to 20% solvent B over 20 minutes, returned to 80% solvent B in 0.5 minutes, and was held at 80% for 7 minutes. (PMID: 28388410). The method was run at a flow rate of 0.150 mL/min. Subsequently, the mass spectrometer was operated in full-scan, polarity-switching mode from 1 to 20 minutes, with the spray voltage set to 3.0 kV, the heated capillary held at 275 °C, and the HESI probe held at 350 °C. The sheath gas flow was set to 40 units, the auxiliary gas flow was set to 15 units, and the sweep gas flow was set to 1 unit. MS data acquisition was performed in a range of m/z 70–1,000, with the resolution set at 70,000, the AGC target at 1 × 106, and the maximum injection time at 200 ms [35]

For data analysis, acquired LC-MS data were processed by Thermo Scientific TraceFinder 4.1 software, and metabolites were identified based on the University of Iowa Metabolomics Core facility standard-confirmed, in house library. NOREVA was used for signal drift correction [26,35]. Data were normalized to total ion signals, and MetaboAnalyst 4.0 was used for further statistical processing and visualization[36,37].

Analyzing Metabolomic Data

We conducted analyses using the MetaboAnalyst 4.0 webservice (http://www.metaboanalyst.ca). Metaboanalyst is a module that uses both statistical and machine learning methods to provide visualization for classifying our data into groups. This program was utilized for heat mapping, enrichment pathways, pathway analysis, and statistical analysis. Unless otherwise noted, all data are reported as mean ± SD. One-way ANOVA followed by Tukey multiple comparison test was utilized. A probability value of $P \leq 0.05$ was considered significantly different. Statistical calculations were performed using the GraphPad Prism software (La Jolla, CA, USA).

Linking genome to metabolome

Here the goal was to identify patterns of SNP differentiation between the A, C, and S populations that best predict key patterns of metabolomic differentiation. For this analysis, efforts were focused on

significantly differentiated metabolites associated with the top ten enriched pathways from our GC-MS and LC-MS metabolomic profiling (see Supplementary Table 1 for list of metabolites, and Figures 2C and 3C for enriched pathways). SNP data came from previously published pool-seq DNA data from the (see Graves et al. 2017[26] for DNA extraction details for the A and C populations, and Kezos et al. 2019[15] for the S populations). Genomic data was reprocessed to since the original studies used different version of the *D. melanogaster* reference genome. However, processing steps were otherwise the same. Briefly, fastq files corresponding to each population were mapped to the *D. melanogaster* reference genome (version 6.14) with BWA[38] using bwa mem with default setting, and SAMtools [38] was used to convert the resulting SAM files to BAM files, remove potential PCR duplicates, and merge all BAM files into a single mpileup. PoPoolation2 [39] was used to convert this mpileup file to a simplified file format that contains counts for all bases in the reference genome and for all populations being analyzed. RepeatMasker (http://www.repeatmasker.org) and PoPoolation2 were then used to identify and remove highly repetitive genomic regions where proper read mapping is difficult. Lastly, SNPs were called based on the following criteria: minimum coverage of 20X and maximum of 200X in each population, and a combined minor allele frequency of 2% across all populations. This resulted in a SNP table with ~781K sites.

Supplementary Figure 1. outlines the analytic approach used to link patterns of metabolomic and genomic differentiation. First, the Generalized Cochran-Mantel-Haenszel (CMH) test [40] was used to identify SNPs that were significantly differentiated between the A-, C-, and S populations. This was done in R using the mantelhaentest function [41] tests were performed for each SNP in the data set, correcting for multiple comparisons was essential. To do this, the "plug in method" (Hastie et al. 2009) was used. Briefly, suppose we have M total hypothesis tests, and we let V be the number of false positives and S be the number of true positives, then the false discovery rate (FDR) is defined as $V/(V+S)$. A critical test statistic, C, can be chosen and the plug-in method computes the FDR for that critical point. $V+S$ is estimated to simply be the total number of test statistics from the M hypothesis tests that exceed C. V can in turn be estimated through permutation. Here that was done by essentially shuffling population labels then performing the CMH test at each polymorphic site in the permuted data set. The number of significant test statistics greater than all values of C was then recorded. This was repeated 100 times to ensure accurate estimates of V for values of C. Once V and $V+S$ were estimated, the FDR was calculated for each value of C. Using these results, we calibrated to a conservative FDR of 0.005 (i.e. significance threshold is the value of C that gives an FDR of 0.005).

After identifying SNPs that were differentiated across our treatments, a statistical learning approach called the "fused lasso additive model" or "FLAM" [42] was used to determine which of the genomic regions these SNPs represent best predict patterns of differentiation for our top candidate metabolites (Note: this was done on a per metabolite basis). Here the assumption is that these genomic regions are the ones most to be casually linked to relevant patterns of metabolite differentiation. Due to genetic linkage, it is not necessary to consider individual significant SNP. Instead, a list of the most significant SNP for every 50KB genomic window and these markers were used in the FLAM analysis. This mimics the implementation of FLAM as described in Muller et al[23] where the approach was validated using simulated and real data sets. It is also worth noting that this implementation strategy differs from the original Petersen et al[42] implementation in one major way. As described in Mueller et al.[23] , a single run of FLAM is limited to finding N casual SNPs, where N is the number of populations in the study. As a result, the order of potential predictor variables can impact results. Here and in Muller et

al. [23] this is accounted for using a permutation procedure where each FLAM analysis is run multiple times and the order of potential predictor variables is randomly shuffled. The final list of "best predictors" consists of genetic loci that occur at commonly identified across permutations. In this study, a total of 100 permutations were run for each metabolite, and the final list of best predictors consisted of loci that showed up in at least 50% of permutations of a given metabolite. Lastly, based on past validation, FLAM itself is particularly well suited for the task of linking genomic and metabolomic differentiation in an E&R context. A major consideration in these studies is the need to distinguish between parallel differences with some underlying relationship with the phenotype of interest, and differences simply due to chance. Muller et al. [23] suggests that FLAM has the power to distinguish between these two groups based on subtle differences among replicate populations within a treatment, which is not always possible with standard linear model approaches.

After determining which genomic regions best predict observed differences in candidate metabolites, a list of genes associated with each region was generated. Given that predictors are markers representing genomic regions, it cannot simply be assumed that the SNP markers are themselves casual. However, given that the SNPs used in the FLAM analyses are the most significant SNPs in their respective genomic region and the and the level of replication featured in this study, we would expect them to be relatively close to the true causative sites given theoretical work on the power of E&R studies to localize candidate genes [43]. Candidate genes were ultimate defined as those in 5KB windows around each SNP marker with this in mind. After generating a list of candidate genes associated with each candidate metabolite, Metascape [44] was used to perform gene ontology (GO) term enrichment analysis, protein network analysis, and Molecular Complex Detection (MCODE) Component analysis. All analyses were run using default settings. Cytoscape [45] was used to visualize results.

Results

Metabolomic results

To identify metabolic changes associated with life history, a detailed characterization of whole-body metabolism was undertaken in A, C, and S populations using both GC-MS and LC-MS.

Characterization of GC-MS Metabolomic Profiling of A, C, and S populations

GC-MS analysis returned 104 metabolites, 48 of which were significantly changed between populations. Principal component analysis (PCA) of metabolites revealed three distinct metabolic populations, with PC1 and PC2 explaining 47.8% and 14.1% of the variability, respectively (Figure 2A). The S and C populations were similar to one another and distant from A along PC1. Visualization of the top 25 most significant metabolites revealed a distinct inverse relationship between A and S metabolomic profiles, with the C profile appearing intermediate (Figure 2B). Enrichment analysis revealed propanoate metabolism, the pentose phosphate pathway, histidine metabolism, and valine, leucine, and isoleucine degradation as the top metabolic hits associated with life history (Figure 2C). Closer inspection of propanoate metabolism revealed that C and S populations had decreased 2-hydroxybuterate and glutamate, and increased beta-alanine compared to A (Figure 2Di). In the pentose phosphate pathway, fructose 6-phosphate, 6-phosphogluconate, and glucose 6-phosphate were elevated in the C and S populations compared to A (Figure 2Dii). Histidine was decreased in C and S populations compared to A (Figure Diii). Isoleucine, leucine, and valine were elevated in C and S compared to A (Figure Div). Components of the tricarboxylic acid (TCA) cycle, a critical pathway for aerobic respiration, were also

significantly associated with life history (Figure 2E). Citrate and isocitrate levels decreased in a stair-step manner from A to S. Interestingly, succinate, fumarate, and malate levels were not significantly different between A and C populations. In the S population, succinate levels were higher while fumarate and malate levels were lower than A and C populations.

Characterization of LC-MS Metabolomic Profiling of A, C, and S populations

To broaden our metabolic characterization of A, C, and S populations, we also used liquid chromatography mass spectrometry (LC-MS) for better characterization of non-volatile, thermally unstable, and higher molecular weight species. LC-MS analysis returned 38 metabolites, of which were significantly changed between populations. PCA revealed a much greater degree of separation among these metabolites than that uncovered by GC-MS. PC1 and PC2 explained 51.2% and 16.6% of variability, respectively (Figure 3A). There was less distinction between C and S populations along PC1, with most variability between the two explained by PC2. The A type population remained separate and distinct (Figure 3A). Accordingly, visualization of the top 25 significant metabolites revealed similar profiles among C and S populations and both appeared to have an inverse relationship to that of the A (Figure 3B). Fatty acid metabolism, beta-oxidation of long and short chain fatty acids, and methionine metabolism were the top enriched metabolic processes identified (Figure 3C). Short and long-chain species of carnitine were significantly lower in the A population compared to the C and S populations (Figure 3D). Similarly, methionine levels were significantly lower in the A population compared to C and S populations (Figure 3D). Key metabolic co-factors, FAD, NAD+, and NADP+, were also identified among the top 25 significant metabolites (Figure 3E). FAD and NAD+ levels among the C and S populations were elevated compared to A, while NADP+ levels were decreased.

Genome to Metabolome

Linking patterns of SNP variation between the A, C, and S populations to top candidate aging metabolites was a multistep process (see Supplementary Figure 1 for an overview). Conventional statistical methods were first used to identify significantly differentiated SNPs between the three groups. This resulted in a list of ~76K SNPs. However, as many of SNPs are likely neutral variants linked to causative sites, this list was reduced to the most significant SNP in every 50KB genomic window with at least 3 significant SNPs which yielded 1827 SNP markers. FLAM, a statistical learning approach, was then used to determine which markers best predicted patterns of differentiation across the A, C, and S populations for each of candidate metabolite. On average, FLAM identified ~11 SNP markers associated each candidate metabolite with values ranging from 3 to 36 hits per metabolite (Supplementary Table 1). Candidates were found across all major chromosome arms, and there was no clear relationship with chromosome length and number of candidates (Supplementary Table 2). Next, unique markers identified by FLAM were converted to a gene list based on genes present in 5-kb windows around each marker. This resulted in a list of 494 candidate genes (Supplementary Table 3 for a list of all genes and associated GO terms).

Metascape, which clusters terms into groups based on similarity, was used to identify enriched GO terms and pathways given our list of candidate genes. For enriched GO term clusters, there was a common trend of terms associated with development and growth, especially with regards to the nervous system (Figure 4A and Supplementary Table 4 for more details). We also identified several GO terms associated with metabolism (e.g. regulation of RNA metabolic processes and glycolytic processes).

Examining the relationships between these clusters, we found a pattern where nodes representing development, growth, and nervous system-related clusters largely grouped together while terms related to metabolic processes were more distal or even unconnected like in the case of the glycolytic processes cluster (Figure 4B). In addition to GO term enrichment, we also used Metascape to perform Protein Network and MCODE Component analyses. Here we again found similar trends with enriched terms from the Protein Network analysis being associated with neurological development, and terms from the MCODE component analysis being associated with metabolic processes (Table 1).

Discussion

Metabolomic characterization

One of the primary objectives of this study was to identify metabolomic profiles associated with longevity differences by incorporating metabolomic characterization into the E&R framework (GC-MS: Figure 2 and LC-MS: Figure 3). GC-MS is the gold standard for identifying compounds that are volatile and thermostable. At the same time, LC-MS is more adept at identifying non-volatile, thermally unstable, and more significant compounds [46]. Metabolites returned from GC-MS analysis were, to a degree, able to differentiate between A and S populations. For most metabolites returned, these two groups were inversely related, suggesting these metabolites are essential for understanding the driving mechanisms for starvation-resistant longevity (Figure 2B). However, care should be taken in the selectivity of these metabolites in association with longevity as the metabolic profile of the long-lived C population oscillated between that of the A and S populations. For example, 60% of TCA metabolites in the C population correlated with that of the short-lived A population (Figure 2E), whereas 67% of branched amino acids correlated with that of the long-lived S population (Figure 2Di-Figure 2Dv).

Deepening the metabolic characterization through LC-MS, we identified profiles that better-reflected lifespan (Figure 3). Metabolites returned from this analysis showed clustering of both long-lived populations (C and S) and greater separation from the short-lived population (A) (Figure 3A). Of the top 25 significant metabolites identified, ~70% correlated between C and S populations, suggesting that further analysis and exploration of these hits could yield greater insights into mechanisms underlying longevity differences in this system. For example, NAD^+ is an abundant coenzyme that acts as the oxidizing-reducing agent inside the cell [31,47,48]. NAD^+ plays a necessary role for hydrogen transfer in redox reactions through accepting hydride from metabolic processes, including glycolysis, the TCA cycle, and fatty acid oxidation (FAO), to form NADH. NADH serves as a central hydride donor to ATP synthesis through mitochondrial oxidative phosphorylation, leading to the generation of reactive oxygen species. Studies have shown that NAD^+ concentrations decline with age in worms, flies, mice, and humans [49–56]. For instance, decreasing NAD^+ levels in *C. elegans* results in a reduction in lifespan [57–59]. The depletion of NAD^+ is linked to mitochondrial dysfunction, mtDNA genomic instability, and deregulated nutrient sensing [53,60,61]. This suggests that the maintenance of higher NAD^+ levels is beneficial, potentially through maintenance of mitochondrial homeostasis.

NAD^+ is also a co-substrate for numerous enzymes including sirtuins[62,63], PARPs, CD38, CD157, CD73, and SARM1[64,65]. NAD^+ serves as a required substrate for the deacetylase activity of the sirtuin family of proteins (SIRTs). Mitochondrial SIRTs, SIRT3, -4, and -5, are connected in metabolism, mitochondrial fidelity, and cell stress. SIRT3, SIRT4, and SIRT5 are found primarily located in the mitochondria, and are implicated in several of the principal processes of this organelle. SIRT3 has been

the subject of serious investigation and is fundamentally a deacetylase estimated to function as a mitochondrial fidelity protein, with roles in mitochondrial substrate metabolism, protection against oxidative stress, and cell survival pathways [66–68]. Less is known about the useful targets of SIRT4, which has deacetylase, ADP-ribosylase, and a newly defined lipoamidase function [69–71]. SIRT5 modulates acyl modifications including succinylation, malonylation, and glutarylation in both mitochondrial and extra-mitochondrial compartments. However, the functional significance of SIRT5 in the regulation of many of its proposed target proteins remains to be discovered[72–74]. Sustained mitochondrial stress leads to mitochondrial unfolded protein stress response stress and ER stress[75–79]. Data suggest that the short-lived flies would have accelerated mtUPR and sustained levels of ER stress, whereas the C and S, longer lived groups would have a blunted effect. Our findings here suggest a potential elevation in NAD^+ consumption or decrease in synthesis in the A population compared to C and/or S (Figure 3E). Paired with a decreased presence of beta oxidation products and elevated TCA cycle substrates, our findings may suggest decreased mitochondrial function in the A population (Figure 2E and Figure 3). Thus, decreased NAD^+ availability may be a contributing and modifiable factor in age-related diseases. The mechanisms controlling its levels in aging, however, remain incompletely understood.

NAD^+, through SIRT1, can regulate the mTOR pathway [74]. Mechanistic target of rapamycin (mTOR), or TOR in *Drosophila*, is a key regulator of metabolism. mTOR is composed of two distinct kinase complexes, mTOR complex 1 (mTORC1) and 2 (mTORC2), which are characterized by the signature components Raptor and Rictor, respectively[74,80–82]. Branched chain amino acids (BCAA), including leucine, isoleucine, and valine, play a crucial role in the activation of the TOR pathway[83]. Mitochondrial stress can be regulated by Tor through activation of ATF4 to induce the integrated stress response and activation of GCN2 to increase amino acid availability [84–92], which we believe to be a protective response. We hypothesize that the lower BCAA levels in the A population (Figure 2Di) indicate that they are not able to effectively activate this pathway. We believe the elevated BCAAs in the long-lived populations (Figure 2iv) suggest greater TOR activation may be associated with longevity [93]. suggests that the A populations may have reduced levels of ATF4 activation which inhibits the cells' ability to regulate stress responses and cellular homeostatic process (Figure 2iv). Equally, under physiological glucose concentrations, mTORC1 is stimulated and leads to a number of proteins and enzymes being altered. These altered proteins are involved in anabolic processes, while restricting the autophagic process. Conversely, when glucose levels are low, mTORC1 is inhibited, in turn leading to the repression of numerous anabolic processes, sparing ATP and antioxidants. However, in the S populations, there are high levels of glucose and high levels of amino acids (Figure 2 and Figure 3). Therefore, suggesting that prolonged activation of mTOR leads to beneficial autophagy.

Genomics to metabolomics

In additional to characterizing metabolomic profiles, another major goal of this study was to explore the relationship between patterns of genetic differentiation and differences in candidate aging metabolites. In our two-step analysis, we sought to filter out uninformative sites and identify regions of the genome that best predict focal patterns of metabolite differentiation (Supplementary Figure 1). This ultimately yielded several hundred candidate genes, which suggests the underlying genetic architecture for observed metabolomic differences is highly polygenic (Supplementary Table 3). This is in keeping with findings from experimental evolution studies in sexually reproducing eukaryotes [21,94], and studies on the genetics of complex traits as a whole [95]. However, it is worth noting that focusing on genomic regions that best predict candidate metabolite differentiation narrowed candidates to hundreds of genes from

thousands if we had relied solely on SNP differentiation. As such, there appears to be some real value in combining different types of omic data when attempting to parse the genetic architecture of complex phenotypes like longevity.

Although longevity is a major axis of differentiation between the A, C, and S populations and our metabolomic analysis was structured to identify longevity associated metabolites, we did not find significant enrichment for genes directedly related to aging looking across our candidate genes (Figure 4A and Table 1). The failure to find an overrepresentation of aging genes is not uncommon in E&R studies focused on longevity [29,96]. This is perhaps due to a lack of segregating genetic variation at related loci as Fabian et al. (2018)[96] suggests, but the context dependent and complex nature of genotype to phenotype mapping may also play a role. That aside, looking across our list of candidates, we do indefinity three major canonical aging genes: *Tor*, *Atg7*, and *Kermit*. *Tor* and *Atg7* in particular have clear links to our metabolomic results; *Tor* through its many well defined aging related function [97–99] and *Atg7* though its effects on autophagic activity and relationship to mitochondrial function and turnover [100–102]. *kermit*, while less obviously connected to our metabolomic results, is a gene required for proper locomotive activity that when inhibited or overexpressed has strong negative impacts on longevity and survival [103,104]. So, while we do not observe an overrepresentation of canonical aging, the ones present seem likely to be biologically meaningful.

Given that our metabolomic results point to towards differences in mitochondrial function as a major force driving longevity differences between the A, C, and S populations, we queried our list of candidate genes for related term (see Supplementary Table 3 for genes and associated terms). As with aging, we do not find enrichment for genes associated with mitochondrial function. However, in addition to the previously mentioned *Atg7* and *Tor*, we find a number of genes directly or predicted to be associated with mitochondrial function (e.g. *ND-B14*, *COX6AL*, and *CG15386*) and maintenance (*mtDNA-helicase*, *mre11*, *CG11975*, and *larp)*. We also find a number of genes directly or predicited to be associated with endoplasmic reticulum stress (CG9934, CG8405, S2P), which as explained above can also impact mitochondrial function. So, while we do not find clear overrepresentation, we nevertheless again find concordance between specific candidate genes and our metabolomic characterization. And the lack of enrichment could once again be due to a general lack of relevant segregating genetic variation at loci related to mitochondrial function. Strong purifying selection at such loci should perhaps be expected given the importance of maintaining mitochondrial function (i.e. strong selection to purge deleterious variants).

Focusing on our enrichment results (Figure 4 and Table 1) and total list of candidate genes (Supplementary Table 3), we find clear overrepresentation of candidates associated with nervous system development and maintenance. With the exception of *Atg7* and *kermit*, the genes in these clusters are not explicitly defined as aging genes based on past studies. However, decades of research has established a clear link between longevity and nervous system development, regulation and function [105–110]. As such, we can reasonably interpret our findings as evidence that longevity differences between the A, C, and S populations are at least in part due to genes impacting the nervous system and these effects are reflected in the populations' metabolomic profiles. Also, while less pronounced, we find evidence of many candidate genes (Supplementary Table 3) and enriched clusters (Figure 4 and Table 1) associated with various forms of carbohydrate metabolism (e.g. glucose metabolism, glycolytic processes, amino sugar metabolic processes, and glutathione metabolic processes). This ties directly into the metabolomic results described above, and such process again have clear links to aging and longevity [111,112].

Lastly, we compared our general findings to those of Barter et al. [29] . Here individuals were sampled for RNA sequencing in the A and C populations at day 21 and a differential gene expression analysis was performed to identify genes associated with longevity and development time. As there was and is no RNA-seq data available for the S populations, the strategy used in the present study to isolate variation related to longevity is not possible. Enriched clusters based on these results differ in a number of ways from enriched clusters found in our genomic analysis (Supplementary Figure 2). For instance, while there is still a theme of development, it comes in the form of somatic muscle development and ecdysteroid metabolic processes instead of neurological development. Enrichment for genes associated ecdysteroid metabolic processes are particularly suggestive given ecdysteroid play an important role in guiding developmental transitions[113]. However, aspects of these results are still perhaps relevant to our metabolomic findings. Namely, we find enrichment for clusters associated with mitochondrial electron transport and sulfur metabolism, also related to mitochondrial function[114]. That being said, we acknowledge that comparisons between these sets of candidate genes are likely confounded by differences in data types and methodology.

Conclusions

Broadly speaking, our effort to incorporate metabolomics into the E&R framework has successfully generated insights into the factors underlying longevity differences in this experimental system. On its own, comparing and contrasting metabolomic profiles between our experimental populations revealed a number of key mechanisms shaping aging and longevity differences across the system. Our findings also suggest that, with regards to longevity, the relationship between the genome and the metabolome is complex and highly polygenic. However, the incorporating metabolomic results into our genomic analysis allowed us to narrow our focus to a tractable, and potentially more impactful, subset of candidate genes. The presence of key established aging genes and enrichment for functional gene clusters and networks related to longevity also suggests we are capturing some meaningful aspect of the genetic mechanisms underlying longevity differences in this system. Taken together, this study serves as a proof of concept that combining different types of omics data in the E&R context may have significant benefits when attempting to parse the physiological and genetic mechanism shaping complex phenotypes like longevity.

Acknowledgments: We would like to thank our undergraduate colleague Kit Neikirk and Taylor A Rodman for helping to optimize the analysis technique. We would also like to thank Dr. Michael R. Rose for access to the experimental populations maintained in his lab at UC Irvine. This work was supported by The Burroughs Wellcome Fund Award Career Awards at the Scientific Interface (CASI), UNCF/BMS EE Just Faculty Fund, the Ford Foundation, and NIH SRP subaward to #5R25HL106365-12 from the NIH PRIDE Program all awarded to AJH. MRM. is supported by the Burroughs Wellcome Fund and Howard Hughes Medical Institute via the PDEP and Hanna H. Gray Fellows Program. Dr. Melanie McReynolds, Dr. Edgar Garza Lopez, and Dr. Derrick J. Morton all provided consulting for this project.

Data Availability: Core data files (metabolite readings, SNP tables, results of statistical analyses, etc.) are available through Dryad (https://doi.org/10.5061/dryad.547d7wm92), and scripts used to carry out

analysis linking genomic results to metabolomic results are available through Github (https://github.com/ttbarter317/Fruit-fly-Genomics-and-Metabolomics).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Author Contributions:
Conceptualization MAP and AJH
Methodology: KRA, MAP, ZV, AJH
Validation: MAP, ZV, HKB, AGM, MRM
Formal Analysis: MAP, HKB, KRA
Investigation: MAP, ZV, HKB, AGM, MRM, AJH,
Resources: MAP, AJH
Writing—Original Draft Preparation, MAP, AGM, HKB, ZV, MR, AJH
Writing—Review & Editing MAP, KRA, ZV, HKB, AGM, DJM, MRM, AJH
Visualization, MAP, ZV, HKB, AGM
Supervision, MAP, AJH
Project Administration MAP, AJH
Funding Acquisition AJH
All authors have read and agreed to the published version of the manuscript.

References

1. Rose, M. R. *Evolutionary biology of aging*. (Oxford Univ. Press, 1994).
2. Stearns, S. C. Life history evolution: successes, limitations, and prospects. *Naturwissenschaften* **87**, 476–486 (2000).
3. Hughes, K. A. & Reynolds, R. M. EVOLUTIONARY AND MECHANISTIC THEORIES OF AGING. *Annu. Rev. Entomol.* **50**, 421–445 (2005).
4. Flatt, T. & Partridge, L. Horizons in the evolution of aging. *BMC Biol* **16**, 93 (2018).
5. Mueller, L. D. & Ayala, F. J. Trade-off between r-selection and K-selection in Drosophila populations. *Proceedings of the National Academy of Sciences* **78**, 1303–1305 (1981).
6. Luckinbill, L. S., Arking, R., Clare, M. J., Cirocco, W. C. & Buck, S. A. SELECTION FOR DELAYED SENESCENCE IN *DROSOPHILA MELANOGASTER*. *Evolution* **38**, 996–1003 (1984).
7. Rose, M. R. LABORATORY EVOLUTION OF POSTPONED SENESCENCE IN *DROSOPHILA MELANOGASTER*. *Evolution* **38**, 1004–1010 (1984).
8. Chippindale, A. K., Alipaz, J. A., Chen, H.-W. & Rose, M. R. EXPERIMENTAL EVOLUTION OF ACCELERATED DEVELOPMENT IN *DROSOPHILA*. 1. DEVELOPMENTAL SPEED AND LARVAL SURVIVAL. *Evolution* **51**, 1536–1551 (1997).
9. Burke, M. K. *et al.* Rapid divergence and convergence of life history in experimentally evolved *Drosophila melanogaster*. *Evolution* **70**, 2085–2098 (2016).
10. Rose, M. R., Passananti, H. B. & Matos, M. *Methuselah flies: a case study in the evoultion of aging*. (World Scientific Pub, 2004).

11. Service, P. M., Hutchinson, E. W., MacKinley, M. D. & Rose, M. R. Resistance to Environmental Stress in Drosophila melanogaster Selected for Postponed Senescence. *Physiological Zoology* **58**, 380–389 (1985).
12. Chippindale, A. K., Chu, T. J. F. & Rose, M. R. COMPLEX TRADE-OFFS AND THE EVOLUTION OF STARVATION RESISTANCE IN *DROSOPHILA MELANOGASTER*. *Evolution* **50**, 753–766 (1996).
13. Harshman, Hoffmann, & Clark. Selection for starvation resistance in *Drosophila melanogaster*: physiological correlates, enzyme activities and multiple stress responses. *Journal of Evolutionary Biology* **12**, 370–379 (1999).
14. Schwasinger-Schmidt, T. E., Kachman, S. D. & Harshman, L. G. Evolution of starvation resistance in Drosophila melanogaster: measurement of direct and correlated responses to artificial selection: Evolution of starvation resistance. *Journal of Evolutionary Biology* **25**, 378–387 (2012).
15. Kezos, J. N. *et al.* Genomics of Early Cardiac Dysfunction and Mortality in Obese *Drosophila melanogaster*. *Physiological and Biochemical Zoology* **92**, 591–611 (2019).
16. Graves, J. L., Toolson, E. C., Jeong, C., Vu, L. N. & Rose, M. R. Desiccation, Flight, Glycogen, and Postponed Senescence in Drosophila melanogaster. *Physiological Zoology* **65**, 268–286 (1992).
17. Gibbs, A. G., Chippindale, A. K. & Rose, M. R. PHYSIOLOGICAL MECHANISMS OF EVOLVED DESICCATION RESISTANCE IN *DROSOPHILA MELANOGASTER*. in *Methuselah Flies* 89–100 (WORLD SCIENTIFIC, 2004). doi:10.1142/9789812567222_0010.
18. Chippindale, A. K. *et al.* RESOURCE ACQUISITION AND THE EVOLUTION OF STRESS RESISTANCE IN *DROSOPHILA MELANOGASTER*. *Evolution* **52**, 1342–1352 (1998).
19. Djawdan, M., Chippindale, A. K., Rose, M. R. & Bradley, T. J. Metabolic Reserves and Evolved Stress Resistance in *Drosophila melanogaster*. *Physiological Zoology* **71**, 584–594 (1998).
20. Schlötterer, C., Kofler, R., Versace, E., Tobler, R. & Franssen, S. U. Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation. *Heredity* **114**, 431–440 (2015).

21. Long, A., Liti, G., Luptak, A. & Tenaillon, O. Elucidating the molecular architecture of adaptation via evolve and resequence experiments. *Nat Rev Genet* **16**, 567–582 (2015).

22. Jha, A. R. *et al.* Whole-Genome Resequencing of Experimental Populations Reveals Polygenic Basis of Egg-Size Variation in *Drosophila melanogaster*. *Mol Biol Evol* **32**, 2616–2632 (2015).

23. Mueller, L. D., Phillips, M. A., Barter, T. T., Greenspan, Z. S. & Rose, M. R. Genome-Wide Mapping of Gene–Phenotype Relationships in Experimentally Evolved Populations. *Molecular Biology and Evolution* **35**, 2085–2095 (2018).

24. Burke, M. K. *et al.* Genome-wide analysis of a long-term evolution experiment with Drosophila. *Nature* **467**, 587–590 (2010).

25. Carnes, M. U. *et al.* The Genomic Basis of Postponed Senescence in Drosophila melanogaster. *PLoS ONE* **10**, e0138569 (2015).

26. Graves, J. L. *et al.* Genomics of Parallel Experimental Evolution in *Drosophila*. *Mol Biol Evol* msw282 (2017) doi:10.1093/molbev/msw282.

27. Hardy, C. M. *et al.* Genome-Wide Analysis of Starvation-Selected Drosophila melanogaster—A Genetic Model of Obesity. *Molecular Biology and Evolution* **35**, 50–65 (2018).

28. Remolina, S. C., Chang, P. L., Leips, J., Nuzhdin, S. V. & Hughes, K. A. GENOMIC BASIS OF AGING AND LIFE-HISTORY EVOLUTION IN *DROSOPHILA MELANOGASTER*: GENOMICS OF LIFE-HISTORY EVOLUTION. *Evolution* **66**, 3390–3403 (2012).

29. Barter, T. T. *et al.* Drosophila transcriptomics with and without ageing. *Biogerontology* **20**, 699–710 (2019).

30. Wishart, D. S. Metabolomics for Investigating Physiological and Pathophysiological Processes. *Physiological Reviews* **99**, 1819–1875 (2019).

31. Srivastava, S. Emerging Insights into the Metabolic Alterations in Aging Using Metabolomics. *Metabolites* **9**, 301 (2019).

32. Cox, J. E., Thummel, C. S. & Tennessen, J. M. Metabolomic Studies in *Drosophila*. *Genetics* **206**, 1169–1185 (2017).

33. Hoffman, J. M. *et al.* Effects of age, sex, and genotype on high-sensitivity metabolomic profiles in the fruit fly, *D rosophila melanogaster* . *Aging Cell* **13**, 596–604 (2014).

34. Laye, M. J., Tran, V., Jones, D. P., Kapahi, P. & Promislow, D. E. L. The effects of age and dietary restriction on the tissue-specific metabolome of *D rosophila* . *Aging Cell* **14**, 797–808 (2015).

35. Li, B. *et al.* NOREVA: normalization and evaluation of MS-based metabolomics data. *Nucleic Acids Research* **45**, W162–W170 (2017).

36. Chong, J. *et al.* MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Research* **46**, W486–W494 (2018).

37. Chong, J. & Xia, J. MetaboAnalystR: an R package for flexible and reproducible analysis of metabolomics data. *Bioinformatics* **34**, 4313–4314 (2018).

38. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

39. Kofler, R. *et al.* PoPoolation: A Toolbox for Population Genetic Analysis of Next Generation Sequencing Data from Pooled Individuals. *PLoS ONE* **6**, e15925 (2011).

40. Landis, J. R., Heyman, E. R. & Koch, G. G. Average Partial Association in Three-Way Contingency Tables: A Review and Discussion of Alternative Tests. *International Statistical Review / Revue Internationale de Statistique* **46**, 237 (1978).

41. *https://www.R-project.org/.*

42. Petersen, A., Witten, D. & Simon, N. Fused Lasso Additive Model. *Journal of Computational and Graphical Statistics* **25**, 1005–1025 (2016).

43. Baldwin-Brown, J. G., Long, A. D. & Thornton, K. R. The Power to Detect Quantitative Trait Loci Using Resequenced, Experimentally Evolved Populations of Diploid, Sexual Organisms. *Molecular Biology and Evolution* **31**, 1040–1055 (2014).

44. Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10**, 1523 (2019).

45. Shannon, P. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research* **13**, 2498–2504 (2003).

46. Hanai, Y. *et al.* Rapid, simple, and clinically applicable high-performance liquid chromatography method for clinical determination of plasma colistin concentrations. *J Pharm Health Care Sci* **4**, 22 (2018).

47. Pollak, N., Dölle, C. & Ziegler, M. The power to reduce: pyridine nucleotides--small molecules with a multitude of functions. *Biochem J* **402**, 205–218 (2007).

48. Yang, Y. & Sauve, A. A. NAD + metabolism: Bioenergetics, signaling and manipulation for therapy. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1864**, 1787–1800 (2016).

49. Chini, C. C. S. *et al.* CD38 ecto-enzyme in immune cells is induced during aging and regulates NAD+ and NMN levels. *Nat Metab* **2**, 1284–1304 (2020).

50. Chini, C. *et al.* The NADase CD38 is induced by factors secreted from senescent cells providing a potential link between senescence and age-related cellular NAD+ decline. *Biochemical and Biophysical Research Communications* **513**, 486–493 (2019).

51. Chini, C. C. S., Tarragó, M. G. & Chini, E. N. NAD and the aging process: Role in life, death and everything in between. *Molecular and Cellular Endocrinology* **455**, 62–74 (2017).

52. McReynolds, M. R., Chellappa, K. & Baur, J. A. Age-related NAD+ decline. *Exp Gerontol* **134**, 110888 (2020).

53. Gomes, A. P. *et al.* Declining NAD+ Induces a Pseudohypoxic State Disrupting Nuclear-Mitochondrial Communication during Aging. *Cell* **155**, 1624–1638 (2013).

54. Yoshino, J., Baur, J. A. & Imai, S. NAD+ Intermediates: The Biology and Therapeutic Potential of NMN and NR. *Cell Metabolism* **27**, 513–528 (2018).

55. Zhang, H. *et al.* NAD+ repletion improves mitochondrial and stem cell function and enhances life span in mice. *Science* **352**, 1436–1443 (2016).

56. McReynolds, M. R. *et al.* NAD+ flux is maintained in aged mice despite lower tissue concentrations. *Cell Systems* S2405471221003380 (2021) doi:10.1016/j.cels.2021.09.001.

57. Lee, S.-J., Murphy, C. T. & Kenyon, C. Glucose Shortens the Life Span of C. elegans by Downregulating DAF-16/FOXO Activity and Aquaporin Gene Expression. *Cell Metabolism* **10**, 379–391 (2009).

58. Mouchiroud, L. *et al.* The NAD+/Sirtuin Pathway Modulates Longevity through Activation of Mitochondrial UPR and FOXO Signaling. *Cell* **154**, 430–441 (2013).

59. Chen, S. *et al.* The conserved NAD(H)-dependent corepressor CTBP-1 regulates Caenorhabditis elegans life span. *Proceedings of the National Academy of Sciences* **106**, 1496–1501 (2009).

60. Salvatori, I., Valle, C., Ferri, A. & Carrì, M. T. SIRT3 and mitochondrial metabolism in neurodegenerative diseases. *Neurochemistry International* **109**, 184–192 (2017).

61. Verdin, E. NAD+ in aging, metabolism, and neurodegeneration. *Science* **350**, 1208–1213 (2015).

62. S. Mohar, D. The Sirtuin System: The Holy Grail of Resveratrol? *J Clin Exp Cardiolog* **03**, (2012).

63. Lombard, D. B., Tishkoff, D. X. & Bao, J. Mitochondrial Sirtuins in the Regulation of Mitochondrial Activity and Metabolic Adaptation. in *Histone Deacetylases: the Biology and Clinical Implication* (eds. Yao, T.-P. & Seto, E.) vol. 206 163–188 (Springer Berlin Heidelberg, 2011).

64. Xie, N. *et al.* NAD+ metabolism: pathophysiologic mechanisms and therapeutic potential. *Sig Transduct Target Ther* **5**, 227 (2020).

65. Covarrubias, A. J., Perrone, R., Grozio, A. & Verdin, E. NAD+ metabolism and its roles in cellular processes during ageing. *Nat Rev Mol Cell Biol* **22**, 119–141 (2021).

66. Zhang, J. *et al.* Mitochondrial Sirtuin 3: New emerging biological function and therapeutic target. *Theranostics* **10**, 8315–8342 (2020).

67. Giralt, A. & Villarroya, F. SIRT3, a pivotal actor in mitochondrial functions: metabolism, cell death and aging. *Biochemical Journal* **444**, 1–10 (2012).

68. Kincaid, B. & Bossy-Wetzel, E. Forever young: SIRT3 a shield against mitochondrial meltdown, aging, and neurodegeneration. *Front. Aging Neurosci.* **5**, (2013).

69. Bonkowski, M. S. & Sinclair, D. A. Slowing ageing by design: the rise of NAD+ and sirtuin-activating compounds. *Nat Rev Mol Cell Biol* **17**, 679–690 (2016).

70. Li, X. & Kazgan, N. Mammalian Sirtuins and Energy Metabolism. *Int. J. Biol. Sci.* **7**, 575–587 (2011).

71. Chong, Z. Z., Shang, Y. C., Wang, S. & Maiese, K. SIRT1: new avenues of discovery for disorders of oxidative stress. *Expert Opinion on Therapeutic Targets* **16**, 167–178 (2012).

72. Rajabi, N. *et al.* Mechanism-Based Inhibitors of the Human Sirtuin 5 Deacylase: Structure-Activity Relationship, Biostructural, and Kinetic Insight. *Angew. Chem. Int. Ed.* **56**, 14836–14841 (2017).

73. Weng, H. *et al.* A New Vision of Mitochondrial Unfolded Protein Response to the Sirtuin Family. *CN* **18**, 613–623 (2020).

74. Sadria, M. & Layton, A. T. Interactions among mTORC, AMPK and SIRT: a computational model for cell energy balance and metabolism. *Cell Commun Signal* **19**, 57 (2021).

75. Jiang, D. *et al.* ATF4 Mediates Mitochondrial Unfolded Protein Response in Alveolar Epithelial Cells. *Am J Respir Cell Mol Biol* **63**, 478–489 (2020).

76. Schröder, M. & Kaufman, R. J. ER stress and the unfolded protein response. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **569**, 29–63 (2005).

77. Rutkowski, D. T. & Kaufman, R. J. A trip to the ER: coping with stress. *Trends in Cell Biology* **14**, 20–28 (2004).

78. Rutkowski, D. T. *et al.* UPR Pathways Combine to Prevent Hepatic Steatosis Caused by ER Stress-Mediated Suppression of Transcriptional Master Regulators. *Developmental Cell* **15**, 829–840 (2008).

79. Rutkowski, D. T. & Kaufman, R. J. All Roads Lead to ATF4. *Developmental Cell* **4**, 442–444 (2003).

80. Yoon, M.-S. The Emerging Role of Branched-Chain Amino Acids in Insulin Resistance and Metabolism. *Nutrients* **8**, E405 (2016).

81. Moberg, M. *et al.* Activation of mTORC1 by leucine is potentiated by branched-chain amino acids and even more so by essential amino acids following resistance exercise. *American Journal of Physiology-Cell Physiology* **310**, C874–C884 (2016).

82. Chen, K. *et al.* Different regulation of branched-chain amino acid on food intake by TOR signaling in Chinese perch (Siniperca chuatsi). *Aquaculture* **530**, 735792 (2021).

83. Zhang, S., Zeng, X., Ren, M., Mao, X. & Qiao, S. Novel metabolic and physiological functions of branched chain amino acids: a review. *J Animal Sci Biotechnol* **8**, 10 (2017).

84. Pakos Zebrucka, K. *et al.* The integrated stress response. *EMBO Rep* **17**, 1374–1395 (2016).

85. Pereira, R. O. *et al.* OPA 1 deficiency promotes secretion of FGF 21 from muscle that prevents obesity and insulin resistance. *EMBO J* **36**, 2126–2145 (2017).

86. Byles, V. *et al.* Hepatic mTORC1 signaling activates ATF4 as part of its metabolic response to feeding and insulin. *Molecular Metabolism* **53**, 101309 (2021).

87. Torrence, M. E. *et al.* The mTORC1-mediated activation of ATF4 promotes protein and glutathione synthesis downstream of growth signals. *eLife* **10**, e63326 (2021).

88. Costa-Mattioli, M. & Walter, P. The integrated stress response: From mechanism to disease. *Science* **368**, eaat5314 (2020).

89. Pereira, R. O. *et al.* OPA1 deletion in brown adipose tissue improves thermoregulation and systemic metabolism via FGF21. *eLife* **10**, e66519 (2021).

90. Averous, J. *et al.* GCN2 contributes to mTORC1 inhibition by leucine deprivation through an ATF4 independent mechanism. *Sci Rep* **6**, 27698 (2016).

91. Longchamp, A. *et al.* Amino Acid Restriction Triggers Angiogenesis via GCN2/ATF4 Regulation of VEGF and H2S Production. *Cell* **173**, 117-129.e14 (2018).

92. Ye, J. *et al.* The GCN2-ATF4 pathway is critical for tumour cell survival and proliferation in response to nutrient deprivation. *EMBO J* **29**, 2082–2096 (2010).

93. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The Hallmarks of Aging. *Cell* **153**, 1194–1217 (2013).

94. Barghi, N., Hermisson, J. & Schlötterer, C. Polygenic adaptation: a unifying framework to understand positive selection. *Nat Rev Genet* **21**, 769–781 (2020).

95. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177–1186 (2017).

96. Fabian, D. K. *et al.* Evolution of longevity improves immunity in *Drosophila*. *Evolution Letters* **2**, 567–579 (2018).

97. Kapahi, P. *et al.* Regulation of Lifespan in Drosophila by Modulation of Genes in the TOR Signaling Pathway. *Current Biology* **14**, 885–890 (2004).

98. Evans, D. S., Kapahi, P., Hsueh, W.-C. & Kockel, L. TOR signaling never gets old: Aging, longevity and TORC1 activity. *Ageing Research Reviews* **10**, 225–237 (2011).

99. Partridge, L., Alic, N., Bjedov, I. & Piper, M. D. W. Ageing in Drosophila: The role of the insulin/Igf and TOR signalling network. *Experimental Gerontology* **46**, 376–381 (2011).

100. Nakamura, S. & Yoshimori, T. Autophagy and Longevity. *Mol Cells* **41**, 65–72 (2018).

101. Vincow, E. S. *et al.* Autophagy accounts for approximately one-third of mitochondrial protein turnover and is protein selective. *Autophagy* **15**, 1592–1605 (2019).

102. Vincow, E. S. *et al.* The PINK1-Parkin pathway promotes both mitophagy and selective respiratory chain turnover in vivo. *Proceedings of the National Academy of Sciences* **110**, 6400–6405 (2013).

103. Kim, J., Lee, S., Ko, S. & Kim-Ha, J. dGIPC is required for the locomotive activity and longevity in Drosophila. *Biochemical and Biophysical Research Communications* **402**, 565–570 (2010).

104. Pereira, G. B., Valente, V., de Queiroz, M. S., Vianna, M. C. B. & Paçó-Larson, M. L. Overexpression of kermit/dGIPC is associated with lethality in Drosophila melanogaster. *Braz J Med Biol Res* **44**, 283–290 (2011).

105. Buchanan, R. L. & Benzer, S. Defective Glia in the Drosophila Brain Degeneration Mutant drop-dead. *Neuron* **10**, 839–850 (1993).

106. Kretzschmar, D., Hasan, G., Sharma, S., Heisenberg, M. & Benzer, S. The *Swiss Cheese* Mutant Causes Glial Hyperwrapping and Brain Degeneration in *Drosophila*. *J. Neurosci.* **17**, 7425–7432 (1997).

107. Lu, B. & Vogel, H. *Drosophila* Models of Neurodegenerative Diseases. *Annu. Rev. Pathol. Mech. Dis.* **4**, 315–342 (2009).

108. Alcedo, J., Flatt, T. & Pasyukova, E. G. Neuronal Inputs and Outputs of Aging and Longevity. *Front. Genet.* **4**, (2013).

109. Haddadi, M. *et al.* Brain aging, memory impairment and oxidative stress: A study in Drosophila melanogaster. *Behavioural Brain Research* **259**, 60–69 (2014).

110. Omelyanchuk, L. V., Shaposhnikov, M. V. & Moskalev, A. A. Drosophila nervous system as a target of aging and anti-aging interventions. *Front. Genet.* **6**, (2015).

111. Haselton, A. T. & Fridell, Y.-W. C. Adult Drosophila melanogaster as a model for the study of glucose homeostasis. *Aging* **2**, 523–526 (2010).

112. Ravera, S. *et al.* Discrete Changes in Glucose Metabolism Define Aging. *Sci Rep* **9**, 10347 (2019).

113. Niwa, R. & Niwa, Y. S. Enzymes for ecdysteroid biosynthesis: their biological functions in insects and beyond. *Bioscience, Biotechnology, and Biochemistry* **78**, 1283–1292 (2014).

114. Francioso, A., Baseggio Conrado, A., Mosca, L. & Fontana, M. Chemistry and Biochemistry of Sulfur Natural Compounds: Key Intermediates of Metabolism and Redox Biology. *Oxidative Medicine and Cellular Longevity* **2020**, 1–27 (2020).

Figures

Figure 1. Lifespan comparison for the A, C, and S populations (A), and PCA plot showing how populations cluster based differences and similarities in SNP frequencies across the genome (B). Plots were generated using data from Burke et al. (2016), Graves et al. (2017), and Kezos et al. (2019).

Figure 2. Figure 2: GC-MS Metabolomic Profiling of A, C, and S populations. PCA plot of A, C, and S metabolite profiles (A). Heatmap comparison of top 25 changed low energy state metabolites among A, C, and S populations (B). Top 25 enriched metabolic pathways (C). Quantification of metabolites from

the top 4 enriched metabolic pathways among A-, C-, and S populations (D). Key components of the TCA cycle (E) were among the top 25 changed metabolites. Mean $\pm$ SD; n = 6; *p-value $\leq$ 0.05.

Figure 3: LC-MS Metabolomic Profiling of A, C, and S populations. PCA plot of A, C, and S populations high energy state metabolite profiles (A). Heatmap comparison of top 25 changed high energy state metabolites among A, C, and S populations (B). Top 25 enriched metabolic pathways (C). Quantification of metabolites in select enriched pathways from A, C, and S populations (D). Select important metabolic co-factors (E). Mean $\pm$ SD; n = 6 runs, 5 flies/run/group; *p-value $\leq$ 0.05.

Figure 4. Heatmap of top 20 enriched GO clusters (A), and a network showing relationships between clusters (B).

Tables

Table 1. Results of protein-protein enrichment analysis and MCODE component analysis.

Supplementary Figures

Supplementary Figure 1. A flowchart outlining methods to identify candidate genes associated with differences in candidate metabolites.

Supplementary Figure 2. Top enriched GO clusters based on the 629 differentially expressed genes identified in Barter et al. (2019) comparing samples from the A and C populations at day 21.

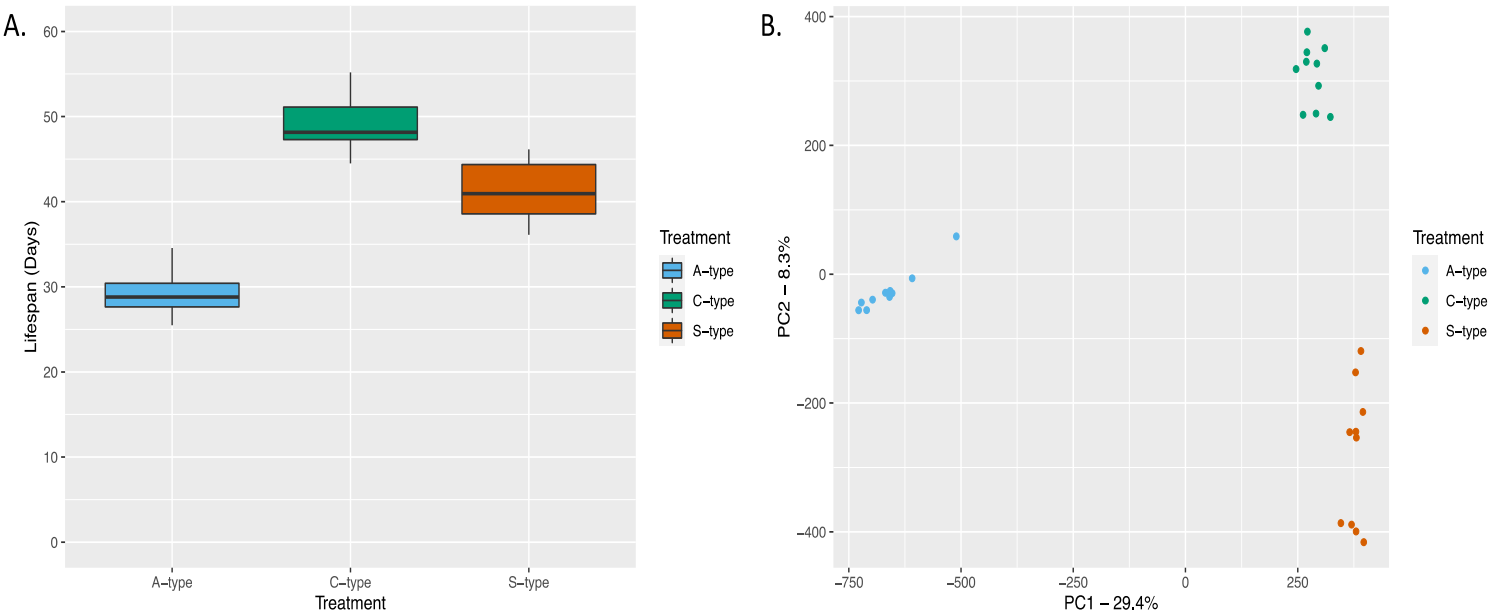


Figure 1. Lifespan comparison for the A, C, and S populations (A), and PCA plot showing how populations cluster-based differences and similarities in SNP frequencies across the genome (B). Plots were generated using data from Burke et al. (2016), Graves et al. (2017), and Kezos et al. (2019).

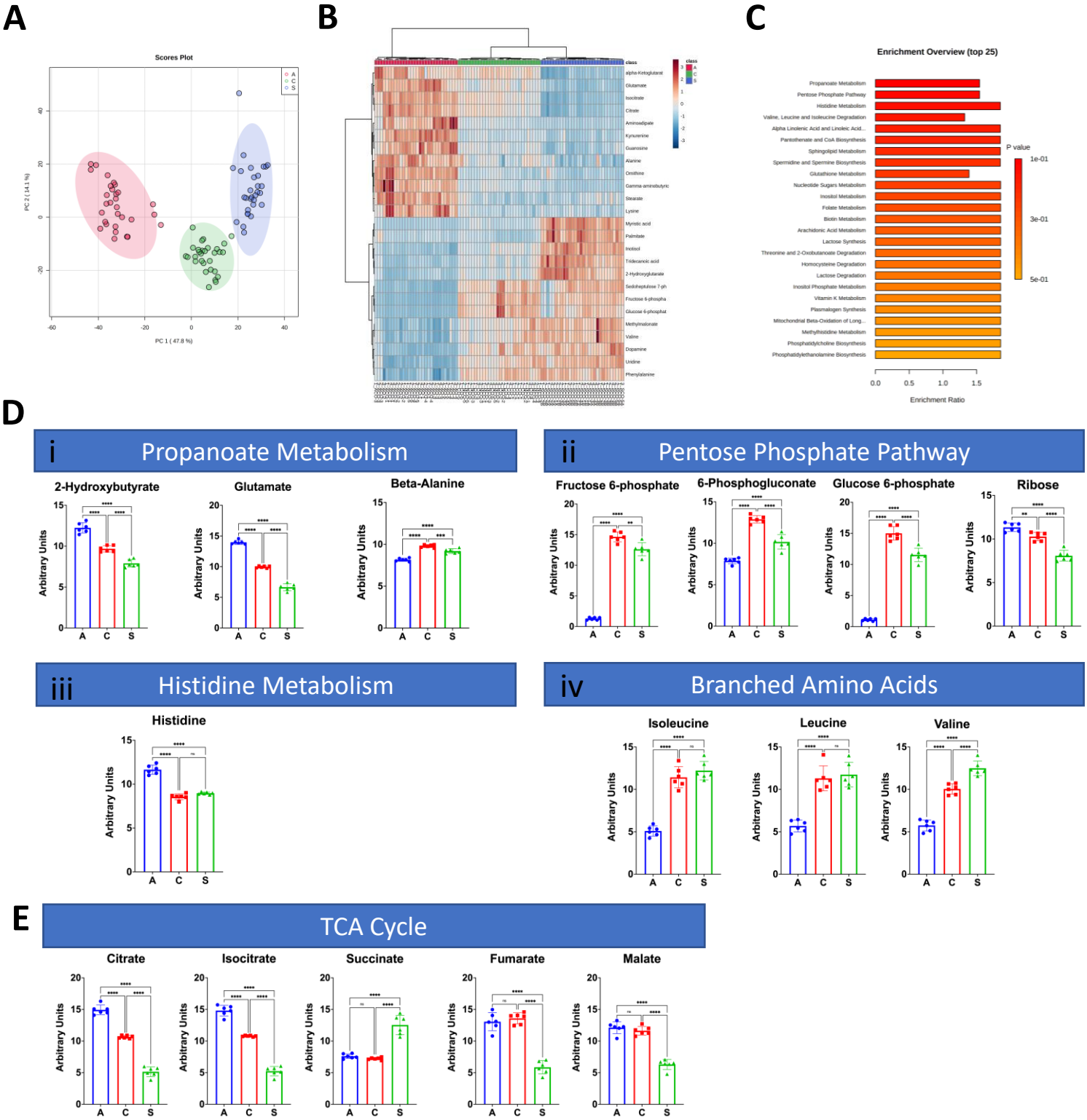


Figure 2. GC-MS Metabolomic Profiling of A-, C-, and S-type whole *D. melanogaster*. PCA plot of A-, C-, and S-type metabolite profiles (A). Heatmap comparison of top 25 changed low energy state metabolites among A-, C-, and S-type populations (B). Top 25 enriched metabolic pathways (C). Quantification of metabolites from the top 4 enriched metabolic pathways among A-, C-, and S-type populations (D). Key components of the TCA cycle (E) were among the top 25 changed metabolites. Mean $\pm$ SD; n = 6; *p-value $\leq$ 0.05.

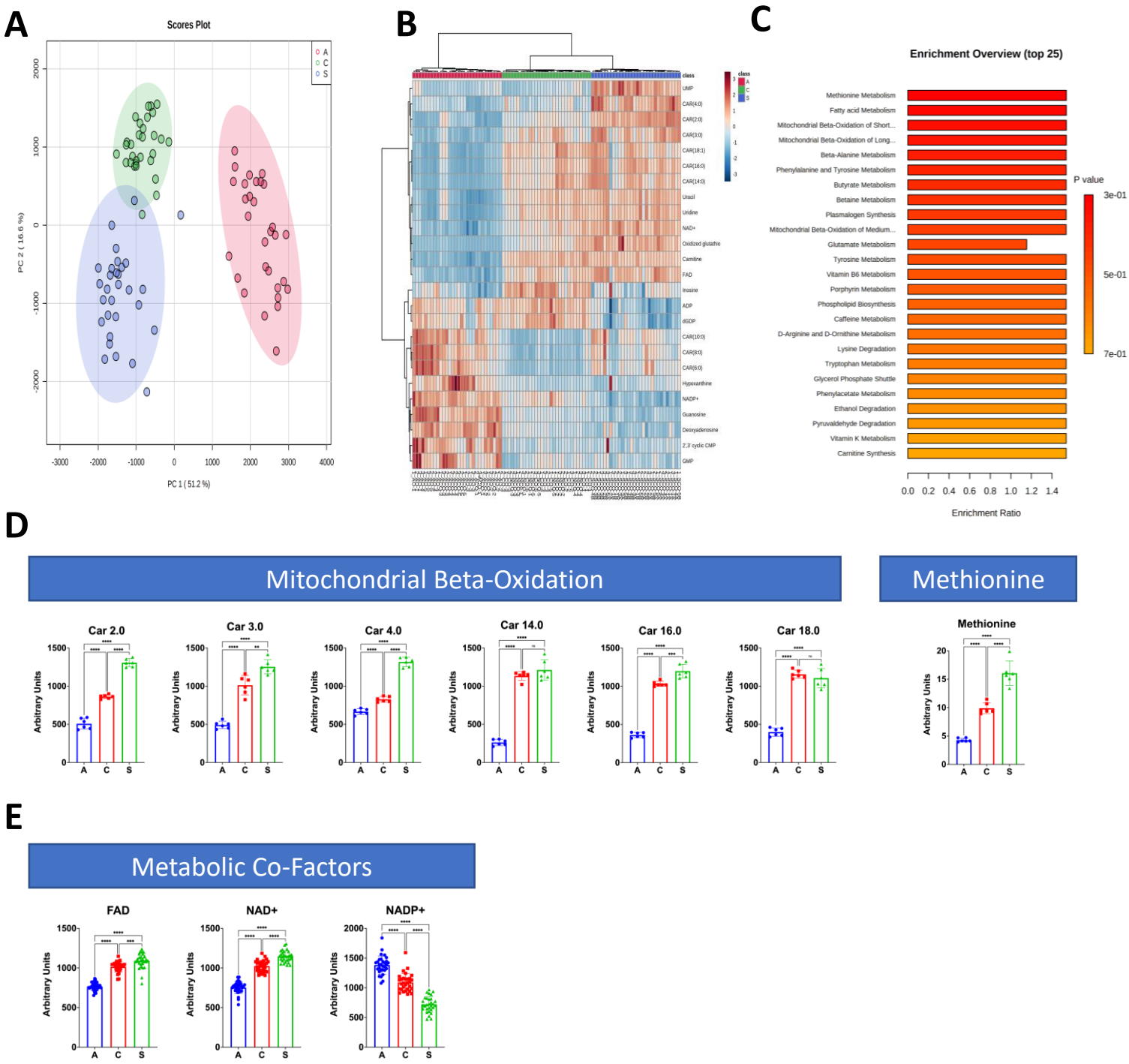
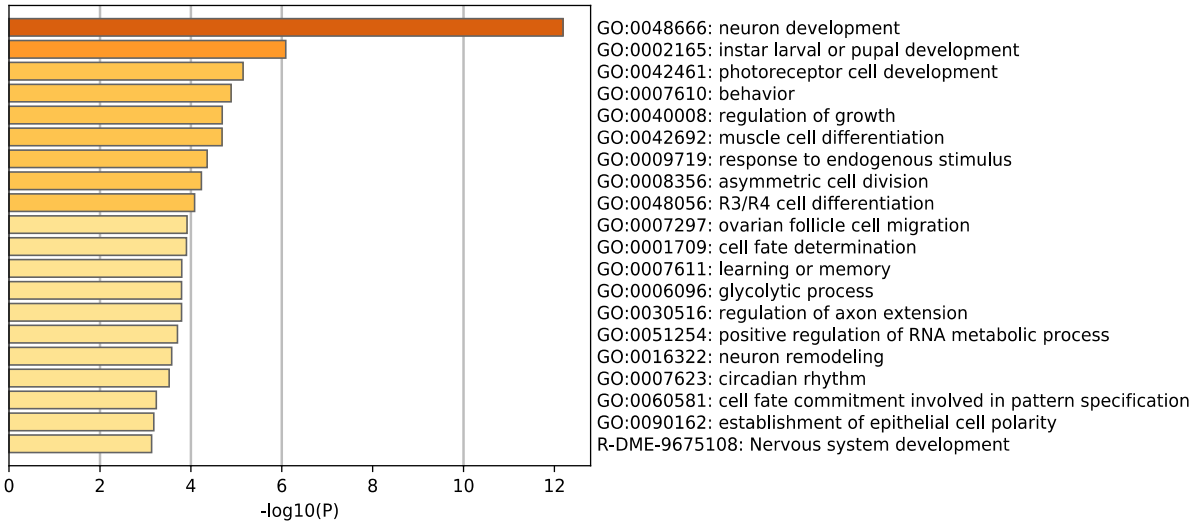


Figure 3. LC-MS Metabolomic Profiling of A-, C-, and S-type whole *D. melanogaster*. PCA plot of A-, C-, and S-type high energy state metabolite profiles (A). Heatmap comparison of top 25 changed high energy state metabolites among A-, C-, and S-type populations (B). Top 25 enriched metabolic pathways (C). Quantification of metabolites in select enriched pathways from A-, C-, and S-type populations (D). Select important metabolic co-factors (E). Mean $\pm$ SD; n = 6 runs, 5 flies/run/group; *p-value $\leq$ 0.05.

A.



B.

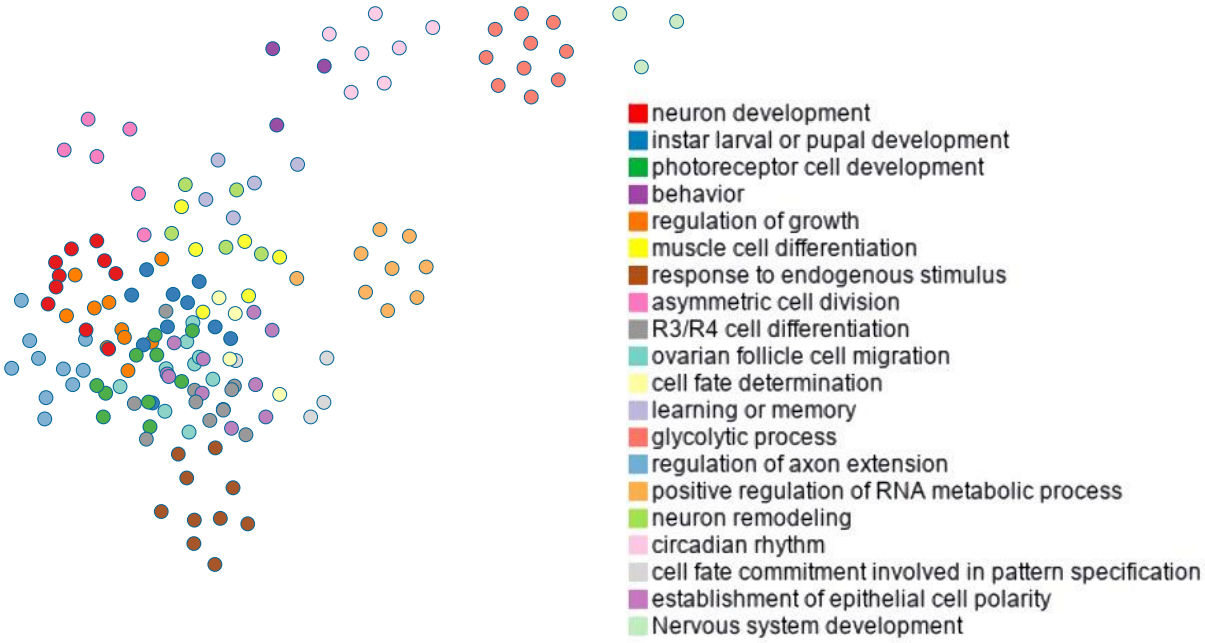


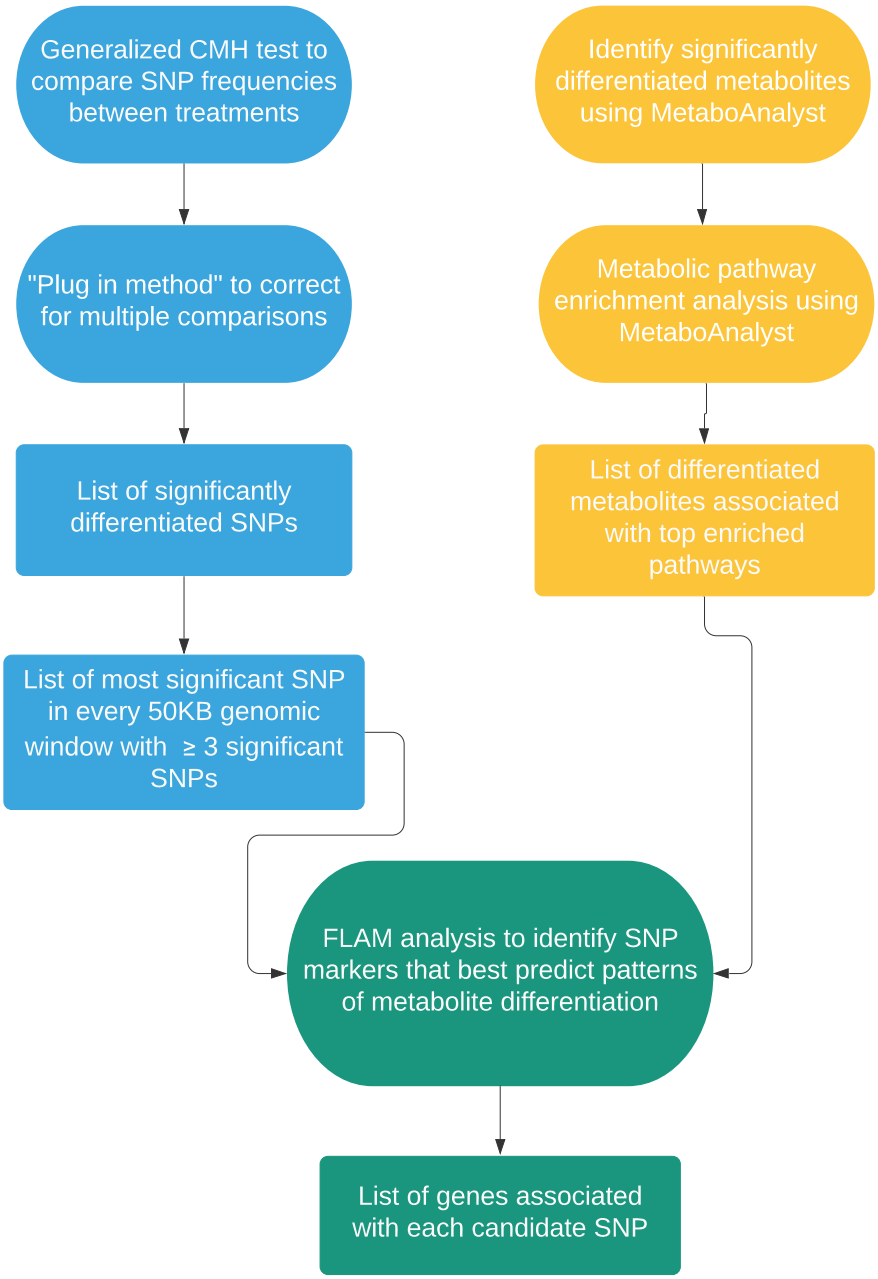
Figure 4. Heatmap of top 20 enriched GO clusters (A), and a network showing relationships between clusters (B).

Tables

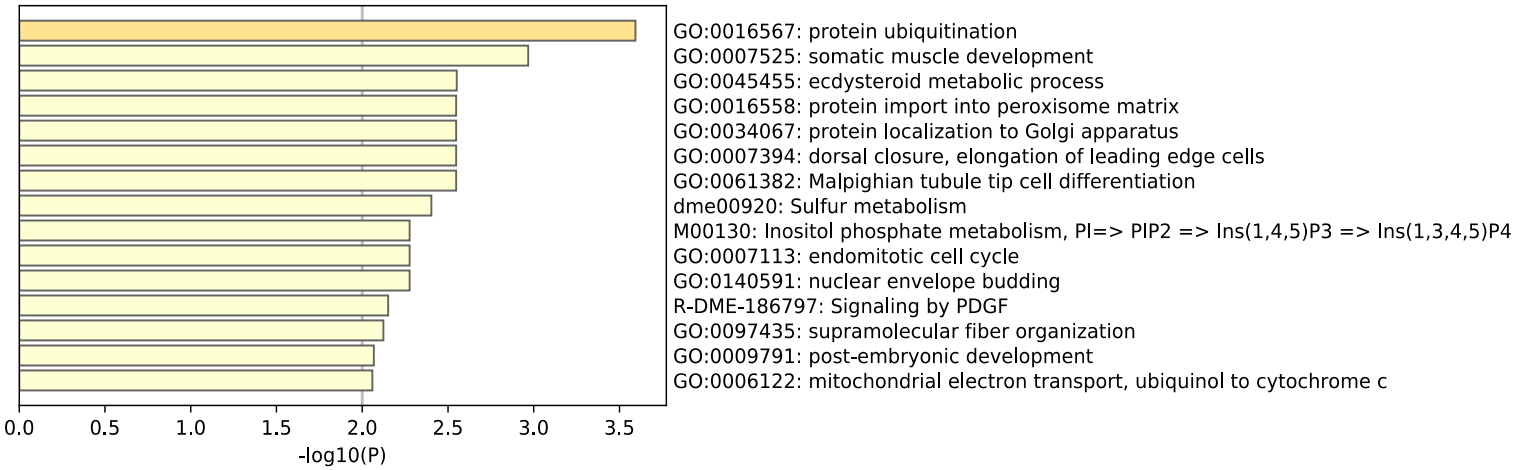
Table 1. Protein Network + MCODE results

	GO	Description	log(pvalue)	
Protein-protein interaction network	GO:0030182	neuron differentiation	-16.5	
	GO:0048666	neuron development	-15.5	
	GO:0048667	cell morphogenesis involved in neuron differentiation	-12.3	
				MCODE Cluster
MCODE networks	R-DME-70171	Glycolysis	-6.6	MCODE_1
	R-DME-70326	Glucose metabolism	-6	MCODE_1
	R-DME-6798695	Neutrophil degranulation	-5.8	MCODE_1
	R-DME-5620924	Intraflagellar transport	-8.8	MCODE_2
	R-DME-5617833	Cilium Assembly	-7.7	MCODE_2
	GO:0007018	microtubule-based movement	-7.7	MCODE_2
	R-DME-983168	Antigen processing: Ubiquitination & Proteasome degradation	-4.7	MCODE_4
	R-DME-983169	Class I MHC mediated antigen processing & presentation	-4.7	MCODE_4
	GO:0043161	proteasome-mediated ubiquitin-dependent protein catabolic process	-4.5	MCODE_4
	R-DME-72312	rRNA processing	-5.9	MCODE_5
	R-DME-8868773	rRNA processing in the nucleus and cytosol	-5.9	MCODE_5
	R-DME-6791226	Major pathway of rRNA processing in the nucleolus and cytosol	-5.9	MCODE_5
	R-DME-156827	L13a-mediated translational silencing of Ceruloplasmin expression	-5.7	MCODE_6
	R-DME-72613	Eukaryotic Translation Initiation	-5.6	MCODE_6
	R-DME-72737	Cap-dependent Translation Initiation	-5.6	MCODE_6
	R-DME-8856828	Clathrin-mediated endocytosis	-8.4	MCODE_7
	GO:0072583	clathrin-dependent endocytosis	-7.4	MCODE_7
	GO:0006898	receptor-mediated endocytosis	-6.4	MCODE_7
	dme03020	RNA polymerase	-7.9	MCODE_8
	ko03020	RNA polymerase	-7.9	MCODE_8
	ko00240	Pyrimidine metabolism	-6.5	MCODE_8
	R-DME-8877627	Vitamin E	-8.4	MCODE_9
	R-DME-6806667	Metabolism of fat-soluble vitamins	-8	MCODE_9
	R-DME-196854	Metabolism of vitamins and cofactors	-5.8	MCODE_9
	GO:0045944	positive regulation of transcription by RNA polymerase II	-4.7	MCODE_10
	GO:1902680	positive regulation of RNA biosynthetic process	-4.3	MCODE_10
	GO:1903508	positive regulation of nucleic acid-templated transcription	-4.3	MCODE_10

Supplemental



Supplementary Figure 1. A flowchart outlining methods to identify candidate genes associated with differences in candidate metabolites.



Supplementary Figure 2. Top enriched GO clusters based on the 629 differentially expressed genes identified in Barter et al. (2019) comparing samples from the A and C populations at day 21.