TITLE

The Molecular Landscape of the Aging Mouse Liver

AUTHOR LIST

Evan G. Williams[1], Niklas Pfister[2], Suheeta Roy[3], Cyril Statzer[4], Jesse Ingels[3], Casey Bohl[3], Moaraj Hasan[5], Jelena Čuklina[5]

Peter Bühlmann[6], Nicola Zamboni[5], Lu Lu[3], Collin Y. Ewald[4], Robert W. Williams[3], Ruedi Aebersold[5,7]

FACILITIES

[1]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

[2]Department of Mathematical Sciences, University of Copenhagen, Denmark

[3]Department of Genetics, Genomics and Informatics, University of Tennessee Health Science Center, Memphis, Tennessee, United States

[4]Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland

[5]Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland

[6]Department of Statistics, ETH Zurich, Zurich, Switzerland

[7]Faculty of Science, University of Zurich, Zurich, Switzerland

MANUSCRIPT STATISTICS

Length:	~8000 words [main text] ~1200 words [main figure legends]
Figures:	5 + 5 supplementary

ABSTRACT

Systems biology approaches often use networks of gene expression and metabolite data to identify regulatory factors and pathways connected with phenotypic variance. Separating upstream causal mechanisms, downstream biomarkers, and incidental correlations remains a significant challenge, yet it is essential for designing mechanistic experiments. To address this, we first designed a population following 2157 individual mice from 89 isogenic strains of BXD mice across their lifespans to identify molecular interactions between genotype, environment, age (GxExA) and metabolic fitness. Each strain was separated into two cohorts, fed low fat (6% cal/fat) or high fat (60% cal/fat) diets. One-third of individuals (662) were sacrificed at ~6, 12, 18, or 24 months-of-age, with the remainder monitored until natural death. Transcriptome, proteome, and metabolome profiles were generated from liver samples. These multi-omic measurements were deconvolved into metabolic networks, where we observed varying network connectivity as a function of GxExA. The multiple independent study variables permitted causal inference analysis for the network variants using stability selection. This calculates the strength and directionality of the interactions between molecular measurements and metabolic networks as a function of age, diet, and genotype, and assigns each gene a score for its relative position to the target pathway. At 1% FDR, 94% of novel connections were stable across age and diet, such as the connection between *Rdh11* with cholesterol biosynthesis and *Mut* with mitochondrial translation. 6% of discovered candidate genes were unstable, indicating a clear causal relationship between the segregating independent variable, the gene, and the pathway. For instance, age drives variation in proteasomal genes (e.g. *Psmb3, Psmb4*), which in turn drive changes in the mitochondrial ribosome. Conversely, COX7A2L malformation drives variation in OXPHOS genes, but both are downstream of changes in mitochondrial translation. Finally, we examined all data for connections with the longevity and known longevity-related pathways, identifying several dozen novel candidate genes. Thus, top two candidates, *Ctsd* and *St7*, had their orthologs knocked down in *C. elegans* and were novelly found to reduce longevity both in wildtype and in mutant long-lived strains.

KEYWORDS

Aging; multi-omics; genetic reference population; liver; time course; gene by environment (GxE); causal inference; network biology; multivariate analysis

INTRODUCTION

Aging is a dynamic and diverse process driven over a lifetime of interactions between genetic variants, environmental factors, and stochastic processes. Despite its complexity, lifespan is a heritable trait, with genotype explaining 30–50% of its variation across laboratory mice [1, 2] and ~25% in humans [1]. Because age is the most prominent "risk factor" for a wide range of metabolic diseases, such as diabetes, heart disease, neurodegeneration, and most cancers [3, 4], it is critical to understand the cellular and molecular changes that accumulate and lead to these diseases. Cells and tissues display common perturbations across age, such as a diminished capacity for proteostasis [5, 6] and the accumulation of mitochondrial defects [7]. These common endpoints are clearly recognized, but there is a dramatic diversity of the timelines connecting these mechanisms to chronological age ("lifespan") and biological age ("healthspan") across individuals, let alone across organisms. It is now possible to measure biomarkers of biological age, such as by DNA methylation signatures [8]. However, these biomarkers do not predict lifespan, and it is unclear whether interventions that directly affect dynamics of aging biomarkers, such as methylation signatures, would causally improve either lifespan or healthspan [9, 10]. Given the inherent challenges of obtaining tissue biopsies longitudinally in human clinical cohorts, model organisms have unique applications for the biomolecular analysis of aging. In particular, isogenic cohorts have been favored because tissue biopsies can be consistently collected across multiple times and environments. Large RNA-seq studies, such as by the *Tabula Muris* consortium [11], have established a baseline resource for how gene expression changes across tissues. Further research is necessary to indicate how these changes occur across time, whether results are congruent between mRNA and protein, the relationships between molecular expression and phenotypes, the causality of these relationships, and how these vary by genetic and environmental interactions (GxE), compounded with age (GxExA).

In this study, we examine the relationships between GxExA, gene expression, metabolic networks, and longevity phenotypes by following a population of 2157 mice from 89 isogenic strains of the BXD family of mice on two different diets: a normal chow diet (CD), with 6% calories from fat, and a high fat diet (HFD) with 60% calories from fat. 1495 individuals lived out their natural lifespans, while 662 were sacrificed at four target timepoints—6, 12, 18, and 24 months-of-age—to collect livers across

strain, age, and diet. To understand the molecular effects of GxExA, we generated transcriptomic, proteomic, and metabolomic profiles of all cohorts' livers. These multi-omics datasets have been combined with other phenotypes collected across the population, namely blood biomarkers, organ weights, longitudinal body weight, and lifespan. These samples provided a solid platform for detecting, evaluating, and testing how biomolecular processes vary as a function of GxExA, and to what extent each tier of molecular data can reveal architectural changes in metabolic networks and consequent phenotypes. We examined 25 molecular pathways in-depth which are known to vary according to diet or age. For instance, the CYP450 gene family is downregulated in HFD-fed individuals corresponding to a reduction in plant-based dietary xenobiotics [12], while oxidative phosphorylation (OXPHOS) subunits are downregulated in aged individuals [13]. However, these general patterns give way to distinct connectivities in gene expression networks when magnified: a strong edge connecting two nodes (genes) in young animals may disappear in old animals. Current machine learning approaches to data-driven network discovery tend to search for consensus across input data for this reason [14]. However, differences in gene networks as a consequence of independent variables can also indicate the causal reason behind pathway variance. In this case, unstable edges need to be retained, examined, and ranked.

The multiple independent study variables in this study (diet, age, genotype) and molecular profiles (mRNA, protein) allowed us to broadly apply a causal inference method that we recently developed [15, 16] called stabilized regression. This starts with supervised learning, using as input a canonical gene set (e.g. a pathway) hypothesized to respond to diet or age. Next, the entire molecular dataset is scanned to identify other molecules that covary with the target network as a function of the multiple independent variables. These correlating molecules are assigned a regression coefficient for each independent variable selected, followed by a stability score that estimates whether the novel target is upstream of the canonical pathway (i.e. causal), downstream (i.e. a biomarker), or ambiguous (e.g. a connection that is not affected by the secondary independent variable). By this means, we identified 4484 associations from 1102 distinct genes which are linked to 25 metabolic pathways but which not among the canonical genes for these pathways. Roughly 12% of these new interactions matched in both proteomics and mRNA data, with the remainder unique to protein (39%) or mRNA (49%). Six percent of novel causal

associations (i.e. 289s) were clearly delineated as upstream or downstream to the set of input metabolic pathways and networks that vary as a function of GxExA.

Finally, we filtered genes discovered through all of our analyses to uncover individual candidate genes related to longevity. Two genes were selected for validation in *C. elegans*: cathepsin D (*Ctsd*) has a consistent decrease in expression due to age and is a regulator of proteostasis, and suppressor of tumorigenicity 7 (*St7*) which was the strongest biomarker of biological age, but which did not have expression variation as a function of age. Together, this study examines how we may take advantage of increased complexity in study designs, advances in measurement technologies, and larger sample sizes in order to developed refined targets for molecular validation and make a leap towards causality in network analysis.

RESULTS

Clinical Analysis of Lifespan as a Function of Genotype and Diet

In this study, we followed 2157 mice from 89 strains of the BXD family across their natural range of lifespan. These individuals—96% female—were placed in the colony around 5 months of age, after which cohorts were evenly segregated into the dietary cohorts fed a standard chow diet (CD; Harlan Teklad 2018, 6% calories from fat) or high fat diet (HFD; Harlan 06414, 60% calories from fat). For 58 strains, selected pairs of individuals from each cohort (i.e. strain and diet matched) were sacrificed at 6, 12, 18, and when possible 24 months of age in order to collect a final biobank from 662 individuals with tissues across time, diet, and genotype (**Figure 1A**). The liver was selected as the initial organ of interest due to its central role in metabolism and the wide range of liver-related clinical and molecular phenotypes known to vary across the BXDs as a function of diet, sex, and genotype [17, 18]. The liver was pulverized in liquid nitrogen and aliquoted for transcriptomic, proteomic, and metabolomic preparations which were executed in parallel (**Figure 1B**).

Earlier studies have demonstrated that lifespan across the BXD family of mice varies by ~3-fold—from 11 to 32 months [19-21]. These differences among strains are consistent across studies, even twenty years apart (rho = 0.61, **Figure S1A** [1, 20, 21]). Here, 1418 individuals lived out their natural lifespans and were used for the calculation of longevity, permitting comparisons across diets (**Figure 1C**) and strain (**Figure 1D**). 47 strains had sufficient data in both diets for this analysis (≥ 6 natural deaths in both cohorts). Genetic variation across the population explained 73% of

variation in mean lifespan, versus 12% by diet and 15% by gene–diet interactions, with an unexplained residual of only 0.5% (**Figure S1B**). Mean strain lifespans vary from 314±37 days (mean ± SEM for BXD13, n = 20) to 820±27 days (BXD91, n = 16) (**Figure 1D, Figure S1C**). While HFD feeding causes a mean 10% decrease in longevity, this varied across strains: BXD9's lifespan is unaffected, while BXD65 lives nearly an additional year longer on CD ($p < 3e\text{-}6$, **Figure S1D**). HFD leads to a significant decrease in lifespan in 40% of the strains ($p \leq 0.05$, Kaplan–Meier curves), and 64% have at least a tendency to live shorter on HFD ($p \leq 0.10$). The effect is relatively consistent effect across strains, with the mean lifespan of strains between diets correlating at r = 0.60 (**Figure S1E**). Although diet has a relatively modest effect on longevity, it has a substantial impact on weight (**Figure 1E–F**), explaining 40% of variance (versus 23% explained by genotype, 5% by genotype-diet interaction, and a 23% unexplained residual, **Figure S1B**). By 18 months of age, individuals had on average a 78% increase in body mass. 89% of strains gained weight significantly upon HFD feeding ($p < 0.05$, AUC). As with lifespans, the effect of HFD on body weight varied depending on genetic background: BXD16s gained the least with an average increase of 11%, while BXD100s gained an average of 133% (**Figure S1F–G**).

Two individuals from each strain and diet were pre-selected for sacrifice and tissue collection at 6, 12, 18, or (when possible) 24 months of age, eventually resulting in 662 collected individuals (**Figure S1H**). In addition to body weight and longevity, we also measured 18 plasma metabolites commonly used in clinical settings (e.g. cholesterol, iron, glucose, and alkaline phosphatase (ALPL) levels; **Supplemental Table 1**). As a function of diet and genotype, we observed that HFD reduces the circulating serum level of ALPL, as reported previously in the BXDs [18] along with an increase in circulating ALPL in old mice (**Figure S1I**), as has been observed in humans [22]. Strains with the B6 allele of *Alpl* are known to have lower ALPL levels than those with the D2 allele [17], which we again observed in typical control conditions (CD, young) ($p = 0.004$, **Figure 1G**). However, this effect is not independent of environment: for old females, the known genetic effect caused by *Alpl* gene variants is masked by environmental interactions between age and diet ($p = 0.54$, **Figure 1G**). These interactions between age, genotype, and diet on ALPL are known [17, 18], but this illustrates the essential challenge of causal discovery: a single circulating metabolite is affected by genotype (*Alpl* allelic variants), diet, and

age. Without this prior knowledge, the p-values here using traditional statistical analyses would quickly be reduced to non-significance after correction for multiple testing.

Molecular Physiology of the Aging Liver

The observed range in clinical phenotypes as a function of diet, genetics, and age were expected to be in part reflected by underlying molecular expression changes in the liver. Both mRNA and protein are key parameters of gene expression, and together make up the most comprehensively quantifiable estimate of gene activity. However, mRNA and protein measurements often diverge and dissociate in response to stimuli, with only ~20-30% of variance of protein levels is explained by variance in their paired mRNA [23]. Furthermore, studies on aging have found that only a small minority of transcripts vary substantially across age, e.g. only 0.4% of transcripts varied by ≥ 2-fold between young and old animals as a function of caloric restriction mouse muscles [24]. Similarly small effect sizes are seen for proteomics, with only perhaps 1% of protein assays changing by more than 2-fold between 5 and 26 months-of-age within a single inbred strain [25]. Conversely, comparatively large effect sizes are seen in *pathway* activity between young and old individuals, e.g. ribosome translation rates drop by as much as 2-fold between young and old individuals [26]. Differential responses of mRNA and protein expression as a function of GxExA for entire pathways led us to search for patterns in these relationships, and the consequences on hypothesis generation and validation. Thus, by comprehensively quantifying these molecular traits in the liver—mRNA, protein, metabolites—we hypothesized that we could examine entire molecular pathways identify and characterize the etiology of hepatic aging, dietary, and genetic interactions in metabolic pathways.

In order to examine this, we first performed a broad examination of the omics data. 20 million read per sample RNA-seq measurements were obtained on a HiSeq PE150 from 291 individuals (i.e. 291 passed all data quality controls (QC)), corresponding to 25394 distinct transcripts, of which 20827 are annotated as protein coding. SWATH total proteomics measurements were obtained for 315 individuals (again after QC), corresponding to 3940 distinct proteins. Untargeted metabolomic data were also measured for 571 individuals, with 402 validated and annotated metabolite features detected after QC. Altogether, 275 individuals have both mRNA

and protein data, and 254 individuals have all three layers of data. We first prioritized individual-omic analysis for genes with both mRNA and protein data to get an overall view of the gene expression data. 3772 genes were considered quantified at both layers (127 transcripts with values of 0 in ≥ 50% of individuals were removed). The remaining 3772 genes include some highly represented ontologies (e.g. mitochondria, cytoplasm, and ribosomal proteins) while others are depleted (e.g. membrane proteins and secreted proteins) (**Supplemental Table 2**). Note that some functional categories are depleted due to tissue type (e.g. olfactory receptors) or selection time (e.g. developmental proteins), while other depletions are due to the proteomics extractions, e.g. membrane-bound proteins are often depleted [27, 28].

Diet had a significant impact on the expression of 850 transcripts and 1356 proteins (**Figure 2A**, adjusted t-test between discrete groups) while 1888 transcripts and 1099 proteins significantly covaried with age (**Figure S2A**, correlation coefficient adjusted p-value). An average of 65% of observed variation in transcripts and proteins was explained by the three independent variables—strain, diet, age—with genotype having by far the largest single average impact on gene expression (~28%) with a similar amount of unexplained residual (~28%) (**Figure 2B**). Conversely, for metabolites only 45% of observed variance was explainable by the independent variables, again with strain playing the strongest individual role (~16%) but with a much larger residual (~55%) (**Figure S2B**). For an individual gene, its transcript and protein tended to be correspondingly affected by diet or age ($r = 0.42$ and $r = 0.49$ respectively; **Figure 2C**), while the effects of diet and age on gene expression were independent overall (**Figure S2C**). We next examined the relationships between mRNA and protein levels. Across all samples, mRNA is reasonably predictive for the abundance of a protein ($r = 0.44$, **Figure 2D**)—that is, the most abundant mRNAs tend to be the most abundant proteins and vice-versa. In this study however, we are primarily focused on how particular genes and pathways respond to perturbations (i.e. genotype, diet, or age). This calculation yielded an average correlation coefficient of rho = 0.14 across all 3772 genes with mRNA and protein measurements, with 33% of pairs covarying significantly (adj.p < 0.05; **Figure 2E**). That is, knowing that a specific mRNA is highly affected by genotype is only a weak predictor that its corresponding protein will be affected similarly.

Despite this low average correlation, additional data can be used to improve the predictive capacity of mRNA variance on its protein product for specific genes. For

instance, 60% of transcripts in the top decile of expression variability (range ≥ 4.1-fold) significantly covary with their protein in response to a perturbation, compared to only 18% of transcripts in the bottom decile of expression variability (range ≤ 1.7-fold) (**Figure 2F** and **Figure S2D**). That is: independent variables with a large effect size on a transcript's expression are far more likely to have a corresponding effect on the protein's expression. Furthermore, more abundant transcripts tended to correlate better with their proteins—60% of the most abundant decile of transcripts covary with their protein, versus only 15% of the least-abundant decile (**Figure S2E)**. While this could indicate higher levels of noise in low-abundance transcripts (and proteins, given that overall abundance correlates at r = 0.44), it is worth noting that abundance and variability are only weakly correlated (rho = 0.1 p = 2e-9, **Figure S2F**). Thus, measurement knowledge of these two (mostly-)independent variables allows some edge cases where protein and transcript measurements can be used as good proxies. For the 60 transcripts that are in the top decile of abundance and variability, 87% correlate with their protein significantly and with an average rho = 0.51—versus an average 33% at an average rho = 0.14 for an average pair across the full set.

Other factors stemming from prior knowledge can also be used to predict mRNA–protein covariation. For instance, genes with significant cis-eQTLs are more likely to have significantly covarying transcript–protein levels (**Figure 2F**). Genetic knowledge can also provide information; transcripts that have cis-eQTLs (LOD ≥ 1.7) correlate with an average rho ~ 0.18, and those with strong cis-eQTLs (LOD ≥ 4) correlate better yet (rho ~ 0.27). While looking at the genetic mapping and transcript expression data can provide clues about protein expression levels, other patterns can be observed based on the literature for particular molecular pathways. For instance, genes that are involved in CORUM-annotated complexes are substantially less likely to have significant mRNA–protein covariance (average rho ~ 0.06, **Figure 2G**). This is in spite of the fact that complex-member mRNAs tend to be slightly more abundant than average (p = 2e-6) and have no difference in their variation (p = 0.08). Moreover, the size of the complex had a significant impact on the expected correlation; genes in dimers had equivalent mRNA–protein correlations to average "solitary" genes (rho ~ 0.13), whereas genes in complexes of ≥ 10 subunits had only marginal correlation between mRNA and protein (rho ~ 0.03) (**Figure S2G**). Together, these findings indicate that the highest-variance genes selected through differential gene expression analysis will *mostly* validate across-layer. However,

caution must be taken for less dramatic effects. This is especially important for the study of complex traits and GxExA, as we recall that expression variation is relatively subdued: < 1% of genes vary by >2-fold as a function of age. That is, overlapping mRNA–protein results are helpful for increasing confidence in the mechanism, but these will be a small percentage of the overall connectivity networks. Critical molecular changes may only be evident at the transcriptome or the proteome level, and thus the discrepancies must be retained (e.g. [18, 29]).

Finally, we next examined the relationships between gene expression and the underlying varying genetics across the BXD population via QTL mapping on all 3772 transcript–protein pairs. 309 genes mapped to a significant cis-eQTL or cis-pQTL at LOD ≥ 4 (~98.4% true positive rate using *discovery* cutoffs; **Figure 2H** and **Figure S2H**). While only 23% of cis-QTLs were observed at this threshold for both mRNA and protein levels concurrently, an additional 20% were observed at a secondary threshold when followed-up with a specific QTL hypothesis (LOD ≥ 2, corresponding to a ~97% true positive rate for *validation* cutoffs). The observation that around half of cis-QTLs (43%) are observed uniquely to transcript or protein levels, with more unique cis-eQTLs observed than cis-pQTLs, is in line with previous results (e.g. [30]) but underscores the utility of separately analyzing both gene products. We next examined the reproducibility of cis-QTLs as a consequence of diet. At discovery cutoffs (i.e. LOD ≥ 4), around half of cis-eQTLs (**Figure 2I**) and cis-pQTLs (**Figure S2I**) were observed in common across diets, while at validation cutoffs, a full 90% of cis-QTLs were observed in both dietary conditions. This indicates that in general the interactions of diet with genetic background have a modulating interaction combined with genotype, but not a uniform effect across all strains — a parallel analysis leading to the same conclusion as the ANCOVA analysis (**Figure 2B**). Similar general trends were observed when comparing cis-QTLs across age instead of diet; only ~40% of cis-QTLs were found to concordantly affect both transcript and protein within age group, while ~75% of cis-eQTLs and cis-pQTLs were found in common across age groups (**Figure S2J**). Together, these results indicate that while genes' mRNA and protein have many common responses to GxE variables, they cannot be *ad hoc* assumed to be proxies for one-another. Hypothesis discovery and molecular analyses are thus expected to yield unique results when the different measurement layers are used separately, or better yet in tandem.

Using gene-environment-age interactions to understand liver physiology

These observations led us to develop two further avenues for analyses of functional expression networks: (1) identify stable molecular signatures which are observed congruently across data type or independent variables, indicating higher confidence for subsequent reproducibility studies, and (2) examine molecular signatures that diverge across data type or independent variable, which may have been overlooked by transcriptomic studies in control diets and young age. To decrease the search space, we pre-selected 25 functional metabolic gene sets from GSEA which we hypothesized to be significantly affected by diet, age, and/or genotypic variants across the BXD population (e.g. mitochondrial ribosome variants [31]) (**Supplemental Table 3**). All pathways had significant levels of genotypically-induced expression variation (as expected, i.e. **Figure 2B**). 24 pathways had significant protein coexpression networks, and 21 pathways had significant coexpression at the mRNA level (**Supplemental Table 3**).

Given the discrepancies between mRNA and protein for complexes (**Figure 2G**), we first examined OXPHOS, as it is driven by large protein complexes, is known to decrease as a function of age [32], and has variant complex assembly across the BXDs due to genetic variants in *Cox7a2l* [18]. Gene expression of both the mRNA and protein corresponded to strong correlation networks (**Figure 3A**), but no correlation was observed between the mRNA and protein expression networks (**Figure 3B** and **Figure S3A**). Furthermore, two key substructural elements of OXPHOS were evident. Uniquely at the mRNA level, the mitochondrially-encoded OXPHOS subunits (red-highlighted triangles, **Figure 3A**) were clearly apart from the nuclear-encoded primary cluster. Uniquely at the protein level, each complex of OXPHOS formed a distinct subnetwork within the overall structure (**Figure S3B**), and with no difference observed between nuclear and mitochondrially-encoded genes (**Figure 3A**).

We next examined the cholesterol biosynthesis process, a comparatively linear molecular pathway of enzymatic reactions driven by individual genes, rather than multi-protein complexes like OXPHOS. Here, while the mRNA and protein formed distinct coexpression networks, the two layers of gene expression were also strongly correlated (**Figure 3C–D**). When we added in the beta oxidation pathway, we observed a strong negative correlation between it and cholesterol biosynthesis, indicative of their complementary underlying functions. For beta oxidation, analogous

networks were also observed with both its mRNA and protein (**Figure 3E–F**). While gene sets of metabolic pathways that are not predominantly protein complexes tend to have significant cross-layer correlation (i.e. **Figure 2G**), this relationship is always weaker than the within-layer correlation (e.g. beta oxidation, TCA cycle, **Figure S3C**). To examine the consequences of divergence between mRNA and protein on subsequent hypothesis discovery and phenotypic deconstruction, we examined the impact of our independent variables of diet and age on each of our selected functional gene networks. Most gene sets were affected at the mRNA and/or protein level by diet or age—in line with expectations, as these gene sets were specifically selected due to known literature connecting them with diet or age (**Figure 3G, Supplemental Table 3**). Although these pathways diverged according to age and diet, the genetic variation due to strain played the largest role and precluded a simple categorization of a given animal into an age or diet purely based on PCA (**Figure 3H**). Moreover, functional correlation networks remained of similar strength regardless of the cohorts selected (**Figure S3D**), suggesting that causal mechanisms driving the networks' diverging responses to the independent variable can lie outside of the canonical gene sets. We thus set out to discover these interactors and effectors outside of the canonical sets.

Data-Driven Approaches to Non-Consensus Networks & Causal Inference

Functional gene categories and ontologies provide a crucial platform for moving between data-driven hypothesis generation and mechanistic molecular studies. Data-driven approaches to expand gene ontologies or to find novel disease-associated genes tend to scan for associations starting with a known target (e.g. a disease phenotype) and then calculate quantitative associations against a large measurement set of unknown relevance (e.g. a full transcriptome) [33]. This type of multi-omics analysis tends to be consensus-driven [34], a critical approach to reduce the perennial issue of false discovery in data-driven studies, where the number of response variables p typically far exceeds the number of observations n (i.e. $p \gg n$). Moreover, functional gene categories frequently necessitate arbitrary cutoffs, as metabolic pathways are subsets of larger sets of interconnected genetic mechanisms. The majority of the genome still remains relatively unexplored in the literature—sometimes dubbed the "ignorome" [35]. We hypothesized that these canonical functional networks may have novel interacting gene partners outside the

canonical gene sets, particularly for unstable genes which are associated only in peculiar study conditions (e.g. aged HFD-fed individuals). Moreover, our study design sought to also examine the causal *why* and *how* genes are associated with core metabolic pathways as a consequence of diet, age, and differences in genetic background.

Since divergent nodes are potentially the causal factors driving the biological differences between networks across study conditions, consensus networks will by definition *not* select such genes as top candidates. Consequently, we applied a machine learning technique which we recently developed for gene expression studies which compares the effects of multiple independent variables on a target network [15]. This method allows for a linear regression-based variable selection (similar to lasso regression [36]) and is combined with stability selection [37] which uses resampling to control for false discoveries. This significantly reduces false discovery compared to correlation networks or hierarchical clustering. Furthermore, the method performs a causal analysis, similar to one we developed previously [16], to assess if strongly-asssociated candidate genes mediate the effect of a secondary independent (in this case, diet or age). Together, the regression analysis should ensure that discovered target genes are proximal to the target pathway, while the stability analysis can indicate whether non-consensus connections are due to the nodes being upstream, downstream, or inside the target network. Consequently, we processed our same 25 canonical metabolic gene sets to identify potential novel associating genes for all 3772 genes measured at both the mRNA and protein level (e.g. for cholesterol biosynthesis, **Figure 4A**). The analysis first uncovers novel interacting genes, which will fit in a correlation network with the input pathway as a function of the segregating independent variable (in this case, diet). Genes such as *Obscn* (**Figure 4A**) which are centrally located in one condition (CD) but peripheral or absent in the other (HFD) can then be further analyzed for directionality, while novel nodes that are similarly positioned in both networks are retained, but without causality inferred (e.g. *Rdh11*). Moreover, causal inference lets us determine that the OBSCN protein is a target of diet, and that it is upstream of the cholesterol biosynthesis pathway in response to HFD (**Supplemental Table 4**).

This analysis was done for all 25 pathways as a function of diet and age, and candidates were retained and considered potential "hits" if they had either prediction or stability scores ≥ 0.50, corresponding to a permutation-based false discovery of

approximately 1 node per gene set and test (e.g. **Figure S4A**, each plot is the overlaid sum of 100 randomized tests). 4484 significant associations were found between canonical gene sets and non-canonical discovered genes, belonging to 1102 distinct genes. Note that candidate genes appear multiple times, as they can be associated with multiple independent gene sets and as a function diet, age, and mRNA/protein measurement type. 289 of the 4484 total associations—around 6%—had clear directional relationships to their target pathway (prediction or stability score of ≥ 0.50 and at least 0.50 units away from the X=Y linear axis). That is, these genes were unstable *and* their relationship between diet or age and the target pathway could be statistically determined. 197 (68%) causal associations are upstream of the target and 91 (32%) are downstream. These 289 associations correspond to 244 unique genes, as one gene can appear multiple times, e.g. APOE appears upstream of the PPAR signalling pathway as a function of both diet and age (**Supplemental Table 4**).

At the cutoffs selecting 4484 associations, we detected an average of 38 gene candidate hits per mRNA network and 34 per protein network (**Supplemental Table 3–4**), with substantial fluctuation in hits across input pathways (min: 10, max: 52) and the separating variable of diet or age. Within mRNA or protein datasets, roughly 44% of hits were associated consistently across diet and age, while the remaining hits shifted according to the independent variable. This also varied by pathway: 77% of the 44 new gene targets associated with the peroxisome mRNA networks were consistently associated independently of age or diet, compared to only 36% for the 46 new CYP450-related mRNAs (**Figure 4B** and **Supplemental Table 3–4**). Comparing mRNA to protein, no more than 18% of discovered genes were consistently associated in both expression layers with the target pathway, even for pathways with significant mRNA–protein correlations, such as cholesterol biosynthesis (**Figure 4C** and **Supplemental Table 4**). Overlap was similarly low for pathways that were strongly affected by the independent variables at both mRNA and protein level, with only 9% of discovered genes being the same for the CYP450 pathway despite both the mRNA and protein networks being similarly strongly affected by diet ($p = 7e\text{-}10$ mRNA, $p = 2e\text{-}19$ protein, **Figure S4B**).

To determine the relevance of these findings, we examined the known relationships between the hit genes and the canonical gene set. We first used DAVID [38] to determine the ontological association of the discovered genes to the target

pathway. For CYP50, 81 distinct genes were picked up across mRNA, protein, and diet/age networks, of which 36 were associated with the "oxidoreductase activity" (adj.p = 2e-30, **Figure 4D**). This includes 7 genes that are CYP450 genes in mice, but which were not included in the canonical GSEA set (based off of human annotations, e.g. murine *Cyp2c29* was not in the reference because its human homolog is *CYP2C18*). While this represented an oversight in creating our reference gene sets, the detection of these genes also improves confidence that the novel gene targets are indeed functionally relevant. Beyond this “leave-one-out” (or leave-7-out) analysis, numerous other targets have clear functional interactions with CYP450, such as two genes in glutathione metabolism (*Gstm6* and *Gstm7*) and six carboxylesterase genes (e.g. *Ces1c*). However, at least two dozen candidate genes have no clear, literature-reported connection with CYP450 or any proximal pathway, such as *Nipsap1*, *Afmid*, or *Tmem205* (**Supplemental Table 4**, sheet “6_CYP450”). Similar patterns were seen with other gene network associations. For the 76 genes found associated to the canonical peroxisome set (**Figure 4B**), 7 were DAVID-annotated peroxisome (adj.p = 2e-4) and 30 were mitochondrial genes (adj.p = 7e-14), and primarily those that are clearly related to the peroxisome's function (e.g. *Acss3*, *Acot12*, *Acadl*, *Acadm;* **Supplemental Table 4**, sheet “7_Peroxisome”). Thus, as with CYP450, most of the detected non-canonical genes fit into clearly adjacent molecular pathways, but again some genes with no clear functional association are confidently identified such as *Ufd1l*, a ubiquitin-related gene.

Given our laboratories' interest in mitochondrial gene sets and in protein complexes, we decided to look at the mitochondrial ribosome and OXPHOS gene sets in greater detail. The mitochondrial translation gene set (Reactome, M27446) yielded 100 associated candidate genes: 57 from the mRNA data, 45 from the protein data, of which only 2 were found in common: *Ict1* and *Cox4i1* (**Figure 4E–F** and **Supplemental Table 4**, sheets “23_” and “26_”). *Ict1* is in fact a canonical mitochondrial ribosome protein, and was missed in the input due to it being renamed to *Mrpl58* in October 2016 (UniProt ID: Q8R035, changed in version 109). Similarly, *Dap3* (aka *Mrps29*) was also picked up, but only in the protein data. Fittingly, ETC genes in general were the most highly enriched associations with mitochondrial translation, with 33 of the 100 associated genes being canonical components of the ETC (e.g. Nduf*, Cox*, Atp5*). To classify the remaining 65 candidate genes (33 from protein data, 32 from mRNA), we entered the candidates into DAVID (**Figure 4G**). 42

of the remaining candidates mapped to the "mitochondrion" keyword, leaving 23 candidate genes with no clearly direct connection to the mitochondria (**Supplemental Table 4**, sheet “26_Mitochondrial_Translation”). These indicate fully novel associations with mitochondrial function of unknown mechanism. To assist with annotating gene candidates beyond literature knowledge, we used GeneBridge, a tool have recently developed that performs literature mining and meta-analysis of GEO datasets [39] (**Supplemental Table 4**, sheet “26_”, columns titled "GMAD"). Notably, one of the top novel candidate genes associated with mitochondrial translation, *Ddt*, was also recently selected and validated as OXPHOS-related by our separate GMAD analysis and was shown to play a role in the electron transport chain [39].

We further considered two other analytical possibilities. First: the target variable does not have to be a gene set; it can be any target trait or set of functionally related traits which are expected to have common causal and response factors. We first examined overall body weight as a function of diet to identify related transcripts or proteins. Diet has a huge impact on body weight ($p = 1e\text{-}32$), and we observed that among the 59 hits, 58% had predictions of causal direction (compared to 6% overall!), and of those, 6 were statistically downstream and 28 upstream (**Figure S4C** and **Supplemental Table 4**, sheet “BodyWeight”). This is logical—dietary feeding in mice has an unambiguous and unidirectional causal effect on body weight. Here, stability inference indicates which factors are most proximal to the separating variable. In this case, targets in the direction diet → gene → body weight should be discovered more frequently than diet → [unknown gene] → body weight → gene. Indeed, upstream candidate genes outnumber downstream candidates by 4:1 for the effect of diet on body weight, compared to 2:1 in the overall inference data.

The last possibility we considered is that the causal independent variable need not be diet or age, but it can be drawn from fundamental aspects of the population, as the genotype axis can be further split (as it is a group of 47 independent variables, i.e. strains). This cannot be unbiased due to the exponential size of the potential search space of allelic combinations: prior knowledge is required as for the grouping of selected strains, and the underlying reason for their grouping must be causal, and the expected effect should be known. For selecting strains, Mendelian expression variants in the population, e.g. SNPs leading to very strong cis-QTLs, can be used as a causal segregating variable in the same vein as diet, similar to Mendelian

randomization experiments. Here, we separated the BXD strains as a function of their *Cox7a2l* allele. A *cis*-acting variant in this gene is known to lead to major protein expression variants (but not mRNA) which causally leads to large differences in the total protein organization of OXPHOS complexes (but again, no change for the corresponding mRNA) [18]. Here, we examined how separating the mitochondrial translation networks as a function of *Cox7a2l* allele would affect the discovery of its interactions. As expected, both the mRNA and protein analysis discovered logical interacting genes with the mitochondrial ribosome: predominantly OXPHOS subunits and known UPR^{mt} regulators (**Figure 4H** and **Supplemental Table 4**, sheet "26_MitoTranslation _vs_Cox7a2lAllele"). More strikingly was the observation that these hits had very low stability scores—indicating that they are downstream of the target pathway. This shows that we can statistically infer that changes in OXPHOS proteins due to COX7A2L are downstream of the mitochondrial ribosome, while there is no causal effect for mRNA. That is, COX7A2L variants do not directly affect the mitochondrial ribosome. Note that this effect is specific to the COX7A2L allele. If strains are segregated due to strong allelic variants unrelated to OXPHOS and mitochondrial ribosome (such as HMGCS2), all hits will be found near the X~Y axis (**Figure S4D**), indicating that all the hits are stable. That is, regardless of the HMGCS2 allele, the correlation networks of the mitochondrial ribosome are both unperturbed and also cause no perturbations. Note that the input gene set and discovery gene sets are identical in **Figure 4E**, **4H**, and **S4E**: the only difference is the separation axis based on the independent variable (of age, COX7A2L allele, or HMGCS2 allele).

Metabolic Characteristics of Age

Last, we examined all of the data for candidate genes related to longevity. We first examined top hits from the differential expression analysis across aging (i.e. **Figure S2A**) but now expanded to include all transcripts, and not only those with protein measurements (**Supplemental Table 5**). We separated each dietary cohort and observed which traits coexpressed with longevity consistently in both diets, and selected the 100 with the most consistent cross-diet longevity correlation for DAVID analysis (corresponding to $p < 1e\text{-}5$ in both CD and HFD, roughly a correlation coefficient of rho > |0.33|). Only 9 genes correlated significantly in both diets at both the mRNA and protein level: *Anxa1, Ces3a, Ces3b, Coro1a, Nudt7, Rbp1, S100a9,*

Trf, and *Ugt2a3.* While no KEGG pathways were enriched among these top correlates, there was a striking enrichment of genes in the extracellular exosome, with more than half of the hits falling into this gene set (GO category #0070062, 51 genes, adj.p = 1e-18), while the next most significant fraction, cytosol, was far down in significance (21 genes, adj.p = 4e-3). This finding is in line with previously-observed relationships between longevity and the matrisome (genes that form or regulate the extracellular matrix) [40]. Notably, despite the distinct separation of mRNA and protein hits, the extracellular exosome was by far the most enriched GO category for both mRNA and protein even when considered entirely independently (36 genes, adj.p = 1e-10 mRNA; n = 21 genes, adj.p = 5e-11 for protein).

With these gene candidates and the enriched extracellular set in mind, we took a look at the individual candidates more carefully, particularly for those which we can examine in *C. elegans*. Roughly 30 genes had clear orthologs of either individual genes or small sets of related genes; e.g. cathepsin D (*Ctsd*) matched to homologs *asp-3* and *asp-4*. Conversely, genes such as *Slc22a30* and *Cyp4f13* matched to dozens of orthologs—essentially the entire family of solute carriers and cytochrome P450 genes—and were thus considered too non-specific for reasonable cross-species analysis. *Ctsd* was highlighted as a gene of interest due to its involvement in protein degradation, especially extracellular matrix proteins. More generally, loss of proteostasis has been recognized as one of the hallmarks of aging, a mechanism with which the lysosomes (and thus the cathepsins) have a direct relationship [41]. For instance, the *Ctsb* homolog *W07B8.4* is known to increase with age and interact with longevity extension in *daf-2* mutants [42]. However, its reduction improves reproductive aging, but it does not affect lifespan. Age-related increases of *Ctsd* have been generally observed across organs [43], although literature has been largely focused on brain tissues [44].

In our study, we observe significant increases in *Ctsd* as a function of age in mRNA and protein data both categorically (**Figure 5A**) and in a quantitative correlation with age (**Figure 5B**). We thus decided to check the top homolog, *asp-4* (BLAST p = 1e-107, score=386), and for which no *C. elegans* longevity studies have been published. We hypothesized *asp-4* may affect longevity, and that it may have an interactive effect with *daf-2*, a gene whose knockdown leads to decreased protein turnover (around 30%) and large increases in lifespan (more than 50%) [45]. RB2035, the *asp-4* mutant *C. elegans* line, had a significant decrease in lifespan

compared to wildtype when both were on empty vector treatment (L4440), with a median lifespan of 17.2 days, versus 23.7 days for control (p = 7e-27, **Figure 5C, S5A**). Reducing insulin/IGF-1 receptor signalling via adulthood-specific *daf-2* treatment resulted in the expected doubling of the lifespan (45.1 days, p = 2e-95), on the RB2035 background this effect was much reduced, with the same median lifespan as controls (23.7 days) although with a far longer lifespan tail, resulting in an overall lifespan extension (p = 1e-15).

Our analyses thus far focused on *chronological* age, based on the days since birth for each animal when its tissue was collected. However, given the highly heritable roughly 2-fold range in lifespan, patterns related to *biological* age were expected to be visible. To examine associations between biological age and gene expression, we first correlated all data against the median lifespan for each strain and cohort (**Supplemental Table 1**, row "Pheno_StrainMedianLifespan"). Using the same cutoffs as before (i.e. p < 1e-5 in both CD and HFD), only a single gene constitutively correlated with expected lifespan: suppressor of tumorigenicity 7 (*St7*), which was only measured at the mRNA level (**Figure 5D** and **Supplemental Table 5**). Fortuitously, *St7* has a single and clear homolog in *C. elegans*, called *F11A10.5* (referred to in this paper as *st-7*, BLAST p = 1e-112, score = 403), which also has no published longevity data. As there was no mutant for this gene available, we used RNAi to inhibit gene expression. For the temperature-sensitive sterile control *C. elegans* (*spe-9*), adulthood-specific *st-7* inhibition caused a small reduction in lifespan (17.7 vs 18.8 days, p = 1e-3; **Figure 5E, S5B**). However, in the normally long-lived *glp-1* mutant model which loses germline stem cells, *st-7* inhibition completely suppressed the lifespan extension effect (15.7 vs 30.5 days, p=5e-48). While mouse strains with shorter lifespans had higher *St7* expression (**Figure S5C**), there was no change in *St7* expression over time, unlike *Ctsd* (**Figure 5F**). It is also worth noting that *St7* has significant *cis*-eQTLs in the BXDs (**Figure S5D**) and correspondingly, expression is strongly correlated across dietary conditions (rho = 0.66, p = 2e-6). Furthermore, a meta-analysis of *St7* expression across BXD tissues indicates that it is relatively ubiquitously and constantly expressed across tissues. Additionally, its expression correlates robustly by across both tissues and across independent studies using the same BXD strains (median r ~ 0.50, **Figure S5E**), supporting a global relationship between genotype and the expression of this gene, regardless of tissue and environmental factors. Together, this demonstrates that

there is still a wealth of novel hypotheses related to aging that can be made through biostatistical data mining as well.

DISCUSSION

Aging is a dynamic process driven by a complex longitudinal mixture of genetic predestination, environmental effects, stochastic processes, and their interactions. Despite the relatively high heritability of longevity, little is known about the underlying molecular changes—and even less about the molecular causes. We measured the transcriptional, proteomic, and metabolomic landscapes of livers from 662 individuals from the BXD mouse population under two diets across their natural lifespans. Genetic differences explained the plurality of variation for all molecular measurement layers, dwarfing the solitary effects of diet and age. However, the interactions between genotype and diet and age explained an equally large share of variance as did genotype independently of environmental factors. Consequently, it stands to reason that in more complex population studies (e.g. humans), the unitary effect of genotype will be progressively diminished with the addition of further environmental variables, while the interaction effect will become increasingly important. That is, once a population is diverse enough, even genes that are strongly modulated by environmental factors (e.g. *CYP450* genes by diet) will have more of their variation explained by genetic background or GxE than by a uniform effect of diet on all genotypes. This raises a critical conclusion: studies using single, fixed environmental variables will only rarely be able to extrapolate their findings into a general population—especially if this is compounded by the hidden presence of additional unknown environmental variables. Consequently, in order to demystify the heritable factors of complex traits, we must consider study designs with multiple, simultaneous, independent variables to measure, and then decipher, the interaction effects.

Beyond the importance of GxE for identifying the causes of heritability, over the past decade, it has become increasingly recognized that analyses using mRNA or protein measurements often yield substantially different conclusions, even when measured at the same time in the same samples. That is, transcripts are not merely an intermediate proxy for protein. We observe here that a typical transcript and protein respond to typical GxE stimuli with only moderate concomitance (rho ~ 0.14). This general conclusion has numerous exceptions, particularly that factors leading to large mRNA changes are far more likely to replicate at the protein level and

conversely, genes in complexes are far less likely to have shared differentiation in response to a stimulus. However, it raises an additional critical point for understanding variation in natural populations: the vast majority of gene expression variants in mouse populations (or humans) will be of relatively modest effect size. A 2-fold change in OXPHOS gene expression should be considered just as dramatic as would be a 2-fold change in someone's daily exercise, insulin response, or lifespan. When studying natural populations, the effect sizes will be comparatively small, barring for extreme outliers and a handful of relatively bimodal phenotypes such as those driven by Mendelian diseases.

It is now possible to quantify the transcriptome and proteome across hundreds of samples. This has opened up new avenues in biostatistics. While differential analysis by volcano plots, gene set enrichment, and correlation networks benefit from larger sample sizes, in practical terms beyond a certain sample size, this results in diminishing returns. That is, a higher *n* means that smaller effect sizes can be determined "significant", but the significant detection of vanishingly small effect sizes quickly becomes of negligible real-world utility. Consequently, scientists are obliged to find new statistical rationales for generating these larger datasets. Recent advances in machine learning to biology have become popular in unsupervised analysis of genome data and images, but these algorithms tend to require sample sizes that are still orders of magnitude larger than are realistic for gene expression studies (e.g. frequently n > 100,000). This has highlighted a challenge for biostatistics, as the *n* of samples measured will inevitably continue to be far less than *p* the number of observations for multi-omics studies (e.g. measured transcripts). Addressing this "p >> n" issue requires intermediate biostatistical steps. That is, how to take advantage of an increased number of sample measurements when it still falls far short of that necessary for now "classical" machine learning algorithms popularized in recent computer science research. Here, we have applied a stability inference algorithm that we recently developed [15] which takes advantages of two aspects of this study design. First, the study's three independent variables (i.e. genetics, diet, age) permit stability analysis. That is, correlation networks and regression analyses can be performed across the primary axis of genetics, and discrepancies in network connectivity as a function of diet or age can be not only quantified as by node centrality analysis, but also it can be inferred whether the discrepancy is *due to* the intervention, or if it is a response. Second, the study's

acquisition of both mRNA and protein data allow a second “type” of stability analysis: results that are consistent across mRNA and protein gain improved confidence, while results that are inconsistent can be flagged according to certain criteria (e.g. presence of target gene in a protein complex).

Finally, we examined all of the analyses regarding novel gene connections with the aging phenotype, regardless of the biostatistical approach used to generate the hypothesis. Here, we selected two novel gene associations with aging which had clear *C. elegans* orthologs. One gene, *Ctsd*, was found to track strongly with age, independently of condition, in the BXDs. *C. elegans* mutants for the ortholog *asp-4* had decreased lifespan. The other candidate gene, *St7*, has strong divergence in expression across the BXDs as a function of genotype across tissue and even study. Its expression is stable with age, but the expression is constitutively lower in strains with higher *expected* lifespans. Given the consistency of *St7* expression, which was by far the most strongly co-expressed gene with age, we examined the *st-7 C. elegans* ortholog *F11A10.5*. Here, RNAi treatment in fact *decreased* lifespan, with a particularly striking effect on long-lived *glp-1* mutants: reducing its lifespan even further than in control conditions.

Aging research is a burgeoning field with a great wealth of knowledge generated every year as patterns are observed in gene expression, pathway activity, phenotypes, and both chronological and biological age. Due to the challenges of studying age—it is slow and difficult to control in humans, slow and expensive in mice, and it has additional translational challenges for simpler model organisms—it is imperative that improved bioinformatics tools are developed for developing targeted causal hypotheses. Even for simple organisms such as *C. elegans*, it is not feasible to perform longevity assays on every gene, let alone at different levels of knockdown or overexpression or on different genetic backgrounds or with epistasic interactions. Studying natural variation in populations allows the ready identification of clinically relevant *associations* between lifespan and molecular mechanisms, but separating the causal factors driving aging compared to downstream biomarkers has remained a challenge. That is, both greying hair and shortening telomeres have strong, clear associations with age—but it remains a challenge to causally determine whether (e.g.) a telomere-lengthening treatment would have any more effect on extending lifespan than does black hair dye. Moreover, even when causality has been shown such as for caloric restriction (CR) on lifespan, it is essential to deconvolute the

effects of GxE, given that many genetic backgrounds indeed have causally *shortened* life on CR. Developments in causal inference techniques now make such differentiations possible, and it is now possible for biologists to rationally design experiments that can statistically address these questions. Here, we show how experimental designs can be used to examine longstanding questions in aging, complex traits, and GxE in order to develop targeted hypotheses for mechanistic follow-up.

METHODS

Supplementary figures, tables, and further methods are available in the online version of the paper.

Mouse Care and Handling

All animal care was handled according to the NIH's *Guidelines for the Care and Use of Laboratory Animals* and was also approved by the Animal Care and Use Committee of the University of Tennessee Health Science Center (UTHSC). 2157 mice from 89 strains of the BXD family (including parents and both F1s) were followed across their lifespan. Animals were maintained in the UTHSC vivarium in Specific Pathogen-Free (SPF) housing throughout the longevity experiment. The housing environment was a 12-hour day/night cycle in 20–24°C temperature with housing cages of 145 in^2 with up to 10 animals per cage. Diets were either Harlan Teklad 2018 (CD; 18.6% protein, 6.2% fat, 75.2% carbohydrates) or Harlan Teklad 06414 (HFD; 18.4% protein, 60.3% fat, 21.3% carbohydrates). Water was Memphis city municipal tap water. Food and water were *ad libitum*. All animals were followed from their point of entry into the colony (typically around 4 months of age) until death. Animals were checked daily for morbidity and were weighed approximately every 2-3 months throughout their lives. 662 animals were sacrificed at specific ages for tissue collection across cohort (i.e. diet, strain, and age) while all other animals lived out their natural lifespans. For animals living out their natural lifespan, ~90% died naturally while the remaining ~10% were euthanized according to AAALAC guidelines and made by an independent veterinarian at the UTHSC facility. Euthanized animals were retained for lifespan calculations, with the expectation that they would have otherwise died shortly thereafter. 159 animals were males and 1998

animals were females. Males were pulled from only 11 lines: C57BL6, DBA2, both F1s, BXD9, 50, 68, 70, 73, 91, and 102.

Of note: on April 28, 2016 all mice were moved from the study's major housing facility ("Nash", which was slated for demolition) to a new building ("TSRB"). By this point, 94% of sacrificed individuals had been born, raised, and sacrificed in the Nash facility, so only 6% of individuals processed for omics analysis were moved, all of which were sacrificed between 1 September and 26 October 2016, i.e. after 4 to 5 months of acclimatization.

Aging Calculations

Lifespan calculations and significance tests were made using the "survival" package on R with the survfit and survdiff functions. The output longevity data were retained for strains with ≥ 5 recorded lifespans within a cohort. On average, there were 23 natural deaths per strain—11.5 on CD and 11.4 on CD. To compare lifespan across diet, a minimum of 10 natural deaths were considered necessary, with at least 5 on CD and 5 on HFD. Animals had their lifespan data suppressed from all calculations for the following reasons: (1) 48 deaths due to an "other" cause (e.g. 7 animals died due to a flooded cage, 27 were sacrificed for tissues for an urgent paper revision); (2) 29 deaths that were significant outliers (e.g. 8 mice that died within a week of entry into the aging colony, and 2 animals were far too old (> 1.5 years) when they were entered into the aging colony); and (3) the 662 animals that were sacrificed for tissue collection were not used for lifespan calculations.

Cohort Sacrifice Selection

Animals were selected for tissue harvest with the following aims: 2 animals per strain, diet, and age, for a target of 4 age points, i.e. up to a maximum of 16 sacrificed animals per strain (2 replicates * 2 diets * 4 ages). The target ages were 6, 12, 18, and 24 months of age. Roughly every 3 months for the duration of the experiment, ~40 animals were selected for sacrifice, with approximately 15 animals sacrificed per day over the course of 3 or 4 continuous days. Animals were removed from the aging colony the night prior to sacrifice, but retained access to food and water. Sacrifices started at approximately 9am with the anesthetic Avertin used via intraperitoneal injection of 0.2 mL per 10 g of body weight. Animals were perfused with ice cold phosphate-buffered saline. The liver was the first organ harvested. The gall bladder

was removed, the liver weighed, and then immediately frozen in liquid nitrogen in 20 mL scintillator vials.

Time Points for Aging Calculations

While CD and HFD comparisons were binary, comparisons across age were somewhat more challenging. For most analysis, linear regression was used for time at sacrifice (or measurement) against the target variable. However, for certain analyses of the omics data, particularly GSEA and age-QTLs, discrete bins were used. For binning animals based on age group, animals were binned by absolute age, not normalized by strain lifespan, with young mice considered those sacrificed before 419 days of age, and old mice considered those beyond 431 days of age, with a mean±σ of 293±80 vs 615±97, respectively. Note that the bins are more distinct than this cutoff may suggest; only 12 animals were sacrificed between the ages of 400 and 500 days; see **Supplemental Table 1** or **Supplemental Figure 1D** for more details.

Transcriptomics

RNA was prepared using TRIzol reagent and cleaned up with RNEasy mini kits. RNA-seq and RNA integrity (RIN) checks were performed by NovoGene on a HiSeq PE150. Samples with RIN ≥ 6.0 were retained for RNA-seq. All samples were run with 20 million reads. Normalization was performed by ComBat. All RNA-seq data were scaled by adding 1 to the normalized counts and then taking the log2. Transcripts with more than half counts of 0 were removed from analyses.

Proteomics

A detailed step-by-step protocol has been published for the sample preparation (Wu et al., 2017). In brief: liver samples were first entirely pulverized by mortar and pestle in liquid nitrogen, proteins were then extracted from ~20-50 mg of powered liver in 750 µL of RIPA-M buffer. The remaining cell pellet was then lysed fully in 8M of urea. The fractions were combined and 100 µg of each sample precipitated with acetone overnight at -20°C. The precipitated sample was resuspended in urea, reduced with dithiothreitol, and alkylated with iodoacetamide. Samples were diluted to 1.5M urea and then digested overnight (22 hours) using modified porcine trypsin. Peptides were cleaned C18 MACROSpin plates (Nest Group). Roughly 1.5 µg of each peptide

sample was loaded onto a PicoFrit emitter on an Eksigent LC system coupled to an AB Sciex 6600 TripleTOF mass spectrometer and acquired in SWATH data-independent acquisition mode (DIA) (Gillet et al., 2012) with 100 variable windows in a 60 minute gradient. A recent review also provides more detail on the full DIA pipeline [46]. In brief, the resulting .wiff files were converted to mzXML using Proteowizard 3.0.5533 before being run through the OpenSWATH pipeline v2.4.0 [47]. The library used was merged from our prior mouse library [27] together with part of the PanHuman library [48]. All peptides in the PanHuman library were BLASTed against the canonical mouse proteome from UniProt, version downloaded July 2017. Peptides which were found to be proteotypic in mouse, and not already extant in the mouse library, were then merged with the mouse library. Note that all runs for both human and mouse library generation were acquired with on the same machine (a TripleTOF 5600+) with the same settings (detailed in [48]). This merged "PanMouse" library contains 103,644 proteotypic peptides corresponding to 8219 unique proteins. This library was used to search the mzXML files for the OpenSWATH pipeline using the msproteomicstools package available on GitHub. Scoring and filtering were done by PyProphet at 1% peptide FDR, followed by cross-run alignment with TRIC using a max retention time difference of 60 seconds and a target 1% FDR, 29935 proteotypic peptides were identified which corresponding to 3694 unique proteins, were quantified across the 375 retained MS injections. Proteome data were segregated into batches based on noted changes during mass spectrometry (LC column change, MS tuning and cleaning). Within batch, samples were normalized by LOESS followed by across-batch normalization with ComBat, a synthetic approach based off of similar concerns with other large omics datasets, and which we have validated separately [49]. The batches are noted in the proteomics metadata under **Supplemental Table 6**. All data files, including the library file and both processed and unprocessed peptide quantifications, are available on PRIDE under PXD009160. *Reviewer access can be obtained by using the login* ***reviewer08961@ebi.ac.uk*** *and password* ***MuFdMZtp***.

Metabolomics

Pre-homogenized liver samples of were precisely weighed (to a target of ~20 mg) and then extracted in ~7 mL of a solution of 40% acetonitrile, 40% methanol, and 20% water, incubated for 24 hours at -20°C. The suspension was then centrifuged,

the supernatant transferred to a new tube, and then lyophilized. Dried samples were kept at -80°C until ready to be injected on the mass spectrometer, when they were resuspended in water according to the weight of the input tissue sample to a target of 5 mg/mL. The same extraction of all 621 samples was done in duplicate, starting from the same pre-homogenized liver sample, approximately one month apart, which is detailed by the “Run1” and “Run2” suffix in the data (**Supplemental Table 1**). Untargeted metabolomics analysis was performed on an Agilent 6550 QTOF instrument in negative mode at 4 GHz, for a mass range of 50–1000 Da using a previously-described protocol [50]. All samples were injected in technical duplicates in both experiments, and nearly all samples were injected in biological duplicates, i.e. each liver sample was injected 4 times to allow measurement stability to be calculated, so most cohorts had 8 measurements (2 full-process replicates * 2 technical replicates * 2 biological replicates (usually) per age-strain-diet cohort). An average of 19,000 features were detected in the runs, of which about 400 could be tentatively annotated as deprotonated metabolites listed in the Human Metabolome Database. Both experimental runs were run approximately one month apart, with all samples in each run being extracted and run in technical duplicate the same sequence both times. The full data and raw mass spectrometry are available on the MassIVE resource under accession ID MSV000081441 at **ftp://MSV000081441@massive.ucsd.edu.** The login is "Guest", and no password is required.

Strain-Averaged Data

For certain analyses—comparisons with historical BXD data on a per-strain basis (e.g. Figure S5C)—the individual animal data were averaged per cohort. That is, all female individuals for each strain were averaged on the following subsets: (1) taking all such individuals; (2) taking only those CD-fed; (3) only those HFD-fed; (4) only those which are young (≤ 419 days of age); (5) only those which are old (≥ 431 days of age); (5) CD young; (6) CD old; (7) HFD young; (8) HFD old. To reduce excessive multiple testing, not all of these 8 subsets were examined for each analysis. Cohorts 2 (CD, all ages), 3 (HFD, all ages), 4 (young, all diets), and 5 (old, all diets) were the primary four considered.

Hypothesis Discovery, Causal Inference, and Machine Learning Methods

For our network expansion analysis we employed a novel regression and variable selection technique [15], which is optimized for gene expression studies and explicitly allows incorporating causal reasoning similar to a method we recently published [16]. Our preliminary goal was to determine which genes are functionally related to a given pathway. The secondary goal was to determine if the gene was varying as a function of a secondary independent variable (e.g. diet, age), and if it was, then the causal directionality of this association with the target pathway and its sign (positive or negative, i.e. promotive or inhibitory). To this end, we compute the average expression level of the given pathway and use it as a response variable. Then, we randomly sample subsets of the pre-selected predictors and regress each subset onto the response resulting in a single regression coefficient. The individual regression coefficients are finally combined by a weighted average, where the weights are selected to ensure both good predictive performance and stability across the independent variable (diet, age, or a Mendelian QTL). By assessing the predictive performance of these regressions, we then rank the genes by their functional relation to the response (large values indicate a strong functional relation, small values a weak functional relation). For Mendelian separation using COX7A2L and HMGCS2, the 20 heterozygous F1 animals were removed. The selection procedure also uses stability selection [37] in order to control false discoveries and improve reliability. Furthermore, to empirically benchmark the false discovery rate, we perform a permutation based analysis as follows (Figure S4 B): We apply the entire selection procedure 100 times by permuting the observations of the response variable in each iteration and keeping everything else fixed (and hence preserving the correlation structure between the predictors). This analysis was performed for all the pathways we considered.

C. elegans **Testing**

C. elegans populations were maintained on NGM plates seeded with *Escherichia coli* bacteria at room temperature, with the exception that temperature-sensitive mutants were incubated at 15°C. For this work the *C. elegans* strains wildtype N2, TJ1060 *spe-9(hc88) I; rrf-3(b26) II*, CB4037 *glp-1(e2141)* III, and RB2035 *asp-4(ok2693)* were used as well as the RNAi clones *L4440* (empty vector control) and *F11A10.5* (ortholog of mouse *St7*, Vidal library [51]) and *daf-2* (RNAi). Animals were age-synchronized using population lysis [52] and then transferred at the late L4 stage to

plates seeded with the selected RNAi clone and containing 50 µM 5-Fluoro-2'deoxyuridine (FUDR). Depending on the genotype, animals were placed either at 15°C (N2 and RB2035) or at 25°C (TJ1060 and CB4037) to complete their development. After the L4 stage all animals were shifted permanently to 20°C and their survival was quantified. Manual survival scoring (by hand) was conducted as described previously [53]. Briefly, individuals which did not move in response to being prodded were classified as dead. Automated lifespan measurements were conducted using air-cooled Epson V800 flatbed scanners at scanning intervals of 30 minutes as described [54]. For the automated measurements the animals were transferred to fresh plates (BD Falcon Petri Dishes, 50x9mm) at day four of adulthood to facilitate the image detection process by removing as many eggs as possible. The survival data was analyzed using R in combination with the survival (v3.1-12) and survminer (v0.3.1) libraries. In the analysis, all animals which were observed to burrow, undergo bagging, explode or have escaped the agar surface were censored, and the L4 stage was set as the timepoint zero. The survival function was estimated utilizing the product-limit (Kaplan-Meier) approach and the null hypothesis was tested using the log-rank (Mantel-Cox) method.

Miscellaneous Bioinformatics and Statistical Tests

All QTL calculations were performed with r/QTL v1.42-8 (2018-02-19) [55] using the 2019 build of the BXD genotypes from GeneNetwork [56]. Transcript QTLs are referred to as "eQTLs" ("expression QTLs"). Protein QTLs are referred to as pQTLs, and metabolite QTLs as mQTLs. The blood serum measurements are considered as clinical phenotypes rather than as a part of metabolomics or proteomics due partly for historical categorization standards, and also due to differences in the measurement technologies and (in this particular study) source tissue. QTLs were declared as *cis* if their peak LOD region was within ±10 MB of the gene location.

Distributions (e.g. **Figure 3B**) were compared using a chi-squared test or Kalmogorov–Smirnov, as indicated. The R package “corrgram” v 0.84 was used for generating correlation matrices. Two group comparisons were made by Welch's t-test. Multiple group comparisons were made by ANOVA with Tukey post-hoc tests. The contribution of each independent variable to trait expression was calculated using ANCOVA. Correlations were performed using Pearson (r) or Spearman (rho), as indicated. To determine community structure in correlation networks, the package

ggbiplot v0.55 was used. Adjusted p-values were calculated using the Benjamini-Hochberg method of the p.adjust function in the R library *stats* (v 3.5.1). Spearman correlation networks were plotted using the imsbInfer v0.2.4 library in R (itself based off of the R library called iGraph). Lifespan calculations for mice and significance tests were made using the "survival" v2.43-3 package on R with the survfit and survdiff functions. Lifespan calculations for *C. elegans* are detailed in the *C. elegans* section. The output longevity data were retained for strains with ≥ 5 recorded lifespans within a cohort. To compare lifespan across diet, a minimum of 10 natural deaths were considered necessary. Outliers were removed with the R library “outlier” using the “rm.outlier” function.

Reference gene sets for the functional analyses in Figures 3 and 4 were taken via the "C2" curated gene set lists on the GSEA website. The exact reference names of all 25 gene sets are in **Supplemental Table 3**. Note that this analysis only considers the 3772 genes with both mRNA and protein data. Canonical functional gene assignments were curated either from GSEA (for pathway membership) or for CORUM (for complex membership). To determine the significance of correlation networks, random networks were permuted using the same number of nodes as in the target network, but randomly selected from amongst the 3772 other proteins and mRNA. For networks with multiple categories of gene sets, the target and random networks were computed both on a per-set basis as well as the total set and the interaction between the sets. 10,000 random networks were permuted for each comparison, and networks were assigned p-values based on this permutation, according to how many random gene sets had at least as many edges as the target set at the given cutoff, or assigned $p < 0.0001$ if no random network had as many edges as the input network at the cutoff. All figures were generated either in R, or were hand-drawn (e.g. **Figure 1A**).

BXD Epoch Notes

The inbred BXD strains were generated in seven distinct clusters of time: (1) BXD1–30 were generated in the 1970s; (2) BXD31 and 32 were generated around 1980. BXD33–42 were created in the 1990s. BXD43 through 102 were created in the early 2000s. BXD104–157 were created in the late 2000s. BXDs160–186 were created around 2010. BXD187–220 were created in the mid-2010s. Multi-omics data were generated only for strains BXD1–102 as the later strains which were included in the

aging colony were not always above F20 (i.e. >99.5% inbred, the canonical cutoff threshold for inbred nomenclature). The following 9 BXD strains are technically DXB as they have a lineage from a D2 mother and B6 father: 32, 61, 74, 82, 89, 90, 91, 95, and 99. When reporting the number of strains analyzed for each part of this study, we count the two F1s—B6D2 and D2B6—and the parental strains. Although F1 hybrids are not "inbred strains", they can be reliably and reproducibly generated to provide biological replicates, and thus can be used as reliably as inbred strains for studies on gene-by-environment interactions. Further details are available on GeneNetwork: http://www.genenetwork.org/mouseCross.html

SUPPLEMENTARY INFORMATION

Supplementary figures and tables are included. Raw mass spectrometry data for metabolomics are available on the MassIVE resource under accession ID MSV000081441 with reviewer access **ftp://MSV000081441@massive.ucsd.edu** (login is "Guest" with no password). Raw mass spectrometry data for proteomics are available on ProteomeXchange [57] under accession ID PXD009160 with reviewer login **reviewer08961@ebi.ac.uk** and password **MuFdMZtp**. The full transcriptomics data are available on GeneNetwork.org under Species: mouse, Group: BXD NIA Longevity Study, Type: Liver mRNA, and Dataset: UTHSC BXD Harvested Liver RNA-Seq. All processed omics data are also available in **Supplemental Table 1**.

ACKNOWLEDGEMENTS

Research was further supported by the EPFL, ETHZ, the NIH (R01AG043930 to RW), the AgingX program of the Swiss Initiative for Systems Biology (51RTP0-151019, 2013/134, and 2013/153), the European Research Council (Proteomics4D (AdvG grant 670821 and Proteomics v3.0; AdvG-233226 to RA), and SNSF (31003A-140780, 31003A-143914 and CSRII3-136201 to RA, PP00P3_163898 to CS and CYE). NP and PB were supported by ERC No 786461 (CausalStats - ERC-2017-AdvG). EGW was supported by an NIH F32 Ruth Kirchstein Fellowship (F32GM119190). Thanks to Lorne Rose at the University of Tennessee Center of Excellence Sequencing Facility for transcriptomics, and to Sebastien Lamy at the EPFL phenotyping unit (UDP) for clinical blood analysis, to Casey Chapman for BXD colony maintenance, and to Ludovic Gillet, Patrick Pedrioli, Özgen Eren, Chloe Lee, and Yansheng Liu for proteomics discussions.

AUTHOR CONTRIBUTIONS

EGW and RW established the BXD aging project and secured funding. SR, EGW, CB, JI, RWW, and LL managed the colony and the tissue collections. JI and CB managed the daily needs of the BXD colony. JI, SR, and RWW performed the transcriptomics. RA and EGW performed the proteomics. MH, NZ, and EGW performed the metabolomics. JC normalized the proteomics data. NP and PB developed and performed the causality/stability analyses. EGW performed all other statistical analyses on the omics data. CS and CYE performed the *C. elegans* experiments. The paper was written by EGW, with assistance from the other authors for their specific contributions.

REFERENCES

1. Hook, M., et al., *Genetic cartography of longevity in humans and mice: Current landscape and horizons.* Biochim Biophys Acta Mol Basis Dis, 2018. **1864**(9 Pt A): p. 2718-2732.
2. Belknap, J.K., *Effect of within-strain sample size on QTL detection and mapping using recombinant inbred mouse strains.* Behav Genet, 1998. **28**(1): p. 29-38.
3. Kraja, A.T., et al., *Longitudinal and age trends of metabolic syndrome and its risk factors: the Family Heart Study.* Nutr Metab (Lond), 2006. **3**: p. 41.
4. White, M.C., et al., *Age and cancer risk: a potentially modifiable relationship.* Am J Prev Med, 2014. **46**(3 Suppl 1): p. S7-15.
5. Labbadia, J. and R.I. Morimoto, *The biology of proteostasis in aging and disease.* Annu Rev Biochem, 2015. **84**: p. 435-64.
6. Santra, M., K.A. Dill, and A.M.R. de Graff, *Proteostasis collapse is a driver of cell aging and death.* Proc Natl Acad Sci U S A, 2019. **116**(44): p. 22173-22178.
7. Srivastava, S., *The Mitochondrial Basis of Aging and Age-Related Disorders.* Genes (Basel), 2017. **8**(12).
8. Bell, C.G., et al., *DNA methylation aging clocks: challenges and recommendations.* Genome Biol, 2019. **20**(1): p. 249.
9. Levine, M.E., et al., *An epigenetic biomarker of aging for lifespan and healthspan.* Aging (Albany NY), 2018. **10**(4): p. 573-591.
10. Jazwinski, S.M. and S. Kim, *Examination of the Dimensions of Biological Age.* Front Genet, 2019. **10**: p. 263.
11. Tabula Muris, C., *A single-cell transcriptomic atlas characterizes ageing tissues in the mouse.* Nature, 2020. **583**(7817): p. 590-595.
12. Sadler, N.C., et al., *High-Fat Diets Alter the Modulatory Effects of Xenobiotics on Cytochrome P450 Activities.* Chem Res Toxicol, 2018. **31**(5): p. 308-318.
13. Kruse, S.E., et al., *Age modifies respiratory complex I and protein homeostasis in a muscle type-specific manner.* Aging Cell, 2016. **15**(1): p. 89-99.
14. Marbach, D., et al., *Wisdom of crowds for robust gene network inference.* Nat Methods, 2012. **9**(8): p. 796-804.

15. Pfister, N., et al., *Stabilizing Variable Selection and Regression.* arXiv, 2019.
16. Meinshausen, N., et al., *Methods for causal inference from gene perturbation experiments and validation.* PNAS, 2016. **113**(27): p. 7361-8.
17. Andreux, P.A., et al., *Systems genetics of metabolism: the use of the BXD murine reference panel for multiscalar integration of traits.* Cell, 2012. **150**(6): p. 1287-99.
18. Williams, E.G., et al., *Systems proteomics of liver mitochondria function.* Science, 2016. **352**(6291): p. aad0189.
19. De Haan, G. and G. Van Zant, *Genetic analysis of hematopoietic cell cycling in mice suggests its involvement in organismal life span.* FASEB J, 1999. **13**(6): p. 707-13.
20. Gelman, R., et al., *Murine chromosomal regions correlated with longevity.* Genetics, 1988. **118**(4): p. 693-704.
21. Lang, D., et al., *Quantitative Trait Loci (QTL) Analysis of Longevity in C57BL/6J by DBA/2J (BXD) Recombinant Inbred Mice.* Aging Clinical and Experimental Research, 2010. **22**(1): p. 8-19.
22. Frisch, F., et al., *Age-Related Serum Alkaline Phosphatase Trends.* Faseb Journal, 2013. **27**.
23. Liu, Y., A. Beyer, and R. Aebersold, *On the Dependency of Cellular Protein Levels on mRNA Abundance.* Cell, 2016. **165**(3): p. 535-50.
24. Edwards, M.G., et al., *Gene expression profiling of aging reveals activation of a p53-mediated transcriptional program.* BMC Genomics, 2007. **8**: p. 80.
25. Walther, D.M. and M. Mann, *Accurate Quantification of More Than 4000 Mouse Tissue Proteins Reveals Minimal Proteome Changes During Aging.* Molecular & Cellular Proteomics, 2011. **10**(2).
26. Gonskikh, Y. and N. Polacek, *Alterations of the translation apparatus during aging and stress response.* Mech Ageing Dev, 2017. **168**: p. 30-36.
27. Williams, E.G., et al., *Quantifying and Localizing the Mitochondrial Proteome Across Five Tissues in A Mouse Population.* MCP, 2018. **17**(9): p. 1766-1777.
28. Whitelegge, J.P., *Integral membrane proteins and bilayer proteomics.* Anal Chem, 2013. **85**(5): p. 2558-68.
29. Liu, Y.S., et al., *Multi-omic measurements of heterogeneity in HeLa cells across laboratories.* Nat Biotechnol, 2019. **37**(3): p. 314-322.
30. Chick, J.M., et al., *Defining the consequences of genetic variation on a proteome-wide scale.* Nature, 2016. **534**(7608): p. 500-5.
31. Houtkooper, R.H., et al., *Mitonuclear protein imbalance as a conserved longevity mechanism.* Nature, 2013. **497**(7450): p. 451-7.
32. Houtkooper, R.H., et al., *The metabolic footprint of aging in mice.* Sci Rep, 2011. **1**: p. 134.
33. Lee, I., et al., *Prioritizing candidate disease genes by network-based boosting of genome-wide association data.* Genome Res, 2011. **21**(7): p. 1109-21.
34. Subramanian, I., et al., *Multi-omics Data Integration, Interpretation, and Its Application.* Bioinform Biol Insights, 2020. **14**: p. 1177932219899051.
35. Stoeger, T., et al., *Large-scale investigation of the reasons why potentially important genes are ignored.* PLoS Biol, 2018. **16**(9): p. e2006643.
36. Meinshausen, N. and P. Buhlmann, *High-dimensional graphs and variable selection with the Lasso.* Annals of Statistics, 2006. **34**(3): p. 1436-1462.
37. Meinshausen, N. and P. Buhlmann, *Stability selection.* Journal of the Royal Statistical Society Series B-Statistical Methodology, 2010. **72**: p. 417-473.

38. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nat Protoc, 2009. **4**(1): p. 44-57.
39. Li, H., et al., *Identifying gene function and module connections by the integration of multi-species expression compendia.* Biorxiv, 2019.
40. Ewald, C.Y., *The Matrisome during Aging and Longevity: A Systems-Level Approach toward Defining Matreotypes Promoting Healthy Aging.* Gerontology, 2020. **66**(3): p. 266-274.
41. Stoka, V., V. Turk, and B. Turk, *Lysosomal cathepsins and their regulation in aging and neurodegeneration.* Ageing Res Rev, 2016. **32**: p. 22-37.
42. Templeman, N.M., et al., *Insulin Signaling Regulates Oocyte Quality Maintenance with Age via Cathepsin B Activity.* Curr Biol, 2018. **28**(5): p. 753-760 e4.
43. Wiederanders, B. and B. Oelke, *Accumulation of inactive cathepsin D in old rats.* Mech Ageing Dev, 1984. **24**(3): p. 265-71.
44. Sato, Y., et al., *Identification and characterization of an increased glycoprotein in aging: age-associated translocation of cathepsin D.* Mech Ageing Dev, 2006. **127**(10): p. 771-8.
45. Visscher, M., et al., *Proteome-wide Changes in Protein Turnover Rates in C. elegans Models of Longevity and Age-Related Disease.* Cell Rep, 2016. **16**(11): p. 3041-3051.
46. Ludwig, C., et al., *Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial.* Mol Syst Biol, 2018. **14**(8): p. e8126.
47. Rost, H.L., et al., *OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data.* Nat Biotechnol, 2014. **32**(3): p. 219-23.
48. Rosenberger, G., et al., *A repository of assays to quantify 10,000 human proteins by SWATH-MS.* Sci Data, 2014. **1**: p. 140031.
49. Cuklina, J., *Computational challenges in biomarker discovery from high-throughput proteomic data.* ETH Zurich - Thesis, 2018. **25390**.
50. Fuhrer, T., et al., *High-throughput, accurate mass metabolome profiling of cellular extracts by flow injection-time-of-flight mass spectrometry.* Anal Chem, 2011. **83**(18): p. 7074-80.
51. Rual, J.F., et al., *Toward improving Caenorhabditis elegans phenome mapping with an ORFeome-based RNAi library.* Genome Res, 2004. **14**(10B): p. 2162-8.
52. Teuscher, A.C., et al., *Assessing Collagen Deposition During Aging in Mammalian Tissue and in Caenorhabditis elegans.* Methods Mol Biol, 2019. **1944**: p. 169-188.
53. Ewald, C.Y., et al., *NADPH oxidase-mediated redox signaling promotes oxidative stress resistance and longevity through memo-1 in C. elegans.* Elife, 2017. **6**.
54. Stroustrup, N., et al., *The Caenorhabditis elegans Lifespan Machine.* Nat Methods, 2013. **10**(7): p. 665-70.
55. Arends, D., et al., *R/qtl: high-throughput multiple QTL mapping.* Bioinformatics, 2010. **26**(23): p. 2990-2.
56. Mulligan, M.K., et al., *GeneNetwork: A Toolbox for Systems Genetics.* Methods Mol Biol, 2017. **1488**: p. 75-120.
57. Vizcaino, J.A., et al., *2016 update of the PRIDE database and its related tools.* Nucleic Acids Res, 2016. **44**(D1): p. D447-56.
58. Balsa, E., et al., *NDUFA4 is a subunit of complex IV of the mammalian electron transport chain.* Cell Metab, 2012. **16**(3): p. 378-86.

FIGURE AND TABLE LEGENDS

Figure 1. Overview of Aging Colony

(**A**) Study overview. 2157 individuals from 89 strains entered the aging program at around 151 days of age and were set into dietary cohorts. Individuals from each cohort were sacrificed at 7, 12, 18, and/or 24 months of age. (**B**) Workflow of the lifespan and tissue collection cohorts. The 662 sacrificed individuals fall into 303 cohorts based on age, sex, strain, and diet. (**C**) Left: Violin plot of individuals' lifespans on CD or HFD, independent of strain. Right: The same data plotted as a Kaplan–Meier survival curve. (**D**) Kaplan–Meier survival curves for all 53 BXD strains with at least 8 natural deaths; 49 strains had at least 5 deaths in both diets. Some extreme strains are highlighted. (**E**) Basic phenotype data of weight-over-time for 216 CD (top) and 103 HFD (bottom) individuals which reached ≥ 800 days of age. All animals were weighed approximately bimonthly. (**F**) Violin plots of body weight at 18±1.8 months of age (corresponding to 400±90 days of HFD feeding). (**G**) ALPL serum metabolite levels across BXD strains, simplified to separate only as a factor of whether the strain has the B6 or D2 allele of the gene, which is known to drive variance in circulating metabolite levels. The canonical allelic effect is clear for young, CD animals ($p = 0.004$) but is masked by GxE interactions in old and mixed-diet cohorts ($p = 0.54$).

Figure 2. Multi-omics Overview of mRNA, Protein, and Metabolite Liver Expression

(**A**) Variation explained for all 3772 transcripts and proteins as a function of the independent variables and their interactions, and the residual. (**B**) Around 22% of transcripts and 35% of proteins being significantly affected by diet. (**C**) Correlation density plot of the relationship between the effects of diet (left) and age (right) on genes' product mRNAs and proteins. Brighter colors represent higher density data. (**D**) Correlation of average mRNA and protein levels across all samples and all genes. (**E**) Roughly 33% of transcripts had significant covariation with their protein, across all samples. (**F**) (Left) Density plot of mRNA–protein correlations as a function of the mRNA expression variance; (right) density plot showing correlation as a function of cis-QTL presence; (center-right) as a function of diet (**G**) Density plot of correlation as a function of the gene existing in a complex. Reported p-values are for

the comparisons marked with an asterisk (*). (**H**) (Left) 309 genes reported highly-significant cis-QTLs, with ~23% overlapping (LOD ≥ 4) at the mRNA and protein level. (Middle) Strict cutoffs in Venn diagrams can obscure some underlying patterns; (Right) if taking LOD ≥ 4 for discovery and LOD ≥ 2 for validation, closer to 43% of cis-QTLs were congruent across layers, with cis-eQTLs more common than cis-pQTLs. (**I**) Across diet at the mRNA level, ~47% of cis-eQTLs are found congruently in CD and HF conditions. Nearly all cis-eQTLs are congruent across diet when slightly lowering the threshold for reproducibility: 89% overlap at LOD ≥ 4 for discovery followed by LOD ≥ 2 for validation.

Figure 3. Functional Gene Networks of Transcripts and Their Proteins

(**A**) OXPHOS Spearman correlation networks for the 75 genes with both mRNA (left) and protein (right) measurements. Node color represents to which component of OXPHOS the gene belongs. Red-highlighted nodes are mitochondrially-encoded. NDUFA4 is a Complex IV member, despite its gene symbol [58]. (**B**) Spearman correlation density plot corresponding to panel A, but showing all 5550 correlations for mRNA and protein networks (i.e. 75^2 minus identity) and “Across” for the 11000 correlations in the mRNA–protein correlation network (i.e. all possible connections between nodes where one node is an mRNA and the other is a protein). (**C**) Spearman correlation network for both mRNA (blue) and protein (red) in the cholesterol biosynthesis network. (**D**) Spearman correlation density plot corresponding to panel C, but showing all 272 correlations within mRNA and protein networks, and all 544 correlations across. Of all 1088 possible edges, 409 are significant at $p < 0.0001$ (i.e. are drawn in panel C). (**E**) Spearman correlation network for the cholesterol biosynthesis and beta oxidation gene networks for mRNA (left) and protein (right) drawn simultaneously. (**F**) Spearman correlation density plot for the above graphs. “Across” in this context means within mRNA and protein, but across beta oxidation to cholesterol nodes. (**G**) GSEA enrichment for the OXPHOS and cholesterol gene sets as a function of age for mRNA and protein levels. (**H**) PCA plot of the first two principal components of the OXPHOS and cholesterol protein pathways as a function of age and diet, respectively, visualizing the moderate (and significant) separation by these two variables.

Figure 4. Network Expansion and Functional Gene Discovery

(**A**) Spearman correlation plot of the core cholesterol biosynthesis network (17 red nodes) along with the top 37 candidates selected by regression analysis in protein data, performed separately for CD (left) and HFD (right) cohorts (dark red nodes). (**B**) Prediction–Stability plots for the CYP450 and peroxisome pathway using mRNA data, separating by diet and age. The x-axis indicates the probability that a putative node has a direct functional relationship with the target network via linear regression. The y-axis indicates the probability that the target gene is either upstream or inside the target pathway (higher) or downstream of the target pathway (lower) via stability analysis, i.e. that it segregates based on the independent variable of diet or age. When the secondary independent variable does not have a significant effect on the overall network, it is expected that candidate genes will near the X=Y axis (e.g. peroxisome as a function of diet). Randomized permutations of the input data yield roughly 0.1 nodes by chance in the green-highlighted sections of the graph. (**C**) Prediction–Stability plot for the cholesterol biosynthesis pathway, indicating that many of the significant genes found in panel A for the protein network are statistically upstream (upper-left quadrant, e.g. *Fabp2*, *Acat1*, *Aldh1l1*). (**D**) DAVID enrichment analysis of the top candidate genes found through stability analysis of the CYP450 pathway, indicating that most are canonically known to have strong oxidoreductase activity. (**E**) Prediction–Stability plot for the mitochondrial ribosomal pathway for mRNA and protein as a function of age. (**F**) Prediction–Stability plot for the OXPHOS pathway for mRNA and protein as a function of age. (**G**) DAVID enrichment analysis of the top candidate genes found through stability analysis of the mitochondrial ribosome pathway, showing that the plurality are related to OXPHOS.

Figure 5. Aging Candidate Discovery & *C. elegans*

(**A**) *Ctsd* gene expression as a function of age, using a bimodal cutoff. Note that the age groups are relatively discrete. (**B**) Correlation plots of age versus *Ctsd* and CTSD, showing a significant correlation with age regardless of dietary conditions. (**C**) Longevity analysis of RB2035, a mutant *C. elegans* with the removal of the *Ctsd* homolog *asp-4*, compared to wildtype N2 *C. elegans*. (**D**) Correlation of *St7* as a function of *expected* lifespan (i.e. cohort median lifespan). (**E**) Survival curves of *C. elegans* as a function of *st-7* inhibition, on both a wildtype *spe-9* mutant background and in the *glp-1* long-lived mutant. (**F**) *St7* correlates with expected longevity in BXD mice, but it does not change in expression with actual recorded age.

SUPPLEMENTAL FIGURE AND TABLE LEGENDS

Figure S1. Supplemental Overview of Phenotype Data

(**A**) Lifespan correlations between BXD strains in two previous lifespan studies, conducted in 1988 [20] and 2010 [21], with 17 overlapping strains measured in both studies. (**B**) Proportion of variance explained by genotype, diet, age, or the interactions between these three variables, as calculated by ANCOVA. (**C**) Lifespan calculations as a function of diet for two extreme strains: BXD9, whose lifespan is unaffected by HFD feeding, and BXD65, which lives 55% longer on CD than HFD. (**D**) Kaplan–Meier curves for two strains with extreme effects of diet on lifespan; BXD9 which is unaffected, while BXD65 which lives almost a year shorter on HFD. (**E**) Lifespan correlation by strain between CD and HFD fed cohorts in this study. (**F–G**) Example body weight over time chart for two strains with extreme effects of diet on weight; BXD16 which is obese in either case, while BXD100 which more than doubles its weight on HFD. (**H**) Histogram and dot plot of the animals with tissues collected at each time and diet. (**I**) Serum alkaline phosphatase (ALPL) levels, segregated as a function of diet, age, or diet and age.

Figure S2. Supplemental Patterns in Multifactorial, Multi-omic Analysis

(**A**) Volcano plot showing the effect of age on mRNA and protein levels (using a binary cutoff between cohort 1 & 2 = young; cohort 3 & 4 = old). (**B**) Variation explained for all 402 metabolites a function of the independent variables and their interactions, and the residual. (**C**) Correlation density plot of the impact of diet and age on mRNA and protein levels. (**D**) Histogram of the percentage of significantly correlated mRNA–protein pairs as a function of the variance of mRNA expression across the population. (**E**) Density plot of mRNA–protein correlation as a function of transcript abundance. (**F**) Spearman correlation plot between expression variance and abundance for mRNA, where higher-expressed genes also have more variance (and thus are more likely to correlate with their protein). (**G**) Histogram of the percentage of significantly correlated mRNA–protein pairs as a function of the size of the protein complex to which the gene belongs (using CORUM annotations). (**H**) Empirical calculations of the false discovery rate of cis-eQTLs as a function of LOD score depending on if the gene is a novel target (“discovery”) or the validation of expectations from independent cis-QTL data. (**I**) Left: Venn diagram of the overlap of

cis-pQTLs for CD or HFD cohorts using a strict cutoff of LOD ≥ 4. Middle: The actual LOD scores of all 144 genes included in the Venn diagram for both CD and HFD. Right: The same Venn diagram, but now using one diet as the "discovery" cohort and the other diet as the "validation" cohort. (**J**) cis-QTL consistency on a gene-level basis between mRNA and protein levels (left) and as a function of age for protein levels (right).

Figure S3. Supplemental Functional Network Analysis and False Discovery

(**A**) Spearman correlation network for the 75 OXPHOS genes measured at both the mRNA and protein level, showing the low level of connectivity between the two layers (~3% of edges are across mRNA to protein). (**B**) Top: PCA biplots for the 75 OXPHOS genes as a function of complex membership. mRNA has no differentiation between complexes, while the proteins are distinct per complex, except for CIV. Bottom: Negative control PCA plots for a random selection of 75 genes, showing no differentiation between random sets of the same size. (**C**) Correlation density plots of correlation networks for the ribosome, mitochondrial ribosome, beta oxidation, and TCA cycle as a function of mRNA level, protein level, or across the two. Even in cases where mRNA and protein both have strong correlation networks (e.g. ribosome, TCA cycle), there is little or no correlation between mRNA and protein. (**D**) Correlation density plot of OXPHOS versus age and cholesterol versus diet: despite age and diet having a significant effect on these networks, the overall connectivity does not significantly change.

Figure S4. Supplemental Stability Inference and False Discovery

(**A**) The overlap of 100 negative control permutation networks using the 68-member mitochondrial translation gene set (top) and the 42-member PPAR signalling pathway gene set. The theoretical upper-bound false discovery in this area is 1 discovery per test in the green area regardless of gene set. Here, for this randomized gene set, the sum of 200 permutation tests discovers only 4 nodes in the mRNA and 2 in the protein network, empirically indicating a false positive of approximately 0.03 per test averaged across all runs. Permuted false discovery in the white area is approximately 1 node per test. Note that the exact cutoff values for the white, green, and red areas can change slightly depending on input parameters. (**B**) Prediction–Stability plot for the CYP450 network discovery at the mRNA (left) and protein (right) level. Significant

hits are labeled. (**C**) The target "pathway" can also be any trait, such as overall animal body weight. In this case, nearly all genes identified are in the "causal" section of the plot, as HFD feeding has a very strong (p = 1e-32) and directionally uniform effect on body weight. (**D**) Other causal variables can be used as the axis of separation. Here, the known-variant allele of the gene *Cox7a2l* is used to separate the mitochondrial translation network. The discovered protein targets are in the bottom-right quadrant, signifying they are downstream of differences in mitochondrial translation. *Cox7a2l* variants in the BXDs are known to have a causal effect on the expression (and organization) of mitochondrial proteins, but not on transcript levels [18].

Figure S5. Supplemental Analysis of Aging Candidates

(**A-B**) Replication experiments of *st-7* and *asp-4* lifespan experiments. (**C**) Mouse *St7* expression as a function of expected lifespan shows major significant changes, although the gene expression does not track with chronological age. (**D**) *St7* has a significant cis-eQTL in the liver, and (**E**) *St7* also has significant gene expression correlation as a function of genotype across many tissues and even independent studies, suggesting that its links to lifespan are stable and generic.

Table S1. Processed Phenotype, mRNA, Protein, and Metabolite Data

This table contains a full list of normalized measurements for every sacrificed individual at the mRNA, protein, and metabolite level. Note that the same metabolomics extraction was done twice in complete experimental replicate; the two runs are fully independent (other than that they are on the same source tissue).

Table S2. GO Enrichment Categories of 3772 Paired mRNA & Protein Quantified

This is a list of the enriched GO terms for all 3772 genes that were measured at the mRNA and protein level, showing that certain categories are well-covered (e.g. 37% of cytoplasm genes, p = 4e-180) while other categories are relatively poorly represented (e.g. only 4% of lipoprotein genes, p = 2e-6).

Table S3. Candidate Gene Networks for Functional Analysis

This table contains a list of all 25 functional gene networks analyzed at the mRNA and protein level. The gene sets were pre-selected by the authors for the core role

they play in energy metabolism, and with prior hypotheses in mind that gene sets may be affected by diet (based on our prior experience [18]) or age (based on general literature searches). Basic summary data about each gene set are available, e.g. number of members, GSEA results, network significance, and summary of stability analysis.

Table S4. Stability Analysis: Gene Hits

This table contains all of the results from the stability analysis on all 25 candidate gene sets as a function of diet, age, or expression type (protein/mRNA). The gene sets that were highlighted in specific figure panels get a second sheet inside the table containing only the subset of genes that matched the stricter criteria (i.e. x- or y- axis values of ≥ 0.50) along with basic information about each gene (e.g. description, Entrez ID, GMAD result).

Table S5. Stability Analysis: Gene Hits

This table contains the differential expression analysis for all data in Table 1 as a function of diet and age. (Note that Figure 2–4 only used the overlapping genes where both mRNA and protein were measured, whereas Figure 5 and this table also include transcripts for which no paired protein was quantified.)

Table S6. Proteomics Metadata

Metadata for the proteomics measurements, e.g. preparation times and acquisition dates, along with information such as when columns on the liquid chromatographer were changed.

Figure 1. Study Overview

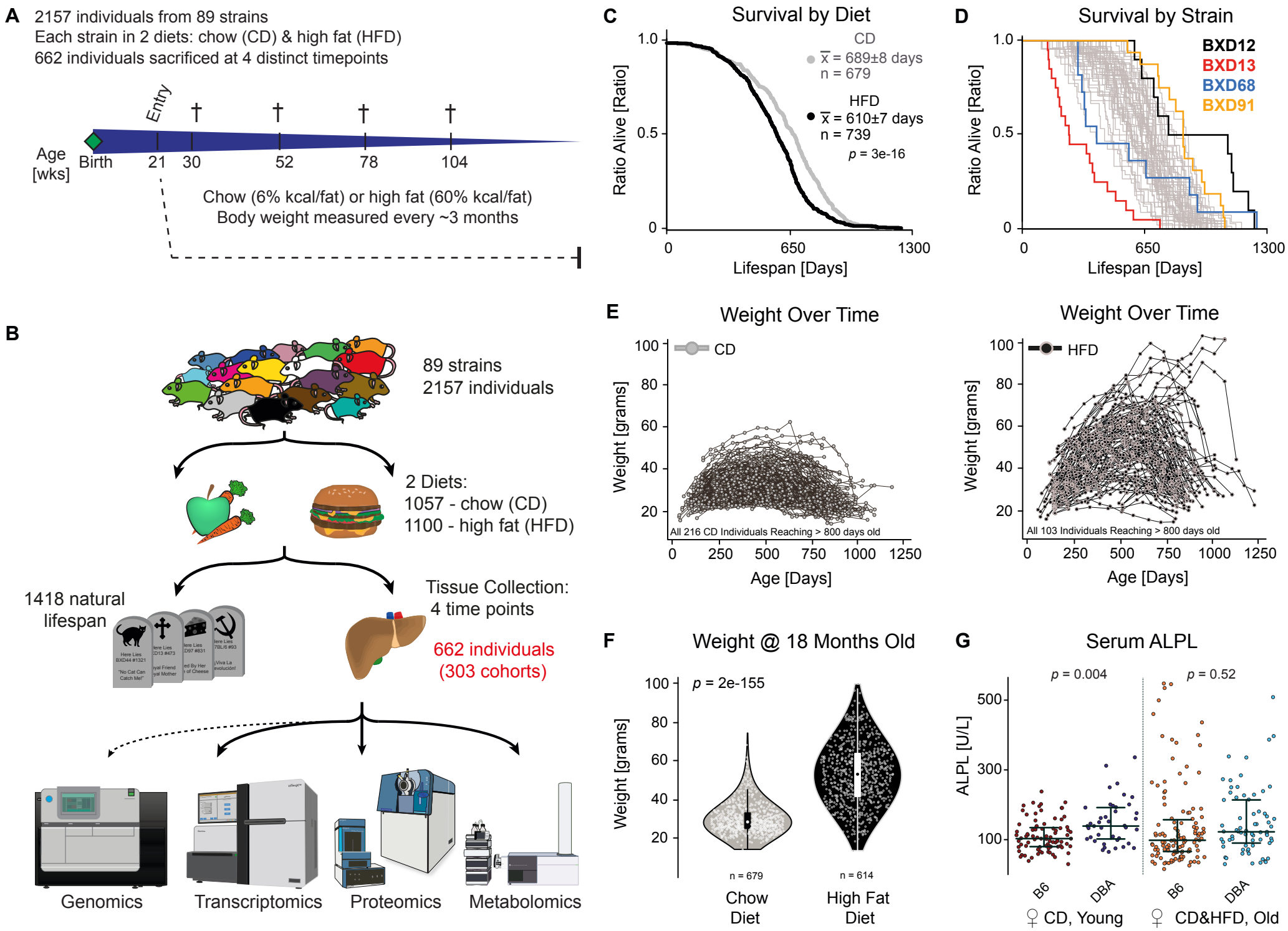


Figure 2. Multi-omics Overview of Gene Expression

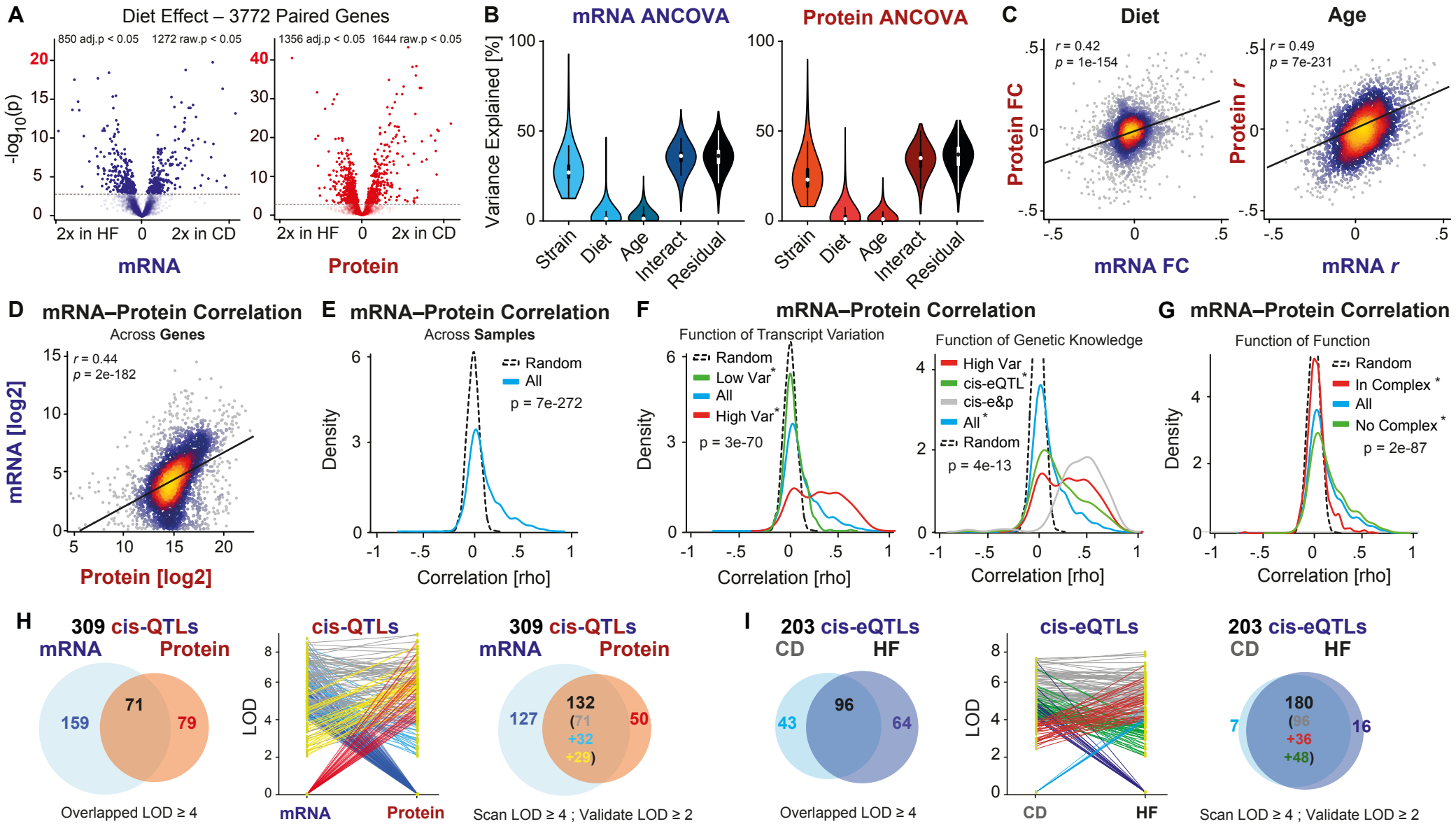


Figure 3. Functional Gene Networks of Transcripts and Their Proteins

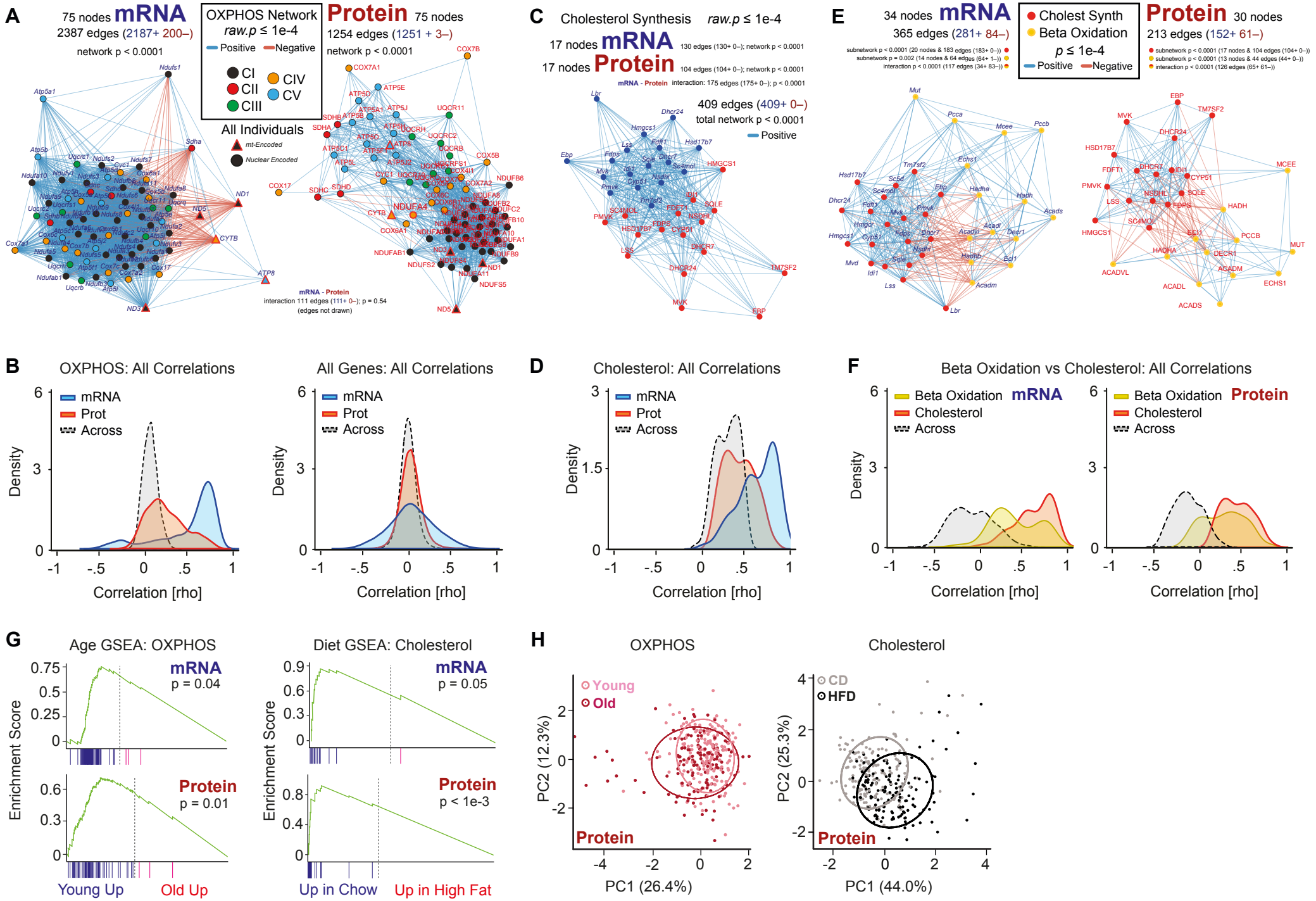
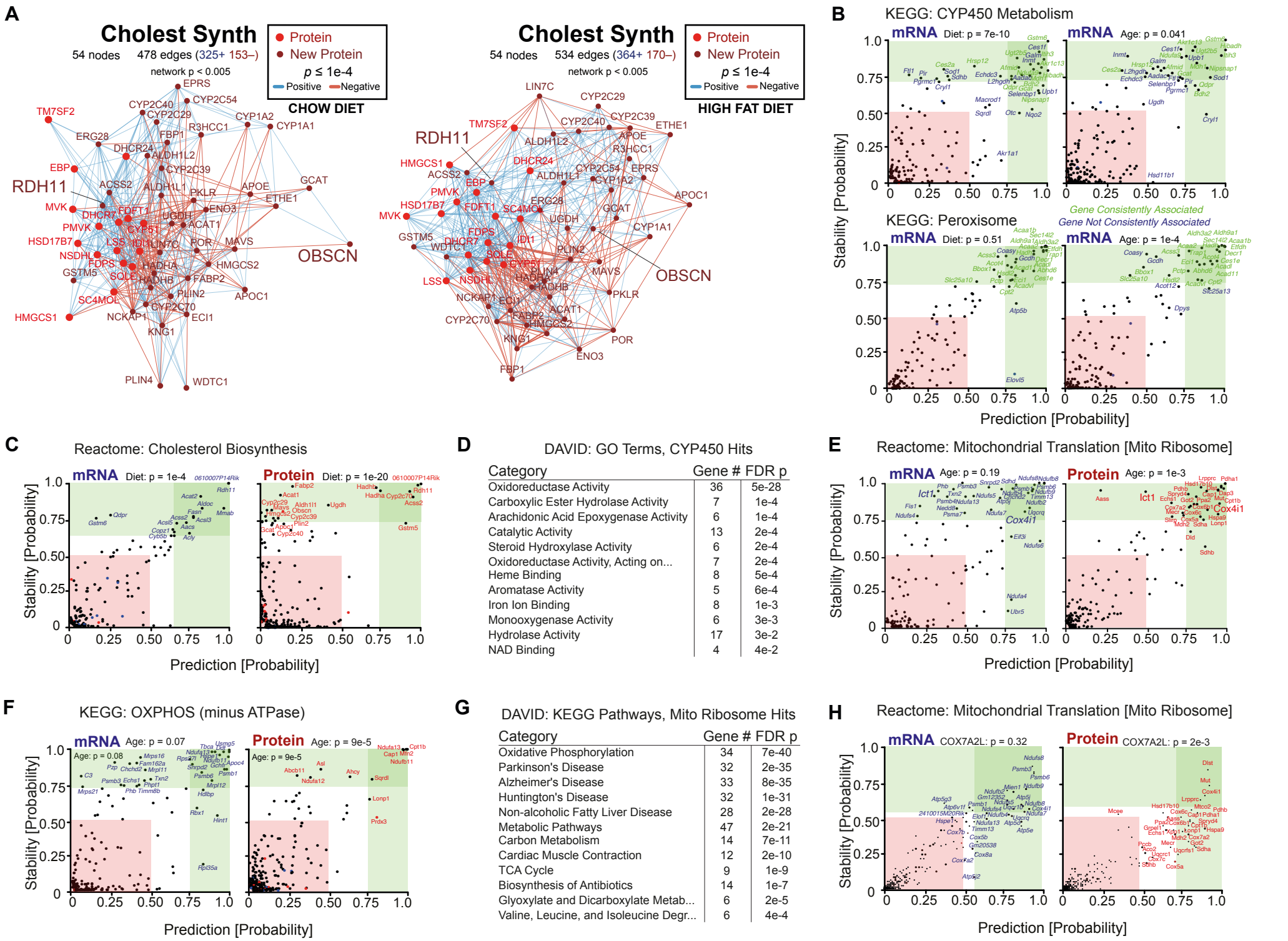


Figure 4. Network Expansion and Functional Gene Discovery



Category	Gene #	FDR p
Oxidoreductase Activity	36	5e-28
Carboxylic Ester Hydrolase Activity	7	1e-4
Arachidonic Acid Epoxygenase Activity	6	1e-4
Catalytic Activity	13	2e-4
Steroid Hydroxylase Activity	6	2e-4
Oxidoreductase Activity, Acting on...	7	2e-4
Heme Binding	8	5e-4
Aromatase Activity	5	6e-4
Iron Ion Binding	8	1e-3
Monooxygenase Activity	6	3e-3
Hydrolase Activity	17	3e-2
NAD Binding	4	4e-2

Category	Gene #	FDR p
Oxidative Phosphorylation	34	7e-40
Parkinson's Disease	32	2e-35
Alzheimer's Disease	33	8e-35
Huntington's Disease	32	1e-31
Non-alcoholic Fatty Liver Disease	28	2e-28
Metabolic Pathways	47	2e-21
Carbon Metabolism	14	7e-11
Cardiac Muscle Contraction	12	2e-10
TCA Cycle	9	1e-9
Biosynthesis of Antibiotics	14	1e-7
Glyoxylate and Dicarboxylate Metab...	6	2e-5
Valine, Leucine, and Isoleucine Degr...	6	4e-4

Figure 5. Aging Candidate Discovery & *C. elegans*

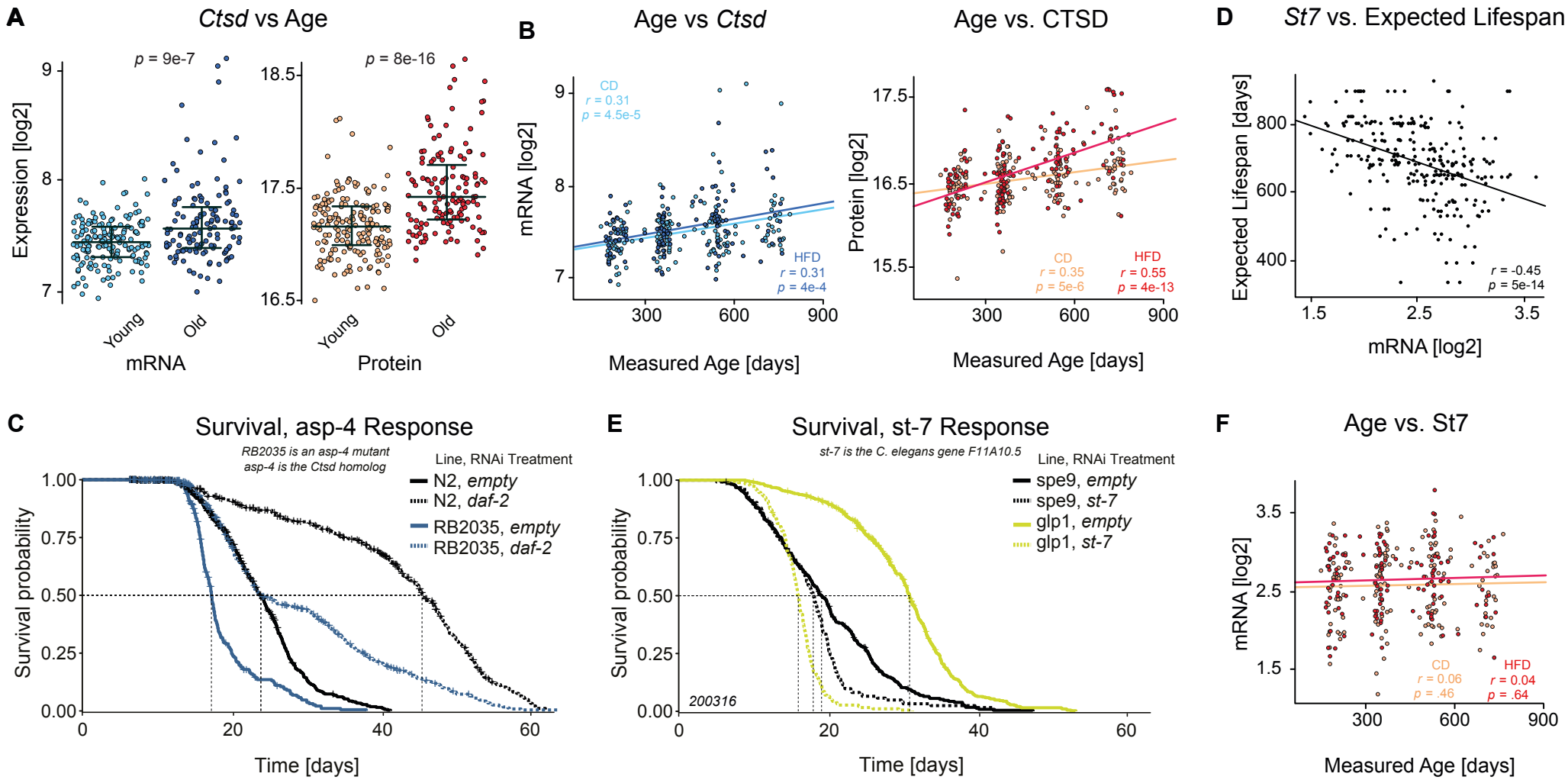


Figure S1. Supplemental Overview of Phenotype Data

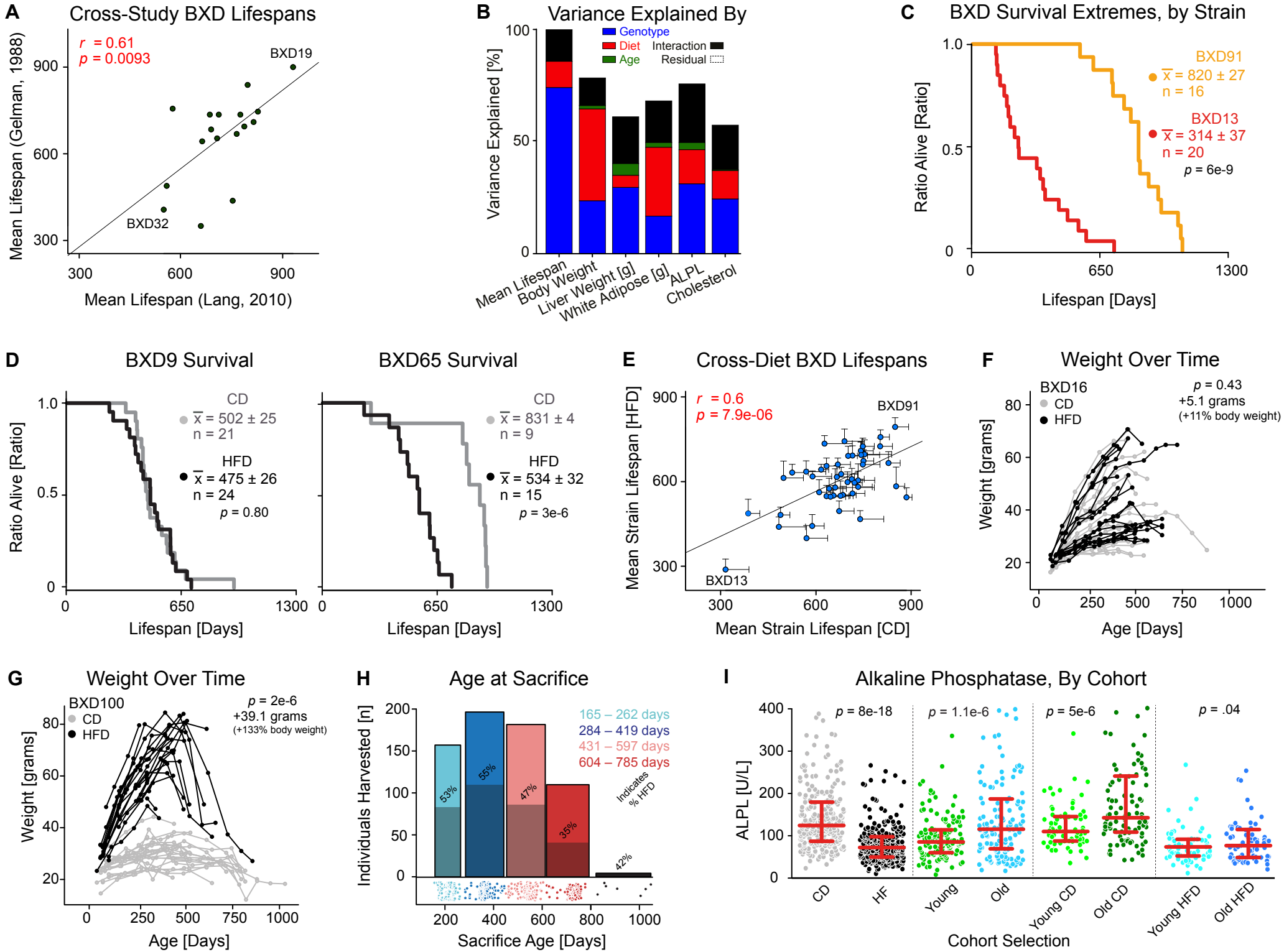


Figure S2. Supplemental Patterns in Multifactorial, Multi-omic Analysis

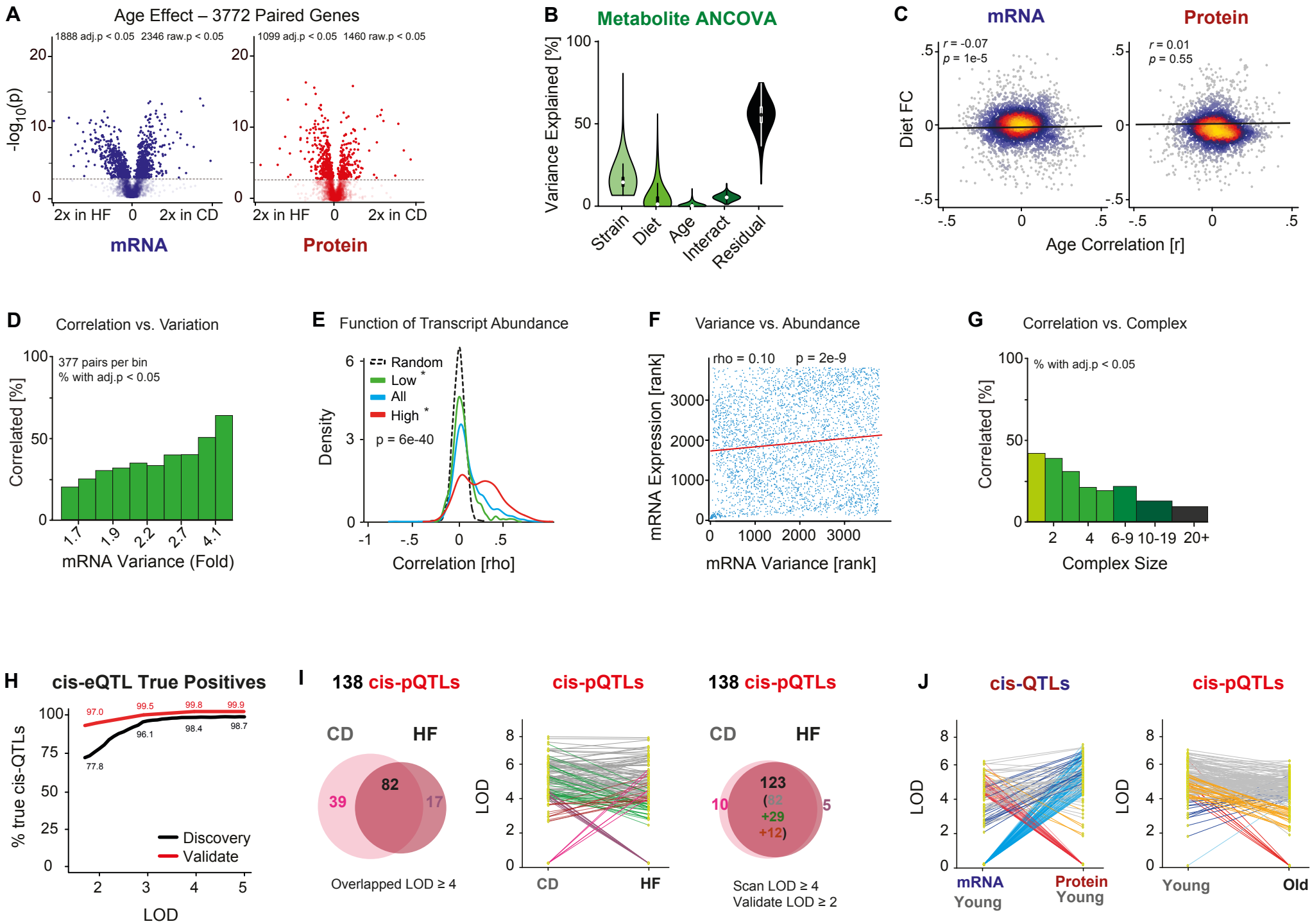


Figure S3. Supplemental Functional Network Analysis and False Discovery

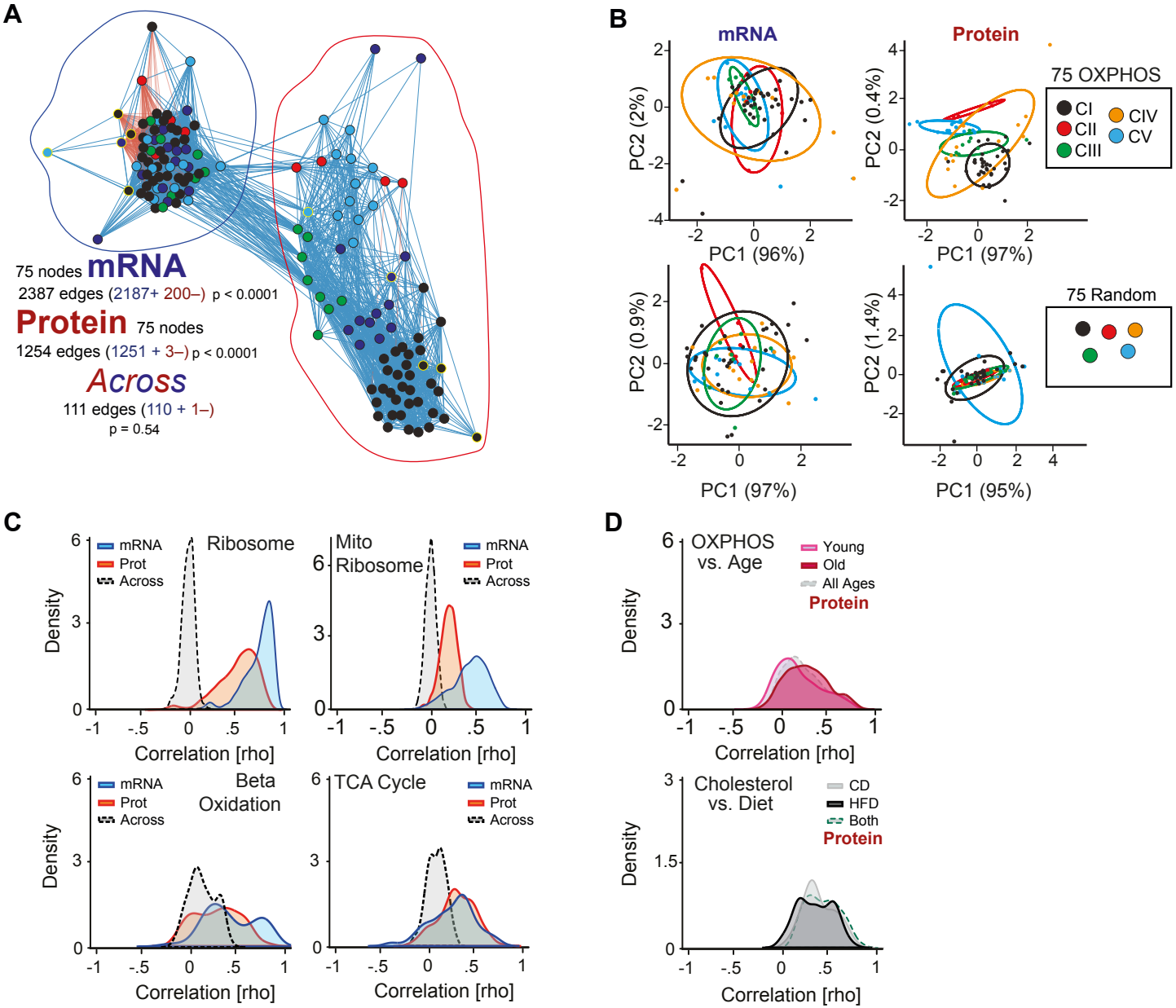


Figure S4. Supplemental Stability Inference and False Discovery

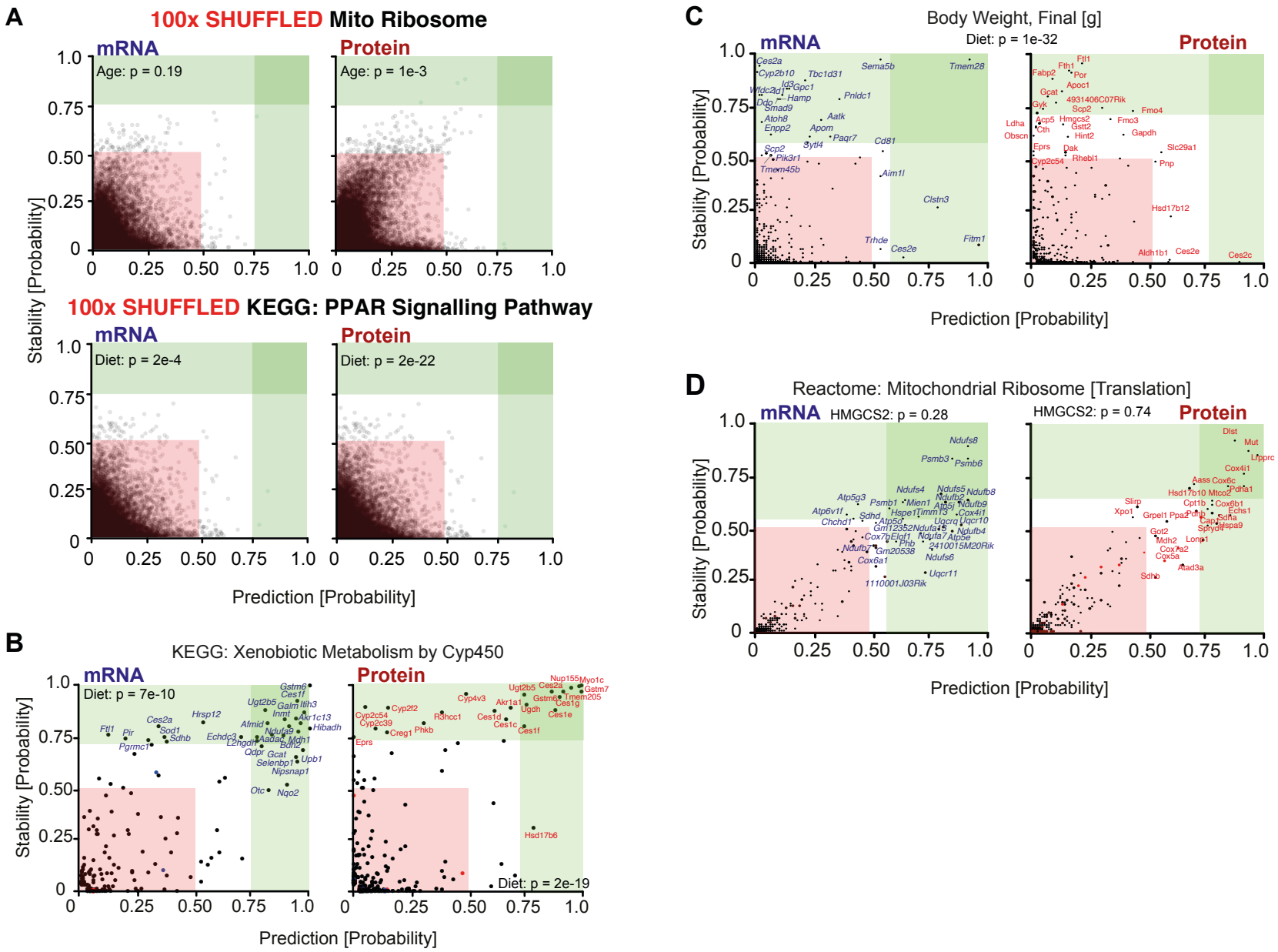
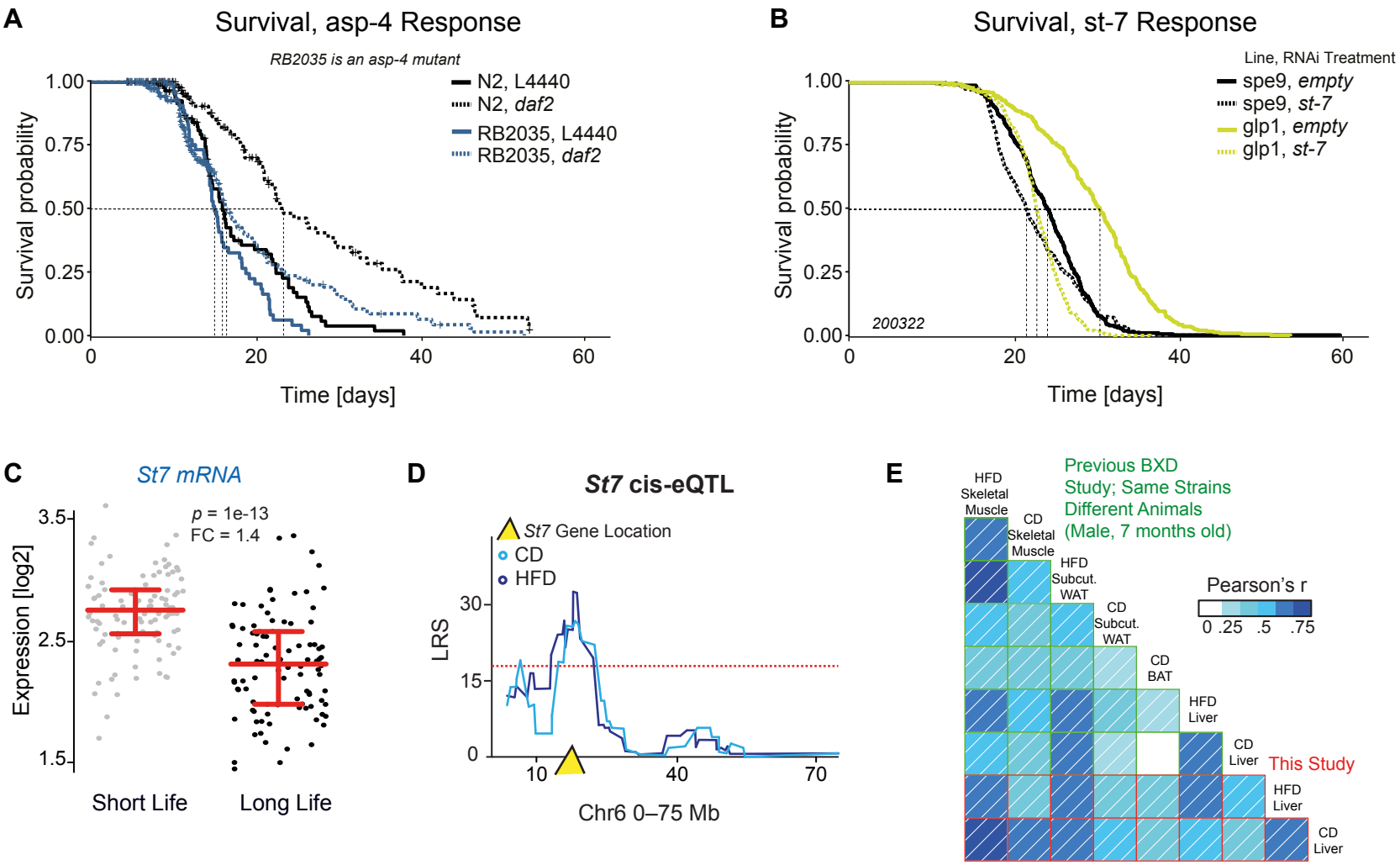



Figure S5. Supplemental Analysis of Aging Candidates