A computational framework identifying concordant gene expression-neuropathology associations reveals Complex I as a potential Alzheimer's disease therapeutic target

Safiye Celik[1], Josh C. Russell[2], Shubhabrata Mukherjee[3], Paul K. Crane[3], C. Dirk Keene[4], Jennifer F. Bobb[5], Matt Kaeberlein[2,*], Su-In Lee[1,6,*]

[1] Paul G. Allen School of Computer Science & Engineering, University of Washington, 185 E Stevens Way NE, Seattle, WA 98195

[2] Department of Pathology, School of Medicine, University of Washington, 1959 NE Pacific St, Seattle, WA 98195

[3] Division of General Internal Medicine, School of Medicine, University of Washington, 325 9th Ave, Seattle, WA 98104

[4] Division of Neuropathology, Department of Pathology, School of Medicine, University of Washington, 325 9th Ave, Seattle, WA 98104

[5] Kaiser Permanente Washington Health Research Institute, 1730 Minor Ave, Suite 1600, WA 98101

[6] Department of Genome Sciences, University of Washington, 3720 15th Ave NE, Seattle, WA 98195

[*] Corresponding authors (suinlee@cs.washington.edu and kaeber@uw.edu)

Abstract

We examine, across nine human brain regions, the spectrum of genome-wide gene expression associations with Alzheimer's disease (AD) neuropathology using 1,746 human individuals from three AD studies. We introduce a new computational approach, DECODER, that leverages discrepancies across different brain regions or different studies in order to identify robust expression markers for complex neuropathological phenotypes. Our computational evaluation experiments demonstrate: (1) the possibility of performing meta-analysis in the highly challenging AD setting where datasets involve study-specific confounders or brain region-specific biological processes, (2) DECODER's potential as a general meta-analysis framework widely applicable to various diseases (e.g., AD, cancer) or phenotypes (e.g., neuropathology, survival), and (3) provide new insights into the similarity of brain regions in terms of expression associations with AD hallmarks. We further extend these computational advances through *in vivo* validation of novel genes using a transgenic *Caenorhabditis elegans* model expressing AD-associated Aβ. Our approach yields several novel genetic modifiers of Aβ toxicity and pinpoints Complex I of the mitochondrial electron transport chain (mETC) as a critical mediator of proteostasis and a promising potential pharmacological avenue toward treating AD.

Introduction

Alzheimer's disease (AD) is the most common cause of dementia and the 6th most common cause of death in the United States. The prevalence of AD is expected to double by mid-century[1] due in large part to increasing life expectancies. No effective therapy exists to delay or prevent AD onset and progression.[2] Neuritic plaques and neurofibrillary tangles are two neuropathological hallmarks of AD. They are pathologically represented by Amyloid-β (Aβ) peptide and tau protein, respectively. Aβ is a 40 to 42-amino acid peptide, which is generated by proteolytic cleavages of the amyloid precursor protein (*APP*) located on chromosome 21.[3] While the precise function of *APP* is not known, mutations in this gene have been suggested to lead to familial susceptibility to AD[4]. The amyloid cascade hypothesis[5] posits that aggregation and extracellular deposition of the misfolded Aβ peptide (specifically, the $A\beta_{1\text{-}42}$ isoform) plays a causal role in several AD-related events, including neurofibrillary tangle formation, resulting in neurodegeneration and neuronal death.[6] However, causality remains unproven.[7] The amyloid cascade hypothesis[5] has given rise to several Aβ-centric treatment approaches

for AD such as blocking *APP* cleavage by β-secretase inhibitors. This approach has been attempted as a strategy for reducing Aβ formation and deposition, but several concerns have been raised about the efficacy of this strategy.[8] Tau, the other hallmark of AD, is a brain-specific, axon-enriched, microtubule-associated protein[9] encoded by a gene located on chromosome 17. Tau is phosphorylated post-translationally and has phosphorylation levels that are significantly higher in brains from AD patients, suggesting that hyperphosphorylation may lead to tau's pathogenic role in AD.[10]

In recent years, significant research has focused on identifying genes that influence the risk of developing, age of onset, and progression of AD. Mutations in genes *APP*, *PSEN1*, and *PSEN2* are known to cause familial early-onset AD.[11] In addition, a handful of bona fide genes for late-onset sporadic AD have been identified by genome-wide association studies (GWAS), including *APOE*, *CLU*, *ABCA7*, *SORL1*, *PICALM*, *PLD3*, and *ADAM10*.[12] However, those genetic modifiers account for only a small fraction of AD risk[13,14] and provide no clear picture of AD neuropathology mechanisms. In addition to genetic markers, identifying reliable *gene expression markers* may help identify molecular mechanisms of AD pathogenesis. At present, we lack well-founded knowledge of the set of genes whose expression affects the formation of neuritic plaques and neurofibrillary tangles and the protective and pathological responses to these purportedly toxic lesions: until recently, we have had limited access to gene expression data and neuropathological phenotypes from post-mortem brain tissues. Biologists have only now begun gathering both gene expression data and Aβ and tau measures from human brain tissues[15], providing a new paradigm of system-level AD data for information mining. These novel bioinformatics resources may provide complementary approaches that permit the development of new hypotheses about fundamental AD mechanisms.

The standard computational approach to identifying expression markers focuses on genes whose expression levels are statistically associated with a phenotype of interest[3–6] (e.g. neuropathological phenotypes of neuritic plaques or neurofibrillary tangles). Unfortunately, false positive findings are very common in these sorts of biomarker analyses, as evidenced by the low success rates (less than 1%) of replication in independent datasets and translation to clinical practice.[3–6] The high false positive rate indicates two challenges posed by the current approach. First, high dimensionality, hidden variables, and feature correlations create a discrepancy between predictability (i.e., statistical associations) and true biological interactions; we need new feature selection criteria to filter out false positive associations. Second, correlational results from observational data without supporting results from interventional experiments do not prove causal associations. In this paper, we present an integrative approach that resolves both challenges.

The increase in the number of studies that attempt to obtain brain gene expression and neuropathology data makes meta-analysis an increasingly powerful approach for reducing false positive findings. As part of the Accelerating Medicines Partnership (AMP)-AD project, Sage Bionetworks recently began sharing and integrating multi-dimensional human 'omic' data from more than 2,000 human brains autopsied in several studies through the AMP-AD Knowledge Portal.[16] Of those studies, the Religious Orders Study[17] and the Memory and Aging Project[18] (ROSMAP) are longitudinal, clinical-pathologic cohort studies of aging and AD; they provide detailed neuropathology quantifications and several types of omics data, including RNA-Seq, miRNA, DNA methylation, and histone modification measured in the dorsolateral prefrontal cortex of more than 700 brains. The Mount Sinai Brain Bank (MSBB) study provides microarray data and partial neuropathology information from 19 regions of more than 60 brains[19] and RNA-Seq data from four regions of more than 200 brains. The Adult Changes in Thought (ACT) study —a longitudinal, population-based prospective cohort study of brain aging and dementia— provides detailed neuropathology quantifications and RNA-Seq data from four regions of around 100 autopsied brains.[20]

Identifying expression markers for neuropathological phenotypes through the meta-analysis of these studies involves three major challenges. First, expression datasets obtained from different AD studies often present data

from different brain regions that contain different cell type compositions, have different functions, and differ in their relevance to AD. While we have a general knowledge of how AD progresses across the brain[21], it is not of an extent that lets us assign a specific weight value to each brain region in a meta-analysis setting to identify AD neuropathology mechanisms. Second, datasets from different studies involve confounding factors due to differences in subject compositions or experimental procedures used to generate data (i.e., severe batch effects). Finally, the brain tissues used for expression and phenotype profiling do not always match, and each study uses slightly different methods to measure Aβ or tau levels. These factors lead to different sources and levels of noise across studies. Natural questions that arise are (1) whether we can leverage discrepancies across studies or regions to filter out false positive expression markers by focusing on concordant expression associations across different sources, and (2) what the implications of the regional spectrum of associations between expression and neuropathological phenotypes are.

We present DECODER (discovering concordant expression markers), an integrative computational framework to identify robust expression markers for a phenotype (e.g., Aβ or tau levels in AD brains) using three concordance-based computational approaches. These approaches work on the gene expression-phenotype associations space rather than the actual expression and neuropathology spaces, which helps us address discrepancies across different studies, tissues, or brain regions. In other words, the Pearson's correlation coefficient between expression and phenotype behaves as a kernel that projects the original expression and phenotype data to a new space in which different datasets become more comparable. We show that concordant expression markers substantially improve statistical and biological consistency compared to expression markers from individual studies, tissues, or brain regions considered separately.

Our experimental results can be divided into three main categories. First, using DECODER, we identify genes that are consistently associated with quantified levels of Aβ across multiple human brain regions. This demonstrates the possibility of performing meta-analysis in the highly challenging AD setting where datasets involve study-specific confounders or brain region-specific biological processes. Second, by applying DECODER to expression-tau associations in AD and expression-patient survival associations in cancer, we show that DECODER is a general meta-analysis framework widely applicable to various phenotypes or diseases. Third, we experimentally validate the identified expression markers for Aβ in an animal model of Aβ proteotoxicity *in vivo*. We show that a mitochondrial Complex I gene identified by DECODER and 13 additional Complex I genes attenuate disease in a nematode genetically modified to express human $A\beta_{1-42}$ which causes an age-related paralysis phenotype. Our experimental results suggest that Complex I may be a promising therapeutic target for AD.

Results

This work seeks to identify robust expression markers for disease phenotypes, specifically expression markers for Aβ neuropathological phenotype which could be used as therapeutic targets for individuals with AD. We leverage concordance of expression associations with neuropathological phenotypes to filter out noise and extract information likely to be enriched for true signals from "big", high-throughput brain tissue data. We hypothesized that even if different brain regions are affected by AD in different stages of the disease or in different severities across individuals in the same stage of AD, the basic molecular mechanisms driving the development of and response to neuropathology may be common across regions. Thus, we believe that genes whose associations with neuropathology are concordant in multiple brain regions and studies are more likely to be true molecular markers, and we can use association concordance to reduce the dimensionality of gene expression data into a smaller set of highly informative genes.

In our experiments, we employ the ACT study and the AMP-AD RNA-Seq studies that provide quantifications of neuropathology findings.[16] Thus, the data we use comes from a total of nine brain regions and 1,746 post-mortem brain tissue samples described in three different studies — ROSMAP,[17,18] ACT,[20,22] and MSBB. Our analysis included 14,912 protein-coding genes that have a nonzero RNA-Seq read count in at least one-third of the samples in each study and overlapping across the three studies. The nine regions we used and the number of samples that have both gene expression levels and Aβ quantification are listed in

Table *1* by brain region. The Methods section provides details about our data collection.

Table 1. The nine regions from three studies we used in this work, and the corresponding number of samples that have both gene expression and Aβ quantifications and hence were used to compute gene expression-Aβ associations.

Region acronym	**Region description**	**Sample size**
ROSMAP-DPFCx	Dorsolateral prefrontal cortex from the ROSMAP study	542
ACT-TCx	Temporal cortex from the ACT study	85
ACT-PCx	Parietal cortex from the ACT study	79
ACT-HIP	Hippocampus from the ACT study	80
ACT-FWM	Forebrain white matter from the ACT study	81
MSBB-BM10	Brodmann area 10 (frontopolar cortex) from the MSBB study	244
MSBB-BM36	Brodmann area 36 (ectorhinal cortex) from the MSBB study	200
MSBB-BM44	Brodmann area 44 (opercular part of the inferior frontal gyrus) from the MSBB study	208
MSBB-BM22	Brodmann area 22 (posterior part of the first temporal gyrus) from the MSBB study	227

Figure 1 summarizes the DECODER methodology. We introduce three different methods to score the genes based on their concordant associations with neuropathology levels in different brain regions. We refer to highly ranked genes by those three concordance-based methods as "concordant expression markers" abbreviated as "CEMs". We investigate the performance of each new scoring system using multiple statistical and biological evaluation metrics. We demonstrate that all three concordance-based methods are statistically more robust than individual region-based methods. Further, the CEMs are biologically more relevant to AD than the top genes from individual region-based methods. Based on our computational evaluations, we obtained a small subset of genes that are potential drivers for Aβ pathology and conducted biological experiments to validate our findings.

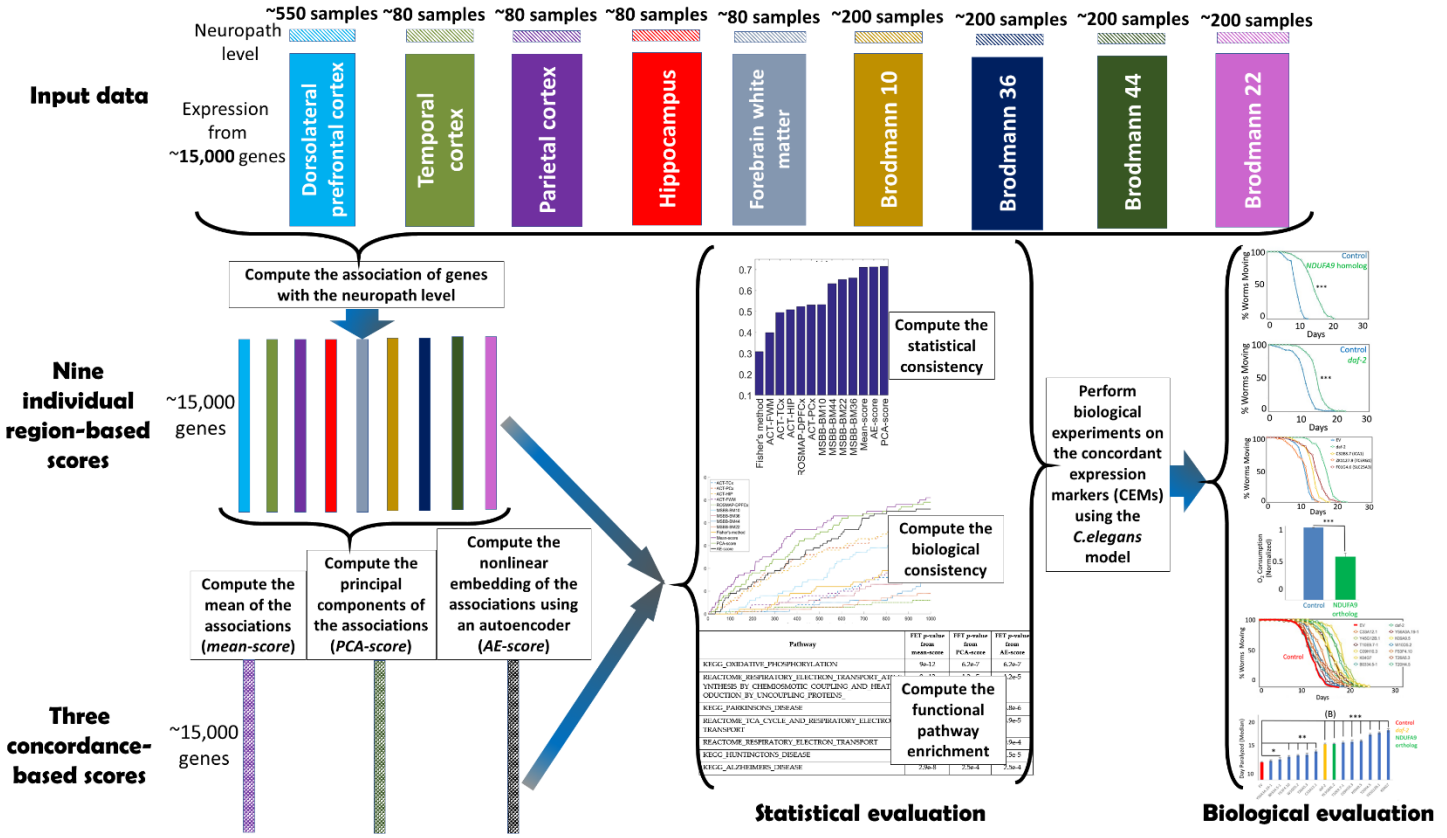

Figure 1. DECODER methodology. We introduce three different methods to score the genes based on their concordant associations with neuropathology levels in different brain regions and evaluate them using our statistical and biological evaluation experiments.

Our results can be summarized in the below ten results which we discuss in the next ten sections:

Result 1: Genes associated with neuropathology tend to be concordant across regions and studies.

Result 2: Brain regions relevant to AD are highly weighted by concordance-based gene ranking approach.

Result 3: Concordance-based methods identify statistically robust markers of AD pathology.

Result 4: CEMs are highly enriched for genes known to be relevant to AD.

Result 5: CEMs are highly enriched for pathways relevant to AD.

Result 6: Concordance-based methods are robust to different settings including AD tau phenotype or cancer survival.

Result 7: The *NDUFA9* gene is biologically validated as a modifier of Aβ toxicity.

Result 8: RNAi knockdown of *NDUFA9*'s nematode homolog strongly decreases whole animal oxygen consumption.

Result 9: *In vivo* validation of *NDUFA9* identifies mitochondrial Complex I as a potential AD target.

Result 10: Multi-omic module subnetwork of Complex I highlights mechanisms relevant to AD.

Result 1. Genes associated with neuropathology tend to be concordant across regions and studies.

We first computed the association of each gene's expression with Aβ levels in each study. Interestingly, for most pairs of the nine regions, we observed a significant overlap in the top genes whose expression is significantly associated with Aβ levels in that region (Table 2). ACT-FWM was relatively less concordant with the other eight regions, which is expected since these regions contain mostly gray matter expression and neuropathology, and gray and white matter are known to be biologically and functionally distinct.

Table 2. Fisher's exact test p-value associated with the overlap between the top 1000 Aβ-associated genes in each region pair (1st column). We highlight in red all pairs that have a statistically significant overlap at $p < 0.1$.

	ACT-TCx	**ACT-PCx**	**ACT-HIP**	**ACT-FWM**	**MSBB-BM10**	**MSBB-BM36**	**MSBB-BM44**	**MSBB-BM22**
ROSMAP-DPFCx	4.1e-37	1.6e-3	0.01	1	0.09	1.6e-13	2.5e-9	7.3e-14
ACT-TCx		2.8e-81	7.8e-17	1.2e-7	3.4e-3	8.1e-13	1.1e-5	1.8e-8
ACT-PCx			5.1e-36	2.2e-43	9.2e-29	3.7e-45	1.8e-35	2.1e-34
ACT-HIP				3.4e-3	8.7e-56	2.4e-46	1.4e-47	2.4e-26
ACT-FWM					0.37	1	1	1
MSBB-BM10						1.4e-117	7.1e-236	4.7e-114
MSBB-BM36							< 2.3e-308	< 2.3e-308
MSBB-BM44								< 2.3e-308

Figure 2 shows a heatmap of Aβ associations for the ~15,000 genes we included in our analysis. The genes and the regions are hierarchically clustered based on their similarities in terms of the strength of association of gene expression with the Aβ phenotype. The rows and columns are rearranged so that similar regions or genes are grouped together. Bright red and green portions of the heatmaps correspond to the clusters. Although regions from the same study tend to group together, the figure shows that MSBB-BM10 connects to ROSMAP-DPFCx in the Aβ associations dendrogram before it connects to other regions from the same study. This observation is consistent with the fact that Brodmann area 10 is in close proximity to the dorsolateral prefrontal cortex in the brain.

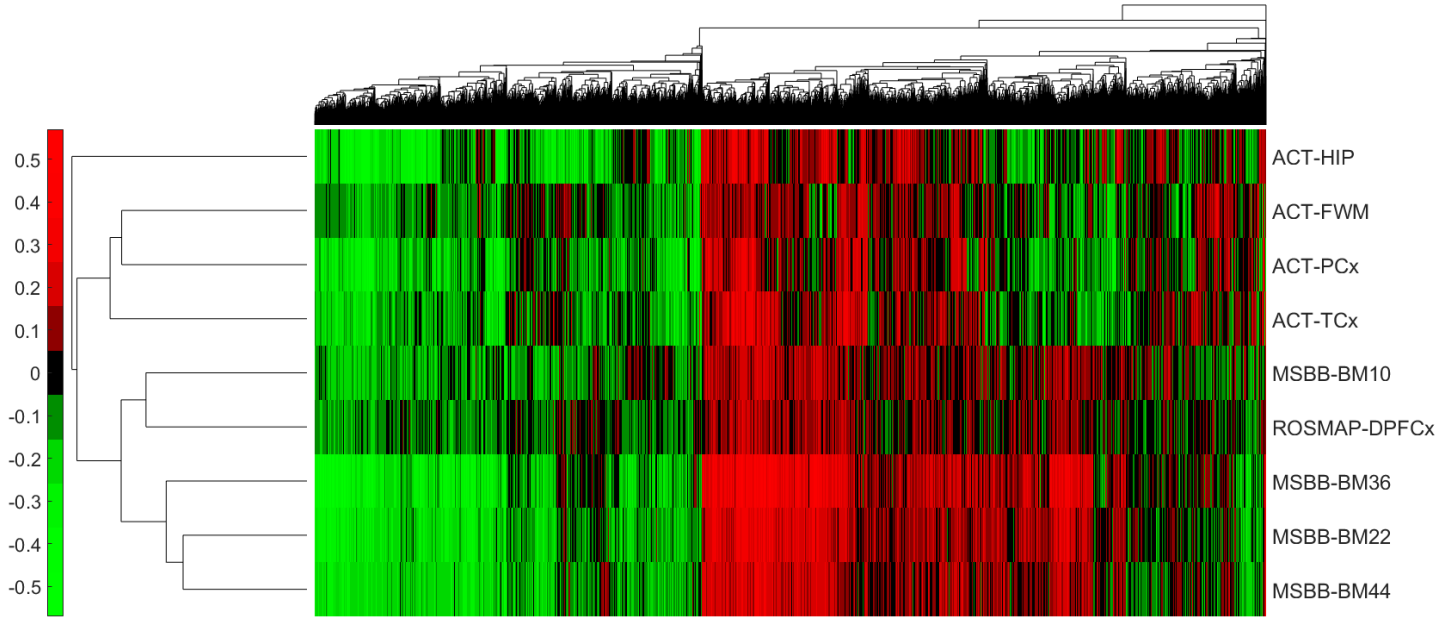

Figure 2. A heatmap of the associations between Aβ levels and gene expression across different regions. Each row corresponds to a region, and each column corresponds to a gene. Green indicates a negative association and red a positive one.

Result 2. Brain regions relevant to AD are highly weighted by concordance-based gene ranking approach.

We hypothesize that the development of neuropathology is driven by similar molecular mechanisms in different brain regions, which is also supported by the results displayed in Table 2 and Figure 2. Therefore, focusing on the concordant molecular markers across different regions could lead to the discovery of reliable genes driving Aβ neuropathology. We developed three different approaches to generate a single quantitative score for each gene based on that gene's association with neuropathology in multiple regions from different studies. We: (1) computed the mean of the expression-neuropathology associations across regions and used these mean values as gene scores, (2) applied principal component analysis (PCA) to the associations to project all genes to the direction of the maximum variance in the data and used as gene scores the first principal component (PC1), which is a weighted linear combination of individual regions' associations, and (3) learned an auto-encoder (AE) network from the associations to identify a one-dimensional latent embedding as a nonlinear combination of the individual regions, and we used the embedded values as gene scores (Methods). We name the new gene scores identified by each of the three methods as "mean-score", "PCA-score", and "AE-score", respectively.

Mean-score and PCA-score both linearly combine regional associations (Methods). The two methods differ in that the mean-score equally weights the regions, while the PCA-score learns the weights associated with each different region from the input data. We observed that, when applied to the gene expression-Aβ associations, PCA highly weights regions previously known to be relevant to AD (Figure 3). The regions MSBB-BM36 and MSBB-BM22 that coincide with the medial temporal lobe, the brain part where AD-related cellular and structural alterations begin and have a more severe effect,[23] are the top two regions with the highest weights. White matter is suggested to be less vulnerable to AD than gray matter,[24] and interestingly, the white matter region ACT-FWM is assigned the lowest weight. Similarly, dorsolateral prefrontal cortex region is affected in later stages of AD progression,[21] and ROSMAP-DPFCx was also assigned a relatively lower weight.

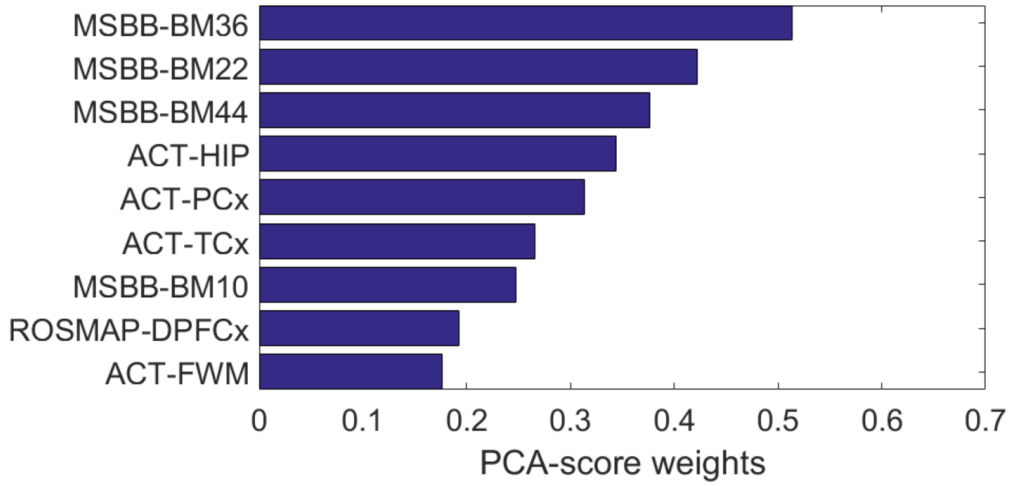

Figure 3. PC1 loadings (i.e., PCA-score weights) associated with each of the nine brain regions for the gene expression-Aβ associations. Brain regions are weighted consistently with their relevance to AD neuropathology.

AE is an artificial neural network-based method that models latent space using a nonlinear combination of input variables. Thus, different than the mean-score and PCA-score, AE-score nonlinearly combines the individual regions (Methods). We conjecture that nonlinear embeddings could be useful for complex interactions among different regions that linear methods cannot capture. Note that how we employ PCA or AE differs from common usage of those methods: we treat the genes as the samples and the regions as the variables (i.e., features). Scores computed by the three concordance-based methods are highly consistent with each other, as shown in Supplementary Figure 1.

Result 3. Concordance-based methods identify statistically robust markers of AD pathology.

We performed a cross-validation (CV) experiment to test whether the gene scores identified based on concordance across multiple regions (i.e., mean-score, PCA-score, or AE-score) are statistically more robust and informative than scores from individual region-based methods (i.e., the association of gene expression with Aβ levels in each individual region). We also compared our concordance-based methods with Fisher's method, a widely used approach for deriving a single p-value from multiple associations with different significance levels (i.e., p-values). Fisher's method assigns a score to each gene that is proportional to the mean negative logarithm of the individual p-values.

In our CV experiment to test statistical consistency, for each of the nine regions as a test fold, we used the other eight regions for training. We computed, for each test fold, how well the gene scores from each of the eight individual regions, Fisher's method, the mean-score, the PCA-score, and the AE-score can estimate the test fold scores. Note that all methods except individual region-based methods computed scores based on the eight training regions. We used a Spearman rank correlation between the actual vs. estimated gene scores to measure each method's performance. We observed that concordance-based methods outperformed all other methods in statistical consistency averaged across all nine regions used as a test fold (Figure 4). Thus, the gene rankings are statistically more robust (i.e. likely to be replicated in an independent, unobserved dataset) when obtained based on the concordance of expression-neuropathology associations across regions compared to when determined based on an individual region or Fisher's method. While the three concordance-based methods performed similarly in terms of statistical consistency, we observed a slight increase in performance with an increase in

complexity of the method (Methods) that models the concordance (i.e., AE-score performed slightly better than PCA-score, which also performed slightly better than mean-score).

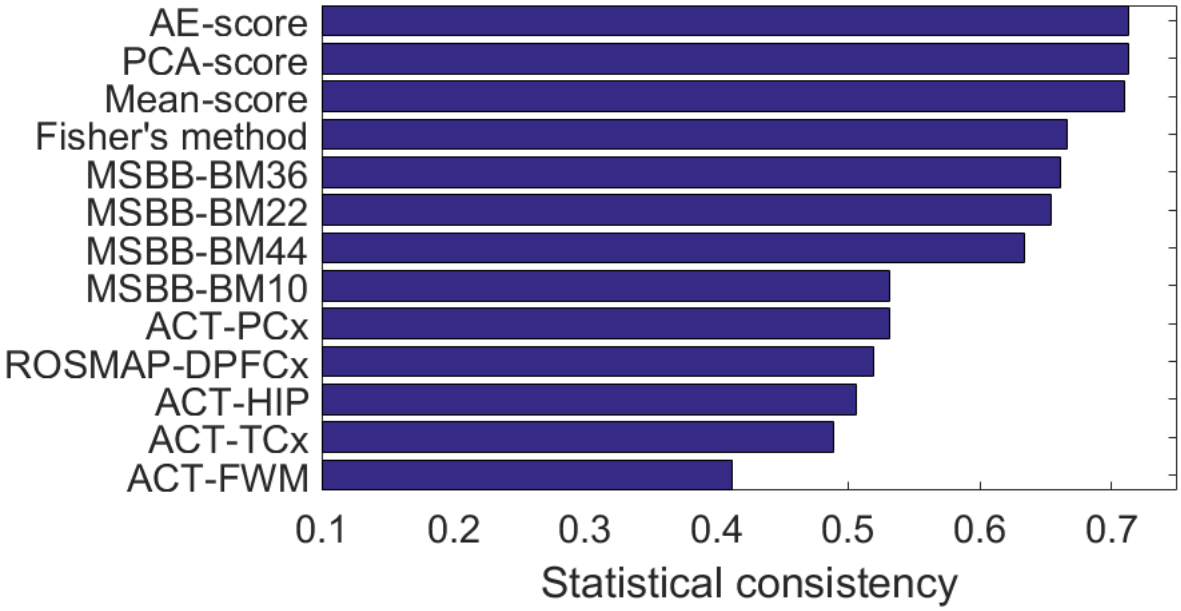

Figure 4. A comparison of the statistical consistency achieved by three concordance-based methods we introduce (mean-score, PCA-score, and AE-score) to the commonly used Fisher's method and nine individual region scores for Aβ.

Different than other methods in comparison, Fisher's method works with p-values rather than the actual associations. Therefore, to compute Fisher's method consistency in a way comparable to other methods, we averaged the consistency results over two different choices of Fisher's method. One choice combines p-values from a one-sided correlation test with alternative hypothesis of negative correlation, and the other with alternative hypothesis of positive correlation. Supplementary Figure 4 includes Fisher's method bars separately from those two alternatives.

The relatively poor performance of Fisher's method is likely because Fisher's method by its nature puts a higher weight on highly sampled regions since smaller association p-values are often obtained from those; however, highly sampled regions may not be the most relevant to AD. In our experimental setting of nine regions, ROSMAP-DPFCx includes the highest number of samples (

Table *1*). However, the dorsolateral prefrontal cortex is usually affected later as AD progresses[21] and is not as relevant to AD as several other regions with a smaller sample size in our data.

Result 4. CEMs are highly enriched for genes known to be relevant to AD.

In addition to statistical relevance, we also assessed the biological relevance of the top-scoring genes from the mean-score, PCA-score, and AE-score methods. To do this, we checked whether the CEMs identified by concordance-based methods are more likely to be relevant to AD compared to the top genes identified by individual region-based methods and Fisher's method. The biological relevance metric we use is the number of genes overlapping with the KEGG_ALZHEIMERS_DISEASE pathway (AD pathway) from the C2 collection (curated gene sets from online pathway databases) of the Molecular Signatures Database (MSigDB).[25] This

pathway contains 169 genes in total, 144 of which exist in all gene expression datasets from the nine brain regions we use in our experiments (Supplementary Table 1).

For each of the nine individual region-based methods, Fisher's method, mean-score, PCA-score, and AE-score, we selected top N genes where genes are sorted based on their score. Then, for each $n = \{1, \ldots, N\}$, we analyzed the number of genes that overlap with the 144 AD pathway genes. We repeated the analysis for the "negative tail" and "positive tail" from each method where those terms refer to genes that are assigned by the method highly negative scores and highly positive scores, respectively. Figure 5 presents the results averaged over the two tails for each method. In Figure 5, the curves corresponding to the mean-score, PCA-score and AE-score (purple, green and black solid curves, respectively) are above all other curves. That is, concordance-based methods identify more true positives earlier than alternative methods (i.e., individual region-based methods represented by nine non-solid curves and Fisher's method represented by the yellow solid curve).

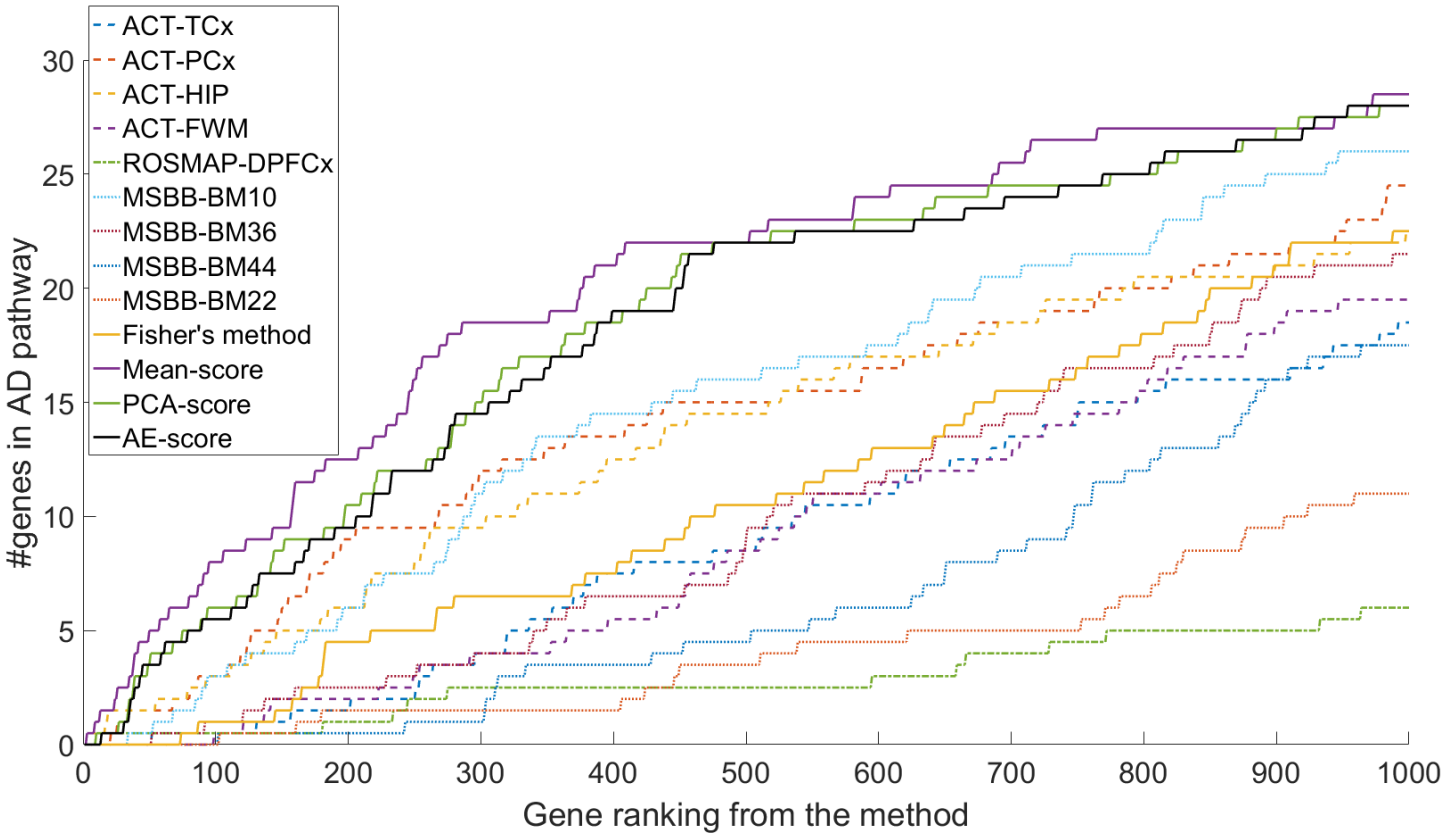

Figure 5. A comparison of the biological consistency achieved by three concordance-based methods we introduce (mean-score, PCA-score, and AE-score) to the commonly used Fisher's method and nine individual region scores for Aβ. Consistency is measured by the overlap between the top $n = \{1, \ldots, 1000\}$ genes from each method (x-axis) and the 144 genes in the KEGG AD pathway.

The negative and positive tails for the Fisher's method refer to genes with smallest p-values from the choice of Fisher's method that combines p-values from the correlation test with alternative hypothesis of negative and with alternative hypothesis of positive correlation, respectively.

Supplementary Figure 5 shows the biological consistency results for up to 1000 genes ($N = 1000$) separately for the positive and negative tails for each method in comparison. Interestingly, the positive tail is not enriched for the AD pathway genes; only a few of the top 1000 positive tail CEMs exist in the AD pathway (Supplementary

Figure 5B) while several dozens of the top 1000 negative tail CEMs exist in the same pathway (Supplementary Figure 5A).

Result 5. CEMs are highly enriched for pathways relevant to AD.

To further investigate the biological relevance of CEMs, we examined functional enrichment of the 50 CEMs from the negative and positive tails of the mean-score, PCA-score, and AE-score where those tails correspond to genes that are assigned highly negative and highly positive scores, respectively. We considered 1,077 Reactome, BioCarta, and KEGG GeneSets (canonical pathways) from the C2 collection (curated gene sets from online pathway databases) of the MSigDB.[25] We computed the significance of the overlap between each GeneSet and the 50 CEMs measured by Fisher's exact test (FET) p-value and then applied false discovery rate (FDR) correction for multiple hypotheses testing for the 1,077 pathways.

The 50 CEMs from the Aβ negative tail of the mean-score, PCA-score and AE-score were all significantly enriched for seven pathways (FET $p \leq 0.05$), as shown in Table 3. These include known pathways for neurodegenerative diseases, including Parkinson's disease and Huntington's disease in addition to AD. Other pathways in the table involved the electron transport chain or oxidative phosphorylation, implicated in a variety of neurodegenerative processes.[26–28]. Interestingly, the positive tail did not exhibit significant pathway enrichment for any of the three methods. Since the mean-score generally achieved a more significant pathway enrichment than the PCA-score and AE-score (Table 3), we decided to pay particular attention to CEMs on the negative tail of the mean-score for our further experiments with CEMs.

Table 3. Fisher's exact test p-values for the seven Reactome, BioCarta, and KEGG pathways for which the top 50 CEMs from the negative tail (i.e., top 50 genes with highly negative scores) of the mean-score (2nd column), PCA-score (3rd column), or AE-score (4th column) are significantly enriched ($p \leq 0.05$). All seven pathways contain the NDUFA9 gene.

Pathway	FET *p*-value from mean-score	FET *p*-value from PCA-score	FET *p*-value from AE-score
KEGG_OXIDATIVE_PHOSPHORYLATION	9.0e-12	6.2e-7	6.2e-7
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT_ATP_SYNTHESIS_BY_CHEMIOSMOTIC_COUPLING_AND_HEAT_PRODUCTION_BY_UNCOUPLING_PROTEINS_	9.0e-12	1.2e-5	1.2e-5
KEGG_PARKINSONS_DISEASE	1.9e-10	6.8e-6	6.8e-6
REACTOME_TCA_CYCLE_AND_RESPIRATORY_ELECTRON_TRANSPORT	2.7e-10	8.9e-5	8.9e-5
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT	9.3e-10	9.9e-4	9.9e-4
KEGG_HUNTINGTONS_DISEASE	3.5e-9	4.5e-5	4.5e-5
KEGG_ALZHEIMERS_DISEASE	2.9e-8	2.5e-4	2.5e-4

Result 6. Concordance-based methods are robust to different settings including AD tau phenotype or cancer survival.

To test the applicability of the DECODER approach to other phenotypes or diseases, we also applied it to (1) gene expression-tau associations from the nine brain regions from the AD studies and (2) gene expression-survival associations from 33 cancer types from The Cancer Genome Atlas (TCGA) study. Through these

experiments, we aimed to obtain the gene markers whose expression levels are predictive of tau pathology across different brain regions and cancer patient survival across different cancers, respectively.

Figure 6 shows a heatmap of gene expression-tau neuropathology associations and the statistical and biological consistency results from applying concordance-based methods and the alternative methods to these associations. Each row in the heatmap (Figure 6A) represents one of the nine brain regions in

Table *1*, and each column represents one of the 14,912 protein-coding genes whose expression was measured in all nine regions. As was the case in the heatmap for expression-Aβ associations (Figure 2), regions from the same study tend to group together in the tau associations dendrogram. Still, MSBB-BM44, a frontal cortex region from the MSBB study, connects to another frontal cortex region from a different study, ROSMAP-DPFCx, before it connects to other cortical regions from the MSBB study. Concordance-based methods estimated unobserved test region scores more accurately than individual region-based methods and equally accurately as Fisher's method (Figure 6B). Moreover, of 14,912 total genes, 144 genes that are in the KEGG_ALZHEIMERS_DISEASE pathway exhibited a large overlap with tau CEMs earlier than the highly ranked genes by individual region-based methods or Fisher's method (Figure 6C).

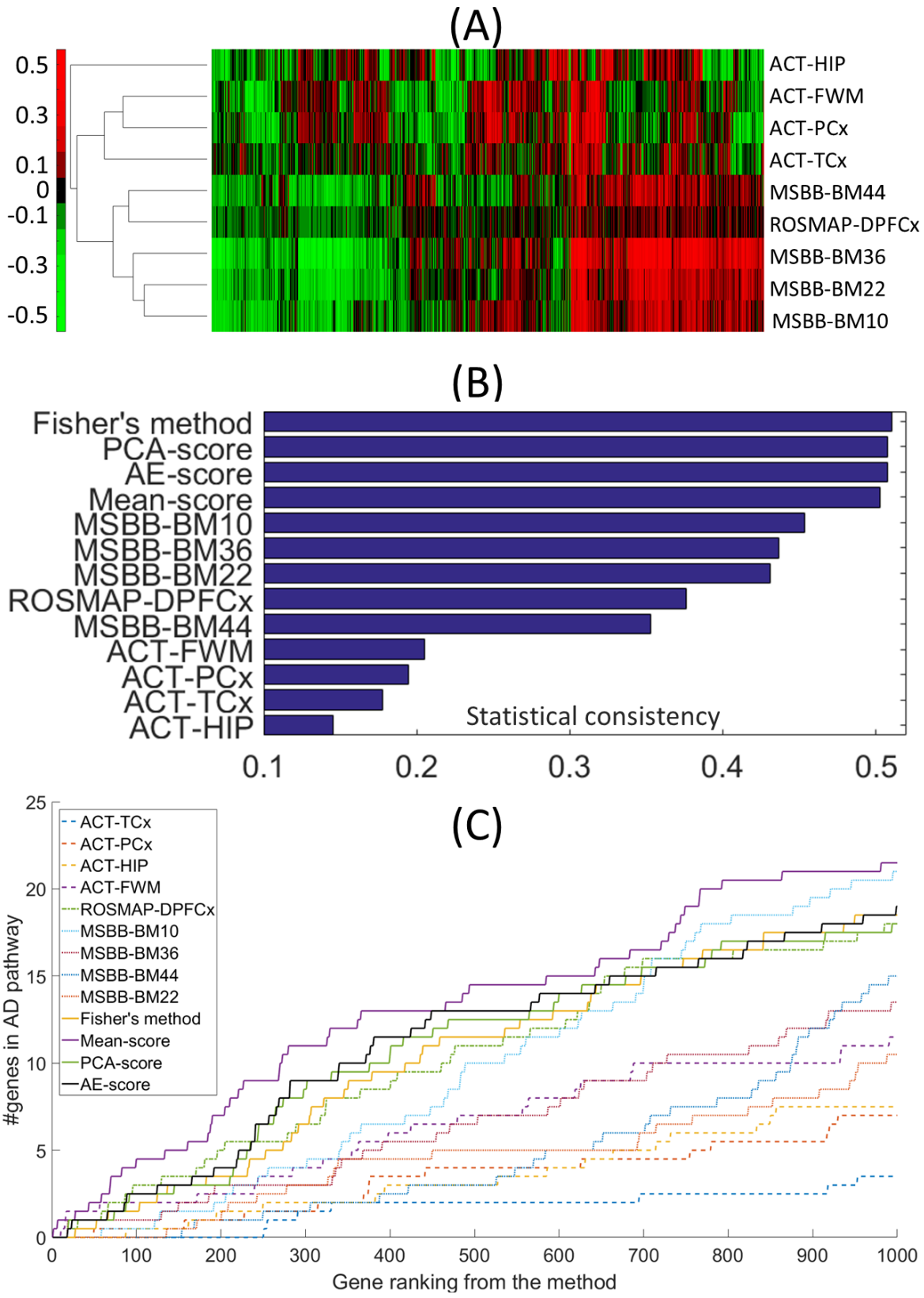

Figure 6. Testing of concordance-based methods for gene expression-tau neuropathology associations. ***(A)*** *A heatmap of the associations between tau levels and gene expression across different regions.* ***(B)*** *Comparison of concordance-based methods to individual region-based methods and Fisher's method in terms of the estimation of unobserved test region scores.* ***(C)*** *Comparison of concordance-based methods to individual region-based methods and Fisher's method in terms of overlap between KEGG AD pathway genes and the highly ranked genes by each method.*

Figure 7 shows a heatmap of TCGA gene expression-survival associations and the statistical and biological consistency results from applying concordance-based methods and the alternative methods to these associations. Each row in the heatmap (Figure 7A) represents one of the 33 cancer types, and each column represents one of the 15,097 genes included in the TCGA RNA-Seq data from all cancer types. Cancer is a heterogeneous disease, and survival is a phenotype that could be affected by a diverse set of factors. So, not surprisingly, we did not observe a visible clustering structure in the dendrogram of gene expression-phenotype associations. Still, concordance-based methods, specifically PCA-score and AE-score, estimated unobserved gene scores from the test cancer type (Figure 7B) more accurately than the alternative methods. Furthermore, of 15,097 total genes, 118 genes that are in the KEGG_CELL_CYCLE pathway exhibited a large overlap with survival CEMs earlier than the highly ranked genes by individual region-based methods or Fisher's method (Figure 7C). This is not surprising given the dysregulation of cell cycle-regulated genes is common in many cancers and has been commonly associated with poor prognosis.[29–31]

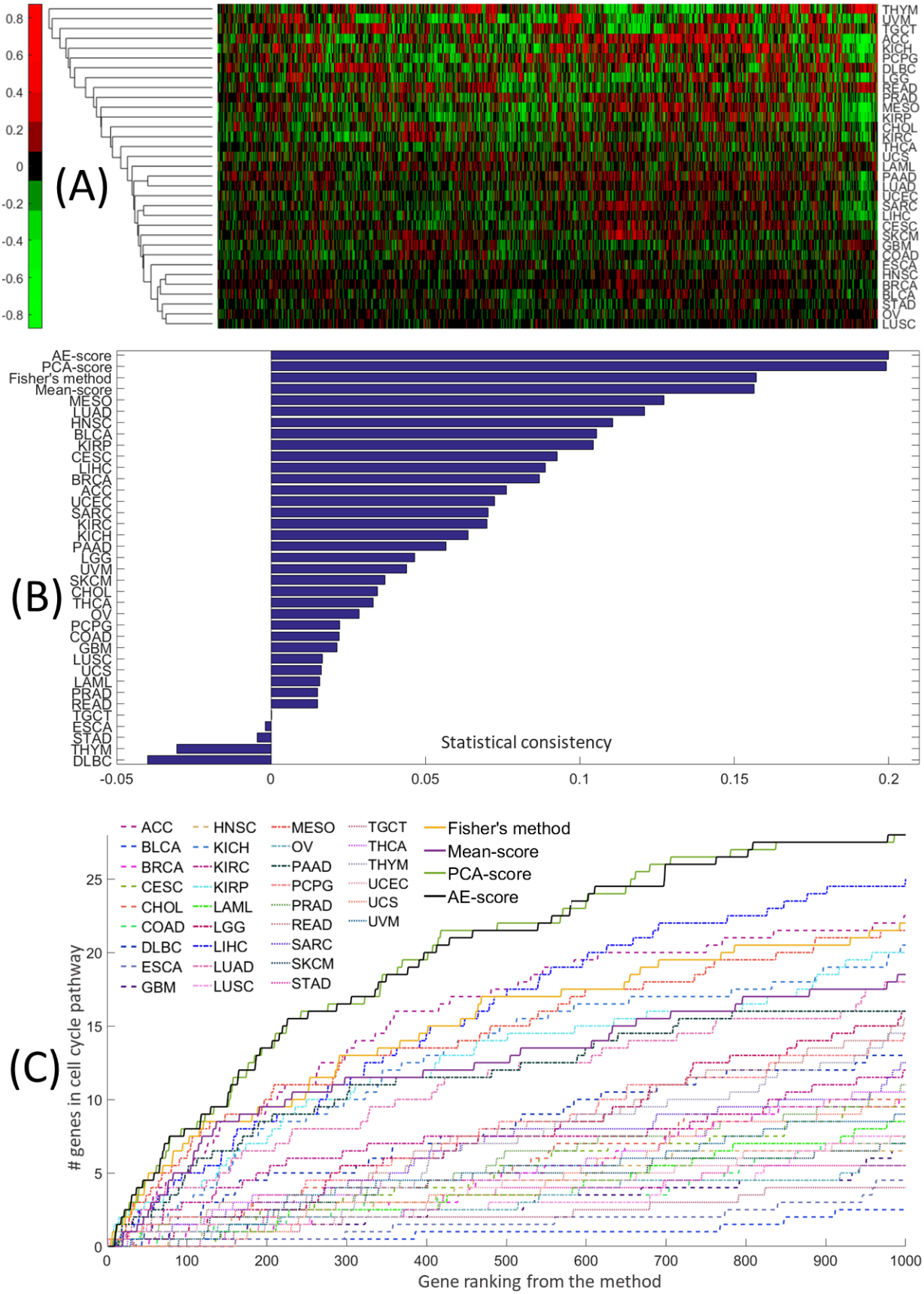

Figure 7. Testing of concordance-based methods for TCGA gene expression-survival associations from 33 cancer types. ***(A)*** *A heatmap of the associations between survival and gene expression across different cancer types. (B) Comparison of*

concordance-based methods to individual region-based methods in terms of the estimation of unobserved test cancer type scores. ***(C)*** *Comparison of concordance-based methods to individual region-based methods in terms of overlap between KEGG AD pathway genes and the highly ranked genes by each method.*

Result 7. The NDUFA9 gene is biologically validated as a modifier of Aβ toxicity.

To gain insight into the relevance of individual CEMs to Aβ toxicity, we utilized the nematode *C. elegans* as an animal model of Aβ proteotoxicity. Over the last two decades, *C. elegans* has been a particularly useful model system for studying genetic pathways that influence proteotoxicity since transgenic strains are easily constructed and maintained, and genes are efficiently knocked down by RNAi.[32,33] We conducted experiments with a transgenic worm line displaying an age-associated aggregation of human $A\beta_{1-42}$ peptide in their body wall muscle cells.[34] The proteotoxic stress induced by the human transgene results in rapid onset of age-associated paralysis.

To identify CEMs that function as modifiers of Aβ toxicity, we first identified nematode orthologs for the 20 CEMs on the negative tail from the mean-score method using a stringent (BLAST e-value $\leq 10^{-30}$) reciprocal best hits (RBH) approach (Methods). This approach identified seven unique human genes with at least one RBH (Table 4). Interestingly, of those seven genes, only *NDUFA9*, a component of the mitochondrial electron transport chain (mETC) Complex I, was contained in any of the seven canonical pathways for which the top CEMs were enriched (Table 3), and all of those seven canonical pathways contained *NDUFA9*. For these reasons, we focused first on the worm ortholog of *NDUFA9*, encoded by the gene Y53G8AL.2.

Table 4. The 20 negative tail CEMs from the mean-score method (1st column) sorted by their absolute mean-scores, reciprocal best hits of the gene in C. elegans (2nd column), BLAST e-values of human-to-C. elegans and C. elegans-to-human mapping (3rd and 4th columns, respectively).

Human gene	**RBH in *C. elegans***	**BLAST e-value (human-to-*C. elegans*)**	**BLAST e-value (*C. elegans*-to-human)**
ICA1	C32E8.7 (*ric-19*)	4.3e-62	6.9e-62
SPARCL1	-	-	-
COX4I1	-	-	-
TCERG1	ZK1127.9 (*tcer-1*)	4.1e-90	8.6e-106
VSNL1	-	-	-
FAM162A	-	-	-
JAKMIP1	-	-	-
STAT4	-	-	-
NDUFB5	-	-	-
PCYOX1L	-	-	-
SLC25A3	F01G4.6	1.8e-157	6.1e-160
VDAC1	-	-	-
NDUFA9	Y53G8AL.2	6.4e-63	6.7e-63
POLR3A	C42D4.8 (*rpc-1*)	< 2.3e-308	< 2.3e-308
TASP1	-	-	-
LAMB1	W03F8.5 (*lam-1*)	< 2.3e-308	< 2.3e-308
RIC3	-	-	-

Human gene	RBH in *C. elegans*	BLAST e-value (human-to-*C. elegans*)	BLAST e-value (*C. elegans*-to-human)
ATAD1	K04D7.2 (*mspn-1*)	1.1e-113	7.3e-116
RNF128	-	-	-
PREP	-	-	-

To test candidate modifiers of $A\beta_{1-42}$ toxicity, we used RNAi by bacterial feeding to reduce expression of the selected target gene.[35] Worms expressing the $A\beta_{1-42}$ transgene fed on bacteria expressing Y53G8AL.2 RNAi displayed a significant delay in paralysis compared to isogenic animals fed on bacteria expressing an identical RNAi vector that does not contain a dsRNA-coding region — empty vector (EV) — (Figure 8A). Suppression of paralysis by knockdown of Y53G8AL.2 was comparable to RNAi knockdown of the insulin-like receptor gene *daf-2* (Figure 8B), one of the strongest known suppressors of Aβ toxicity in worms[36].

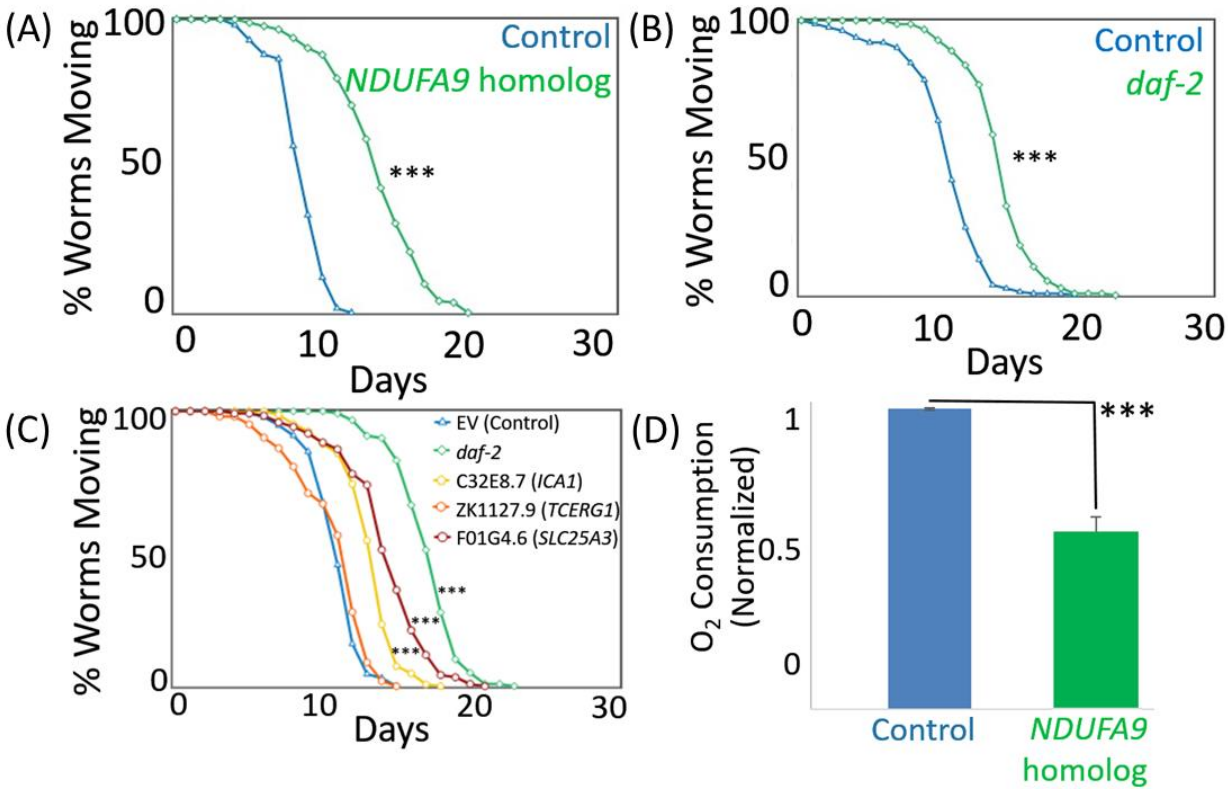

Figure 8. The worm NDUFA9 homolog promotes Aβ toxicity in C. elegans. ***(A)*** *RNAi knockdown of the NDUFA9 homolog Y53G8AL.2 strongly suppresses Aβ-induced paralysis.* ***(B)*** *Y53G8AL.2's suppression of paralysis is comparable to the potent proteotoxicity suppression treatment with daf-2 RNAi.* ***(C)*** *RNAi knockdowns of the C. elegans homologs for two out of three CEMs that precede NDUFA9 in the negative tail of mean-score ranking for Aβ associations also significantly suppress Aβ-induced paralysis.* ***(D)*** *Y53G8AL.2 knockdown also results in greatly reduces whole animal oxygen consumption. O_2 consumption is normalized to the O_2 consumption rate of one animal on control RNAi (EV). *** p-value less than 1e-3.*

We note that identifying the *NDUFA9* gene would not be possible by examining the top genes from a single region, as shown in Table 5. Concordance-based methods led to a high ranking of the *NDUFA9* gene and made it possible to computationally identify its significance, which led to its biological validation in *C. elegans*.

Table 5. NDUFA9 ranking on the negative tail from both individual region-based scores (initial nine columns) and concordance-based scores (last three columns) for Aβ associations.

ROSMAP-DPFCx	ACT-TCx	ACT-PCx	ACT-HIP	ACT-FWM	MSBB-BM10	MSBB-BM36	MSBB-BM44	MSBB-BM22	Mean-score	PCA-score	AE-score
234	320	182	250	433	623	874	1177	1064	13	49	45

We also performed RNAi knockdown experiments to test the *C. elegans* homologs of the three CEMs that precede *NDUFA9* in the negative tail of mean-score ranking (i.e., *ICA1*, *TCERG1*, and *SLC25A3*) (Table 4). We observed that C32E8.7 and F01G4.6 (homologous to human *ICA1* and *SLC25A3*, respectively) RNAi treatments significantly suppressed paralysis ($p < 10^{-3}$) in the *C. elegans* Aβ strain (Figure 8C).

Result 8. RNAi knockdown of NDUFA9's nematode homolog strongly decreases whole animal oxygen consumption.

Y53G8AL.2, the nematode homolog of the human *NDUFA9* gene (Table 4), showed robust suppression of paralysis in our RNAi knockdown experiments in *C. elegans* (Figure 8A). In humans, *NDUFA9* is the alpha subcomplex subunit 9 of the enzyme complex Complex I which is also known as NADH (Nicotinamide Adenine Dinucleotide Hydrate) dehydrogenase or NADH:ubiquinone oxidoreductase. Complex I, the first and largest enzyme complex in the mETC, is located in the inner mitochondrial membrane and catalyzes electron transfer from NADH to ubiquinone.[37] Y53G8AL.2 and *NDUFA9* have a strong sequence conservation (63% protein sequence), and Y53G8AL.2 has been shown to impact oxidative phosphorylation in isolated mitochondria.[38] However, no studies have directly investigated Y53G8AL.2's impact on organismal respiration. Because Y53G8AL.2 has not been extensively studied in *C. elegans*, we sought to confirm that the gene encodes a functional component of the mETC, as predicted by its homology to *NDUFA9*. Consistent with such a role, we observed that RNAi knockdown of Y53G8AL.2 strongly decreases whole animal oxygen consumption (Methods) (Figure 8D). Thus, we concluded that reduced expression of the *NDUFA9* homolog Y53G8AL.2 robustly attenuates Aβ toxicity and impairs the function of Complex I of the mETC.

Result 9. In vivo validation of NDUFA9 identifies mitochondrial Complex I as a potential AD target.

Given that: (1) knockdown of nematode homolog of *NDUFA9*, a subunit of mETC Complex I, strongly suppressed paralysis and significantly decreased whole animal oxygen consumption in *C. elegans*, and (2) the set of CEMs are significantly enriched for functional categories relevant to mitochondrial respiration, ATP synthesis, and mETC all including *NDUFA9* (Table 3), we hypothesized that Complex I may act as a key regulator of Aβ toxicity. We first sought to confirm that the Complex I genes exhibit a cluster structure in the PC1-PC2 space which is the 2-dimensional space which explains the highest amount of variance in the gene expression-Aβ associations from nine regions. Supplementary Figure 2 shows that most of the Complex I genes, represented by blue dots, are projected very closely to each other in the PC1-PC2 space of the expression-Aβ associations. This observation suggests that Complex I genes have similar Aβ association levels, and the other Complex I genes might also function, like Y53G8AL.2, as modifiers of Aβ-induced toxicity. To test this hypothesis, we individually knocked down 13 additional *C. elegans* genes that encode homologs of human

Complex I proteins (Table 6) in the $A\beta_{1-42}$ worms. Strikingly, knockdown of any of these 13 Complex I RNAi clones significantly delayed paralysis, with several exceeding the effect of Y53G8AL.2 (Figure 9). These computational and biological findings pinpoint Complex I of the mETC as a critical mediator of Aβ proteostasis.

Table 6. Human Complex I genes (1st column), the strong ortholog of the NDUF gene in C. elegans (2nd column), BLAST e-values of human-to-C. elegans mapping (3rd column), and the Aβ negative tail ranking of the gene from our concordance-based mean-score method (4th column).

Complex I gene	***C. elegans* homolog**	**BLAST *e*-value**	**Mean-score ranking**
NDUFB5	-	-	9
NDUFA9	Y53G8AL.2	6.4e-63	13
NDUFA12	-	-	31
NDUFB8	-	-	35
NDUFA1	-	-	42
NDUFA5	C33A12.1	1.4e-39	58
NDUFA10	K04G7.4	2.5e-36	65
NDUFV2	F53F4.10	4.0e-113	123
NDUFS3	T10E9.7	1.8e-98	143
NDUFA8	-	-	158
NDUFA4	-	-	160
NDUFB3	-	-	183
NDUFAF2	-	-	234
NDUFAB1	Y56A3A.19	9.8e-40	251
NDUFS1	Y45G12B.1	< 2.3e-308	375
NDUFS4	ZK973.10	1.1e-44	388
NDUFA6	-	-	403
NDUFS5	-	-	409
NDUFAF5	K09E4.3	2.7e-104	499
NDUFA2	-	-	503
NDUFAF1	C50B8.3	8.1e-38	530
NDUFB9	-	-	582
NDUFC1	-	-	609
NDUFA13	-	-	671
NDUFB6	-	-	784
NDUFB2	-	-	944
NDUFB4	-	-	969
NDUFC2	-	-	973
NDUFB10	-	-	1134
NDUFB1	-	-	1149
NDUFAF4	-	-	1370
NDUFAF7	ZK1128.1	1.6e-87	1544
NDUFV1	C09H10.3	< 2.3e-308	1962
NDUFB11	-	-	2320
NDUFA11	-	-	2321

Complex I gene	***C. elegans* homolog**	**BLAST *e*-value**	**Mean-score ranking**
NDUFS8	T20H4.5	4.9e-89	2474
NDUFS2	K09A9.5 T26A5.3	< 2.3e-308 < 2.3e-308	2558
NDUFS6	-	-	2686
NDUFA3	-	-	3050
NDUFV3	-	-	3426
NDUFB7	-	-	4066
NDUFAF6	B0334.5	1.1e-34	4256
NDUFAF3	-	-	4933
NDUFS7	W10D5.2	6.5e-86	8490
NDUFA4L2	-	-	8999

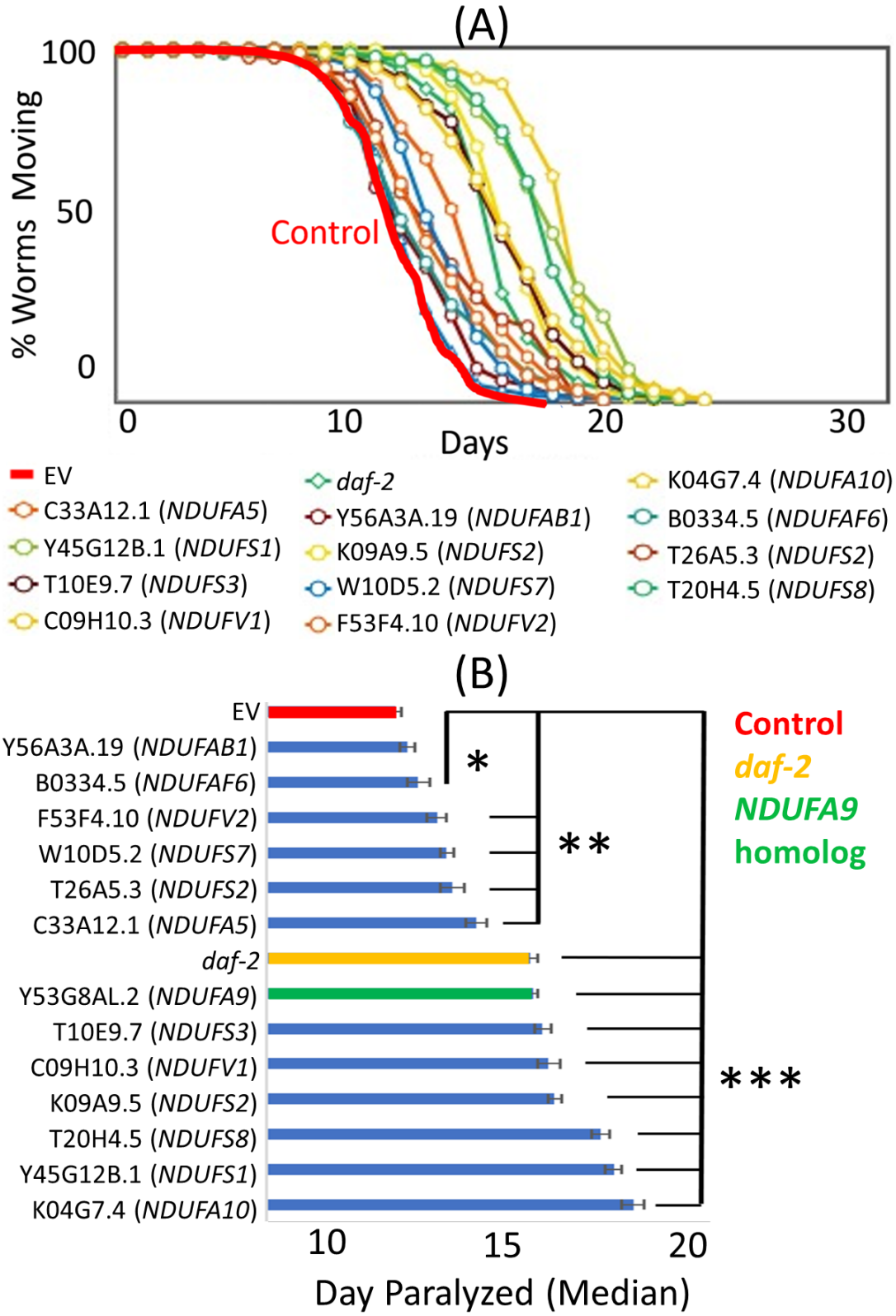

Figure 9. Knockdown of mitochondrial Complex I genes' homologs in C. elegans robustly protects against Aβ toxicity. ***(A)*** *Paralysis curves for the reciprocal best orthologs of human Complex I genes. All tested RNAi conditions significantly suppressed paralysis.* ***(B)*** *The same data as in (A) plotted as the median day of paralysis for each population of worms. Half of the conditions showed even stronger suppression than daf-2 and Y53GAL.2 RNAi conditions. Error bars – S.E.M. * = p=value <0.05 ** = p-value < 0.01 *** = p-value < 0.001.*

Result 10. Multi-omic module subnetwork of Complex I highlights mechanisms relevant to AD.

Molecular processes are orchestrated by a complex interplay between regulatory elements. Given our conclusion that mETC Complex I is a critical mediator of Aβ proteostasis, another important question concerns how Complex I is regulated in this process. The ROSMAP study[17,18] provides DNA methylation and microRNA (miRNA) measurements from the same individuals in addition to mRNA expression data for 542 individuals used in our concordance-based gene ranking framework (

Table *1*). These additional data types let us investigate how Complex I is regulated in a multi-omic setting. We applied the MGL[39] algorithm on a total of 35,354 variables from three types of molecular data – mRNA, DNA methylation, and miRNA – from the ROSMAP study to learn a multi-omic module network that might shed light on the regulatory mechanisms in AD (Methods). DNA methylation and miRNAs are important epigenetic mechanisms that have been suggested to have an important role in AD, especially because environmental factors affect aging through epigenetic modifications in the individual.[40]

Figure 10 shows the subnetwork of the learned module network that represents the modular dependency structure for the Complex I modules. This subnetwork includes ten modules that are: (1) connected to at least three of the eight modules that are significantly enriched (FET $p \leq 0.05$) for Complex I genes, and (2) either significantly enriched for a functional pathway or contain either a DNA methylation probe or a miRNA. Next to each module in Figure 10, we list the top pathway for which this module is significantly enriched. The orange-highlighted four modules are significantly enriched for Complex I genes. Although not the top pathway in their enriched pathway list, each of these four modules is also enriched for the KEGG_ALZHEIMERS_DISEASE pathway. See Supplementary Table 3 for detailed information on each of the 500 modules learned by the MGL algorithm, including the modules that are part of the subnetwork presented in Figure 10 (Methods).

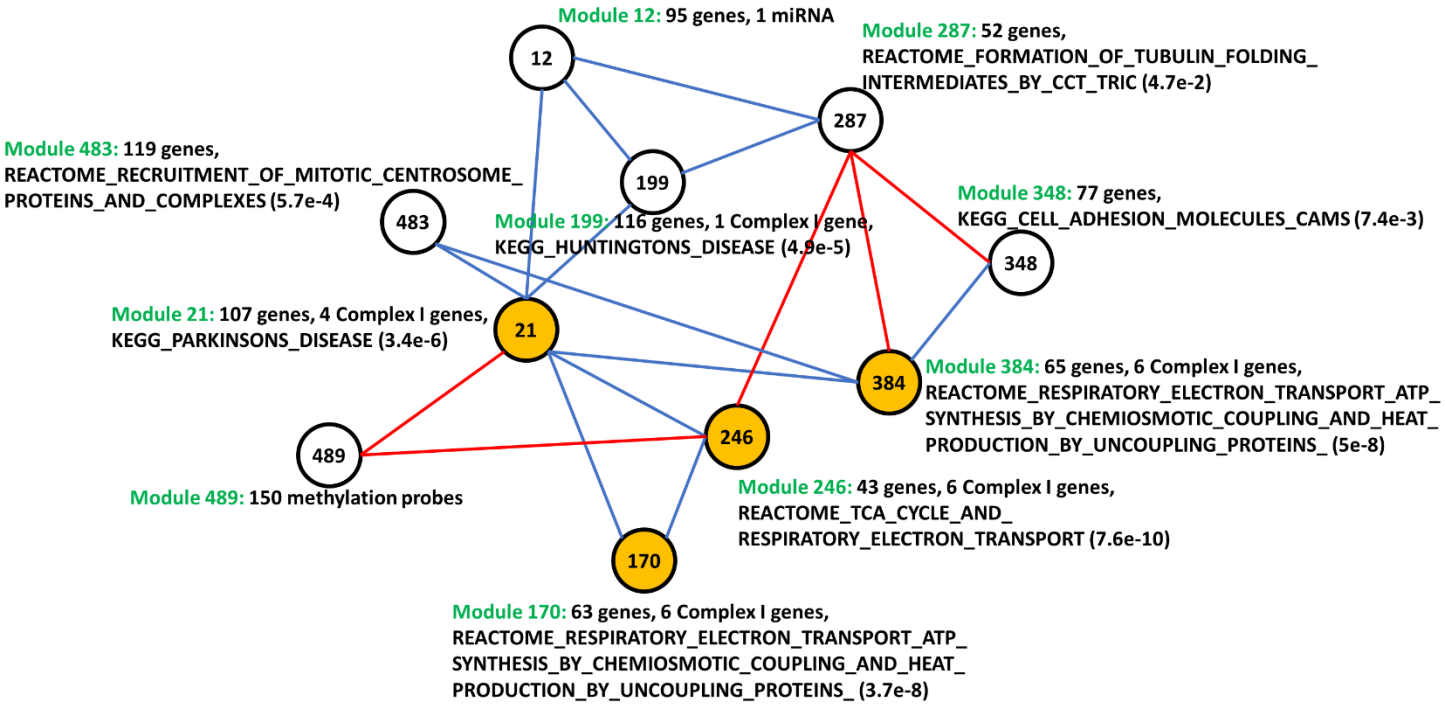

Figure 10. A multi-omic subnetwork of Complex I, learned by the MGL algorithm. Module 489 contains 150 methylation probes and module 12 contains one miRNA (hsa-miR-633) while other modules contain only mRNA variables. Blue edges represent up-regulation, and red edges represent down-regulation, between two connected modules.

Eight modules in the subnetwork presented in Figure 10 – all except 12 and 489 – contain only genes (mRNAs), and each of these modules is significantly enriched for pathways that are relevant to AD. Two of the modules that are connected to module 21, which is the hub in the subnetwork (i.e., connected to many other modules), contain non-gene variables (methylation or miRNA). One of these two modules, module 489, contains 150 methylation probes, not surprisingly since the mitochondrial epigenome has been suggested to have a role in AD.[41,42] Supplementary Table 2 lists the 80 unique genes to which these 150 methylation probes are mapped. Module 12 contains miRNA hsa-miR-633 which was shown to be down-regulated in AD.[43] Given that *NDUFA9* and most other Complex I genes were highly ranked on the negative tail of concordance-based gene scores for Aβ (i.e., down-regulated in AD) (Supplementary Table 4), an up-regulation between module 21 and module 12 is expected. This is represented by a blue edge in Figure 10. We observed that the 95 genes in module 12 significantly overlap ($p < 0.05$) with the 1231 predicted target genes of hsa-miR-633. For miRNA target prediction, we used miRDB[44] database employing MirTarget[45] algorithm version 3.0.

Methods

Gene expression and neuropathology datasets

Our experiments used data from the ROSMAP,[17,18] ACT,[20,22] and MSBB studies. Around half of the people in each cohort had been diagnosed with dementia by the time of death. MSBB neuropathology data was made available by the AMP-AD Knowledge Portal of Sage Bionetworks through https://www.synapse.org/ with Synapse ID syn6101474. We accessed the neuropathology data from ROSMAP and ACT studies through data use agreements. ACT neuropathology data includes four types of Aβ and tau quantifications: (1) immunohistochemistry (IHC) measured on fresh-frozen brain tissue, (2) IHC measured on formalin-fixed, paraffin-embedded (FFPE) brain tissue, (3) histelide[46] quantification from FFPE slides, and (4) Luminex. We

chose histelide for Aβ and IHC FFPE for tau since these two led to the strongest average association with gene expression levels.

ROSMAP RNA-Seq data and MSBB RNA-Seq data were made available by Sage Bionetworks on the AMP-AD Knowledge Portal[16] with Synapse IDs syn3505732 and syn7391833, respectively. The ACT RNA-Seq data[20] was collected by the Allen Institute for Brain Science, Kaiser Permanente Washington Health Research Institute (KPWHRI), and the University of Washington (UW), and was made available on http://aging.brain-map.org. We used normalized and log-transformed RNA-Seq read counts for all datasets.

ACT RNA-Seq data is a part of the bigger ACT project — a collaboration between KPWHRI and UW — that collected clinical and genetic data from thousands of individuals. The approximately 100 ACT individuals from whom we have RNA-Seq measurements were specifically selected for a traumatic brain injury (TBI) study such that half of the selected individuals sustained a TBI with loss of consciousness during life. Therefore, the RNA-Seq cohort does not necessarily reflect the demographics of the entire set of thousands of ACT individuals. Therefore, each RNA-Seq sample was assigned a weight based on the demographic information of each sample in the TBI study. Details of this weighting scheme and the weights themselves are provided on http://aging.brain-map.org. We took the weights into account when computing gene expression-neuropathology associations for the ACT data so that the samples reflect the original ACT cohort demographics.

Algorithmic details on concordance-based methods

One of the three concordance-based gene ranking methods we introduce is the AE-score, which employs an auto-encoder (AE). An auto-encoder is an artificial neural network that is trained to encode the input in a new, usually lower-dimensional representation such that the input can be reconstructed with minimal error from that representation.[47] As opposed to supervised learning, where some specific output features are predicted, an autoencoder is an unsupervised learning framework, where the output to be predicted is the input data itself. Figure 11 displays the AE architecture that we used to compute AE-score values from the nine sets of individual-region associations we used as input. Both the encoder and decoder networks include two hidden layers, and each feature unit (represented by a circle) in each layer contains ~15K genes as samples. We used a fully connected network, and we selected the number of hidden units in each of the encoder's and decoder's intermediate layers to be equal to the number of input features, which is nine (i.e., the number of regions). The three dots in each of the encoder and decoder networks are meant to imply full-connectivity in the intermediate layer. The single hidden unit in the encoder's final layer represents the AE-score. We used no activation in the output layer and used the mean-squared error (MSE) to measure reconstruction quality since gene expression-neuropathology associations are continuous-valued between -1 and 1. We used hyperbolic tangent function (tanh) as the activation function in the hidden layers. Since the cost function of AE is non-convex, the resulting weights may depend on the weight values at initialization. Therefore, we performed 100 AE runs initialized with different random weights and averaged the results across these 100 AE runs.

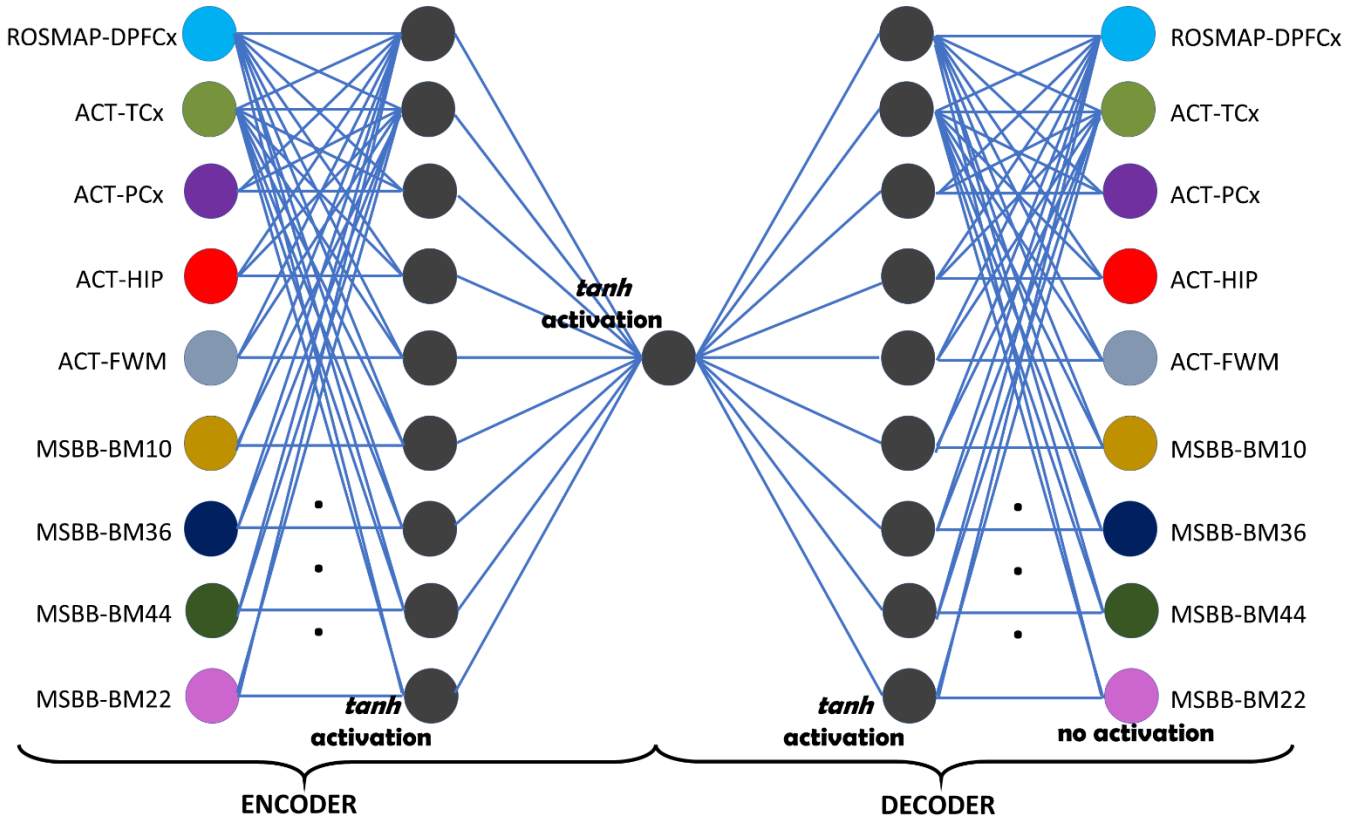

Figure 11.The auto-encoder architecture with two hidden layers that we used to compute AE-score values. Input node region colors are matched to region colors in Figure 1. Hidden units are shown in gray. An edge between two units represents the weight of a feature in computing the linear combination of the feature values to be passed to the activation function (if any). The single hidden unit in the encoder's final hidden layer represents the AE-score.

An auto-encoder with only one hidden layer that linearly encodes the input features in k hidden units with no activation and uses an objective of minimizing MSE between the input and reconstructed data is equivalent to a PCA.[48] Thus, the k hidden units of this AE project the data to a space that is spanned by the first k principal components of the input data. Figure 12A represents such a network, which corresponds to the PCA-score. Figure 12B represents the mean-score, where region weights are assumed to be equal; hence, the final gene scores are computed as the mean across input region associations.

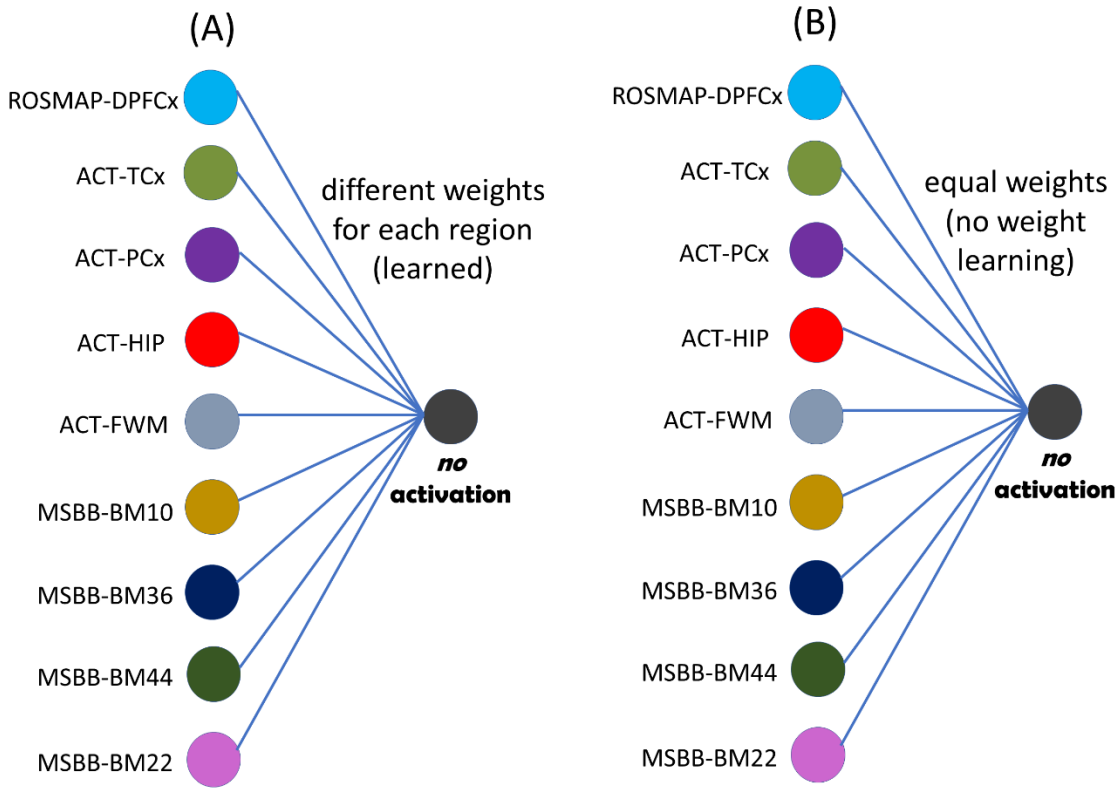

*Figure 12. The single hidden unit represents **(A)** PCA-score and **(B)** mean-score. Input node region colors are matched to region colors in Figure 1. An edge between two units represents the weight of a feature in computing the linear combination of feature values to compute the final concordance-based scores.*

Data collection and association analysis for the experiments with TCGA data

To examine the cancer survival associations, we used clinical data and RNA-Seq data provided by the Broad Institute (http://gdac.broadinstitute.org/). We used the "SurvCorr"[49] R package available on CRAN to estimate a correlation between a right-censored survival time and a gene expression level. Table 7 lists the expansions for the cancer type abbreviations we used in our experiments with TCGA data:

Table 7. Cancer type abbreviations.

ACC	Adrenocortical carcinoma
BLCA	Bladder urothelial carcinoma
BRCA	Breast invasive carcinoma
CESC	Cervical and endocervical cancers
CHOL	Cholangiocarcinoma
COAD	Colon adenocarcinoma
DLBC	Lymphoid neoplasm diffuse large B-cell lymphoma
ESCA	Esophageal carcinoma
GBM	Glioblastoma multiforme
HNSC	Head and neck squamous cell carcinoma
KICH	Kidney chromophobe
KIRC	Kidney renal clear cell carcinoma

KIRP	Kidney renal papillary cell carcinoma
LAML	Acute myeloid leukemia
LGG	Brain lower grade glioma
LIHC	Liver hepatocellular carcinoma
LUAD	Lung adenocarcinoma
LUSC	Lung squamous cell carcinoma
MESO	Mesothelioma
OV	Ovarian serous cystadenocarcinoma
PAAD	Pancreatic adenocarcinoma
PCPG	Pheochromocytoma and paraganglioma
PRAD	Prostate adenocarcinoma
READ	Rectum adenocarcinoma
SARC	Sarcoma
SKCM	Skin cutaneous melanoma
STAD	Stomach adenocarcinoma
TGCT	Testicular germ cell tumors
THCA	Thyroid carcinoma
THYM	Thymoma
UCEC	Uterine corpus endometrial carcinoma
UCS	Uterine carcinosarcoma
UVM	Uveal melanoma

Identification of C. elegans orthologs

To enable biological testing of the human genes identified using our computational analysis, we obtained the RBHs between human and *C. elegans*. To do so, we first identified all unique protein sequences for each potential marker gene using the "biomaRt"[50] R package available on CRAN. Then, we used the latest version of NCBI BLAST tool[51] that we downloaded from ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST, version 2.6.0 as of 8/15/17, to identify the *C. elegans* orthologs for each complete human protein query sequence. We downloaded the *C. elegans* protein sequences from http://www.wormbase.org/species/c_elegans. We took into account only the protein pairs mapped from human to *C. elegans* with a BLAST *e*-value smaller than 10^{-30}. For each of these *C. elegans* isoforms, we identified the corresponding human genes, again using the NCBI BLAST tool, and used only the orthologs that achieved a BLAST *e*-value smaller than 10^{-30}. This process resulted in high-confidence RBHs for us to test in *C. elegans*.

C. elegans cultivation and RNAi treatment

Experimental worm populations of GMC101 animals were obtained from the Caenorhabditis Genetics Center (CGC) and cultivated on NGM plates with OP50 at 15C.[34,52] Care was taken to ensure that the animals were never starved and the plates remained free of contamination.

The gene-specific RNAi clones were obtained from the commercial Ahringer or Vidal *C. elegans* RNAi-feeding libraries (BioScience, Nottingham, UK). Each bacterial clone was struck-out onto LB plates containing carbenicillin (50 ug/ml) and tetracycline (10 ug/ml). Single colonies were then seeded into 5 ml LB + carbenicillin

(50 ug/ml) and tetracycline (10 ug/ml) for growth overnight on a 37C rotator. 100 ul of each overnight culture was then inoculated into 10 ml of LB containing carbenicillin (50 ug/ml) and tetracycline (10 ug/ml) and IPTG (5mM) and incubated on a 37C rotator for 4 hours. Each bacterial growth was then centrifuged at 3500 X G for 25 minutes, decanted, and the pellet resuspended in 0.5 ml of LB containing carbenicillin (50 ug/ml), tetracycline (10 ug/ml), and IPTG (5 mM). To verify RNAi conditions' plasmid DNA, each RNAi clone was purified and assessed through PCR (polymerase chain reaction) with sequence-specific primers or through Sanger sequencing.

Nematode paralysis assays

Paralysis assays were performed by visually inspecting animals daily to determine if they are capable of spontaneous movement or if they are paralyzed. A robotic system equipped with a digital camera was used to obtain images of individual wells of a 12 well-plate at 5-minute intervals over the entire course of the experiment. Each well contained 30-40 individual animals expressing Aβ. Through analysis of serial images from each plate, the age at which each animal stops moving can be easily determined. We applied this system to the transgenic Aβ model line GMC101 to determine the time of paralysis onset for each individual animal. Paralysis data was plotted using Oasis2. Statistical significance of mean paralysis time-points between RNAi conditions was determined by a weighted log-rank[53,54] test.

Prior to loading on the experimental plates, animal populations were amplified on high-growth plates seeded with NA22 bacteria. Worm populations were developmentally synchronized by hypochlorite treatment and the remaining eggs were deposited on unseeded plates overnight. Synchronized larval stage 1 animals were washed off unseeded plates and moved onto standard *C. elegans* RNAi plates containing carbenicillin (50 mg/ml), tetracycline (10 mg/ml), and IPTG (5 mM) 48h at 20C. These developmentally-synchronized late larval stage 4-populations were then washed and transferred to their respective RNAi conditions on 12-well plates. We used standard RNAi conditions plus FuDR (100 ug/ml) to prevent progeny and neomycin (200 mg/ml) to prevent fungal growth.[55] Each RNAi condition was tested in 2-3 wells as technical replicates. At least three biological replicates, each started on different weeks, were conducted for each RNAi clone.

Oxygen consumption assay

We quantified the oxygen consumption rates of day-two adult worms cultivated from L1 on Y53G8AL.2 RNAi or the EV control RNAi at 20C. Oxygen consumption was measured utilizing the Seahorse X24 Bioanalyzer (Seahorse Biosciences, MA, USA).[54] Briefly, young adult worms were washed from NGM plates and rinsed from RNAi bacteria with M9 buffer. Approximately 40-50 worms were pipetted into each well in Seahorse XF24 Cell Culture Microplates with the final volume of 500μL M9. The number of worms in each well was quantified. Basal respiration was analyzed using the average respiration of five technical replicates per condition, and eight recordings over the course of an hour were made of each condition. Three independent biological replicates were conducted on different weeks. For each experiment, we compared the relative rate of oxygen consumption between the paired conditions. In each plate, nematode-free wells were used as control. Oxygen consumption rates were then normalized by the number of worms. All three replicates showed that the animals cultivated on the Y53G8AL.2 RNAi consumed about half as much oxygen as those grown on the control RNAi bacteria (Figure 8D).

Multi-omic module network learning

Given a matrix of values for each variable-sample pair from a number of variables and a number of samples, MGL iteratively learns an assignment of the variables to a user-supplied number of modules and the conditional dependence network among modules. In a conditional dependence module network, two modules that are disconnected are independent of each other given all other modules.

For the MGL algorithm to run in a reasonable amount of time, we selected 20,132 methylation probes with the highest variance across samples, from the 420,132 total probes in the ROSMAP DNA methylation data. Supplementary Figure 3 shows the standard deviation of all methylation probes; the red vertical line represents the standard deviation threshold we used to select the methylation probes to use in MGL. In addition to DNA methylation levels, we used expression levels of 14,912 protein-coding genes and 309 miRNAs from the ROSMAP study. Thus, we learned a network of modules from a total of 35,354 variables. We set the module count to 500 before running the algorithm. This results in an average of 70 genes in a module, while the variable count varied across modules. Since MGL has a non-convex objective function, we performed five runs of the MGL algorithm with different initial parameter values and selected the run with final parameter values leading to the highest likelihood of the training data. Supplementary Table 3 shows information about each of the 500 modules learned by MGL. We note that since the dependence network MGL algorithm[39] learns is based on an inverse covariance matrix, a positive edge weight shown in column (J) corresponds to an up-regulation and a negative edge weight corresponds to a down-regulation.

Discussion

Alzheimer's disease (AD) is a progressive neurodegenerative disorder with no cure. The molecular mechanisms underlying AD neuropathology remain unknown, and it is critical to identify true positive biomarkers for these mechanisms. The number of AD studies that provide different types of brain tissue data including gene expression and neuropathological phenotypes has been increasing. This offers a unique opportunity to perform coordinated analyses leveraging the different samples and data types available.

There have been successful efforts to reduce false positives by focusing on concordant associations across different molecular data types from the same study. White et al. (2017)[56] and Yang et al. (2015)[57] analyzed ROSMAP data, examining SNPs with significant association *p*-values only if their methylation status and/or mRNA levels were also associated with the phenotype. An important prior study that performed multi-regional or multi-study analyses of AD gene expression data is from Wang et al.[19] who employed an integrative network analysis approach using MSBB microarray data from 19 brain regions. Examining the gene expression similarities between regions, the authors detected strong interactions across regions with strong physical interconnectivity. Zhang et al.[15] generated gene regulatory networks from three brain regions using an integrated systems approach and examined the functional organization of the networks and its relation to AD pathology. However, both studies compared findings across different regions; they did not analyze the concordance of pathological molecular mechanisms across regions. Xu et al.[58] compiled publicly available AD expression data from four brain regions measured in a total of 1,246 samples in 20 datasets. They identified genes that are differentially expressed between AD and control samples in all four regions and prioritized potential regulators of those genes by integrating data from GWAS, brain expression quantitative trait loci (eQTL), protein-protein interaction (PPI), and AD mouse models.

We introduced three new concordance-based methods — mean-score, PCA-score, and AE-score — to score and rank genes based on the level of agreement of brain regions on the association between gene expression and neuropathology. We demonstrated that each concordance-based method reduces the rate of false positive associations and, thus, increases the chance that the identified genes are relevant to AD. This process led us to a short list of potential AD neuropathology biomarker genes which we call concordant expression markers

(CEMs), that are highly ranked by our concordance-based methods. CEMs were enriched for several pathways relevant to mitochondrial electron transport chain (mETC). The gene *NDUFA9* which is involved in mETC is a highly ranked CEM. This gene was identified to be differentially down-regulated between cognitively normal controls and AD individuals in four brain regions and significantly correlated with Aβ in mouse models.[58] We tested *NDUFA9* and 13 other Complex I genes in *C. elegans* Aβ proteotoxicity models and showed that knockdown of each Complex I gene significantly suppresses Aβ toxicity.

Our findings suggest that Complex I inhibition could be a new paradigm for developing future AD therapeutics. Consistent with this, prior work indicated that mild inhibition of Complex I with the small molecule CP2 reduced Aβ and tau levels in mouse models of familial AD.[59] Capsaicin is a natural product that can inhibit Complex I, and a capsaicin-rich diet was associated with lower total serum Aβ levels in the elderly.[60] Furthermore, capsaicin reduced AD-associated tau changes in the hippocampus of type 2 diabetes rats.[61] These earlier studies' findings and our novel results from the brain region concordance-based approaches suggest that Complex I is indeed a promising potential pharmacological avenue toward treating AD.

The *C. elegans* gene *K04G7.4* (*nuo-4*) — ortholog of the human Complex I gene *NDUFA10* — was the strongest suppressor of Aβ toxicity, a gene that has been shown to be significantly enriched in serotonergic neurons compared to whole-animal expression.[62] Given our *C. elegans* model expresses human $A\beta_{1-42}$ peptide solely in body wall muscle cells, this suggests the possibility that cell non-autonomous neuron-to-muscle signaling mediates regulation of Aβ toxicity following Complex I inhibition in *C. elegans*. Research studies using *C. elegans* have shown that neuronal signaling and endocrine pathways are integrated into the physiology of other tissues in the organism, including the regulation of heat shock response modulation of protein homeostasis in post-synaptic muscle cells.[63–65] Our hypothesis of Aβ toxicity regulation through neuron-to-muscle signaling can be addressed in the future with tissue-specific RNAi experiments.

One limitation of the computational experiments we performed to test DECODER's performance on identifying neuropathology marker genes is that the neuropathological phenotype quantifications we used for some of the nine brain regions had not been obtained exactly from that phenotype or region. For the MSBB RNA-Seq samples, no tau, tangle, or Aβ quantifications were provided. Therefore, we used Braak stages as a proxy for tau levels since Braak staging is based on the regional distribution pattern of neurofibrillary tangle density across the brain,[66] and we used neuritic plaque density mean across five cortical regions as a proxy for Aβ. Thus, the samples from the four different MSBB regions were assigned the same values for Aβ and the same values for tau. Moreover, histelide[46] Aβ quantification was provided for only two of the four ACT regions — ACT-TCx and ACT-PCx. Therefore, we used ACT-PCx Aβ levels for ACT-FWM and ACT-TCx Aβ levels for ACT-HIP based on regional proximity. We surmise that the reason why regions from the same study tend to group together in the heatmaps in Figure 2 and Figure 6 is likely because of these arrangements we had to make due to limited data available.

DECODER is a general approach applicable to any -omic data type besides transcriptomic data as long as it is possible to get association statistics, such as correlation coefficients, from the data. Therefore, we expect wider availability of AD multi-omic data to let us apply DECODER to identify concordant neuropathology associations of other molecular data types as well.

Another important direction for future work is to develop a systematic approach to iterate between a computational algorithm phase and a biological validation phase. At each iteration, the computational algorithm is improved using causal markers identified by the experiments in the previous biological validation phase, which would lead to a new set of genes to test in *C. elegans* and/or other model organisms in the next biological validation phase. Biological validation experiments enable us to identify causal relationships; however, it is often highly challenging to biologically experiment a large number of genes due to limited lab resources. On the other

hand, computational resources are relatively cheaper, but reliably identifying causal relationships is known to be an NP-hard problem.[67] The described iterative algorithm takes advantage of both sides by improving the set of computationally identified genes by utilizing experimentally-learned causal relations.

We conjecture that our approach will become even more powerful in the near future given the number of AD studies providing brain gene expression and neuropathology data has been increasing. Many of these studies also provide multi-omic data, which will let us utilize a higher sample size for learning regulatory multi-omic module networks for identified biomarker genes.

Acknowledgments

This work was supported by National Science Foundation (NSF) grants DBI-1355899 and CAREER DBI-1552309 to SL, National Institutes of Health (NIH) grant AG049196-01A1 to SL and SC, NIH grants P30AG013280 and P50AG005136 to MK and NIH grant F32AG054098-01S1 to JR. We are grateful to Mount Sinai/JJ Peters VA Medical Center Brain Bank for making the MSBB gene expression and neuropathology data available through the AMP-AD Knowledge Portal, to Caenorhabditis Genetics Center (CGC) for the worm reagents, and to Nick Terzopoulos for preparing the nematode cultivation plates. Generation of the ACT RNA-Seq data was funded by a grant to CD Keene, RG Ellenbogen and Ed Lein from the Paul G. Allen Family Foundation, and supported by the NIH grants U01AG006781 and P50AG005136, and the Nancy and Buster Alvord Endowment. ROSMAP data collection was supported through funding by National Institute on Aging (NIA) grants P30AG10161, R01AG15819, R01AG17917, R01AG30146, R01AG36836, U01AG32984, U01AG46152, the Illinois Department of Public Health, and the Translational Genomics Research Institute (TGen).

References

1. Alzheimer's Association. 2017 Alzheimer's disease facts and figures. *Alzheimer's Dement* **13,** 325–373 (2017).

2. Alzheimer's Association. at <http://www.alz.org/>

3. Kang, J. *et al.* The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface receptor. *Nature* **325,** 733–736 (1987).

4. Goate, A. *et al.* Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature* **349,** 704–706 (1991).

5. Karran, E., Mercken, M. & Strooper, B. De. The amyloid cascade hypothesis for Alzheimer's disease: An appraisal for the development of therapeutics. *Nature Reviews Drug Discovery* **10,** 698–712 (2011).

6. Hardy, J. & Selkoe, D. J. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* **297,** 353–356 (2002).

7. Reitz, C. Alzheimer's disease and the amyloid cascade hypothesis: A critical review. *International Journal of Alzheimer's Disease* (2012). doi:10.1155/2012/369808

8. Vassar, R. BACE1 inhibitor drugs in clinical trials for Alzheimer's disease. *Alzheimer's Research and Therapy* **6,** (2014).

9. Bloom, G. S. Amyloid-β and tau: the trigger and the bullet in Alzheimer's disease pathogenesis. *JAMA*

Neurol. **71,** 505 (2014).

10. Gong, C.-X. & Iqbal, K. Hyperphosphorylation of microtubule-associated protein tau: a promising therapeutic target for Alzheimer disease. *Curr. Med. Chem.* **15,** 2321–8 (2008).

11. Cacace, R., Sleegers, K. & Van Broeckhoven, C. Molecular genetics of early-onset Alzheimer's disease revisited. *Alzheimer's and Dementia* **12,** 733–748 (2016).

12. Giri, M., Zhang, M. & Lü, Y. Genes associated with Alzheimer's disease: An overview and current status. *Clinical Interventions in Aging* **11,** 665–681 (2016).

13. Ridge, P. G., Mukherjee, S., Crane, P. K. & Kauwe, J. S. K. Alzheimer's disease: Analyzing the missing heritability. *PLoS One* **8,** (2013).

14. Ridge, P. G. *et al.* Assessment of the genetic variance of late-onset Alzheimer's disease. *Neurobiol. Aging* **41,** (2016).

15. Zhang, B. *et al.* Integrated Systems Approach Identifies Genetic Nodes and Networks in Late-Onset Alzheimer's Disease. *Cell* **153,** 707–720 (2013).

16. AMP-AD Knowledge Portal. at <https://www.synapse.org/#!Synapse:syn2580853/wiki/409840>

17. Bennett, D. A., Schneider, J. A., Arvanitakis, Z. & Wilson, R. S. Overview and findings from the religious orders study. *Curr. Alzheimer Res.* **9,** 628–45 (2012).

18. Bennett, D. A. *et al.* Overview and findings from the rush Memory and Aging Project. *Curr. Alzheimer Res.* **9,** 646–63 (2012).

19. Wang, M. *et al.* Integrative network analysis of nineteen brain regions identifies molecular signatures and networks underlying selective regional vulnerability to Alzheimer's disease. *Genome Med.* **8,** 104 (2016).

20. Miller, J. A. *et al.* Neuropathological and transcriptomic characteristics of the aged brain. *Elife* **6,** (2017).

21. Masters, C. L. *et al.* Alzheimer's disease. *Nat. Rev. Dis. Prim.* 15056 (2015). at <http://dx.doi.org/10.1038/nrdp.2015.56%5Cn10.1038/nrdp.2015.56>

22. Montine, T. J., Sonnen, J. A., Montine, K. S., Crane, P. K. & Larson, E. B. Adult Changes in Thought study: dementia is an individually varying convergent syndrome with prevalent clinically silent diseases that may be modified by some commonly used therapeutics. *Curr. Alzheimer Res.* **9,** 718–23 (2012).

23. Villain, N. *et al.* Relationships between Hippocampal Atrophy, White Matter Disruption, and Gray Matter Hypometabolism in Alzheimer's Disease. *J. Neurosci.* **28,** 6174–6181 (2008).

24. Douaud, G. *et al.* Preventing Alzheimer's disease-related gray matter atrophy by B-vitamin treatment. *Proc. Natl. Acad. Sci.* **110,** 9523–9528 (2013).

25. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 15545–50 (2005).

26. Federico, A. *et al.* Mitochondria, oxidative stress and neurodegeneration. *J. Neurol. Sci.* **322,** 254–262 (2012).

27. Patten, D. A., Germain, M., Kelly, M. A. & Slack, R. S. Reactive oxygen species: Stuck in the middle of

neurodegeneration. *Journal of Alzheimer's Disease* **20,** (2010).

28. Morán, M. *et al.* Mitochondrial respiratory chain dysfunction: Implications in neurodegeneration. *Free Radical Biology and Medicine* **53,** 595–609 (2012).

29. Rosenwald, A. *et al.* The proliferation gene expression signature is a quantitative integrator of oncogenic events that predicts survival in mantle cell lymphoma. *Cancer Cell* **3,** 185–197 (2003).

30. Mosley, J. D. & Keri, R. a. Cell cycle correlated genes dictate the prognostic power of breast cancer gene lists. *BMC Med. Genomics* **1,** 11 (2008).

31. Rubicz, R. *et al.* Expression of cell cycle-regulated genes and prostate cancer prognosis in a population-based cohort. *Prostate* **75,** (2015).

32. Boisvert, M.-E. L. & Simard, M. J. RNAi pathway in C. elegans: the argonautes and collaborators. *Curr. Top. Microbiol. Immunol.* **320,** 21–36 (2008).

33. Grishok, A. & Mello, C. C. RNAi (Nematodes: Caenorhabditis elegans). *Adv. Genet.* **46,** 339–60 (2002).

34. McColl, G. *et al.* Utility of an improved model of amyloid-beta (Aβ 1-42) toxicity in Caenorhabditis elegans for drug screening for Alzheimer's disease. *Mol. Neurodegener. 2012 71* **17,** 857–872 (2012).

35. Kamath, R. S. & Ahringer, J. Genome-wide RNAi screening in Caenorhabditis elegans. *Methods* **30,** 313–321 (2003).

36. Steinkraus, K. A. *et al.* Dietary restriction suppresses proteotoxicity and enhances longevity by an hsf-1-dependent mechanism in Caenorhabditis elegans. *Aging Cell* **7,** 394–404 (2008).

37. Alberts B, Johnson A, Lewis J, et al. Electron transport chains and their proton pumps. *Mol. Biol. cell 4th Ed.* 1–10 (2002).

38. Falk, M. J. *et al.* Subcomplex Iλ specifically controls integrated mitochondrial functions in Caenorhabditis elegans. *PLoS One* **4,** (2009).

39. Celik, S., Logsdon, B. A. & Lee, S.-I. Efficient Dimensionality Reduction for High-Dimensional Network Estimation. in *International Conference on Machine Learning* (2014).

40. Sen, P., Shah, P. P., Nativio, R. & Berger, S. L. Epigenetic Mechanisms of Longevity and Aging. *Cell* **166,** 822–839 (2016).

41. Devall, M., Mill, J. & Lunnon, K. The mitochondrial epigenome: a role in Alzheimer's disease? *Epigenomics* **6,** 665–675 (2014).

42. Devall, M., Roubroeks, J., Mill, J., Weedon, M. & Lunnon, K. Epigenetic regulation of mitochondrial function in neurodegenerative disease: New insights from advances in genomic technologies. *Neuroscience Letters* **625,** 47–55 (2016).

43. Lau, P. *et al.* Alteration of the microRNA network during the progression of Alzheimer's disease. *EMBO Mol. Med.* **5,** 1613–1634 (2013).

44. Wong, N. & Wang, X. miRDB: An online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res.* **43,** D146–D152 (2015).

45. Wang, X. Improving microRNA target prediction by modeling with unambiguously identified microRNA-target pairs from CLIP-ligation studies. *Bioinformatics* **32,** 1316–1322 (2016).

46. Postupna, N. *et al.* Novel antibody capture assay for paraffin-embedded tissue detects wide-ranging amyloid beta and paired helical filament-tau accumulation in cognitively normal older adults. *Brain Pathol.* **22,** 472–484 (2012).

47. Bengio, Y. Learning Deep Architectures for AI. *Found. Trends® Mach. Learn.* **2,** 1–127 (2009).

48. Bourlard, H. & Kamp, Y. Auto-association by multilayer perceptrons and singular value decomposition. *Biol. Cybern.* **59,** 291–294 (1988).

49. Schemper, M., Kaider, A., Wakounig, S. & Heinze, G. Estimating the correlation of bivariate failure times under censoring. *Stat. Med.* **32,** 4781–4790 (2013).

50. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4,** 1184–1191 (2009).

51. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–10 (1990).

52. Brenner, S. The genetics of Caenorhabditis elegans. *Genetics* **77,** 71–94 (1974).

53. Mantel, N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother. Rep.* **50,** 163–70 (1966).

54. Cox, D. R. Regression models and life tables. *J. R. Stat. Soc. Ser. B* **34,** 187–220 (1972).

55. Sutphin, G. L. & Kaeberlein, M. Measuring Caenorhabditis elegans life span on solid media. *J. Vis. Exp.* 1–7 (2009). doi:10.3791/1152

56. White, C. C. *et al.* Identification of genes associated with dissociation of cognitive performance and neuropathological burden: Multistep analysis of genetic, epigenetic, and transcriptional data. *PLoS Med.* **14,** e1002287 (2017).

57. American Academy of Neurology., H.-S. *et al.* Identification of Chromosomal Loci Influencing Cognitive Resilience to Neuropathology (S21.003). *Neurology* **86,** S21.003 (2016).

58. Xu, M. *et al.* A systematic integrated analysis of brain expression profiles reveals YAP1 and other prioritized hub genes as important upstream regulators in Alzheimer's disease. *Alzheimer's Dement.* **14,** 215–229 (2018).

59. Zhang, L. *et al.* Modulation of Mitochondrial Complex I Activity Averts Cognitive Decline in Multiple Animal Models of Familial Alzheimer's Disease. *EBioMedicine* **2,** 294–305 (2015).

60. Liu, C. H. *et al.* The Associations between a Capsaicin-Rich Diet and Blood Amyloid-β Levels and Cognitive Function. *J. Alzheimer's Dis.* **52,** 1081–1088 (2016).

61. Xu, W. *et al.* Capsaicin reduces Alzheimer-associated tau changes in the hippocampus of type 2 diabetes rats. *PLoS One* **12,** (2017).

62. Spencer, W. C. *et al.* Isolation of specific neurons from C. Elegans larvae for gene expression profiling. *PLoS One* **9,** (2014).

63. van Oosten-Hawle, P. & Morimoto, R. I. Transcellular chaperone signaling: an organismal strategy for integrated cell stress responses. *J. Exp. Biol.* **217,** 129–136 (2014).

64. van Oosten-Hawle, P. & Morimoto, R. I. Organismal proteostasis: Role of cell-nonautonomous

regulation and transcellular chaperone signaling. *Genes and Development* **28,** 1533–1543 (2014).

65. Garcia, S. M., Casanueva, M. O., Silva, M. C., Amara, M. D. & Morimoto, R. I. Neuronal signaling modulates protein homeostasis in Caenorhabditis elegans post-synaptic muscle cells. *Genes Dev.* **21,** 3006–3016 (2007).

66. Braak, H. & Braak, E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol.* **82,** 239–259 (1991).

67. Cooper, G. F. The computational complexity of probabilistic inference using bayesian belief networks. *Artif. Intell.* **42,** 393–405 (1990).