SUPPLEMENTARY MATERIALS AND METHODOLOGY:

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human brain tissue samples. Our KRONOSII series was obtained from 21 National Alzheimer's Coordinating Center (NACC) brain banks and from the Miami Brain Bank as previously described (Corneveaux et al., 2010; Myers et al., 2007; Webster et al., 2009). Additional cohorts were obtained in the same manner as the original US series. Our criteria for inclusion were as follows: self-defined ethnicity of European descent (in an attempt to control for the known allele frequency differences between ethnic groups), neuropathologically confirmed Alzheimer's disease or no neuropathology present, and age of death greater than or equal to 65. Neuropathological diagnosis was defined by board-certified neuropathologists according to the standard NACC protocols (Beekly et al., 2004). Samples derived from subjects with a clinical history of stroke, cerebrovascular disease, Lewy bodies, or comorbidity with any other known neurological disease were excluded. Alzheimer's disease or control neuropathology was confirmed by plaque and tangle assessment with 45% of the entire series undergoing Braak staging (Braak and Braak, 1995).

The RUSH series includes deceased subjects from two large, prospectively followed cohorts maintained by investigators at Rush University Medical Center in Chicago, IL: the Religious Orders Study (ROS) and the Memory and Aging Project (MAP). The ROS cohort, established in 1994, consists of more than 1,100 Catholic priests, nuns, and brothers from 40 groups in 12 states who were at least 55 years of age and free of known dementia at the time of enrollment. The MAP cohort, established in 1997, consists of more than 1600 men and women primarily from retirement facilities in the Chicago area who were at least 53 years of age and free of known dementia at the time of enrollment. All participants in ROS and MAP sign an informed consent agreeing to annual detailed clinical evaluations and cognitive tests, and the rate of follow-up exceeds 90%. Similarly, participants in both cohorts signed an Anatomical Gift Act donating their brains at the time of death. The overall autopsy rate exceeds 85%. The ROS and MAP cohorts were analyzed jointly since they were designed to be combined, are maintained by a single investigative team, and a large set of phenotypes collected are identical in both studies (Bennett et al., 2012b). More detailed information regarding the two cohorts can be found in previously published literature (Bennett et al., 2012a).

In both cohorts, samples were de-identified before receipt, and the study met human studies institutional review board and HIPPA regulations. This work is declared not human-subjects research and is IRB exempt under regulation 45 CFR 46. See the Acknowledgements section for a list of individual sites that contributed samples to this effort. See Supplemental Figure 1 for final counts and exclusions.

Cell Lines. Two different cell lines were used to assess target effects on APP and Tau pathology. The first, HEK293sw, was a human embryonic kidney cell line expressing an Abeta complementary DNA bearing a double mutation ((KM670 and 671NL; HEK293sw; gift from D. Selkoe; reference (Citron et al., 1992)). This was used to examine target effects on Abeta levels. These cells produce ~20 fold more Abeta than cells without the

mutations. The second line was a neuroglioma cell line engineered to overexpress four repeat tau (H4-4R0N, gift from T. Dunkley (Azorsa et al., 2010), Phoenix, AZ). This line was used to examine Tau and Phospho-Tau levels. HEK293sw is maintained in Dulbecco's modification of Eagle's medium/F-12 (DMEM, Sigma-Aldrich, St Louis, MO, USA), supplemented with GlutaMAX™ (Life Technologies, Carlsbad, CA USA), 10% fetal bovine serum (Life Technologies, Carlsbad, CA USA), and 1X penicillin/streptomycin (500U/ml & 500ug/ml respectively; Life Techologies, Carlsbad, CA). H4-4R0N is maintained in DMEM, high glucose with GlutaMAX™ Supplement, and sodium pyruvate (ThermoFisher Scientific, Waltham, MA) plus 10% fetal bovine serum (Life Technologies, Carlsbad, CA) and 250ug/ml G418 sulfate (Corning incorporated, Corning, NY).

METHODS DETAILS

DNA sample processing and data collection. Genomic DNA samples were analyzed on the Genome-Wide Human SNP 6.0 Array (Affymetrix, Inc. Santa Clara, CA) according to the manufacturer's protocols (Affymetrix Genome-Wide Human SNP Nsp/Sty 6.0 User Guide;Rev. 1 2007). Before the initiation of the assay, 50ng of genomic DNA from each sample was examined qualitatively on a 1% Tris-acetate-EDTA agarose gel for visual signs of degradation. Any degraded DNA samples were excluded from further analysis. Samples were quantitated by OD Spectrometry and diluted to 50ng/µl in reduced EDTA TE buffer (10 mM Tris HCL, 0.1 mM EDTA, pH 8.0). 250 ng of DNA was then aliquotted into two 96 well reaction plates and digested in either Sty or Nsp restriction enzymes (New England Biolabs, Inc. Ipswich, MA) for 2 hours at 37°C followed by 65°C for 20 min. Sty and Nsp digested samples were then ligated to either the Sty 1 or the Nsp 1 adaptor (Affymetrix), respectively, with T4 DNA Ligase (New England Biolabs) for 3 hours at 16°C then 20 min at 70°C. The ligated samples were then diluted in molecular-grade water and subaliquotted into 3 (Sty) or 4 (Nsp) 96 well PCR plates. PCR was performed using PCR Primer 002 (Affymetrix) and Titanium Taq DNA Polymerase (Clontech, Mountain View, CA) with the following thermal cycling parameters: 1. 94°C for 3 min., 2. 30 cycles of 94°C for 30 sec., 60°C for 30 sec., and 68°C for 15 sec., and 3. 68°C for 7 min. Replicate samples for all Sty and Nsp reactions were pooled into a single deep well plate, the DNA was bound to Agencourt AMPure beads (Beckman Coulter, Inc. Berea, CA), placed into MultiScreen filter plates (Millipore, Billerica, MA), washed with 75% ethanol and eluted with Buffer EB (QIAGEN, Valencia, CA). Purified samples were then fragmented using Fragmentation Reagent (Affymetrix) and incubated at 37°C for 35 minutes then at 95°C for 15 minutes. Fragmented samples were labeled with DNA Labeling Reagent (Affymetrix) and TdT Enzyme (New England Biolabs) at 37°C for 4 hours followed by 95°C for 15 min. The samples were denatured at 95°C for 10 minutes and held at 49°C until they were loaded on to the arrays. The arrays were placed into the hybridization oven at 50 °C and 60 rpm for 16 to 18 hours. Arrays were then washed, stained and immediately imaged on the GeneChip Scanner 3000 (Affymetrix).

Genotype quality control. Birdsuite (Korn et al., 2008) was used to call SNP genotypes from CEL files. The DNA quality control pipeline was similar to that described in Anderson et al (Anderson et al., 2010). Briefly,

samples were checked for gender-discord with SNP calls using PLINK v1.07 (Purcell et al., 2007). Twenty-eight samples were removed based on gender errors. Samples were then checked for missingness and heterozygosity errors using PLINK v1.07 and R v3.0.2 (2013-09-25). Samples with a failure rate $\geq$ 0.15 (n= 18) and a heterozygosity rate +/- 3 standard deviations from the mean (n=30) were removed. Our datasets were then examined for whether they contained related individuals. Pedigree files were pruned using PLINK v1.07 to exclude SNPs with pairwise LD threshold of r^2>0.5 using a sliding window of 50 SNPs and a shift of 5 SNPs at each step. Additionally, known regions of high LD on chromosomes 1, 2 (two regions), 3 (two regions), 5 (two regions), 6 (three regions), 7, 8 (three regions), 10, 11, 12 and 20 were pruned as in Anderson et al (Anderson et al., 2010). Samples with PIHAT scores > 0.185 were excluded as in Anderson et al (Anderson et al., 2010). Fourty-four samples were excluded by this criterion. Using slightly more stringent PIHAT scores (0.05) does not substantially change the series (final set 161 cases and 166 controls for KRONOSII and 114 cases and 119 controls for RUSH).

Samples were also analyzed for genetic ancestry via EIGENSOFT (Price et al., 2006). SNPs from the pruned pedigree files generated to determine PIHAT scores were used in the analysis of population stratification. The reduced sets of real data (n=1,415 KRONOSII set, N=427 RUSH set) were merged with data (n=395 individuals) from the HapMap Phase 3 project from four ethnic populations(International HapMap et al., 2007) as in Anderson et al (Anderson et al., 2010). To get an idea of the ethnic ancestry structure for each set in comparison to Hapmap 3, the analysis did not exclude any outliers (sigma threshold equal to 100 and the number of principle components along which to remove outliers was set to 0). After Eigensoft analysis, graphs and output files were examined to determine individuals of divergent ancestry. Samples that were 6 standard deviations away from the average of either PC1 or PC2 in the collection of self-reported Caucasian samples were dropped from the analysis. A total of 25 samples were dropped in this analysis. PC1 and PC2 were retained in the remaining cohort and used as variables to correct both RNA and proteome profiles.

SNPS were first assessed for missingness. SNPs where more than 90% of the cohort was not called in either the KRONOSII or RUSH sets were dropped. A total of 31,886 SNPs were dropped from the KRONOSII set and 41,350 SNPs from the RUSH set were removed. Hardy Weinberg equilibrium was examined in both sets. SNPs that deviated significantly from HWE at a p-value cut-off of 1XE-06 in either the affected or unaffected cohorts were excluded from both sets. For the KRONOSII series 1,758 SNPs were excluded and for the RUSH series 29,601 SNPs were excluded. The mishap test function was used to determine whether any variant had a highly significant pattern of non-random missingness. A total of 2,785 SNPs were dropped from the KRONOS set and 888 were dropped from the RUSH set. Finally, SNPs with a minor allele frequency less than 5% were excluded.

Genotypes were phased with SHAPEIT v2.r790 (Delaneau et al., 2011), and missing genotypes were imputed with Impute2 v2.3.2 (Howie et al., 2009) using the reference panel from the 1000 Genomes Project Phase 3 (Genomes Project et al., 2015). Markers with high imputation quality (INFO>0.5 (Howie et al., 2009); minor allele frequency over 1% and Hardy-Weinberg p-value < 1XE-06 were retained for downstream analysis.

RNA sample processing and data collection. Briefly, frozen blocks of cortex were manually dissected from postmortem brain tissue, and total RNA was isolated using the RNeasy lipid tissue kit (Qiagen, Valencia, CA). Prior to runs RNA quality was assessed by gel visualization and RIN values. RNA with RIN values $\leq$ 6 were not used. Two hundred and fifty-four samples failed QC and were not run. For the remaining samples, 250ng of RNA was reverse transcribed into cRNA and biotin-UTP labeled using the Illumina® TotalPrep™ RNA Amplification Kit from Ambion, Inc. (catalogue # L-1755). cRNA was hybridized to Illumina HumanRefseq-HT-12 v3 Expression BeadChip (Illumina, San Diego, CA) using a Scigene Little Dipper® Automated Microarray Processor (Scigene, Sunnyvale, CA) which allows for parallel processing of chips for all post-hybridization procedures. Between 6 and 111 samples were run in parallel for each hybridization. The last channel of each chip included cRNA prepared from a human brain reference RNA (FirstChoice® Human Brain Reference RNA, Ambion catalogue # AM6052, Life Technologies, Grand Island, NY) to account for noise due to technical variability in hybridizations across chips. Chips were imaged using an Illumina BeadStation 500GX according to the manual (revision C). All expression profiles were extracted using the GenomeStudio software available from Illumina (http://www.illumina.com/pages.ilmn?ID=170). Samples that failed the collection QC pipeline were rerun once. After all reruns were completed, a total of 898 RNA samples passed QC.

RNA quality control. Expression profiles were extracted and background was subtracted using the BeadStudio software available from Illumina. Beadstudio was also used to impute missing bead types. Prior to normalization, the RNA series was reduced to only those samples that had protein profiles that passed QC (see Proteomic quality control section below). Of the 898 RNA samples that passed QC, 754 had protein profiles. One hundred and twelve samples were not run due to time and cost factors, 10 RNA QC samples did not pass the peptide calling QC, 21 RNA QC samples had low peptide counts and 1 RNA QC sample was a peptide outlier. In total there were 345 KRONOSII samples (177 controls, 168 cases) and 409 RUSH samples (153 controls, 141 cases, 115 mild cognitive impairment).

In our first expression quantitative trait loci experiment, we noted that hybridization date was one of the strongest noise variables affecting our profiles (Supplemental Figure 2 in Webster et al, 2009). An Ambion control was run in the last channel of each chip and profiles from these samples were used to circumvent this noise variable. Briefly, the geometric mean for each probe from the profiles of all of the Ambion controls used (n= 103) was calculated. The profile of each Ambion control sample was then divided by the vector of the geometric mean values to obtain a per-chip normalization factor. Profiles from the samples of interest were then corrected by dividing each probe profile by the normalization factor for the Ambion control that was run on the same chip. The LumiB function in Lumi (Du et al., 2008) was employed to force all transcript probe profiles to positive values. After hybridization date correction using the Ambion control, probes were filtered according to the following rules: 1. probes with Infinite and NA values were removed (24% of probes removed); 2. probes with low signal-to-noise ratio, i.e. P95/P05<7 were removed (0 probes removed); 3. probes that were not detected (detection p-

value>0.05) in more than 10% of samples in all datasets (control, late onset Alzheimer's disease and mild cognitive impairment) were removed (44% of probes removed). The NEQC function of Limma (Ritchie et al., 2015) was used for background correction and quantile normalization on filtered probe profiles.

To correct for potential biological or technical confounders, the contribution of every covariate of transcriptional variability was characterized using the variancePartition method (Hoffman and Schadt, 2016). This method partitions the total variance into the contribution of each variable in the experimental design (e.g. age, sex, batch), plus the residual variance. The sample data was adjusted for several biological covariates (gender, age at death and cortical region) and several methodological covariates (institute source of sample, post-mortem interval, detection and hybridization date). While *APOE* is an important covariate to consider for these 'omics relationships, *APOE* genotype is collinear with diagnosis; therefore, corrections did not include *APOE* genotypes. The residuals from this correction were then used in downstream analysis. The residual output was analyzed using singular value decomposition of the expression matrices and plots were created of the first two explanatory vectors. None of the technical and or biological potential confounds clustered in the SVD plots, demonstrating that the corrections were sufficient to remove any captured variable that might skew our data.

Proteomic sample processing and data collection. Tissue was placed in a 2 ml 96-deepwell plate and 1ml of homogenization buffer (8M Urea, 10 mM DTT in 50 mM Tris-HCl) was added. Homogenization was performed using a Retsch Mixer Mill MM 400 at 20 Hz for 2 minutes. 100 µL aliquots were taken from homogenized samples and filtered using 1.2 µm low-protein binding filter plates (Pall Life Sciences) to remove particulates. Total protein concentration was determined by coomassie assay. Protein concentrations were then normalized and a subsequent 100 µL aliquot was taken to provide 150 µg of starting material per sample. Samples were then incubated 1 hr at 37 °C for denaturation and reduction. Protein cysteinyl residues were alkylated with 40mM iodoacetamide (Sigma Aldrich) for 1 hr at 37 °C in the dark. Samples were diluted 4-fold with 50 mM NH_4HCO_3 buffer prior to digestion using mass spectrometry grade trypsin (Promega) with a ratio of 1:50 (w/w). Tryptic digests were desalted using C_{18} SPEC tips (Varian). Peptides were eluted in 200 µl 80 % ACN/ 0.1 % TFA and lyophilized. Peptide concentrations were determined by BCA assay and adjusted to 0.5 mg/ml prior to storage at -80 °C until liquid chromatography–mass spectrometry analysis. Sample randomization, denaturation, alkylation, tryptic digestion, SPE, and normalization was carried out using a Biomek FX (Beckman Coulter) liquid-handling robot. For quantitative measurements of relative peptide abundances, isotopic $^{16}O/^{18}O$ technique was applied where each sample was spiked with a reference sample derived by pooling all the samples followed by ^{18}O-labeling.

Samples were analyzed on an in-house built 4-column liquid chromatography system that allowed for analysis of up to 24 samples per day when employing a 60 min separation gradient. The liquid chromatography system was custom built using two Agilent 1200 nanoflow pumps and one 1200 capillary pump (Agilent Technologies, Santa Clara, CA), various Valco valves (Valco Instruments Co., Houston, TX), and a PAL autosampler (Leap Technologies, Carrboro, NC). Full automation was made possible by custom software that

allows for parallel event coordination and therefore near 100% mass spectrometry duty cycle through use of four analytical columns. Reversed-phase columns were prepared in-house by slurry packing 3-µm Jupiter C_{18} (Phenomenex, Torrence, CA) into 35-cm x 360 µm o.d. x 75 µm i.d fused silica (Polymicro Technologies Inc., Phoenix, AZ) using a 1-cm sol-gel frit for media retention (Zimmer et al., 2006). Mobile phases consisted of 0.1% formic acid in water (A) and 0.1% formic acid in 100% acetonitrile (B) with a gradient profile as follows (min:% B/event); 0:0, 1.2:8, 12:12, 45:35, 58:60, 60:85. Sample injection occurred 20 min prior to beginning the gradient while data acquisition lagged the gradient start and end times by 10 min to account for column dead volume that allowed for the tightest overlap possible in multi-column operation. Multi-column operation also allowed for columns to be 'washed' (shortened gradients) and re-generated off-line without any cost to duty cycle.

Mass spectrometry analysis was performed using an LTQ Orbitrap mass spectrometer (Thermo Scientific, San Jose, CA) outfitted with a custom electrospray ionization (ESI) interface. Electrospray emitters were custom made by chemically etching 150 um o.d. x 20 um i.d. fused silica (Wang et al., 2011). The heated capillary temperature and spray voltage were 250ºC and 2.2 kV, respectively. Mass spectrometry spectra (AGC $1x10^6$) were collected from 400-2000 m/z at a resolution of 50k. Instrument cleaning and any necessary maintenance is performed at 24 hr intervals to help control instrument drift.

Proteomic quality control. Identification and quantification of peptides was performed using the accurate mass and time tag approach (Zimmer et al., 2006). The accurate mass and time tag database was populated using a pooled sample that was fractionated offline by high-pH reversed-phase liquid chromatography (bRPLC) prior to Liquid chromatography-tandem mass spectrometry analysis on an LTQ-Orbitrap (Thermo Scientific) (Wang et al., 2011). Tandem mass spectrometry data was searched by MS-GFDB (Kim et al., 2010). Confident identifications were compiled along with the observed elution times to generate the accurate mass and time tag database. For individual study samples, Decon2LS was used for peak-picking and for determining isotopic distributions and charge states (Jaitly et al., 2009). Deisotoped spectral information was loaded into VIPER to find and match liquid chromatography–mass spectrometry features (same monoisotopic mass present in a number of consecutive mass spectrometry scans) to the peptide identifications in the accurate mass and time tag database (Monroe et al., 2007). VIPER provided an intensity report for all detected features, normalized liquid chromatography elution times via alignment to the database, and feature identifications. The relative peptide abundances in samples were assessed based on the $^{16}O/^{18}O$ ratios for detected peptide pairs, where natural ^{16}O peptide abundances from different voxels were compared to the same spiked reference sample with ^{18}O-labeled peptides(Petyuk et al., 2007). This resulted in a list of 3,719 peptides, which mapped to 1,059 proteins, to be used for downstream analysis.

Peptide abundance crosstabs were exported to R and excel for further normalization procedures. All calculations of variance were carried out using peptide-level abundance measurements. This was done because peptides are the chemical entities being measured in a bottom-up proteomics experiment; thus, it was decided that the best measure of variability in protein targets will be peptide abundances. Finally, protein inference (Kall

et al., 2007; Zhang et al., 2007) and peptide rollup to gene based profiles (Milac et al., 2012; Polpitiya et al., 2008) are active areas of research and prone to uncertainties, which could potentially impact our results in unanticipated ways.

As with the transcript expression data, it was important to control for batch variation in our proteomic runs. Each proteomics plate contained 8 controls, for a total of 88 controls throughout the series. These controls were created by pooling across all samples. Peptide profiles from control samples were used to scale abundances per plate via a peptide-specific linear model. Further normalization involved sample-to-sample scaling to account for pipetting error and means-centered normalization. Data was examined to determine the proportion of missing peptides per sample. Samples where the peptide counts were less than twice the standard deviation of the peptide counts for the whole series were dropped (n=20). Peptides where the peptide was detected in less than 45% of controls, cases or mild cognitive impairment samples were dropped. The final set included 754 samples with 1,931 peptides and 618 genes. The same method was employed as was done for the transcript data to correct for biological and methodological confounders. Namely, we regressed expression profiles against the sample data for several biological and methodological covariates (biological: gender, age at death and cortical region; methodological: institute source of sample, post-mortem interval and a covariate based on the proportion of transcripts detected in each sample). The residuals from this correction were then used in downstream analysis.

All processed data for analysis are available through the Laboratory of Functional Genomics website (http://labs.med.miami.edu/myers/LFuN/LFuN.html).

Experimental design. Samples were plated such that runs included cases and controls on each array/run randomizing samples within each set, Additionally, a standard sample was run throughout the entire series for both DNA, RNA and proteomics collections to ensure quality control. See Piehowski et al for further details on randomization methods (Piehowski et al., 2013).

QUANTIFICATION AND STATISTICAL ANALYSIS

Differential expression analysis. The normalized transcript probe and peptide residuals were examined for significant differential expression using the Limma package (Ritchie et al., 2015). Differential expression was calculated in both KRONOSII and RUSH, comparing late onset Alzheimer's disease confirmed samples to pathological controls. For the RUSH series, we also examined differential expression between mild cognitive impairment samples and pathological controls; however, there were no significant effects. The mild cognitive impairment data was not carried forward except for our final targets, where profiles were graphed to ensure that mild cognitive impairment represented a mid-point between control and late onset Alzheimer's disease states for these particular targets. Benjamini-Hochberg correction (5% the false discovery rate) was used to correct for multiple testing.

Expression quantitative trait loci analysis. Expression quantitative trait loci are targets where the downstream expression correlates in a linear fashion with upstream allele dose (Myers, 2012, 2013, 2014; Myers et al., 2007; Webster et al., 2009). The most parsimonious mode of action for these effects is that an allelic variant is changing a downstream transcription factor-binding or enhancer site, which then alters the binding of that transcription factor and results in linear changes in expression; therefore, we focused on recessive models with a dosage effect for our expression quantitative trait loci analysis. Expression quantitative trait locus analysis was conducted using the Matrix eQTL R package (Shabalin, 2012). Briefly, for each microarray probe and each imputed SNP a linear regression was run with the normalized, non-adjusted expression levels as response variable and the genotypes as regressor variables. The following covariates were included in the linear regression: age, sex, PMI, DET, region, site, hybridization date and diagnosis. Both cis (defined as a SNP-probe pairs within 1Mb of each other) and trans analyses were performed. To account both for dependence between individual tests and for multiple testing, the expression quantitative trait loci analysis was repeated 5 times while permuting the sample IDs of the expression data. We then used the fdrci (Millstein and Volfson, 2013) R package to compute a p-value threshold corresponding to a false discovery rate of 5%. Due to the heavy computational load of running an expression quantitative trait loci analysis on several tens of thousands of probes and millions of SNPs we only performed 5 permutation runs. The fdrci package computes a confidence interval for the false discovery rate estimate, so that any excessive uncertainty due to the low number of permutations would be apparent through a large confidence interval. However, none of the confidence intervals were excessively large and the upper bound of the 95% confidence interval was used as the false discovery rate estimate for deciding the p-value threshold.

Co-expression network. For each dataset (KRONOSII and RUSH) six co-expression networks were constructed. Co-expression networks were predicted separately using the case and control data from KRONOSII and RUSH for the transcript profiles only (15,297 transcripts), the peptide profiles only (1,931 peptides), and a multiscale network including both transcripts and peptides (17,228 targets). We built different networks for several reasons. First, the peptide data was ~10 fold sparser than the transcript dataset, which could affect results. Second, co-expression relies on pairwise correlations. Even if transcripts and peptides are acting in concert, due to the nature of the technical differences in collection they may not be linearly correlated. Additionally, peptide and transcript profiles might not be correlated for legitimate biological reasons. For example, peptides might be overexpressed due to a cleavage/processing change, whereas the transcript profiles will remain unchanged due to the fact that there is no difference in expression in late onset Alzheimer's disease. APP is a good example of this effect.

First, transcript and peptide profiles were compared using Spearman's rank correlations. The set of 1,931 peptides was matched to the corresponding transcript probes. For cases where there were multiple transcript probes per peptide or multiple peptides per transcript probe, each combination was considered. The entire

dataset used for comparison included 2,584 transcript probe-peptide pairs. The mean correlation is close to zero confirming a very low linear correlation between transcript and peptides in KRONOSII and RUSH.

Next, coexpression networks were constructed using weighted gene co-expression network analysis (Langfelder and Horvath, 2008). To evaluate the weighted gene co-expression network analysis aggregate multiscale transcript-peptide network, the percentage of transcript probe or peptide target content was determined for each module. Module content was scored on a continuous scale from 0 to 1, with 0 indicating 100% peptide content for that module and 1 indicating 100% transcript probe content. A number in between 0 and 1 indicates hybrid transcript peptide content for that module (See Supplemental Figure 2).

Further evaluation of the networks was performed to determine whether univariate (peptide alone, transcript alone) or multivariate (peptide and transcripts together) should be employed. To evaluate this model, the uncertainty (R, minus log-likelihood ratio) of each dataset under a univariate linear Gaussian model was compared to the uncertainty in the multivariate multinomial model for all pairs of targets. The uncertainty was normalized by dividing its value by the maximal uncertainty across all pairs of targets under each model. R, the minus log-likelihood ratio, was normalized by the same method. Comparisons were made using Wilcoxon-tests. The multivariate/multinomial model yields significantly higher rank and thus lower uncertainty than univariate linear models for the same set of transcript-peptide pairs plotted ($p < 2.2e-16$). There were significant numbers of transcript-peptide pairs with low linear correlation under the univariate/linear model but moderate to high nonlinear correlation under multivariate model indicating that multinomial model can capture nonlinear correlations between gene and protein. We concluded that a multinomial based multivariate model is more suitable than methods based on correlations such as weighted gene co-expression network analysis to model multi-scale network of transcripts and peptides. We carried out causal network analysis based on the multinomial model (see Causal Predictive Network section).

Finally, network structures were evaluated in the aggregate models by comparing module content between KRONOSII and RUSH. First, each module in the four multivariate datasets KRONOSII late onset Alzheimer's disease, KRONOSII control, RUSH late onset Alzheimer's disease, RUSH control was mapped to a gene ontology pathway, by determining gene ontology process enrichment using GoSeq and TopGo in R. Processes were then compared between KRONOSII and RUSH for case datasets separately from control datasets. This was to get a general sense of overlap and what were the gross processes involved in late onset Alzheimer's disease declines. Data is shown in Figure 3. Second, module membership was compared across late onset Alzheimer's disease predictions and control predictions between KRONOSII and RUSH. Each pairwise comparison was considered for the late onset Alzheimer's disease KRONOSII modules and the late onset Alzheimer's disease RUSH modules. Likewise, pairwise comparisons were made between the KRONOSII control modules and the RUSH control modules. Comparisons were made by matching transcript probe or peptide target names, and the proportion of overlap was averaged considering both how many RUSH targets were in the corresponding KRONOSII module and how many KRONOSII targets were in the corresponding

RUSH modules. In this way, perfect overlap was determined rather than just one module being a subset of another.

Module selection. The differential expression KRONOSII and RUSH datasets were used to capture a reduced set of key data for the multi-scale causal predictive network prediction. Reduction was performed because causal predictive network algorithms are computationally inefficient, since each possible 3-way relationship needs to be tested. Reduction was made at the transcript-level only, since the peptide dataset was already sparse and all peptides could be linked to a transcript path. Transcript inputs were reduced by selecting the set of aggregate coexpression network modules that were enriched for differentially expressed transcripts. Enrichment was defined by modules that had a greater proportion of differential expression hits than would be expected by chance, using a fisher exact test. Module selection was done separately in the KRONOSII and RUSH cohorts, so that these datasets remained independent replicates. Only late onset Alzheimer's disease sets were used to reduce the computational space further, since enriching for differential expression targets will capture case-control differences. There were 9 enriched modules from our KRONOSII dataset that were carried forward and 4 enriched modules from our RUSH dataset that were used as our replicate seeding gene list.

Module Expansion. After module selection, the set of targets from the enriched modules was further expanded by including targets within brain-specific signaling pathways that matched to the differential expression lists to be sure we included all key members of pathways of interest. This step was done to maximize the coverage of our final networks on the selected modules from the coexpression network analysis.

Prior to expansion, the module selection transcript data was reduced to a single target probe per gene because the targets from the databases that were used for expansion were compiled at the gene and not at the transcript level; thus, the set of transcript probes had to be reduced to a single list of one target transcript probe per gene. For the differential expression dataset, probes were first filtered by dropping probes matching to repeat sequences, intergenic regions, intronic regions or unlikely to provide specific signal for any transcript. Additionally, probes not matching to known genomic regions or transcripts were dropped. Reduction was then performed using the collapseRows R function (Miller et al., 2011) from Weighted Gene Co-expression Network Analysis, which reduces probes matching to the same Entrez identifier by selecting the probe with the highest average expression across all samples. For the cis expression quantitative trait loci gene lists, the lists of cis expression quantitative trait loci probes were simply mapped to an Entrez ID. If there were more than one cis expression quantitative trait loci probe mapping to the same Entrez ID, they were collapsed to the same Entrez ID, using the same method as the differential expression data. Peptide data was not reduced.

Expansion of the reduced sets was based on an enrichment network that was used to determine additional targets that were in the same biological pathways as the differential expression and expression quantitative trait loci targets lists, but were not differentially expressed or expression quantitative trait loci themselves. MetaCore from Thomson Reuters (v6.24) and the ConsensusPathDB (Kamburov et al., 2011;

Kamburov et al., 2013) databases were used to build this enrichment network. Both MetaCore and ConsensusPathDB contain interactions at the gene-gene, gene-protein or protein-protein level. The nodes in ConsensusPathDB are classified as genes, protein complex and gene families and these sets are indexed solely by Entrez gene IDs. The nodes in MetaCore are classified into protein and genes and these sets are indexed by network object IDs internal to the MetaCore database. To merge MetaCore and the ConsensusPathDB, node nomenclature was unified to a set of common Entrez IDs by mapping proteins in MetaCore to their corresponding genes; thus, while the merged network contained only gene-gene interactions, these gene-gene interactions represented three possibilities: the original direct gene-gene interactions from MetaCore and the ConsensusPathDB, gene-protein interactions from MetaCore, or protein-protein interactions from MetaCore. This merged enrichment network contained interactions across a variety of human tissues. To make the network brain specific, we integrated the Roadmap Epigenomics Project data (Roadmap Epigenomics et al., 2015) from mid frontal lobe and inferior temporal lobe cell lines to predict which genes are active, repressed or bivalent (poised) in tissue regions relevant to this study. All nodes that were predicted to be repressed in mid frontal or inferior temporal cortex as well as all incoming and outgoing edges from these repressed nodes were removed from the enrichment network. Therefore, the remaining networks are brain-specific and involve interactions of all active genes in the mid frontal lobe and inferior temporal lobe respectively.

Our reduced networks of 7,681 KRONOSII probes and 7,198 RUSH probes were expanded by the data from the enrichment network (24,611 targets) using pathFinder. All neighbor nodes in the enrichment network that were within 3-steps of the 7,681 KRONOSII or 7,198 RUSH targets were extracted from the enrichment network. This list of additional hits was used to pull the transcript profiles of additional targets from the KRONOSII and RUSH data and these were added to the 7,681 KRONOSII and 7,198 RUSH transcript sets. The enriched transcript set was combined with all 1,931 peptides to form the seeding set.

Multi-scale Causal Predictive Network. The co-expression networks reflect modules of highly co-regulated genes and proteins operating in coherent biological pathways; however, such network modules do not reflect the probabilistic causal information needed to identify key driver genes. Key driver analysis relies on sifting through layers of target relationships and it is crucial to understand direction within the networks to determine what are the most upstream connected nodes (Zhang et al., 2013; Zhu et al., 2012).

A typical approach to directing networks is to use Bayesian network prediction; however, there are shortcomings with those models. Specifically, Bayesian networks cannot capture opposite causality, since typical Bayesian network approaches do not distinguish between positive and negative correlations. Therefore, for this screen, the Bayesian networks approach was expanded to develop a multiscale causal predictive network. The causal predictive network approach integrates a conventional Bayesian networks in top-down model with recently developed bottom-up causality inference. Causality interference addresses the problem of opposite causality. Recent work has demonstrated that the causal predictive network powered by bottom-up causality inference can deduce significantly more accurate causality than a conventional Bayesian network approach.

Our causal predictive network pipeline incorporated genotypes, transcript profiles, peptide profiles from KRONOSII and RUSH, as well as external epigenomics and signaling pathway data via our expansion procedures. Causal predictive networks were built separately in KRONOSII and RUSH, using only the late onset Alzheimer's disease data to reduce the search space. Again, targets of importance for late onset Alzheimer's disease are enriched in the inputs, since module selection included enrichment for modules that had significantly more differentially expressed transcripts. Two different runs were performed: one containing only module selection/module enrichment selected gene data (transcript set) and the other contained both module selection/module enrichment selected genes and the full peptide set (multi set).

The basic procedure involved building directed edges (relationships) between nodes (allele, transcript and peptide data). For the aggregate builds of genes and peptides, the causal predictive networks considered three types of molecular interactions: 1. gene-to-peptide (gene produces protein), 2. peptide-to-gene (transcriptional regulation) and 3. peptide-to-peptide (protein-protein interaction). The causal predictive network builds were constrained to parent-child relationships (see Figure 1, step 6). The following structural constraints were applied: 1. Edges were initialized and fixed within the run such that parent-child relationships were restricted to matches (i.e. peptides could not have parents that were not their corresponding genes and vis versa), 2. cis-expression quantitative trait loci genes or cis- protein quantitative trait loci peptides were fixed to be top nodes during structure learning. Unmatched peptide-genes pairs were allowed to have any edge relationship, so long as the resulting graph was directed and acyclic. The second group was only constrained such that cis-expression quantitative trait loci genes or cis- protein quantitative trait loci peptides were fixed to be top nodes.

To demonstrate that our multi-scale predictive network captures additional useful information and not random noise, several sanity checks were performed. First, the causal predictive network multi-scale network was approximated into single-scale networks for both transcripts and proteins to test whether integrating entities at the two-scale, peptide-transcript level did not distort the single-scale network structures. This was achieved by omitting the intermediate proteins or transcripts along any pathway. The downgraded transcript-only or peptide-only network structures were then overlapped with the actual multi-scale predictive network. Fisher's exact tests gave no significant differences, indicating the predictions were robust.

Second, to examine whether the three types of molecule interactions (transcript-to-protein, protein-to-protein, and protein-to-transcript) in the predicted multi-scale network reflected known biology pathways, the predicted transcript-only, and transcript-peptide causal structures were compared to the prior late onset Alzheimer's disease specific signaling pathways from our enrichment network built from the MetaCore and ConsensusPathDB. There were no significant differences, indicating the predictions were robust.

Key driver analysis. After the causal networks were predicted independently in the KRONOSII and RUSH datasets, key driver analysis was performed. Key drivers are targets that are predicted to have a significant impact on the regulatory states of other targets in the modules. Our analysis was performed using the package

key driver analysis in R, which takes either directed or undirected networks as well as a projection dataset as inputs. The goal is to determine whether the members of the projection datasets are significantly over represented in the networks compared to other members of the set. This methods are an improvement over link-counting, as is done in hub gene calculations because it considers enrichments from a test set, which allows for hypothesis testing as well as predicting key drivers by counting links in layers of directed data (Zhang, 2013). Key driver analysis was performed in the causal predictive network for both the transcript-only and transcript-plus-peptide sets. Several projection datasets were used: 1. The module enrichment set only including differentially expressed transcripts, 2. Each transcript module in its entirely, 3. The full differentially expressed transcript set, 4. All transcripts and peptides against the transcript-peptide causal network and 5. The entire peptide set was projected onto the transcript-peptide causal network. Analysis was performed separately in KRONOSII and RUSH. Significance was set to p=0.05, the h-layer neighbor expansion to k=3 and 5 permutations were performed to correct for multiple testing.

In vitro validation: APP and MAPT cell lines. Replication-defective lentiviral particles were generated using a 3rd generation lentiviral packaging system by Origene (Rockville, MD). Briefly, around 80% confluent HEK293T cells in 6 wells plates were transiently co-transfected with 1.0 µg of target open reading frame in lenti vector with C-terminal monomeric GFP tag (pLenti-C-mGFP, OriGene Technologies, Rockville, MD USA), and 1.2 µg of 3rd generation mixture of packaging plasmids. Media was replaced 24 hours later and viral supernatants were collected twice at 24 hour intervals. The combined viral supernatant was centrifuged at 2000g for 5 min and filtered with a 0.45 micron syringe filter. Viral particles were pelleted with Peg-it (SBI, Palo Alto, CA USA) and resuspended in 1 ml of DPBS.

For transduction HEK293sw were seeded at around 300,000 cells/well and H4-4R0N at 80,000 cells/well in 12 wells plates. Transduction was performed 24 hours later when cells reached around 80% confluence by adding 100µl of viral particles. Media at transduction was supplemented with Hexadimethrine bromide (Sigma-Aldrich, St Louis, MO) at a final concentration of 8µg/ml. Cells were incubated with the viral particles at 37°C with 5% CO2 for 24 hours. Cells transduced with the empty vector pLenti-C-mGFP (OriGene Technologies, Rockville, MD USA) were used as controls. One day after transduction, cells were visualized under a fluorescent microscope to check for GFP expression. Media was collected every 24 hours for protein quantification. Four days after transduction RNA was extracted for qPCR gene expression testing.

To determine the levels of Abeta40, Abeta42, Tau and Phospho-Tau, a Luminex 100 system was employed. Singleplex beads for the quantification of Tau (Total), Tau[pT181] and the Beta Amyloid peptides Aβ40 and Aβ42 were obtained from Life Technologies (Carlsbad, CA). For each bead assay, internally red and infrared dyed beads, conjugated with capture antibodies for Tau (total), Tau[pT181], Aβ40 and Aβ42 were incubated with the conditioned culture media from the transduction experiments and specific detector antibodies. Beads were then washed and incubated with R-Phycoerythrin conjugate that binds to the detector antibodies. Beads were washed again and analyzed using the Luminex detection system. The Luminex system uses the

spectral properties of the beads and the RPE fluorescence to compare against standards of known concentration and determine the protein concentration in the conditioned culture media. Since protein production is an ongoing process and can be related to cell division and growth, we performed all measurements of Abeta and Tau as a time-course, with multiple samplings of extracellular media. Conditioned media was sampled at 48, 72 and 96 hours post transduction.

ANCILLARY DATASETS LEGENDS:

Ancillary Dataset 1: Differential expression transcripts. Shown are all the transcripts in the KRONOSII (first tab) and RUSH (second tab) that survived multiple testing correction using Benjamini-Hochberg correction (5% false discovery rate). Illumina manifest (HumanHT-12_v3_0-r2-11383641_A) probe names and gene names are given, as well as gene names from GRC38.p13. Probe_missing_from_assembly indicates that probe is not currently mapped to GRC38.p13. Entrez identifiers are from NCBI and log-fold changes and adjusted p-values are given from the limma analysis. Adjustment is done by Benjamini-Hochberg with a 5% false discovery rate. Overlap between the KRONOSII and RUSH sets is given in the last column.

Ancillary Dataset 2: Differential expression Peptides. Shown are all the peptides in the KRONOSII (first tab) and RUSH (second tab) that survived multiple testing correction using Benjamini-Hochberg correction (5% false discovery rate). Peptide names are given by gene name_underscore_peptide sequence. Overlap between the KRONOSII and RUSH sets is given in the last column.

Ancillary Dataset 3: Cis Overlapping expression quantitative trait loci. Listed are the cis expression quantitative trait loci that overlap between KRONOSII and RUSH. Best SNP indicated the SNP with the lowest p-value for that transcript target.

Ancillary Dataset 4: KRONOSII Transcript Co-expression Networks. Listed are the Entrez identifiers, probe identifiers and module labels for the co-expression network analysis for KRONOSII late onset Alzheimer's disease samples (first tab) and KRONOSII controls (second tab) on the set of 15,297 transcripts. Index is a random number just indicating the order of probes.

Ancillary Dataset 5: RUSH Transcript Co-expression Networks. Listed are the peptide identifiers and module labels for the coexpression network analysis for RUSH late onset Alzheimer's disease samples (first tab) and RUSH controls (second tab) on the set of 15,297 transcripts. Index is a random number just indicating the order of probes.

Ancillary Dataset 6: KRONOSII Peptide Co-expression Networks. Listed are the peptide identifiers and module labels for the co-expression network analysis for KRONOSII late onset Alzheimer's disease samples (first tab) and KRONOSII controls (second tab) on the set of 1,931 peptides. Peptide names are given by gene name_underscore_peptide sequence. Index is a random number just indicating the order of probes.

Ancillary Dataset 7: RUSH Peptide Co-expression Networks. Listed are the peptide identifiers and module labels for the co-expression network analysis for KRONOSII late onset Alzheimer's disease samples (first tab)

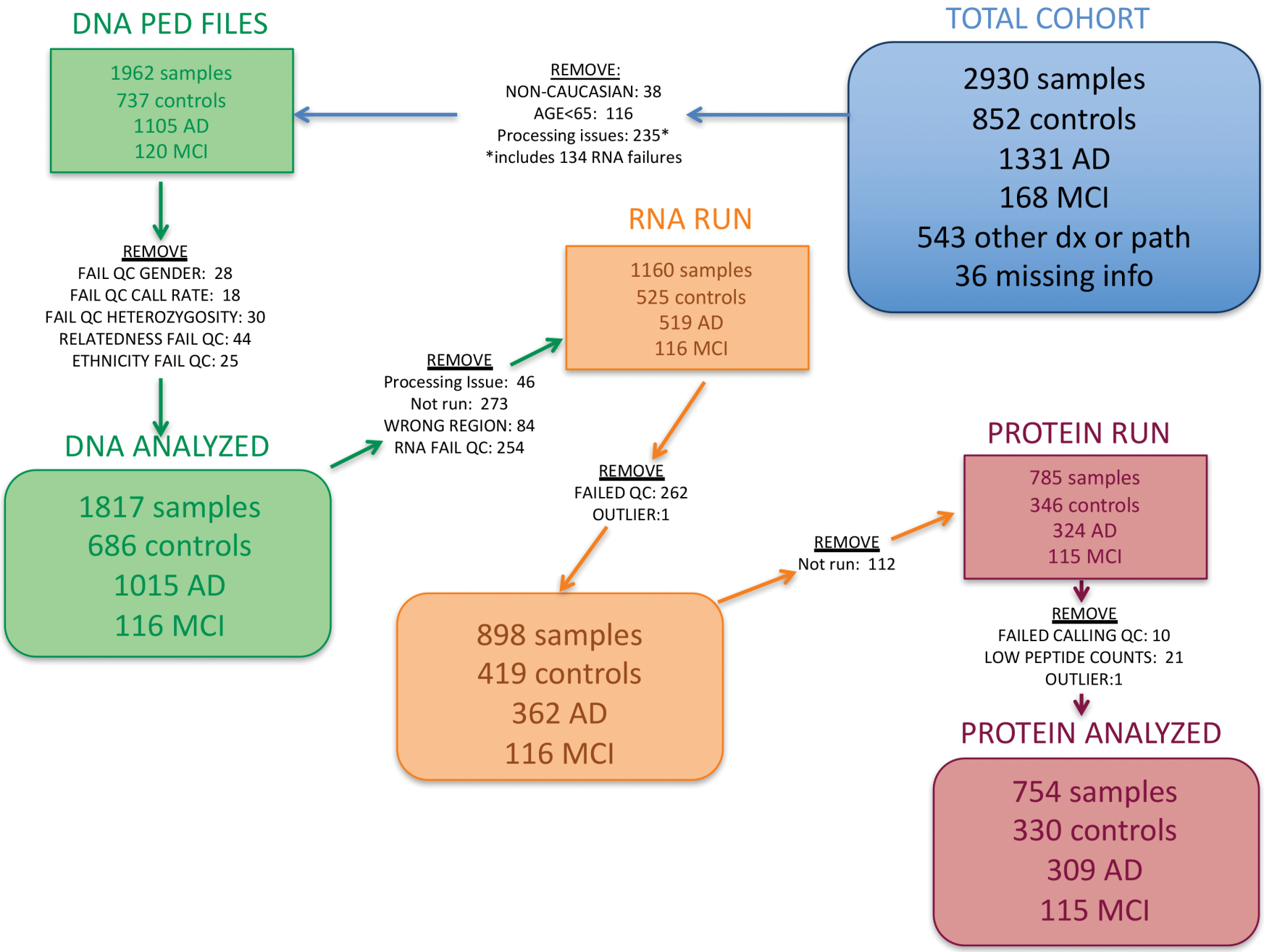
and KRONOSII controls (second tab) on the set of 1,931 peptides. Peptide names are given by gene name_underscore_peptide sequence. Index is a random number just indicating the order of probes.

Ancillary Dataset 8: KRONOSII multiscale Co-expression Networks. Listed are the peptide identifiers and module labels for the co-expression network analysis for KRONOSII late onset Alzheimer's disease samples (first tab) and KRONOSII controls (second tab) on the reduced set of 15,297 transcripts and 1,931 peptides. Peptide names are given by gene name_underscore_peptide sequence. Index is a random number just indicating the order of probes.

Ancillary Dataset 9: RUSH multiscale Co-expression Networks. Listed are the peptide identifiers and module labels for the co-expression network analysis for RUSH late onset Alzheimer's disease samples (first tab) and RUSH controls (second tab) on the reduced set of 15,297 transcripts and 1,931 peptides. Peptide names are given by gene name_underscore_peptide sequence. Index is a random number just indicating the order of probes.

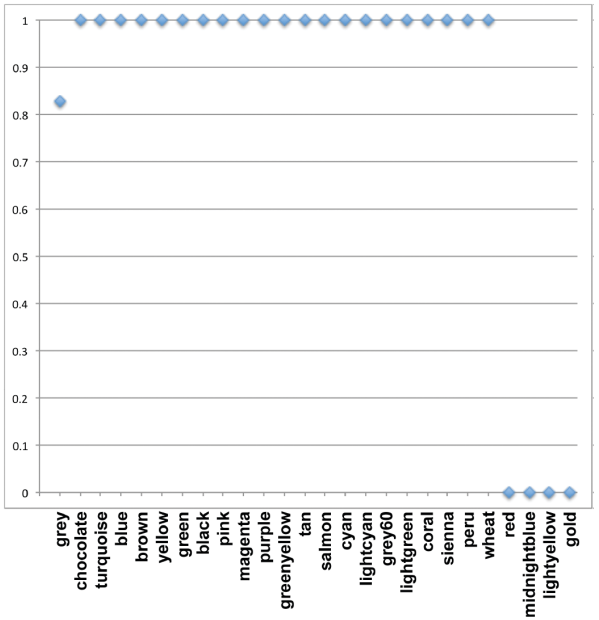
SUPPLEMENTAL FIGURE LEGENDS:

Supplemental Figure 1: Samples. Shown are the sample counts at each step of the data collection process. From a cohort of ~3,000 different individuals, we obtained tissue profiling on ~800 of those individuals for all datasets. Our biggest drops were eliminating samples for co-morbid pathology and diagnostic criteria (19% of the cohort dropped) and there was another major drop when RNA quality was assessed (20% of the cohort dropped, 72% of dropped samples were from affected tissues).

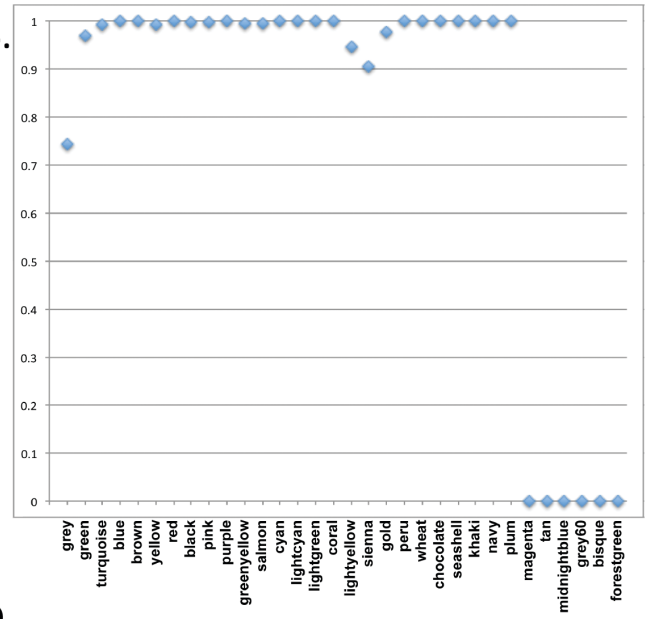

TOTAL COHORT
2930 samples
852 controls
1331 AD
168 MCI
543 other dx or path
36 missing info
REMOVE:
NON-CAUCASIAN: 38
AGE<65: 116
Processing issues: 235*
*includes 134 RNA failures
DNA PED FILES
1962 samples
737 controls
1105 AD
120 MCI
REMOVE
FAIL QC GENDER: 28
FAIL QC CALL RATE: 18
FAIL QC HETEROZYGOSITY: 30
RELATEDNESS FAIL QC: 44
ETHNICITY FAIL QC: 25
DNA ANALYZED
1817 samples
686 controls
1015 AD
116 MCI
REMOVE
Processing Issue: 46
Not run: 273
WRONG REGION: 84
RNA FAIL QC: 254
RNA RUN
1160 samples
525 controls
519 AD
116 MCI
REMOVE
FAILED QC: 262
OUTLIER:1
898 samples
419 controls
362 AD
116 MCI
REMOVE
Not run: 112
PROTEIN RUN
785 samples
346 controls
324 AD
115 MCI
REMOVE
FAILED CALLING QC: 10
LOW PEPTIDE COUNTS: 21
OUTLIER:1
PROTEIN ANALYZED
754 samples
330 controls
309 AD
115 MCI

Supplemental Figure 2: Multi-scale co-expression network and module content comparison. Plotted are the percentage of gene content for each module in (**A**) KRONOSII Alzheimer's disease, (**B**) RUSH Alzheimer's disease, (**C**) KRONOSII controls and (**D**) RUSH Alzheimer's disease. The Y-axis indicates the fraction of the module content that is gene related, with scores of 1 indicating the module is solely comprised of transcripts and a score of 0 indicating the module is solely derived of peptides.

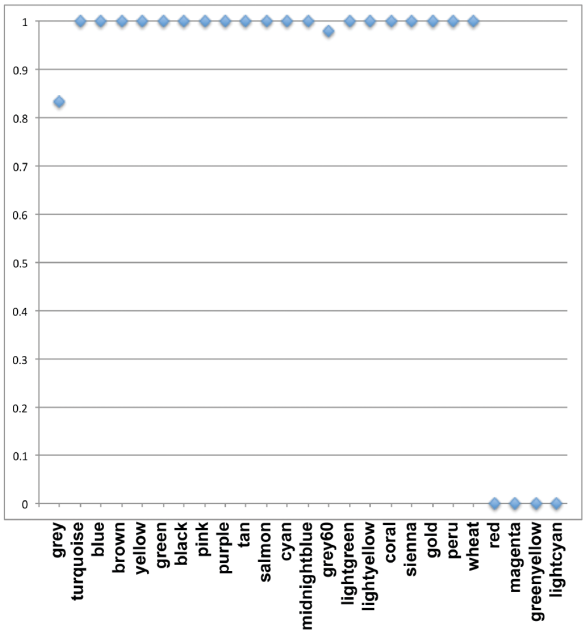
A.


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
grey
chocolate
turquoise
blue
brown
yellow
green
black
pink
magenta
purple
greenyellow
tan
salmon
cyan
lightcyan
grey60
lightgreen
coral
sienna
peru
wheat
red
midnightblue
lightyellow
gold

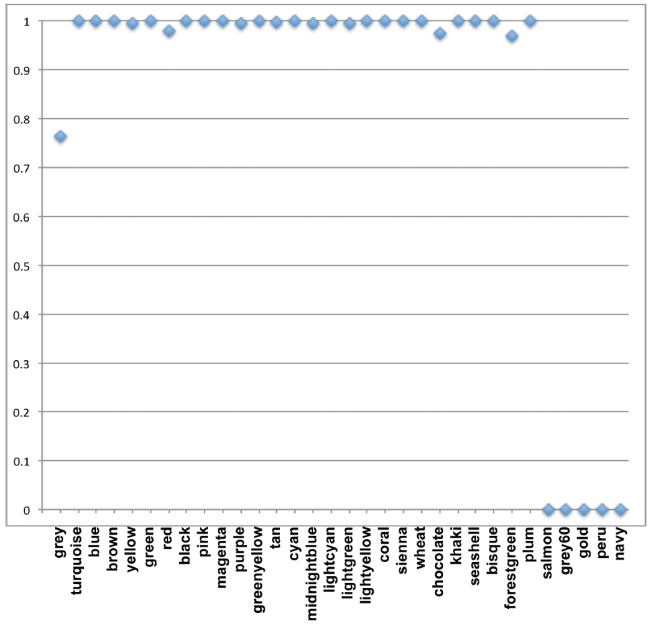
B.


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
grey
green
turquoise
blue
brown
yellow
red
black
pink
purple
greenyellow
salmon
cyan
lightcyan
lightgreen
coral
lightyellow
sienna
gold
peru
wheat
chocolate
seashell
khaki
navy
plum
magenta
tan
midnightblue
grey60
bisque
forestgreen

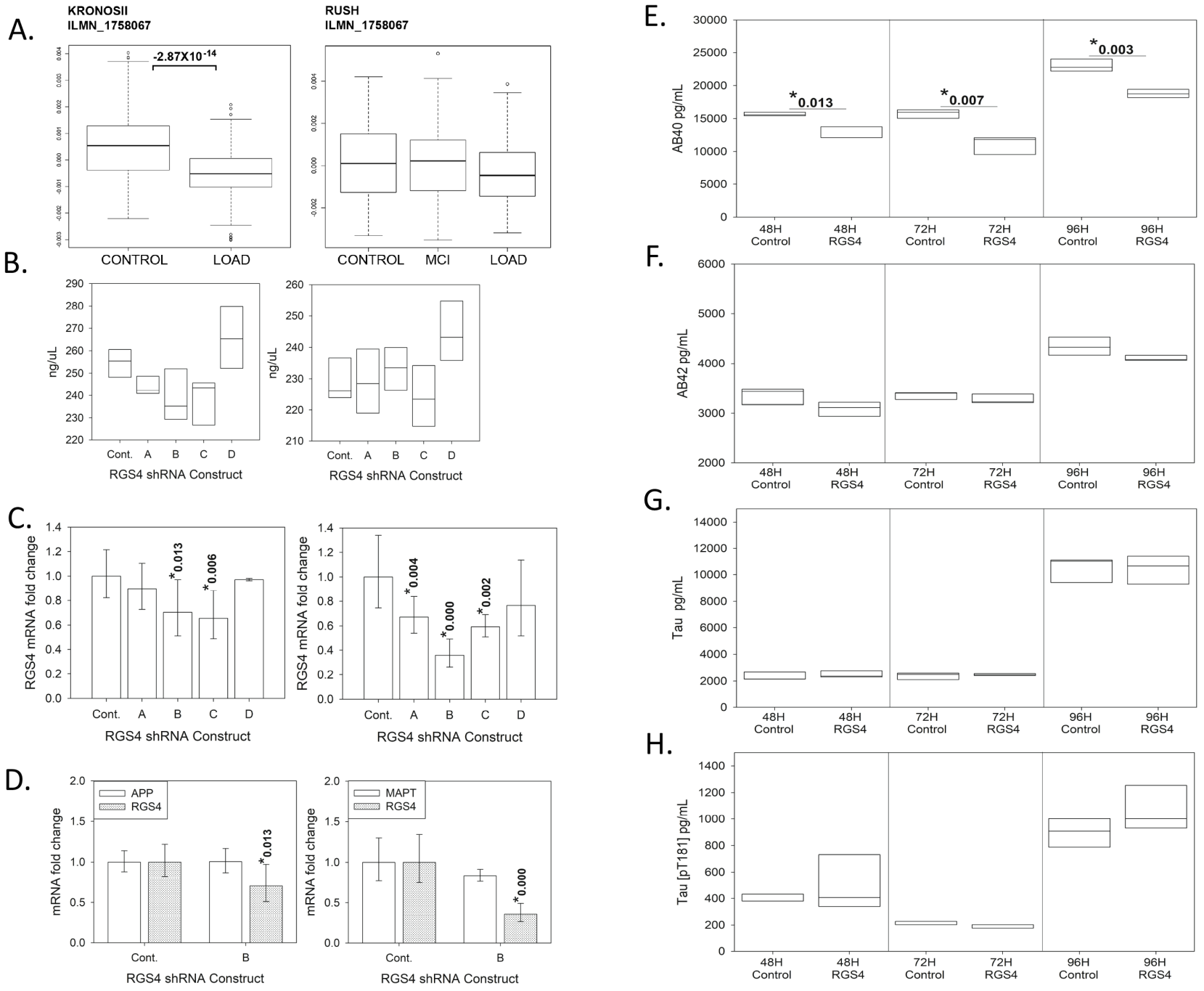
C.


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
grey
turquoise
blue
brown
yellow
green
black
pink
purple
tan
salmon
cyan
midnightblue
grey60
lightgreen
lightyellow
coral
sienna
gold
peru
wheat
red
magenta
greenyellow
lightcyan

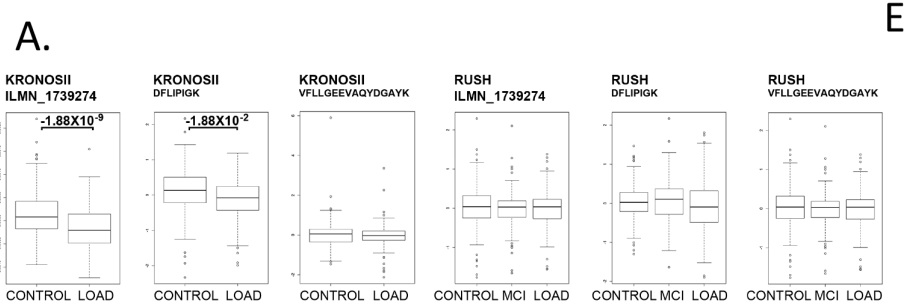
D.

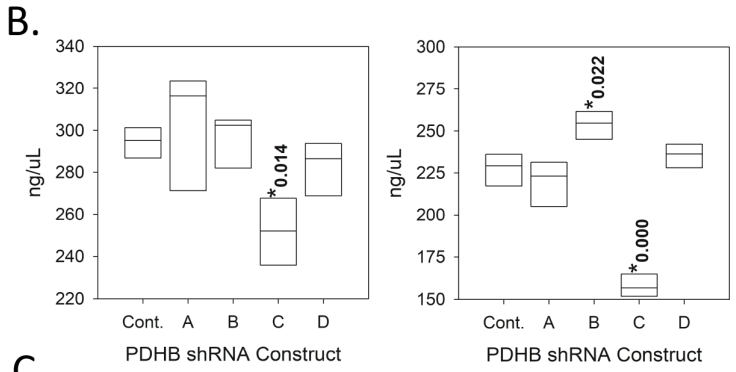

1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
grey
turquoise
blue
brown
yellow
green
red
black
pink
magenta
purple
greenyellow
tan
cyan
midnightblue
lightcyan
lightgreen
lightyellow
coral
sienna
wheat
chocolate
khaki
seashell
bisque
forestgreen
plum
salmon
grey60
gold
peru
navy

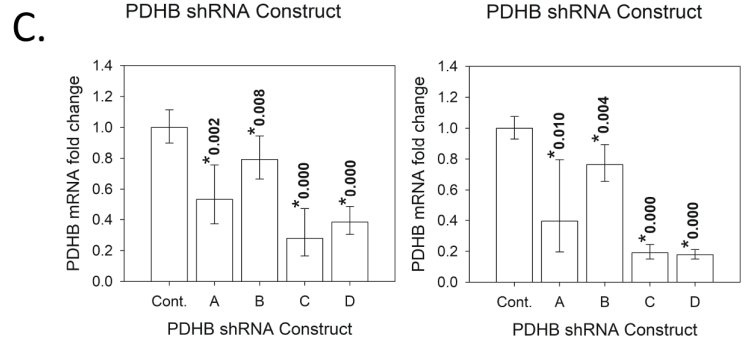
Supplemental Figure 3: ***RGS4*** **results.** Shown in the figure are the differentially expressed transcript probe profiles in KRONOSII and RUSH, including late onset Alzheimer's disease, mild cognitive impairment (RUSH) and controls (**A**), along with the levels of total RNA (**B**, 96 hours shown, measured as a surrogate of the level of cell death, all four constructs shown, left in HEK293sw cells, right in H4-4R0N cells), transcript knockdown for target (**C**, 96 hours shown, all four constructs shown, left in HEK293sw cells, right in H4-4R0N cells) and mRNA (**D**, 96 hours shown, only construct B shown, left HEK293sw cells, right H4-4R0N cells) . Construct B was carried forward as it had the greatest knockdown without affecting total RNA levels in both the HEK293 and H4 lines. (**E**) shows the AB40 levels, (**F**) the AB42 levels, (**G**) the Total Tau levels and (**H**) the Phospho-Tau levels for three repeat measures of conditioned media at 3 different time-points. Measurements are taken at 48, 72 and 96 hours post transduction. Bar in (**A**) gives the limma differential expression p-value. * = t-Test p value.

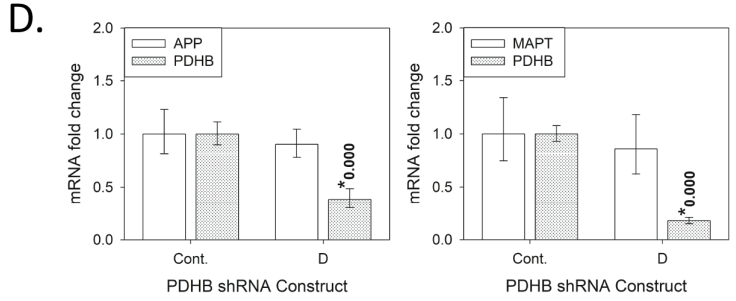

A.
KRONOSII
ILMN_1758067
-2.87X10-14
CONTROL
LOAD
RUSH
ILMN_1758067
CONTROL
MCI
LOAD
B.
ng/uL
Cont. A B C D
RGS4 shRNA Construct
C.
RGS4 mRNA fold change
*0.013
*0.006
*0.004
*0.000
*0.002
D.
mRNA fold change
APP
RGS4
MAPT
RGS4
*0.013
*0.000
Cont.
B
E.
AB40 pg/mL
*0.013
*0.007
*0.003
F.
AB42 pg/mL
G.
Tau pg/mL
H.
Tau [pT181] pg/mL
48H Control
48H RGS4
72H Control
72H RGS4
96H Control
96H RGS4

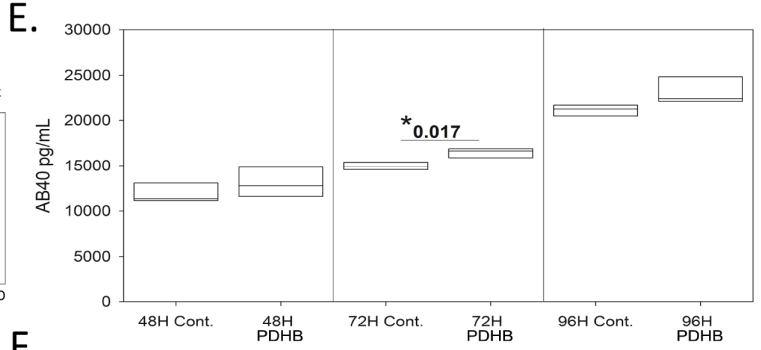
Supplemental Figure 4: ***PDHB*** **results.** Shown in the figure are the differentially expressed transcript probe profiles in KRONOSII and RUSH, including late onset Alzheimer's disease, mild cognitive impairment (RUSH) and controls (**A**), along with the levels of total RNA (**B**, 96 hours shown, measured as a surrogate of the level of cell death, all four constructs shown, left in HEK293sw cells, right in H4-4R0N cells), transcript knockdown for target (**C**, 96 hours shown, all four constructs shown , left in HEK293sw cells, right in H4-4R0N cells) and mRNA (**D**, 96 hours shown, only construct D shown, left HEK293sw cells, right H4-4R0N cells). Construct D was carried forward as it had the greatest knockdown without affecting total RNA levels in both the HEK293 and H4 lines. (**E**) shows the AB40 levels, (**F**) the AB42 levels, (**G**) the Total Tau levels and (**H**) the Phospho-Tau levels for three repeat measures of conditioned media at 3 different timepoints. Measurements are taken at 48, 72 and 96 hours post transduction. Bar in (**A**) gives the limma differential expression p-value. * = t-Test p value.

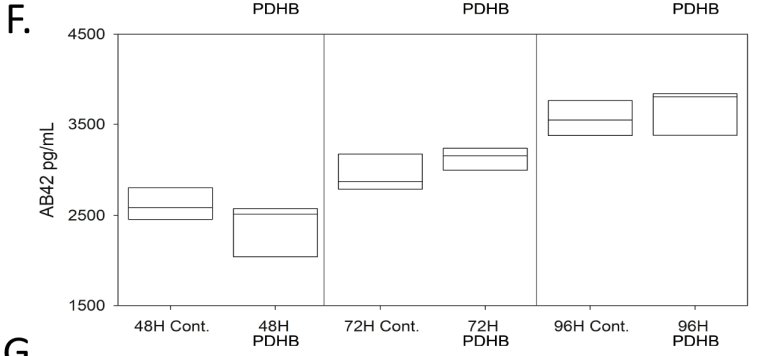

A.
KRONOSII
ILMN_1739274
-1.88X10-9
KRONOSII
DFLIPIGK
-1.88X10-2
KRONOSII
VFLLGEEVAQYDGAYK
RUSH
ILMN_1739274
RUSH
DFLIPIGK
RUSH
VFLLGEEVAQYDGAYK
CONTROL LOAD
CONTROL LOAD
CONTROL LOAD
CONTROL MCI LOAD
CONTROL MCI LOAD
CONTROL MCI LOAD

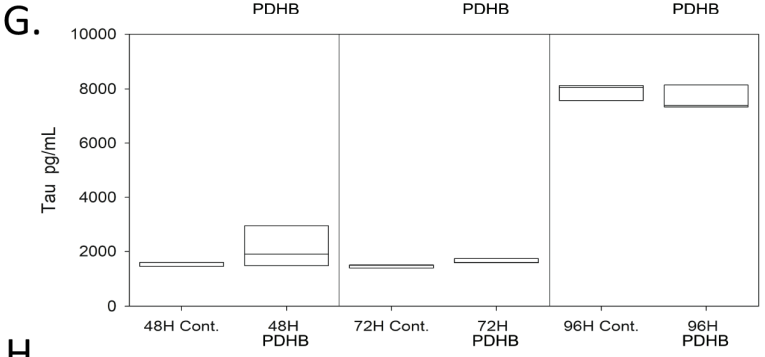

B.
ng/uL
*0.014
*0.022
*0.000
Cont. A B C D
PDHB shRNA Construct

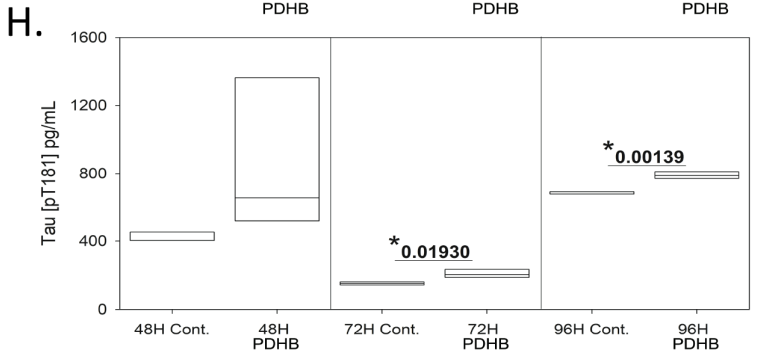

C.
PDHB mRNA fold change
*0.002
*0.008
*0.000
*0.000
*0.010
*0.004
*0.000
*0.000
Cont. A B C D
PDHB shRNA Construct


D.
APP
PDHB
MAPT
PDHB
mRNA fold change
*0.000
*0.000
Cont. D
PDHB shRNA Construct

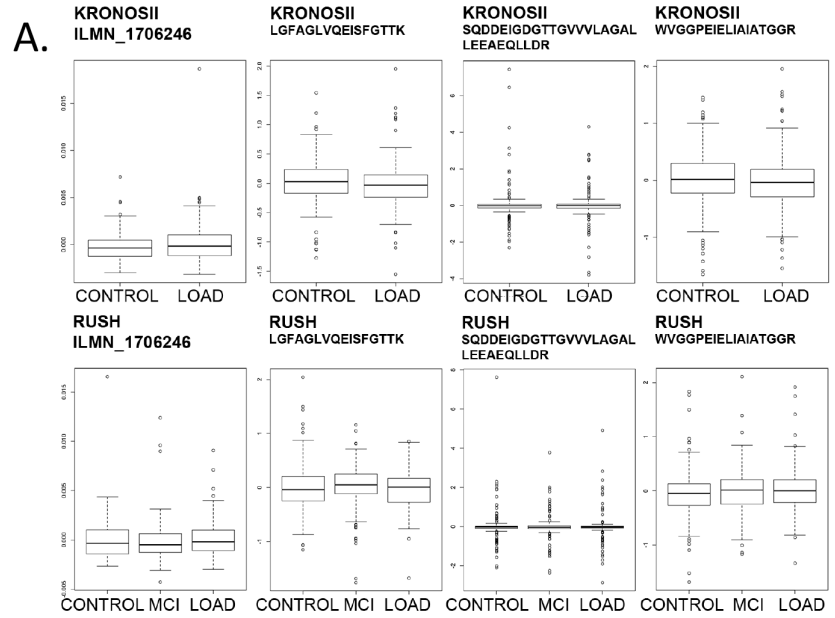

E.
AB40 pg/mL
*0.017
48H Cont.
48H PDHB
72H Cont.
72H PDHB
96H Cont.
96H PDHB

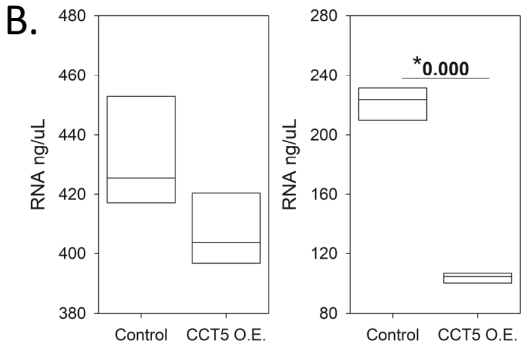

F.
AB42 pg/mL
48H Cont.
48H PDHB
72H Cont.
72H PDHB
96H Cont.
96H PDHB

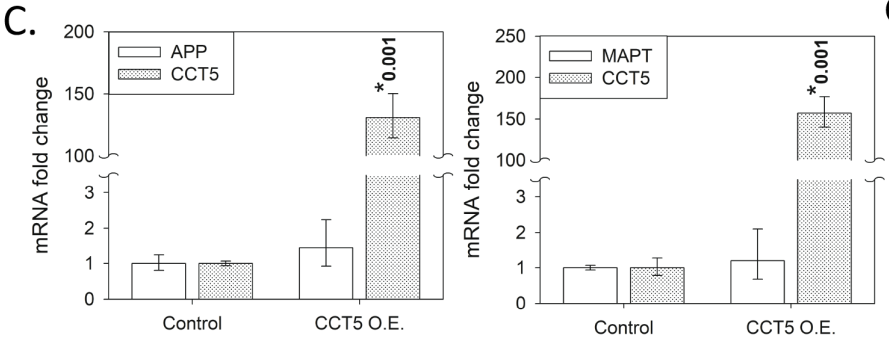

G.
Tau pg/mL
48H Cont.
48H PDHB
72H Cont.
72H PDHB
96H Cont.
96H PDHB

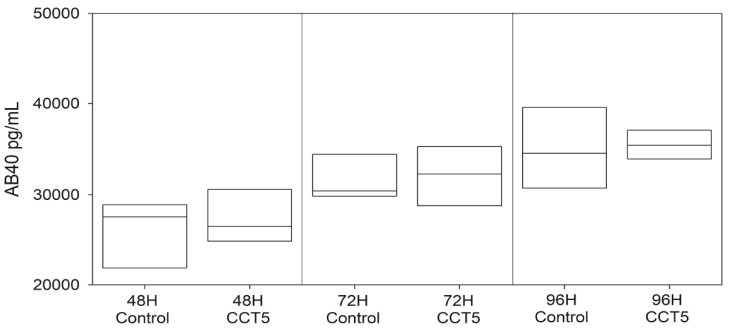

H.
Tau [pT181] pg/mL
*0.01930
*0.00139
48H Cont.
48H PDHB
72H Cont.
72H PDHB
96H Cont.
96H PDHB

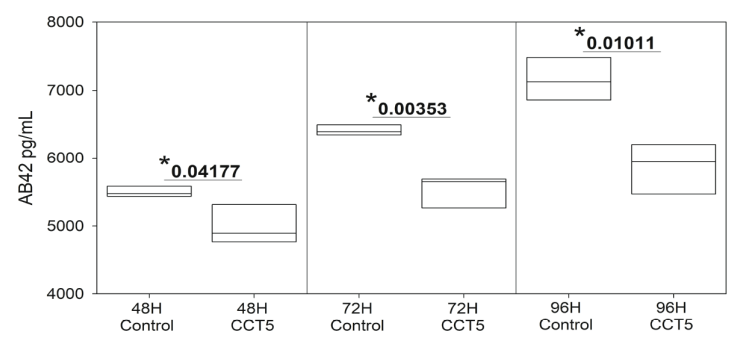
Supplemental Figure 5: ***CCT5*** **results.** Shown in the figure are the differentially expressed transcript probe profiles in KRONOSII and RUSH, including late onset Alzheimer's disease, mild cognitive impairment (RUSH) and controls (**A**), along with the levels of total RNA (**B**, 96 hours shown, measured as a surrogate of the level of cell death, left in HEK293sw cells, right in H4-4R0N cells), and mRNA overexpression for target (**C**, 96 hours shown, left in HEK293sw cells, right in H4-4R0N cells). (**D**) shows the AB40 levels, (**E**) the AB42 levels, (**F**) the Total Tau levels and (**G**) the Phospho-Tau levels for three repeat measures of conditioned media at 3 different timepoints. Measurements are taken at 48, 72 and 96 hours post transduction. * = t-Test p value.

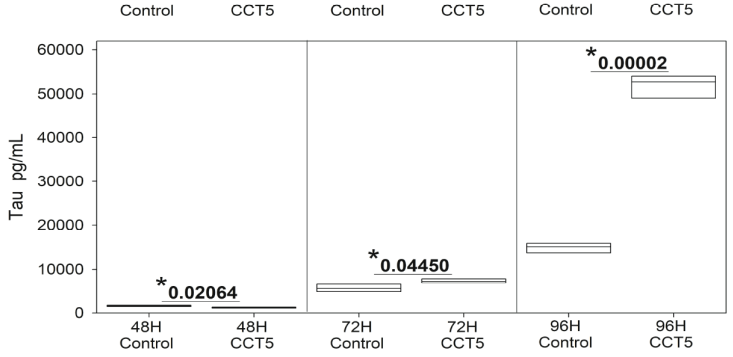

A.
KRONOSII
ILMN_1706246
KRONOSII
LGFAGLVQEISFGTTK
KRONOSII
SQDDEIGDGTTGVVVLAGAL
LEEAEQLLDR
KRONOSII
WVGGPEIELIAIATGGR
CONTROL LOAD
RUSH
ILMN_1706246
RUSH
LGFAGLVQEISFGTTK
RUSH
SQDDEIGDGTTGVVVLAGAL
LEEAEQLLDR
RUSH
WVGGPEIELIAIATGGR
CONTROL MCI LOAD

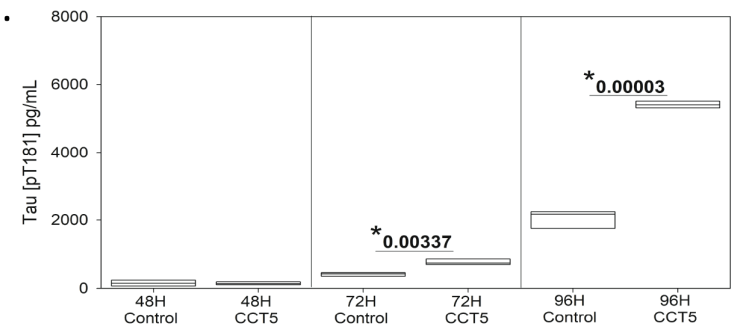

B.
RNA ng/uL
Control
CCT5 O.E.
*0.000


C.
APP
CCT5
MAPT
CCT5
mRNA fold change
*0.001
Control
CCT5 O.E.

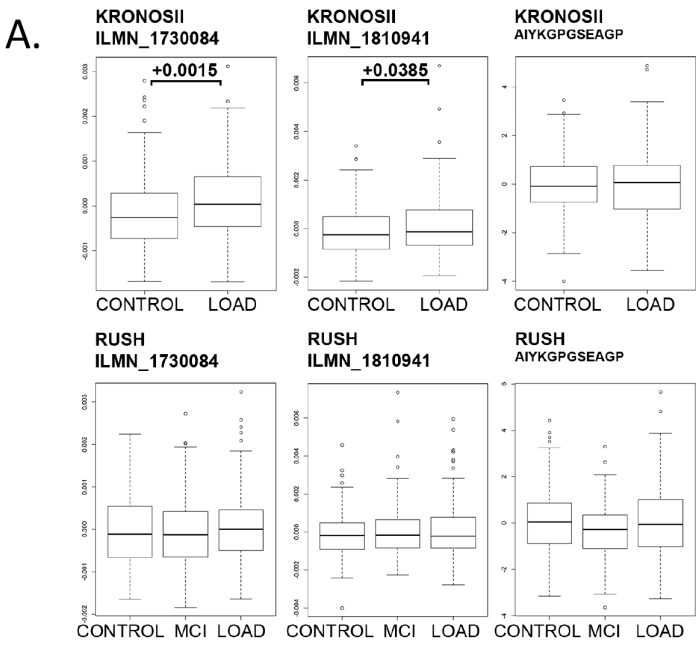

D.
AB40 pg/mL
48H Control
48H CCT5
72H Control
72H CCT5
96H Control
96H CCT5

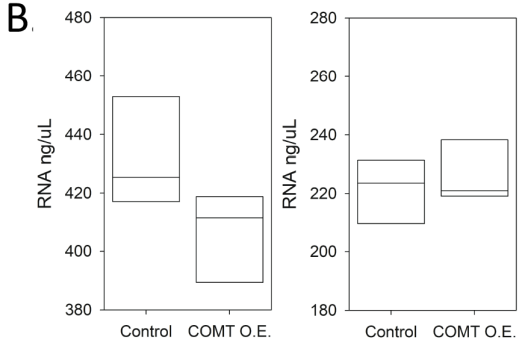

E.
AB42 pg/mL
*0.04177
*0.00353
*0.01011

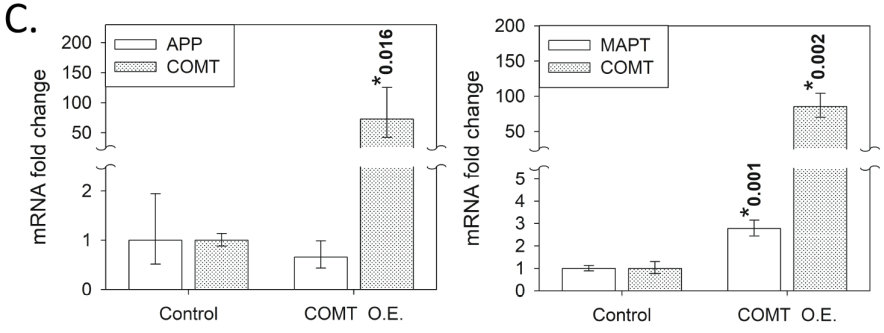

F.
Tau pg/mL
*0.02064
*0.04450
*0.00002

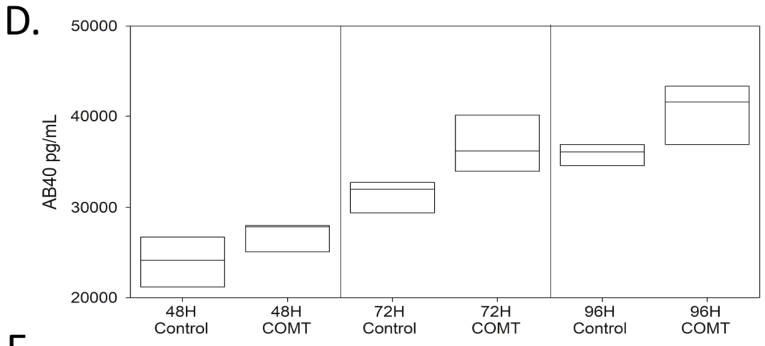

G.
Tau [pT181] pg/mL
*0.00337
*0.00003

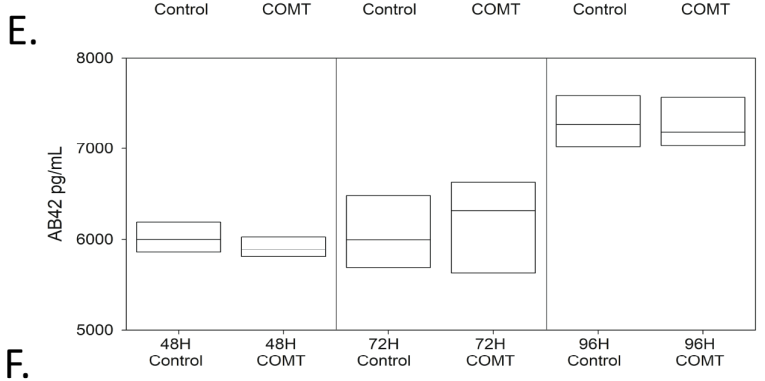
Supplemental Figure 6: ***COMT*** **results.** Shown in the figure are the differentially expressed transcript probe profiles in KRONOSII and RUSH, including late onset Alzheimer's disease, mild cognitive impairment (RUSH) and controls (**A**), along with the levels of total RNA (**B**, 96 hours shown, measured as a surrogate of the level of cell death, left in HEK293sw cells, right in H4-4R0N cells), and mRNA overexpression for target (**C**, 96 hours shown, left in HEK293sw cells, right in H4-4R0N cells). (**D**) shows the AB40 levels, (**E**) the AB42 levels, (**F**) the Total Tau levels and (**G**) the Phospho-Tau levels for three repeat measures of conditioned media at 3 different timepoints. Measurements are taken at 48, 72 and 96 hours post transduction. * = t-Test p value.

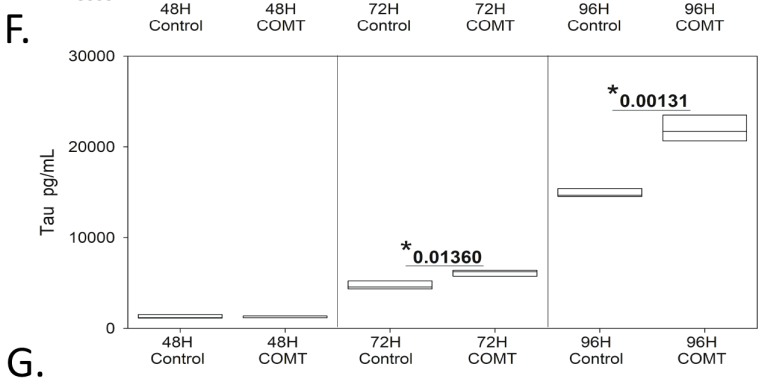

A.
KRONOSII
ILMN_1730084
+0.0015
CONTROL LOAD
KRONOSII
ILMN_1810941
+0.0385
CONTROL LOAD
KRONOSII
AIYKGPGSEAGP
CONTROL LOAD
RUSH
ILMN_1730084
CONTROL MCI LOAD
RUSH
ILMN_1810941
CONTROL MCI LOAD
RUSH
AIYKGPGSEAGP
CONTROL MCI LOAD

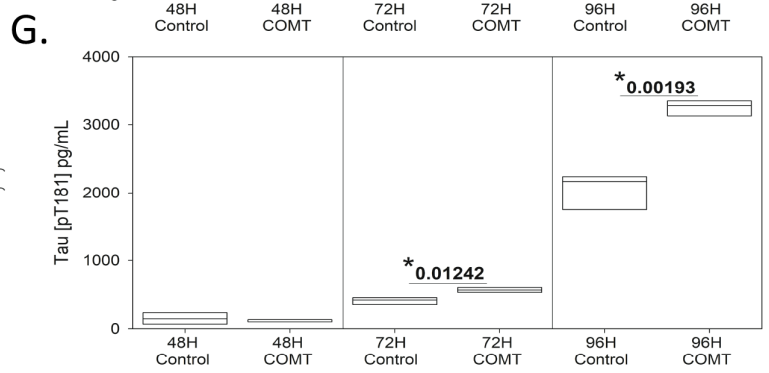

B.
RNA ng/uL
Control COMT O.E.
RNA ng/uL
Control COMT O.E.


C.
APP
COMT
mRNA fold change
*0.016
Control COMT O.E.
MAPT
COMT
mRNA fold change
*0.001
*0.002
Control COMT O.E.

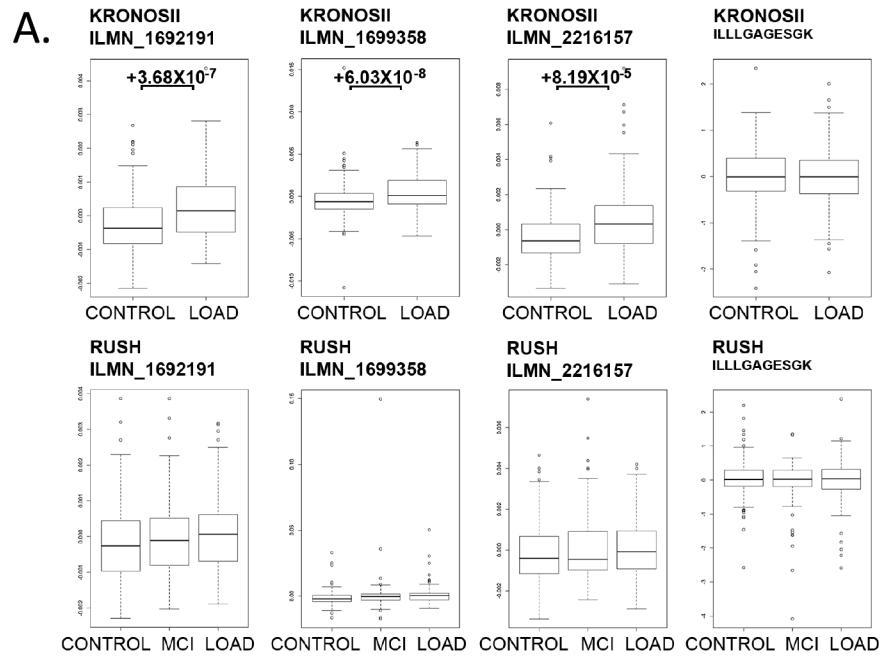

D.
AB40 pg/mL
48H Control
48H COMT
72H Control
72H COMT
96H Control
96H COMT

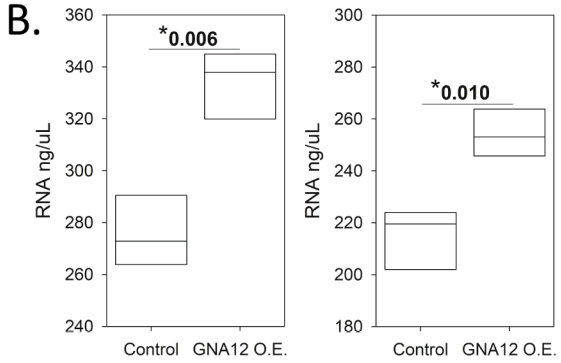

E.
AB42 pg/mL
48H Control
48H COMT
72H Control
72H COMT
96H Control
96H COMT

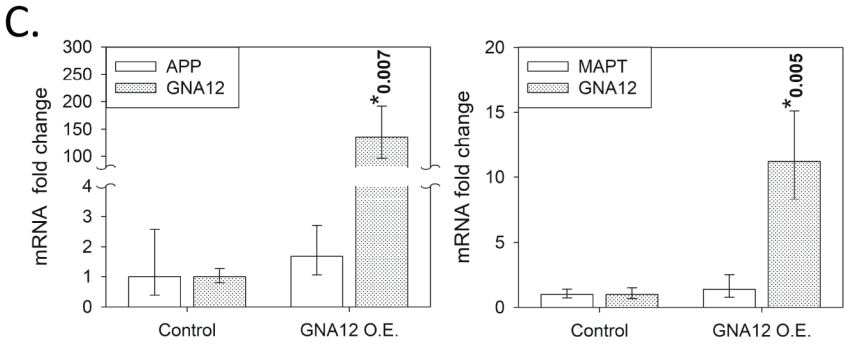

F.
Tau pg/mL
*0.00131
*0.01360
48H Control
48H COMT
72H Control
72H COMT
96H Control
96H COMT

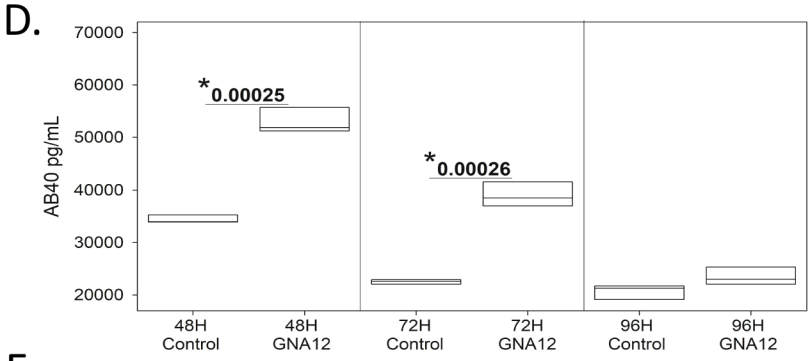

G.
Tau [pT181] pg/mL
*0.00193
*0.01242
48H Control
48H COMT
72H Control
72H COMT
96H Control
96H COMT

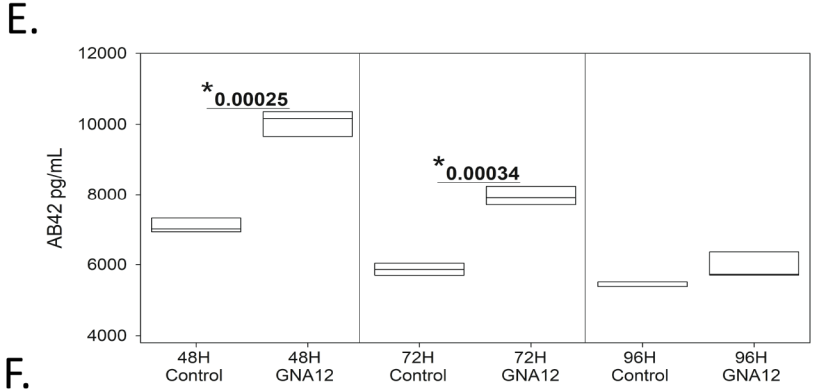
Supplemental Figure 7: ***GNA12*** **results.** Shown in the figure are the differentially expressed transcript probe profiles in KRONOSII and RUSH, including late onset Alzheimer's disease, mild cognitive impairment (RUSH) and controls (**A**), along with the levels of total RNA (**B**, 96 hours shown, measured as a surrogate of the level of cell death, left in HEK293sw cells, right in H4-4R0N cells), and mRNA overexpression for target (**C**, 96 hours shown, left in HEK293sw cells, right in H4-4R0N cells). (**D**) shows the AB40 levels, (**E**) the AB42 levels, (**F**) the Total Tau levels and (**G**) the Phospho-Tau levels for three repeat measures of conditioned media at 3 different timepoints. Measurements are taken at 48, 72 and 96 hours post transduction. * = t-Test p value.

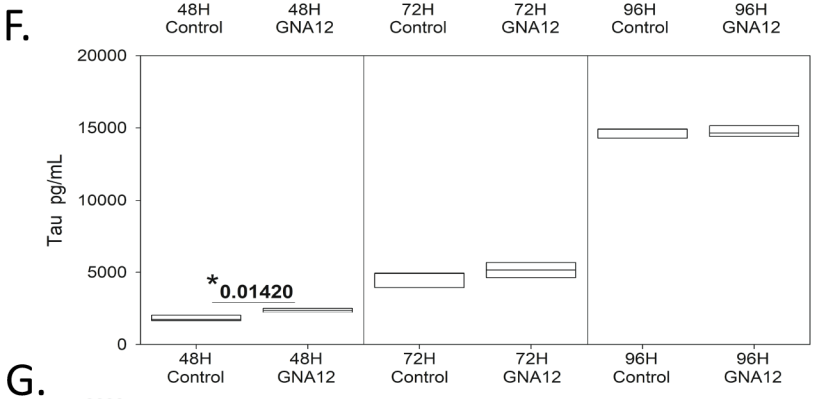

A.
KRONOSII
ILMN_1692191
+3.68X10-7
KRONOSII
ILMN_1699358
+6.03X10-8
KRONOSII
ILMN_2216157
+8.19X10-5
KRONOSII
ILLLGAGESGK
CONTROL LOAD
RUSH
ILMN_1692191
RUSH
ILMN_1699358
RUSH
ILMN_2216157
RUSH
ILLLGAGESGK
CONTROL MCI LOAD

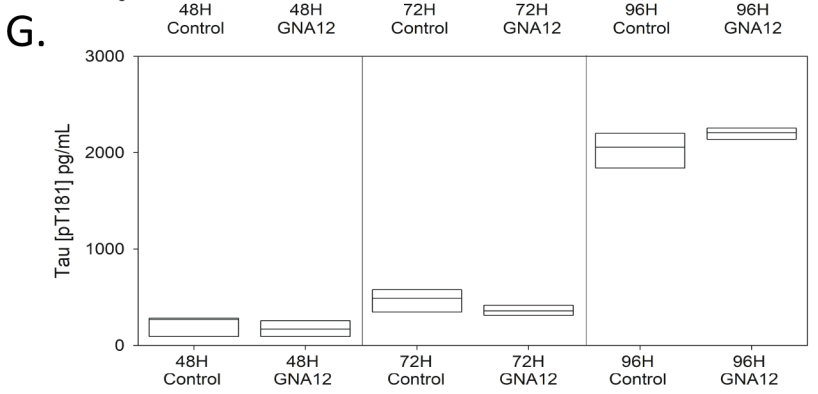

B.
*0.006
RNA ng/uL
Control GNA12 O.E.
*0.010
RNA ng/uL
Control GNA12 O.E.


C.
APP
GNA12
*0.007
mRNA fold change
Control
GNA12 O.E.
MAPT
GNA12
*0.005
mRNA fold change
Control
GNA12 O.E.


D.
*0.00025
*0.00026
AB40 pg/mL
48H Control
48H GNA12
72H Control
72H GNA12
96H Control
96H GNA12


E.
*0.00025
*0.00034
AB42 pg/mL
48H Control
48H GNA12
72H Control
72H GNA12
96H Control
96H GNA12


F.
*0.01420
Tau pg/mL
48H Control
48H GNA12
72H Control
72H GNA12
96H Control
96H GNA12


G.
Tau [pT181] pg/mL
48H Control
48H GNA12
72H Control
72H GNA12
96H Control
96H GNA12

REFERENCES:

Anderson, C.A., Pettersson, F.H., Clarke, G.M., Cardon, L.R., Morris, A.P., and Zondervan, K.T. (2010). Data quality control in genetic case-control association studies. Nat Protoc *5*, 1564-1573.
Azorsa, D.O., Robeson, R.H., Frost, D., Meec hoovet, B., Brautigam, G.R., Dickey, C., Beaudry, C., Basu, G.D., Holz, D.R., Hernandez, J.A.*, et al.* (2010). High-content siRNA screening of the kinome identifies kinases involved in Alzheimer's disease-related tau hyperphosphorylation. BMC Genomics *11*, 25.
Beekly, D.L., Ramos, E.M., van Belle, G., Deitrich, W., Clark, A.D., Jacka, M.E., and Kukull, W.A. (2004). The National Alzheimer's Coordinating Center (NACC) Database: an Alzheimer disease database. Alzheimer Dis Assoc Disord *18*, 270-277.
Bennett, D.A., Schneider, J.A., Buchman, A.S., Barnes, L.L., Boyle, P.A., and Wilson, R.S. (2012a). Overview and findings from the rush Memory and Aging Project. Curr Alzheimer Res *9*, 646-663.
Bennett, D.A., Wilson, R.S., Boyle, P.A., Buchman, A.S., and Schneider, J.A. (2012b). Relation of neuropathology to cognition in persons without cognitive impairment. Ann Neurol *72*, 599-609.
Braak, H., and Braak, E. (1995). Staging of Alzheimer's disease-related neurofibrillary changes. Neurobiol Aging *16*, 271-278; discussion 278-284.
Citron, M., Oltersdorf, T., Haass, C., McConlogue, L., Hung, A.Y., Seubert, P., Vigo-Pelfrey, C., Lieberburg, I., and Selkoe, D.J. (1992). Mutation of the beta-amyloid precursor protein in familial Alzheimer's disease increases beta-protein production. Nature *360*, 672-674.
Corneveaux, J.J., Myers, A.J., Allen, A.N., Pruzin, J.J., Ramirez, M., Engel, A., Nalls, M.A., Chen, K., Lee, W., Chewning, K.*, et al.* (2010). Association of CR1, CLU and PICALM with Alzheimer's disease in a cohort of clinically characterized and neuropathologically verified individuals. Hum Mol Genet *19*, 3295-3301.
Delaneau, O., Marchini, J., and Zagury, J.F. (2011). A linear complexity phasing method for thousands of genomes. Nature methods *9*, 179-181.
Du, P., Kibbe, W.A., and Lin, S.M. (2008). lumi: a pipeline for processing Illumina microarray. Bioinformatics *24*, 1547-1548.
Genomes Project, C., Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A.*, et al.* (2015). A global reference for human genetic variation. Nature *526*, 68-74.
Hoffman, G.E., and Schadt, E.E. (2016). variancePartition: Interpreting drivers of variation in complex gene expression studies. bioRxiv.
Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS genetics *5*, e1000529.
International HapMap, C., Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P.*, et al.* (2007). A second generation human haplotype map of over 3.1 million SNPs. Nature *449*, 851-861.
Jaitly, N., Mayampurath, A., Littlefield, K., Adkins, J.N., Anderson, G.A., and Smith, R.D. (2009). Decon2LS: An open-source software package for automated processing and visualization of high resolution mass spectrometry data. BMC Bioinformatics *10*, 87.
Kall, L., Canterbury, J.D., Weston, J., Noble, W.S., and MacCoss, M.J. (2007). Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nat Methods *4*, 923-925.
Kamburov, A., Pentchev, K., Galicka, H., Wierling, C., Lehrach, H., and Herwig, R. (2011). ConsensusPathDB: toward a more complete picture of cell biology. Nucleic Acids Res *39*, D712-717.
Kamburov, A., Stelzl, U., Lehrach, H., and Herwig, R. (2013). The ConsensusPathDB interaction database: 2013 update. Nucleic Acids Res *41*, D793-800.
Kim, S., Mischerikow, N., Bandeira, N., Navarro, J.D., Wich, L., Mohammed, S., Heck, A.J., and Pevzner, P.A. (2010). The generating function of CID, ETD, and CID/ETD pairs of tandem mass spectra: applications to database search. Mol Cell Proteomics *9*, 2840-2852.

Korn, J.M., Kuruvilla, F.G., McCarroll, S.A., Wysoker, A., Nemesh, J., Cawley, S., Hubbell, E., Veitch, J., Collins, P.J., Darvishi, K.*, et al.* (2008). Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. Nat Genet *40*, 1253-1260.
Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.
Milac, T.I., Randolph, T.W., and Wang, P. (2012). Analyzing LC-MS/MS data by spectral count and ion abundance: two case studies. Stat Interface *5*, 75-87.
Miller, J.A., Cai, C., Langfelder, P., Geschwind, D.H., Kurian, S.M., Salomon, D.R., and Horvath, S. (2011). Strategies for aggregating gene expression data: the collapseRows R function. BMC Bioinformatics *12*, 322.
Millstein, J., and Volfson, D. (2013). Computationally efficient permutation-based confidence interval estimation for tail-area FDR. Front Genet *4*, 179.
Monroe, M.E., Tolic, N., Jaitly, N., Shaw, J.L., Adkins, J.N., and Smith, R.D. (2007). VIPER: an advanced software package to support high-throughput LC-MS peptide identification. Bioinformatics *23*, 2021-2023.
Myers, A.J. (2012). The age of the "ome": genome, transcriptome and proteome data set collection and analysis. Brain Res Bull *88*, 294-301.
Myers, A.J. (2013). AD gene 3-D: moving past single layer genetic information to map novel loci involved in Alzheimer's disease. J Alzheimers Dis *33 Suppl 1*, S15-22.
Myers, A.J. (2014). The Genetics Of Gene Expression: Multiple Layers and Multiple Players. In The OMICs: Applications in Neuroscience
, G. Coppola, ed. (New York, NY: Oxford University Press), pp. 132-152.
Myers, A.J., Gibbs, J.R., Webster, J.A., Rohrer, K., Zhao, A., Marlowe, L., Kaleem, M., Leung, D., Bryden, L., Nath, P.*, et al.* (2007). A survey of genetic human cortical gene expression. Nat Genet *39*, 1494-1499.
Petyuk, V.A., Qian, W.J., Chin, M.H., Wang, H., Livesay, E.A., Monroe, M.E., Adkins, J.N., Jaitly, N., Anderson, D.J., Camp, D.G., 2nd*, et al.* (2007). Spatial mapping of protein abundances in the mouse brain by voxelation integrated with high-throughput liquid chromatography-mass spectrometry. Genome Res *17*, 328-336.
Piehowski, P.D., Petyuk, V.A., Orton, D.J., Xie, F., Moore, R.J., Ramirez-Restrepo, M., Engel, A., Lieberman, A.P., Albin, R.L., Camp, D.G.*, et al.* (2013). Sources of technical variability in quantitative LC-MS proteomics: human brain tissue sample analysis. J Proteome Res *12*, 2128-2137.
Polpitiya, A.D., Qian, W.J., Jaitly, N., Petyuk, V.A., Adkins, J.N., Camp, D.G., 2nd, Anderson, G.A., and Smith, R.D. (2008). DAnTE: a statistical tool for quantitative analysis of -omics data. Bioinformatics *24*, 1556-1558.
Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet *38*, 904-909.
Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J.*, et al.* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet *81*, 559-575.
Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res *43*, e47.
Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J.*, et al.* (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317-330.
Shabalin, A.A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics *28*, 1353-1358.

Wang, Y., Yang, F., Gritsenko, M.A., Wang, Y., Clauss, T., Liu, T., Shen, Y., Monroe, M.E., Lopez-Ferrer, D., Reno, T.*, et al.* (2011). Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. Proteomics *11*, 2019-2026.
Webster, J.A., Gibbs, J.R., Clarke, J., Ray, M., Zhang, W., Holmans, P., Rohrer, K., Zhao, A., Marlowe, L., Kaleem, M.*, et al.* (2009). Genetic control of human brain transcript expression in Alzheimer disease. Am J Hum Genet *84*, 445-458.
Zhang, B., Chambers, M.C., and Tabb, D.L. (2007). Proteomic parsimony through bipartite graph analysis improves accuracy and transparency. J Proteome Res *6*, 3549-3557.
Zhang, B.Z., J. (2013). Identification of Key Causal Regulators in Gene Networks. Paper presented at: Proceedings of the International MultiConference of Engineers and Computer Scientists 2013 (London: International Association of Engineers).
Zimmer, J.S., Monroe, M.E., Qian, W.J., and Smith, R.D. (2006). Advances in proteomics data analysis and display using an accurate mass and time tag approach. Mass Spectrom Rev *25*, 450-482.