Leveraging selective hippocampal vulnerability among Alzheimer's disease subtypes reveals a novel tau binding partner SERPINA5

Angela M. Crist¹, Kelly M. Hinkle¹, Xue Wang², Christina M. Moloney¹, Billie J. Matchett¹, Sydney A. Labuzan¹, Isabelle Frankenhauser^{1,3}, Nkem O. Azu¹, Amanda M. Liesinger¹, Elizabeth R. Lesser², Daniel J. Serie², Zachary S. Quicksall², Tulsi A. Patel Ph.D¹, Troy P. Carnwath¹, Michael DeTure¹, Xiaojia Tang⁴, Ronald C. Petersen⁵, Ranjan Duara⁶, Neill R. Graff-Radford⁷, Mariet Allen¹, Minerva M. Carrasquillo¹, Hu Li⁸, Owen A. Ross¹, Nilufer Ertekin-Taner^{1,7}, Dennis W. Dickson¹, Yan W. Asmann², Rickey E. Carter², and Melissa E. Murray^{1*}

¹ Department of Neuroscience, Mayo Clinic, Jacksonville, FL, USA

² Department of Health Sciences Research, Mayo Clinic, Jacksonville, FL, USA

³ Paracelsus Medical Private University, Salzburg, Austria

⁴ Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA

⁵ Department of Neurology, Mayo Clinic, Rochester, MN, USA

⁶ Wien Center for Alzheimer's Disease and Memory Disorders, Mount Sinai Medical Center, Miami Beach, FL, USA

⁷ Department of Neurology, Mayo Clinic, Jacksonville, FL, USA

⁸ Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, MN, USA

Keywords:

Alzheimer's disease; Neuropathology; Hippocampus; Gene Expression; Extreme phenotype; Hippocampal sparing; Limbic predominant

Summary

Selective vulnerability is a central concept to the myriad of devastating neurodegenerative disorders. Although hippocampus and cortex are selectively vulnerable in Alzheimer's disease (AD), the degree of involvement lies along a spectrum that we previously defined as AD subtypes revealing distinct clinical correlates. To operationalize heterogeneity of disease spectrum, we classified corticolimbic patterns of neurofibrillary tangles to capture extreme and representative phenotypes. We combined bulk RNA sequencing with digital pathology to examine hippocampal vulnerability in AD. Using a multidisciplinary approach, we uncovered disease-relevant hippocampal gene expression changes. Biological relevance was prioritized using machine learning and several levels of human validation. This resulted in five genes highly predictive of neuropathologically diagnosed AD: *SERPINA5*, *RYBP*, *SLC38A2*, *FEM1B*, and *PYDC1*. Deeper investigation revealed *SERPINA5* to be a novel tau binding partner that may represent a "tipping point" in the dynamic maturity of neurofibrillary tangles. Our study highlights the importance of embracing heterogeneity of the human brain to yield promising gene candidates as exemplified by *SERPINA5*.

Introduction

Alzheimer's disease (AD) is a multi-factorial neurodegenerative disorder characterized under the microscope by the abnormal accumulation of two misfolded proteins, tau and amyloid- β ($A\beta$)¹. While both are hallmarks of the disease, they accumulate in very different fashions. $A\beta$ accumulation is the result of enzymatic cleavage of the amyloid precursor protein^{2,3}, whereas abnormal tau accumulation occurs following post-translational modification that disrupts its utility in microtubule stabilization^{4,5}. $A\beta$ plaque deposits accumulate in the extracellular space of isocortical convexities before descending into the limbic system, diencephalon, brainstem, and lastly cerebellum according to Thal amyloid phase⁶. In contrast, abnormally folded tau accumulates intracellularly as neurofibrillary tangles in subcortical nuclei and entorhinal cortex before ascending to limbic structures, association cortices, and lastly primary cortices according to Braak tangle stage⁷. Thus, two partners in the devastation of the AD brain with divergent topographic origins intersect in the limbic system, and more specifically the hippocampus, before continuing on their unique neuroanatomic journey.

Studies demonstrate A β plaque pathology as predictable and rarely deviating from the well-established Thal amyloid phase^{8,9}. However, we identified three AD subtypes using an objective mathematical algorithm to assess corticolimbic patterns of neurofibrillary tangle pathology: hippocampal sparing AD, typical AD, and limbic predominant AD¹⁰⁻¹³. Typical AD brains were representative of the expected patterns of hippocampal and cortical involvement as outlined by Braak tangle stage. In contrast, we discovered two extremes that exist outside of the Braakian-concept of neurofibrillary tangle patterns. Hippocampal sparing AD cases demonstrate selective resilience of the hippocampus relative to severely involved association cortices, whereas limbic predominant AD cases demonstrate inundation of the hippocampus relative to mildly involved association cortices (Fig. 1a). Further examination of their clinical relevance revealed a constellation of clinical and demographic differences that were monotonically-directed among AD subtypes, including striking differences in age at onset, sex, and APOE ϵ 4 status¹⁰⁻¹³. Given the neuropathologic and clinical differences observed among AD subtypes, we sought to test the hypothesis that the objective classification of disease spectrum could be leveraged to uncover transcriptomic changes that underlie selective vulnerability of the hippocampus in AD. Herein, we outline a novel translational neuropathology approach for gene prioritization that, coupled with machine learning, can be used to identify biologically relevant gene expression changes throughout a disease spectrum.

Fig. 1 | Case selection criteria and neuropathologic subtyping of AD. **a**, Flowchart describing selection criteria of AD cases to obtain frozen hippocampal tissue used to perform RNA-Seq and NanoString studies. AD was subtyped using a validated algorithm based on topographic distribution of hippocampal and cortical neurofibrillary tangles¹³. Brain cartoons illustrate corticolimbic patterns of neurofibrillary tangle burden throughout the brain in each AD subtype, where red indicates a high tangle burden and yellow/tan indicates a low tangle burden. **b**, Overview of sample size for RNA-Seq and NanoString cohorts organized by AD subtype and control classification. Note: Cohort expansion for NanoString included all RNA-Seq cohort samples. Acronyms: AD=Alzheimer's disease. DV300=distribution value over 300

Results

Bulk transcriptional profiling of hippocampal vulnerability in AD

Postmortem human brain samples of neuropathologically diagnosed AD cases lacking co-existing neuropathologies and nondemented controls were derived from the Mayo Clinic brain bank (Fig. 1a, Extended Data Fig. 1). We obtained bulk transcriptomic data from the hippocampi of 40 AD cases and 15 controls (Fig. 1b) before employing a five step multi-disciplinary approach toward gene prioritization with the overall goal of selecting disease-relevant, protein-coding genes (Fig. 2). **Our first step** was to select genes from the literature associated with clinical expression of AD. We selected 25 genes located at loci identified in genome-wide association studies (GWAS) of late-onset AD¹⁴⁻¹⁹, 3 genes associated with early-onset AD^{20,21}, and 3 additional genes identified to modify the phenotype of AD²²⁻²⁴ (Fig. 2 [Step 1]). **Our second step** was to utilize our prior

Fig. 2 | Workflow depicting novel multidisciplinary method to identify genes involved in hippocampal vulnerability in AD. Controls and typical AD cases were grouped into the representative phenotype (green-black), while hippocampal sparing AD and limbic predominant AD were grouped into the extreme phenotype (red-blue). Frozen human hippocampal tissue was first dissected, followed by RNA extraction, purification, and lastly prepped for RNA-Seq. A multi-step approach was implemented for gene prioritization, which began with genes nominated from the AD literature (**Step 1**). Next, we examined differentially expressed genes derived from the representative phenotype and extreme phenotype (**Step 2a**) before further bioinformatic prioritization (**Step 2b**). The differentially expressed genes identified in Step 2a were further prioritized based upon process network interactions (**Step 3**). We then combined the 339 unique genes from **Steps 1-3** to apply a novel translational neuropathology approach (**Step 4**) whereby protein-coding genes needed to exhibit monotonic directionality and associate with local neuropathologic markers (tau and amyloid- β) or global measures of AD pathology (Braak stage and Thal phase). Once genes were prioritized, biological significance was investigated first by expanding the RNA-Seq cohort to a total of n=182 in NanoString analyses. Machine learning was applied to NanoString data using the random forest algorithm (**Step 5**). Random forest was applied to the phenotypic groups, in addition to controls versus all AD subtypes (depicted by green vs white bar) to identify the top 5 genes predictive of hippocampal vulnerability in AD as a whole. This enabled us to utilize objective classification of disease spectrum to uncover transcriptomic changes that underlie selective vulnerability of the hippocampus in AD. **Note:** Numbers in boxes represent total gene number from each part of workflow. Steps 1-3 culminated in 339 unique genes, of which 19 overlapped in one or more of the steps. *In Step 4, one gene appeared in both groups. Acronyms: AD=Alzheimer's disease, DE=differential expression, HpSp=Hippocampal sparing, Limbic=limbic predominant, RNA-Seq=RNA sequencing.

knowledge of disease spectrum classified by the AD subtype algorithm¹³ (**Fig. 2 [Step 2]**). Therefore, we divided our RNA sequencing (RNA-Seq) cohort into two groups: representative phenotype (typical AD compared to control) and extreme phenotype (limbic predominant AD compared to hippocampal sparing AD) (**Fig. 2**). Given the constellation of clinicopathologic features found to differ among AD subtypes, we did not adjust RNA-Seq data by age, sex, or *APOE* $\epsilon 4$ status. Differential expression analysis of the representative phenotype identified 1,407 genes, among 19,720 genes that were expressed (**Fig. 2 [Step 2a]**). Of the 1,407 differentially expressed genes in the representative phenotype, 141 genes met bioinformatic prioritization criteria and remained for further analyses of the representative phenotype (**Fig. 2 [Step 2b]**). Differential expression analysis of the extreme phenotype identified 349 genes, among 19,747 genes that were expressed (**Fig. 2 [Step 2a]**). Following bioinformatic prioritization, 73 genes remained for further analyses of the extreme phenotype (**Fig. 2 [Step 2b]**). Of the 186 unique genes prioritized in **Step 2b**, only 28 (15%) overlapped between the representative phenotype and extreme phenotype.

In both the representative phenotype and extreme phenotype, a majority of the genes were downregulated (**Fig. 3a-b**). There were 106/141 (75%) downregulated genes identified in the representative phenotype and 60/73 (82%) downregulated genes identified in the extreme phenotype. No upregulated genes were observed to overlap (**Fig. 3b**). These results suggest that within the spectrum of a disease, the pattern of upregulation or downregulation observed between disease versus controls may be recapitulated in the extremes of the disease state. In the current study, the degree of $\log_2$ fold change ($\log_2FC$) observed in the downregulated genes was nearly twice that observed in the upregulated genes (**Fig. 3c**).

Differentially expressed genes in isolation may not be representative of their own importance to the phenotype, but they could instead reflect the indirect relationship with a gene network²⁵. Thus, our third step was to identify differentially expressed genes from Step 2a that were enriched in pre-built process networks using a curated database created by MetaCore™ to model main cellular processes (**Fig. 2 [Step 3]**). All genes from the top process networks, that were selected for the representative phenotype and separately for the extreme phenotype, were then included as candidate genes. The representative phenotype was highly enriched for genes associated with cell growth and proliferation, whereas the extreme phenotype was highly enriched for genes associated with inflammation and immune response (**Fig. 3d-e**). A total of 111 genes were selected from the top 10 process networks identified in the representative phenotype and 34 genes were selected from those identified in the extreme phenotype (**Fig. 2 [Step 3]**).

Translational neuropathology approach

Our overall goal was to identify biologically meaningful gene expression changes that relate to AD neuropathology and cognition, therefore we employed a novel translational neuropathology approach. Following steps 1-3, we had narrowed down our gene set to a total of 339 unique genes to be reviewed in our fourth step toward final gene prioritization (**Fig. 2 [Step 4]**). The disease spectrum was separated back into four distinct groups to enforce monotonic directionality of gene expression changes in an expected direction of upregulation (control < hippocampal sparing AD < typical AD < limbic predominant AD) as exemplified by *SIRT1* or downregulation (limbic predominant AD < typical AD < hippocampal sparing AD < control) as exemplified by *PSEN2* (**Fig. 3f-g, Extended Data Fig. 2**). Of the 182 genes that showed group-wise differences in gene expression ($p < 0.05$), 79 were monotonically directed.

Next, we investigated the relationship of the remaining 79 genes with neuropathologic burden of tau and A β across serial hippocampal sections for each AD case and control using digital pathology to quantify disease burden^{10,11,13,26-28}. Tau accumulation inside the neuron does not occur in isolation as a static entity, but instead it accumulates along the lifespan of a neurofibrillary tangle^{29,30}. To quantitatively capture the dynamic range of neurofibrillary tangle maturity, we used an early marker of maturity (CP13, **Fig. 3h**)³¹ and an advanced marker of maturity (Ab39, **Fig. 3i**)³². Data are presented in **Extended Data Fig. 1**. The early marker of tangle maturity in the hippocampus was not found to differ among AD subtypes (AD specific $p = 0.11$); however the advanced marker of tangle maturity was found to differ (AD specific $p = 0.018$). This was not

Fig. 3 | Transcriptomic analysis reveals an abundance of downregulated genes across the disease spectrum of hippocampal vulnerability in AD. a-b, Venn diagram depicting the number of genes that were downregulated (a) or upregulated (b) in the representative phenotype (black) or extreme phenotype (purple). There was no overlap in upregulated genes. c, Waterfall plot of the top ten upregulated or downregulated genes from each phenotype. d, Process networks in the representative phenotype (black) and extreme phenotype (purple) reveal non-overlapping functions. f-g, Examples of genes selected from the literature that exhibit monotonic directionality. *SIRT1* is monotonically upregulated (f) while *PSEN2* is monotonically downregulated (Statistical values can be found in **Extended Data Fig. 2**) (g). h-j, Digital pathology examples of immunohistochemistry and burden analysis markup using (h) CP13, (i) Ab39, and (j) 33.1.1 antibodies in CA1 subsector of the hippocampus of typical AD brains. i-j, Volcano plots showing fold change versus significance in the representative phenotype (i) and the extreme phenotype (j). Volcano plots

surprising given the abundance of pretangle and neuritic pathology recognized by the CP13 antibody³¹; in comparison to the Ab39 antibody which recognizes advanced tangle pathology (i.e., mature tangles and extracellular tangles) that better reflects tau-induced neurodegeneration. A β burden was quantified with a pan-A β marker (33.1.1, **Fig. 3j**) to ensure the capture of all plaque types. We did not observe a difference in A β burden among AD subtypes (AD specific $p=0.94$). To complement these local measures of pathology in the hippocampus, we additionally investigated the relationship of the prioritized genes with Braak tangle stage⁷ and Thal amyloid phase⁶ as global measures of AD pathology (**Extended Data Fig. 3**).

The remaining genes to be evaluated in **Step 4** were highly enriched for protein-coding genes (72/79 [91%]), which we focused on for evaluation of neuropathologic association. Of the 72 protein-coding genes, 44 robustly associated with a measure of tau or A β pathology (**Extended Data Fig. 3**). We did not identify any genes in this final set that associated with the pan-A β marker (33.1.1), but 23/44 (52%) were associated with Thal amyloid phase. A majority of the genes were associated with tau pathology: advanced marker of tangle maturity (Ab39, 35/44 [80%]) or Braak tangle stage (13/44 [30%]). Gene selection from **Steps 1-4** are visually represented using volcano plots of log₂FC versus significance for differential gene expression in the hippocampus (**Fig. 3k-l**).

Machine learning to identify associations with AD neuropathology

To examine biological significance of the 44 prioritized genes, we expanded our gene expression studies from 55 hippocampi evaluated using RNA-Seq (**Extended Data Fig. 1**) to 190 hippocampi to be evaluated using the NanoString nCounter platform (**Extended Data Fig. 4**). The NanoString platform provided an affordable approach to multiplexing samples by allowing the user to design a custom codeset³³. Following log₂ transformation of normalized NanoString gene expression, we identified seven AD cases that would require a 2.5 fold adjustment with respect to housekeeping gene correction (**Extended Data Fig. 5**). An additional case was excluded due to low gene expression counts overall. These eight cases were excluded from further analyses, resulting in a final NanoString cohort of 182 AD cases and controls (**Fig. 1b**). As expected, gene expression levels derived from RNA-Seq and NanoString platforms showed a high degree of correlation (**Extended Data Fig. 6**).

Machine learning was then employed in our fifth step to assess the top genes predictive of hippocampal vulnerability in AD (**Fig. 2 [Step 5]**). Our first two in silico experiments examined the top genes in the representative phenotype and in the extreme phenotype. Specifically, we employed a classification and regression method known as random forest to identify the top genes predictive for the representative phenotype and which were predictive for the extreme phenotype (see example in **Extended Data Fig. 7**). Random forest uses an ensemble of simple decision trees to classify data³⁴. To facilitate prioritization and feasibility of evaluating top genes, mean minimum depth was used to rank the genes and a cutoff at the top 5 was applied (**Extended Results 1**). Out of the top 5 genes predictive when the representative phenotype and the extreme phenotype were separately analyzed, only *SERPINA5* overlapped between phenotypes (**Extended Data Fig. 8**).

The aforementioned in silico experiments provide insight into the molecular differences among AD subtypes. To maximize applicability to selective hippocampal vulnerability in AD as a whole, we combined all AD subtypes together and generated a predictive model for subsequent analysis (**Fig. 4**). Through inspection of the variable importance summaries (**Fig. 4a-b**), the following genes were considered for best discrimination of AD from control: *SERPINA5*, *RYBP*, *SLC38A2*, *FEM1B*, and *PYDC1*. A logistic regression model was fit to the data to estimate the ability for our top 5 genes, in concert with age and sex, to discriminate between neuropathologically diagnosed AD cases and controls. The model provided strong discrimination with an area under the curve of 92.6% (95% confidence interval [CI]: 87.9% - 97.2%; **Fig. 4c**). To represent the model visually (**Fig. 4d**), a nomogram was created^{35,36}. The utility of the nomogram enables each individual's

demographic and gene expression data to be converted to a point-value (Fig. 4d, top), which is summarized into a total number of points that directly relates to a predicted value (Fig. 4d, bottom). The higher the predicted value, the more likely the data is representative of a neuropathologically diagnosed AD case. The nomogram visually provides a quantitative contribution of key factors like sex and age compared to gene expression values from the hippocampus (Extended Data Fig. 9). Examples of the utility of the nomogram in a nondemented control and an AD case are provided in Extended Data Fig. 10. The summarized predictive value was visually distinct between Braak tangle stage 0-III versus IV-VI and Thal amyloid phase 0-2 versus 3-5 when plotted (Fig. 4 e-f). On a clinical level, we observed a non-linear relationship between the last Mini Mental State Examination (MMSE³⁷) score and nomogram predictive value, with higher MMSE scores (i.e. better cognition) corresponding to lower predicted value (Fig. 4g).

To examine the relationship between gene expression changes in each of the top 5 genes with local measures of neuropathology in the hippocampus, we performed a series of multiple linear regression models. Gene expression levels along with selected covariates (age, sex, and APOE ε4) were plotted against the following measures of neuropathologic burden: early marker of tangle

Fig. 4 | Implementation of machine learning to prioritize disease-relevant genes that associate with AD pathology and cognitive decline. **a**, Minimum depth plot generated using random forest algorithm³⁴ identified the top 5 genes predictive of neuropathologically diagnosed AD based on summary of gene position within AD versus control generated trees: *SERPINA5*, *RYBP*, *SLC38A2*, *FEM1B*, *PYDC1*. **b**, Multiway importance plot depicting all genes used in random forest analysis. The top 5 are highlighted in coral, demonstrating the importance of low minimum depth (x-axis) and high root frequency (y-axis). **c**, Gene expression values were built into a logistic regression model along with age and sex to estimate the ability of the top 5 genes to discriminate between neuropathologically diagnosed AD and controls. Receiver operating characteristic reveals high discrimination with an area under the curve of 92.6%. **d**, Nomogram showing the predictive value of neuropathologically diagnosed AD for each of the top 5 genes along with age and sex. The final predictive value is determined by adding up points from each variable. See Extended Data Fig. 10 for example of utilization. **e**, Predictive value of each case and control compared to Braak tangle stage demonstrates higher predictive values corresponding to higher stages. **f**, Predictive value of each case and control compared to Thal amyloid phase demonstrates higher predictive values corresponding to higher phases. **g**, Predictive value of each case and control with corresponding last MMSE score shows higher cognitive function corresponds with lower predictive value from nomogram. **h**, Radar plot depicting the overlay of each of the five genes regressed on demographics and digital pathology measures of tau, Aβ, and cell specific markers. Radar plot axes correspond to coefficient of partial determination for each of these regressors in the individual models. Additional statistics for radar plot can be found in Extended Data Fig. 11. Acronym: AD=Alzheimer's disease, AUC=area under the curve,

maturity (CP13), advanced marker of tangle maturity (Ab39), pan-A β pathology (33.1.1), and cell-type specific markers to adjust for contribution of microgliosis (CD68), astrogliosis (GFAP) and endothelial burden (CD34). A neuronal marker was not included in the models to avoid regressing out genes relevant to tangle-bearing neurons. A radar plot visually displays the unique relationship between each of the prioritized genes and their degree of association with markers of neuropathologic burden (**Fig. 4h**). Each of our top 5 genes correlated with different aspects of AD pathology (results summarized in **Extended Data Fig. 11**).

Serine protease inhibitor SERPINA5 identified as tau interactor

To validate the utility of our translational neuropathology approach to identify disease-relevant gene expression changes, we assessed the biological relevance of our top gene *SERPINA5*. To put into context of the above-mentioned steps, *SERPINA5* was first prioritized in the differential expression analysis of the representative phenotype (Step 2a: $p=0.00012$; Step 2b: false discovery rate[FDR]=0.0062, $\log_2(\text{foldchange}[FC])=1.5$). The translational neuropathology approach (Step 4) further identified *SERPINA5* gene expression levels to be monotonically directed ($p=0.00422$, **Fig. 5a**) and associate robustly with neuropathologic measures (robust association: Braak tangle stage $p<0.0001$ and $\text{FDR}<0.10$; Ab39 burden $p<0.002$ and $\text{FDR}<0.22$, and Thal amyloid phase $p<0.001$ and $\text{FDR}<0.14$, **Extended Data Fig. 3**). Random Forest analyses (Step 5, **Fig. 4a-b**, **Extended Data Fig. 8**) identified *SERPINA5* as the 2nd top predictor in the representative phenotype (mean minimum depth=2.32, $p<0.0001$), 3rd top predictor in the extreme phenotype (mean minimum depth=2.97, $p<0.0001$), and as the top predictor in the overall discrimination of AD from control (mean minimum depth=2.18, $p<0.0001$). Examination of three validation datasets from AMP-AD^{25,38,39} (**Extended Results 2**) revealed *SERPINA5* to be significantly upregulated in AD compared to controls in the temporal cortex (Mayo-TCX: $\text{FDR}=0.000040$,

Fig. 5 | *SERPINA5* gene expression changes differ among AD subtypes and track with advanced neurofibrillary tangle pathology. a-b, *SERPINA5* expression increases in a monotonic fashion among AD subtypes in the RNA-Seq cohort (a), which was replicated in the larger NanoString cohort (b). c, *SERPINA5* gene expression from NanoString cohort associates with % burden of advanced tangle maturity marker Ab39. d, Immunohistochemical investigation of *SERPINA5* in controls and among AD subtypes reveals a monotonic increase in hippocampal subsectors and a monotonic decrease in the association cortices. Pair-wise comparisons for graphs can be found in **Extended Data Fig. 2**. Control cases=green, Hippocampal sparing AD=red, Typical AD=grey, Limbic predominant AD=blue. Scale bar represents 100 μm . Acronyms: AD=Alzheimer's disease, A β =Amyloid- β , HpSp=hippocampal sparing, Limbic=limbic predominant, RNA-Seq=RNA sequencing.

log2FC=2.1), superior temporal gyrus (Mount Sinai brain bank [MSBB]-Brodmann area [BM]22: FDR=0.095, log2FC=1.0), and parahippocampal gyrus (MSBB-BM36: FDR=0.0016, log2FC=2.0).

SERPINA5, also commonly known as protein C inhibitor, is a glycoprotein that inhibits serine proteases. SERPINA5 interacts with activated protein C, thrombin, and factor Xa to play a major role in hemostasis/blood coagulation and thrombosis⁴⁰⁻⁴². Our studies revealed that in addition to its conventional role in blood coagulation, it may also play a functional role in AD pathogenesis. In both our RNA-Seq and NanoString datasets of hippocampal gene expression, we saw an increase in *SERPINA5* levels among all AD subtypes (**Fig. 5a-b**). We did not observe differences in *SERPINA5* gene expression levels when males and females were stratified within controls ($p=0.92$), hippocampal sparing AD ($p=0.61$), typical AD ($p=0.83$), or limbic predominant AD ($p=0.87$) (**Extended Data Fig. 12**). Linear regression analysis revealed *SERPINA5* expression associated with the advanced marker of tangle maturity ($p=0.010$) and approached significance with microgliosis ($p=0.075$) (**Fig. 4h, Fig. 5c, Extended Data Fig. 11**). To assess SERPINA5 localization within the brain, we utilized immunofluorescence microscopy to examine colocalization. We examined SERPINA5 immunostaining patterns across several brain cell types including astrocytes (GFAP), neurons (MAP2), microglia (resting [IBA1] and activated [CD68]), endothelial cells (CD34), and oligodendrocytes (OLIG2) (**Extended Data Fig. 13**). SERPINA5 was observed in MAP2-positive neurons in a typical AD hippocampus, whereas SERPINA5 was absent in control brain (**Extended Data Fig. 13b**). Interestingly, gene expression levels of *SERPINA5* and the neuronal marker *ENO2* inversely associate in both controls ($R=-0.55$, $p=0.001$) and AD cases ($R=-0.35$, $p<0.001$) (**Extended Data Fig. 14**).

Given SERPINA5's association with neurofibrillary tangle pathology (**Fig. 4h**) and its predominantly neuronal expression (**Extended Data Fig. 13b**), we hypothesized that SERPINA5 would follow the corticolimbic patterns inherent to each of the AD subtypes. We employed digital pathology to quantify SERPINA5 immunohistochemical burden in hippocampal subsectors (CA1 and subiculum) and association cortices (temporal, parietal and frontal regions) vulnerable to AD neuropathologic change^{7,13} (**Fig. 5d**). Controls lacked or had minimal SERPINA5 immunopositive lesions (all medians $<0.020\%$ burden) and differed in all regions from AD cases ($p\leq 0.002$) (**Extended Data Fig. 15**). As predicted by hippocampal gene expression studies, we observed increasing immunohistochemical burden of SERPINA5 among AD subtypes in the hippocampus (AD-specific $p=0.011$). Based upon our observations of tangle patterns in the cortex of AD subtypes, we hypothesized we would observe an inverse monotonic directionality in the cortex. Thus, we extended our findings to investigate immunohistochemical burden of SERPINA5 across superior temporal, inferior parietal, and mid-frontal cortices (**Fig. 5d**). The average cortical burden of SERPINA5 demonstrated an overall decreasing monotonic directionality from hippocampal sparing AD > typical AD > limbic predominant AD (AD-specific $p<0.001$).

Next, we sought to investigate SERPINA5's relationship with neurofibrillary tangle maturity, as more advanced tangle pathology closely associates with neuronal death^{9,43}. Neurofibrillary tangles are not static entities^{29,44}. They develop progressively from early pretangles with punctate tau staining and perinuclear accumulation to advanced mature tangles with fibrillar structure that form into ghost tangles with diffuse fibrils lacking a nucleus (**Fig. 6a**)²⁹. To confirm the association between advanced tangle pathology (Ab39 burden) and *SERPINA5* gene expression from regression modeling, we performed immunofluorescence microscopy. In pretangles, SERPINA5, like tau, was punctate in nature and found localized to the cytoplasm and perinuclear space (**Fig. 6b**). A marked increase in SERPINA5 immunostaining was observed in mature tangles aligning with more fibrillar aspects of tau (**Fig. 6c**). Of note, areas relatively devoid of tau pathology in mature tangles appeared to be occupied by SERPINA5 accumulation. SERPINA5 immunopositive ghost tangles ranged from robustly to lightly stained in their burnt-out form as remnants of tau fibrils (**Fig. 6d**). SERPINA5 staining was predominantly found in the neuronal soma with only sparse labeling of extrasomal neuritic pathology. SERPINA5 immunopositive neurites could be found proximal to an affected neuron, which represented a subset of tau-positive neurites (**Fig. 6b-d**). Based upon our findings from immunofluorescence microscopy, we hypothesized that SERPINA5-immunopositive tangles would be in greater abundance in mature tangles and ghost tangles. The number and type of tangle were manually counted in the CA1 subsector of the hippocampus in 60 independent AD cases (**Fig. 6e**) on digitized brightfield images of SERPINA5 immunostaining. The proportion of SERPINA5-positive tangles dramatically increased in frequency from

pretangles (4%) to mature tangles (52%) and ghost tangles (42%) ($p < 0.001$). In addition to immunopositive tangles, SERPINA5 pathology was found interspersed between and colocalized with tau-positive and thioflavin-S-positive neurites in neuritic plaques (**Fig. 6f-g**).

Immunofluorescent colocalization of SERPINA5-positive/tau-positive neurofibrillary tangle pathology suggested that SERPINA5 may interact with tau. To investigate this further, we performed a co-immunoprecipitation (co-IP) experiment to determine if we detected a SERPINA5-tau protein-protein

Fig. 6 | SERPINA5 is a tau binding partner expressed in neurofibrillary tangles and neuritic plaques. **a**, Cartoon depiction of the lifespan of neurofibrillary tangle maturity ranging from early to advanced forms: pretangle with punctate tau staining, mature tangle with fibrillar tau, and ghost tangle in which only the remnants of the mature tangle remain once the neuron has died. **b**, SERPINA5 accumulates alongside, as well as colocalizes with punctate tau (Tau E178, red) in pretangles. **c**, SERPINA5 accumulates in areas fairly devoid of tau and colocalizes along the fibrillar aspect of tau (Tau E178, red) in mature tangles. **d**, SERPINA5 accumulates in the extracellular space and to a lesser extent colocalizes with tau (Tau pS396, green) in ghost tangles. SERPINA5 is also observed in a subset of dystrophic neurites (arrows, b-d). **e**, Manual quantification of SERPINA5-positive neurofibrillary tangle counts in CA1 region of the posterior hippocampus shows a greater frequency of mature tangles and ghost tangles. Pretangles, mature tangles, and ghost tangles were manually quantified. **f-g**, SERPINA5 accumulates independently and colocalizes with tau in neuritic plaques as shown by pan-tau (E1) staining and thioflavin-S. **h**, Immunoblot of pan-tau (E1) using immunoprecipitated SERPINA5 demonstrates a tau-SERPINA5 protein complex in the AD brain, but not in control brain. (Left) Tissue was sampled from frozen hippocampi of a 73 year old male control (Braak=I, Thal=0) and an 86 year old male AD case (Braak=V and Thal=5). (Right) Tissue was sampled from frozen frontal cortices of a 75 year old female control (Braak=I, Thal=0) and a 68 year old male AD case (Braak=VI and Thal=5). Note that uncropped Western blots can be found in **Extended Data Fig. 19-20**. Acronyms: AD=Alzheimer's disease. Co-IP=Co-immunoprecipitation. Thio-S=Thioflavin-S.

interaction. Bulk hippocampal tissue was taken from an AD case and control and used to immunoprecipitate SERPINA5, which was followed by immunoblotting using the pan-tau marker, E1⁴⁵. We observed a strong tau signal in the total homogenate (input) of both AD cases and controls, but only detected a tau band in the SERPINA5 co-IP of the AD case (**Fig. 6h**). This initial study provided supportive evidence that SERPINA5 acts as a tau protein binding partner. To provide stronger evidence to support our hypothesis of a SERPINA5-tau interaction, we expanded our studies to the frontal cortex where larger tissue volumes can be dissected for greater yields. Western blot analysis of co-immunoprecipitation experiment revealed that tau was successfully pulled down by SERPINA5 in all three AD cases, but not in controls, suggesting that SERPINA5-tau interaction is disease-specific (**Fig. 6h**).

Discussion

In this study, we leveraged the Alzheimer's disease spectrum using both phenotypic differences (extreme phenotype vs. representative phenotype)¹⁰⁻¹³ and quantitative measures of severity (digital pathology)^{10,11,26-28} to identify key proteins involved in the molecular pathogenesis underlying selective hippocampal vulnerability. To our knowledge, this study represents the largest RNA-Seq experiment performed using bulk hippocampal tissue to date⁴⁶⁻⁴⁸. Using a novel approach that combined next-generation transcriptome sequencing with quantitative digital pathology and machine learning, we performed several levels of disease-relevant validation in humans that included: 1) RNA-Seq cohort expansion using NanoString technology³³, 2) machine learning³⁴ to predict utility of gene expression levels working in concert to discriminate neuropathologic classification, 3) evaluation of the relationship between gene expression levels with macroscopic scales of neuropathology (Braak⁷ and Thal⁶) and cognition (MMSE³⁷), 4) digital pathology to assess biological relevance of corticolimbic distribution of SERPINA5, and 5) co-IP to examine interaction between tau and SERPINA5. In all, these tools provided invaluable insight into the complexity of hippocampal vulnerability both within the disease spectrum and in AD as a whole, highlighting the role of *SERPINA5*, *RYBP*, *SLC38A2*, *FEM1B*, and *PYDC1*.

Upon reviewing the published literature, our top 5 proteins were found to have a multitude of diverse functions within the living system. In addition to SERPINA5's canonical serine protease inhibition function, it exhibits broad protease activity and interacts with phospholipids⁴⁹⁻⁵³, glycosaminoglycans⁵⁴, proteins⁵⁵⁻⁵⁸, and cysteine proteases⁵⁹. These properties make SERPINA5 extremely unique among serine protease inhibitors. RYBP is a multifunctional protein that binds several transcriptional factors⁶⁰, mediates histone H2A monoubiquitination⁶¹ and plays roles in development^{62,63}, apoptosis⁶⁴⁻⁶⁶ and cancer⁶⁰. SLC38A2 functions as a sodium-dependent amino acid transporter that mediates both the efflux of neutral α -amino acids across the blood-brain barrier and their uptake into neurons⁶⁷⁻⁶⁹. FEM1B is an ankyrin repeat protein that induces replication stress signaling through its interaction with checkpoint kinase 1⁷⁰, regulates ubiquitin-protein transferase activity⁷¹ and promotes apoptosis⁷². PYDC1 (also known as ASC2/POP1) regulates the innate immune response by suppressing NF κ B transcription factor activity and pro-caspase-1 activation⁷³.

This work underscores the importance of capturing heterogeneity as a phenotype when investigating disease-relevant gene expression changes using bulk RNA-Seq. Our ability to identify the most relevant genes was enhanced by the judicious application of the random forest algorithm³⁴. This application of machine learning has wide-ranging utility outside the context of AD, as exemplified by its use to uncover transcriptional signatures underlying heterogeneity of smooth muscle cells using single-cell RNA-Seq⁷⁴. To further validate the functional relevance of our findings, we performed regression modeling to assess whether the driver of disease-relevant gene expression changes was underlying neuropathology or cellular admixture. Gene expression levels from our top gene, *SERPINA5*, were upregulated in AD and monotonically directed among AD subtypes. This is in line with a previous study that identified upregulation of *SERPINA5* in the hippocampus of late onset AD brains⁴⁶. In addition, *SERPINA5* expression was significantly upregulated in AD in all three AMP-AD validation datasets from temporal cortex and parahippocampal gyrus^{25,38,39}. The current study provides novel insight demonstrating a strong association between SERPINA5 immunopositive neurons and neuropathologic severity of advanced tangle pathology.

We sought to further investigate cellular distribution of SERPINA5 based upon a previous study describing SERPINA5 immunoreactivity in microglia⁷⁵ and our regression model suggestive of a relationship with microgliosis. Co-immunofluorescence confirmed localization of SERPINA5 to neurofibrillary tangle-bearing neurons in AD brains. In contrast to previous findings in incidental Lewy body disease and Parkinson's disease⁷⁵, we did not observe protein expression in non-neural cells. This difference may be due to methodology (immunohistochemical brightfield staining versus co-immunofluorescence) or regional differences (i.e. hippocampus vs. substantia nigra). Neurons can survive the brunt of early neurofibrillary tangle pathology^{76,77}, which may represent an invaluable window of time to arrest the neurodegenerative aspects of neurofibrillary tangle progression. Although, SERPINA5-immunopositive structures were only observed in the context of neurofibrillary tangle pathology, gene expression studies revealed an inverse association between *SERPINA5* and the neuronal marker *ENO2* in hippocampi of both controls and AD. This may suggest that upregulation of *SERPINA5* represents a repair process in neurons. Currently, it is unclear if SERPINA5 upregulation is a consequence of failed adaptation to neuronal dysfunction, or a contributor to neurofibrillary tangle formation. Our co-IP experiments provide supportive evidence that SERPINA5 and tau interact in the hippocampus and cortex of AD brains. This was complemented by our microscopy experiments that suggest SERPINA5 could be a tipping point between pretangles and mature tangles. Future studies will be directed at investigating the role of SERPINA5 in neurons with a focus on identifying whether SERPINA5 acts in a casual or non-causal fashion. This information could aid in identifying a mechanism to prevent or slow neuronal death.

Several limitations in the context of this research should be discussed. At the discovery stage, age and sex were not used to adjust as the AD subtypes exhibit a constellation of demographic and clinicopathologic features. We hypothesize that adjustment of these biological variables may limit identification of important gene expression changes that track with selective hippocampal vulnerability observed in the AD subtypes. As adjustment for age and sex is commonplace in bioinformatic prioritization, future studies lacking distinct clinicopathologic phenotypes should consider adjustment as appropriate to their study. The applied assumption of monotonic directionality was a reflection of the biology of the AD subtypes observed when examining clinicopathologic and demographic differences. It is important to note that the utility of this assumption may have excluded genes relevant to hippocampal vulnerability in AD that did not track with AD subtypes. The monotonic directionality of hippocampal involvement facilitated evaluation of gene expression changes from selective resilience in hippocampal sparing AD to selective vulnerability in limbic predominant AD. Although this range enabled deeper investigation into late-stage AD cases, more work is needed to determine the relevance of these gene expression changes as epiphenomenal versus active players in the pathogenesis of hippocampal vulnerability. Furthermore, the selection of literature-based genes does not include several contemporary genes, which may dramatically reduce the value of including known genes that can be leveraged using our novel approach. To provide a contemporary examination, we applied our translational neuropathology approach from Step 4 to a recently summarized gene set from Neuner et al.⁷⁸ to identify genes that may associate with hippocampal vulnerability in AD. We identified three additional protein-coding genes that exhibited monotonic directionality among AD subtypes and robustly associated with neuropathologic measures: *ADAM10*, *ANKMY2*, *ATP5F1* (**Extended Results 3**). We encourage the research community to investigate these and other genes from our NanoString gene list in future studies. Future studies will strive to validate these genes and others prioritized for NanoString in an ethnographically diverse series to ensure generalizability to all individuals suffering from this devastating disorder.

By embracing heterogeneity of the disease spectrum of AD^{10-13,79} and pursuing our hypotheses in the context of the human brain, we have uncovered a novel binding partner of tau – SERPINA5. Our study highlights the utility of a multidisciplinary, team science approach that should be encouraged in all disease-relevant fields. This approach may be especially useful in progressive disorders that accumulate aberrant proteins (e.g. motor neuron disease⁸⁰, Parkinson's disease⁸¹) and spectrum disorders such as diabetes that range in severity⁸². Through the use of sophisticated neuropathologic techniques, we were able to leverage our understanding of heterogeneity from the perspective of selective vulnerability in AD subtypes¹⁰⁻¹³. Coupled with machine learning, our study highlights the importance of multi-step prioritization efforts to yield promising gene candidates. In this era of big data generation, scientists need advanced methods to prioritize biologically relevant targets from large, heterogeneous human datasets. Here we outline a novel approach that offers several aspects of human-relevant validation and provides intriguing implications into the dynamic aspect of

accumulating neurofibrillary tangle pathology that could be targeted to prevent further devastation of the AD brain.

Acknowledgements

This study was supported by the National Institute on Aging (R01 AG054449, P01 AG003949, P30 AG062677, P50 AG047266, U01 AG006786, and R01 AG034676), the Florida Department of Health, Ed and Ethel Moore Alzheimer's Disease Research Program (6AZ01, 8AZ06, 20A22), and the Alzheimer's Association (AARG-17-533458). We would like to thank Peter Davies for providing the CP13 antibody, Shu-Hui Yen for Ab39 antibody, Leonard Petrucelli and Casey Cook for the E1 antibody, and Pritam Das for the 33.1.1 antibody. We are grateful to Ariston L. Librero, Jo Landino, and Virginia Phillips for histological support; to Monica Castanedes-Casey for immunohistochemical support; and to our brain bank study coordinators Rachel LaPaille-Harwood and Jessica F. Tranovich. The Translational Neuropathology laboratory is especially grateful for the time and dedication of Samantha Davis. This work would not be possible without the generosity of the Gerstner Family Career Development Award, Center of Individualized Medicine at Mayo Clinic, and a kind gift from David and Frances Strawn. We thank the patients and their families for their generous brain donations to help further our knowledge of Alzheimer's disease.

AMP-AD Data Acknowledgements:

This work was supported by National Institute on Aging [U01 AG046139, RF1 AG051504 to N.E.T.]; National Institute of Neurological Disorders and Stroke [R01 NS080820 to N.E.T.]. Study data were provided by the following sources: The Mayo Clinic Alzheimer's Disease Genetic Studies, led by Dr. Nilüfer Ertekin-Taner and Dr. Steven G. Younkin, Mayo Clinic, Jacksonville, FL using samples from the Mayo Clinic Study of Aging, the Mayo Clinic Alzheimer's Disease Research Center, and the Mayo Clinic Brain Bank. Data collection was supported through funding by NIA grants P30 AG062677, R01 AG032990, U01 AG046139, R01 AG018023, U01 AG006576, U01 AG006786, R01 AG025711, R01 AG017216, R01 AG003949, NINDS grant R01 NS080820, CurePSP Foundation, and support from Mayo Foundation. Study data includes samples collected through the Sun Health Research Institute Brain and Body Donation Program of Sun City, Arizona. The Brain and Body Donation Program is supported by the National Institute of Neurological Disorders and Stroke (U24 NS072026 National Brain and Tissue Resource for Parkinson's Disease and Related Disorders), the National Institute on Aging (P30 AG19610 Arizona Alzheimer's Disease Core Center), the Arizona Department of Health Services (contract 211002, Arizona Alzheimer's Research Center), the Arizona Biomedical Research Commission (contracts 4001, 0011, 05-901 and 1001 to the Arizona Parkinson's Disease Consortium) and the Michael J. Fox Foundation for Parkinson's Research. The results published here are in whole or in part based on data obtained from the AD Knowledge Portal (<https://adknowledgeportal.synapse.org/>). These data were generated from postmortem brain tissue collected through the Mount Sinai VA Medical Center Brain Bank and were provided by Dr. Eric Schadt from Mount Sinai School of Medicine.

Author Contributions

M.E.M. designed and supervised the entire study. K.M.H., A.M.L., M.D.T., and M.E.M. dissected frozen brains. K.M.H. performed RNA extraction and quality control. D.W.D. performed neuropathologic evaluation and quantitative thioflavin-S counts. R.C.P., R.D., N.R.G., N.E.T. performed neurologic evaluation. A.M.L. abstracted clinical information. X.W., Y.W.A., X.T., and H.L. performed all bioinformatic analysis. D.S. performed NanoString Analysis. E.R.L. and R.E.C. performed all statistical analysis. K.M.H., N.O.A., B.J.M. evaluated and validated antibodies. I.F., S.A.L. and A.M.L. performed digital pathology. C.M.M. performed SERPINA5-positive tangle counts. A.M.C. designed and performed SERPINA5 experiments and data analysis.

M.A., M.M.C., O.A.R. performed genotyping and provided substantial intellectual contribution. N.E.T provided funding, direction and supervision for the generation of Mayo Clinic AMP-AD data and analysis of all AMP-AD data in this manuscript. X.W., Z.S.Q., T.A.P., T.P.C., M.A. analyzed all AMP-AD data. M.E.M. and A.M.C. wrote the manuscript. All authors edited and reviewed the manuscript.

Author Disclosures

R.C.P. is a consultant for Hoffman-La Roche, Merck, Biogen, Eisai. R.C.P. is on the Data safety and monitoring board for Genentech. N.G.R. takes part in multi-center studies funded by Lilly, Biogen and Abbvie. M.E.M is a consultant for AVID Radiopharmaceuticals.

Data Availability Statement

All requests for raw and analyzed data and related materials, excluding programming code, will be reviewed by the Mayo Clinic legal department and Mayo Clinic Ventures to verify whether the request is subject to any intellectual property or confidentiality obligations. Requests for patient-related data not included in the paper will not be considered. Any data and materials that can be shared will be released via a Material Transfer Agreement. The Accelerating Medicines Partnership (AMP-AD) data in this manuscript are available via the AD Knowledge Portal (<https://adknowledgeportal.synapse.org>).

Code Availability Statement

Programming code related to the data preprocessing will be made available under the GNU General Public License version 3 upon request to Z.I.A. (attia.itzhak@mayo.edu)

List of Abbreviations

AD: Alzheimer's Disease
 ANOVA: Analysis of variance
 AUC: Area under the curve
 A β : Amyloid- β
 Bp: Basepairs
 BME: beta-Mercaptoethanol
 CI: Confidence interval
 DAB: 3,3'-Diaminobenzidine
 DAPI: 4',6-diamidino-2-phenylindole
 FC: Fold change
 FDR: False discovery rate
 GO: Gene ontology
 HpSp AD: Hippocampal sparing subtype of Alzheimer's Disease
 H&E: Hematoxylin and eosin
 Limbic AD: Limbic predominant subtype of Alzheimer's Disease
 MSBB: Mount Sinai VA Medical Center Brain Bank
 MMSE: Mini mental state exam
 NFT: Neurofibrillary tangle
 OR: Odds ratio
 PBS: Phosphate-buffered saline
 PMSF: phenylmethylsulfonyl fluoride
 RIN: RNA integrity number
 RNA-Seq: RNA sequencing
 ROC: receiver operating characteristic
 TBST: tris-buffered saline with Tween-20
 Thio-S: Thioflavin-S
 Typical AD: Typical Alzheimer's Disease

References

- 1 Montine, T. J. *et al.* National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease: a practical approach. *Acta Neuropathol* **123**, 1-11, doi:10.1007/s00401-011-0910-3 (2012).
- 2 Thinakaran, G. & Koo, E. H. Amyloid precursor protein trafficking, processing, and function. *J Biol Chem* **283**, 29615-29619, doi:10.1074/jbc.R800019200 (2008).
- 3 O'Brien, R. J. & Wong, P. C. Amyloid precursor protein processing and Alzheimer's disease. *Annu Rev Neurosci* **34**, 185-204, doi:10.1146/annurev-neuro-061010-113613 (2011).
- 4 Alonso, A., Zaidi, T., Novak, M., Grundke-Iqbal, I. & Iqbal, K. Hyperphosphorylation induces self-assembly of tau into tangles of paired helical filaments/straight filaments. *Proc Natl Acad Sci U S A* **98**, 6923-6928, doi:10.1073/pnas.121119298 (2001).
- 5 Ballatore, C., Lee, V. M. & Trojanowski, J. Q. Tau-mediated neurodegeneration in Alzheimer's disease and related disorders. *Nat Rev Neurosci* **8**, 663-672, doi:10.1038/nrn2194 (2007).
- 6 Thal, D. R., Rub, U., Orantes, M. & Braak, H. Phases of A beta-deposition in the human brain and its relevance for the development of AD. *Neurology* **58**, 1791-1800, doi:10.1212/wnl.58.12.1791 (2002).
- 7 Braak, H. & Braak, E. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol* **82**, 239-259 (1991).
- 8 Ossenkoppele, R. *et al.* Tau PET patterns mirror clinical and neuroanatomical variability in Alzheimer's disease. *Brain* **139**, 1551-1567, doi:10.1093/brain/aww027 (2016).
- 9 Murray, M. E. *et al.* Clinicopathologic and 11C-Pittsburgh compound B implications of Thal amyloid phase across the Alzheimer's disease spectrum. *Brain* **138**, 1370-1381, doi:10.1093/brain/aww050 (2015).
- 10 Hanna Al-Shaikh, F. S. *et al.* Selective Vulnerability of the Nucleus Basalis of Meynert Among Neuropathologic Subtypes of Alzheimer Disease. *JAMA Neurol*, doi:10.1001/jamaneurol.2019.3606 (2019).
- 11 Janocko, N. J. *et al.* Neuropathologically defined subtypes of Alzheimer's disease differ significantly from neurofibrillary tangle-predominant dementia. *Acta Neuropathol* **124**, 681-692, doi:10.1007/s00401-012-1044-y (2012).
- 12 Murray, M. E. *et al.* Differential clinicopathologic and genetic features of late-onset amnesic dementias. *Acta Neuropathol* **128**, 411-421, doi:10.1007/s00401-014-1302-2 (2014).
- 13 Murray, M. E. *et al.* Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: a retrospective study. *Lancet Neurol* **10**, 785-796 (2011).
- 14 Lambert, J. C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* **45**, 1452-1458, doi:10.1038/ng.2802 (2013).
- 15 Harold, D. *et al.* Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet* **41**, 1088-1093, doi:10.1038/ng.440 (2009).
- 16 Seshadri, S. *et al.* Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA* **303**, 1832-1840, doi:10.1001/jama.2010.574 (2010).
- 17 Naj, A. C. *et al.* Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet* **43**, 436-441, doi:10.1038/ng.801 (2011).
- 18 Hollingworth, P. *et al.* Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet* **43**, 429-435, doi:10.1038/ng.803 (2011).
- 19 Lambert, J. C. *et al.* Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet* **41**, 1094-1099, doi:10.1038/ng.439 (2009).
- 20 Goate, A. *et al.* Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature* **349**, 704-706, doi:10.1038/349704a0 (1991).
- 21 Sherrington, R. *et al.* Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature* **375**, 754-760, doi:10.1038/375754a0 (1995).
- 22 Bonda, D. J. *et al.* The sirtuin pathway in ageing and Alzheimer disease: mechanistic and therapeutic considerations. *Lancet Neurol* **10**, 275-279, doi:10.1016/S1474-4422(11)70013-8 (2011).
- 23 Roses, A. D. *et al.* A TOMM40 variable-length polymorphism predicts the age of late-onset Alzheimer's disease. *Pharmacogenomics J* **10**, 375-384, doi:10.1038/tpj.2009.69 (2010).
- 24 Jin, S. C. *et al.* Pooled-DNA sequencing identifies novel causative variants in PSEN1, GRN and MAPT in a clinical early-onset and familial Alzheimer's disease Ibero-American cohort. *Alzheimers Res Ther* **4**, 34, doi:10.1186/alzrt137 (2012).
- 25 Allen, M. *et al.* Conserved brain myelination networks are altered in Alzheimer's and other neurodegenerative diseases. *Alzheimers Dement* **14**, 352-366, doi:10.1016/j.jalz.2017.09.012 (2018).
- 26 Murray, M. E. *et al.* Early Alzheimer's disease neuropathology detected by proton MR spectroscopy. *J Neurosci* **34**, 16247-16255, doi:10.1523/JNEUROSCI.2027-14.2014 (2014).

- 27 Attems, J., Neltner, J. H. & Nelson, P. T. Quantitative neuropathological assessment to investigate cerebral multi-morbidity. *Alzheimers Res Ther* **6**, 85, doi:10.1186/s13195-014-0085-y (2014).
- 28 Neltner, J. H. et al. Digital pathology and image analysis for robust high-throughput quantitative assessment of Alzheimer disease neuropathologic changes. *J Neuropathol Exp Neurol* **71**, 1075-1085, doi:10.1097/NEN.0b013e3182768de4 (2012).
- 29 Bancher, C. et al. Accumulation of abnormally phosphorylated tau precedes the formation of neurofibrillary tangles in Alzheimer's disease. *Brain Res* **477**, 90-99, doi:10.1016/0006-8993(89)91396-6 (1989).
- 30 Lowe, V. J. et al. An autoradiographic evaluation of AV-1451 Tau PET in dementia. *Acta Neuropathol Commun* **4**, 58, doi:10.1186/s40478-016-0315-6 (2016).
- 31 Jicha, G. A. et al. cAMP-Dependent Protein Kinase Phosphorylations on Tau in Alzheimer's Disease. *J. Neurosci.* **19**, 7486-7494, doi:10.1523/jneurosci.19-17-07486.1999 (1999).
- 32 Dickson, D. W. et al. Immunocytochemistry of neurofibrillary tangles with antibodies to subregions of tau protein: identification of hidden and cleaved tau epitopes and a new phosphorylation site. *Acta Neuropathol.* **84**, doi:10.1007/bf00227736 (1992).
- 33 Norton, N. et al. Gene expression, single nucleotide variant and fusion transcript discovery in archival material from breast tumors. *PLoS One* **8**, e81925, doi:10.1371/journal.pone.0081925 (2013).
- 34 Archer, K. J. & Kirnes, R. V. Empirical characterization of random forest variable importance measures. *Comput Stat Data An* **52**, 2249-2260, doi:10.1016/j.csda.2007.08.015 (2008).
- 35 Abu-Rustum, N. R. et al. A nomogram for predicting overall survival of women with endometrial cancer following primary therapy: toward improving individualized cancer care. *Gynecol Oncol* **116**, 399-403, doi:10.1016/j.ygyno.2009.11.027 (2010).
- 36 Kim, S. E. et al. A Nomogram for Predicting Amyloid PET Positivity in Amnesic Mild Cognitive Impairment. *J Alzheimers Dis* **66**, 681-691, doi:10.3233/JAD-180048 (2018).
- 37 Folstein, M. F., Folstein, S. E. & McHugh, P. R. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* **12**, 189-198, doi:[https://doi.org/10.1016/0022-3956\(75\)90026-6](https://doi.org/10.1016/0022-3956(75)90026-6) (1975).
- 38 Allen, M. et al. Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. *Sci Data* **3**, 160089, doi:10.1038/sdata.2016.89 (2016).
- 39 Wang, M. et al. The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease. *Sci Data* **5**, 180185, doi:10.1038/sdata.2018.185 (2018).
- 40 Marlar, R. A. & Griffin, J. H. Deficiency of protein C inhibitor in combined factor V/VIII deficiency disease. *J Clin Invest* **66**, 1186-1189, doi:10.1172/JCI109952 (1980).
- 41 Suzuki, K., Nishioka, J. & Hashimoto, S. Protein C inhibitor. Purification from human plasma and characterization. *J Biol Chem* **258**, 163-168 (1983).
- 42 Huntington, J. A. Thrombin inhibition by the serpins. *J Thromb Haemost* **11 Suppl 1**, 254-264, doi:10.1111/jth.12252 (2013).
- 43 Nelson, P. T. et al. Correlation of Alzheimer disease neuropathologic changes with cognitive status: a review of the literature. *J Neuropathol Exp Neurol* **71**, 362-381, doi:10.1097/NEN.0b013e31825018f7 (2012).
- 44 Uchihara, T. Pretangles and neurofibrillary changes: similarities and differences between AD and CBD based on molecular and morphological evolution. *Neuropathology* **34**, 571-577, doi:10.1111/neup.12108 (2014).
- 45 Cook, C. et al. Tau deposition drives neuropathological, inflammatory and behavioral abnormalities independently of neuronal loss in a novel mouse model. *Hum Mol Genet* **24**, 6198-6212, doi:10.1093/hmg/ddv336 (2015).
- 46 Annese, A. et al. Whole transcriptome profiling of Late-Onset Alzheimer's Disease patients provides insights into the molecular changes involved in the disease. *Sci Rep* **8**, 4282, doi:10.1038/s41598-018-22701-2 (2018).
- 47 van Rooij, J. G. J. et al. Hippocampal transcriptome profiling combined with protein-protein interaction analysis elucidates Alzheimer's disease pathways and genes. *Neurobiol Aging* **74**, 225-233, doi:10.1016/j.neurobiolaging.2018.10.023 (2019).
- 48 Magistri, M., Velmeshev, D., Makhmutova, M. & Faghihi, M. A. Transcriptomics Profiling of Alzheimer's Disease Reveal Neurovascular Defects, Altered Amyloid-beta Homeostasis, and Deregulated Expression of Long Noncoding RNAs. *J Alzheimers Dis* **48**, 647-665, doi:10.3233/JAD-150398 (2015).
- 49 Malleier, J. M. et al. Regulation of protein C inhibitor (PCI) activity by specific oxidized and negatively charged phospholipids. *Blood* **109**, 4769-4776, doi:10.1182/blood-2006-09-046953 (2007).
- 50 Rieger, D., Assinger, A., Einfinger, K., Sokolikova, B. & Geiger, M. Protein C inhibitor (PCI) binds to phosphatidylserine exposing cells with implications in the phagocytosis of apoptotic cells and activated platelets. *PLoS One* **9**, e101794, doi:10.1371/journal.pone.0101794 (2014).
- 51 Wahlmuller, F. C., Sokolikova, B., Rieger, D. & Geiger, M. New lipid interaction partners stimulate the inhibition of activated protein C by cell-penetrating protein C inhibitor. *Thrombosis and haemostasis* **111**, 41-52, doi:10.1160/th13-06-0478 (2014).

- 52 Nishioka, J., Ning, M., Hayashi, T. & Suzuki, K. Protein C inhibitor secreted from activated platelets efficiently inhibits activated protein C on phosphatidylethanolamine of platelet membrane and microvesicles. *J Biol Chem* **273**, 11281-11287, doi:10.1074/jbc.273.18.11281 (1998).
- 53 Baumgartner, P. *et al.* Phosphatidylethanolamine critically supports internalization of cell-penetrating protein C inhibitor. *The Journal of cell biology* **179**, 793-804, doi:10.1083/jcb.200707165 (2007).
- 54 Pratt, C. W. & Church, F. C. Heparin binding to protein C inhibitor. *J Biol Chem* **267**, 8789-8794 (1992).
- 55 Yang, H., Wahlmuller, F. C., Sarg, B., Furtmuller, M. & Geiger, M. A α -helix of protein C inhibitor (PCI) is a cell-penetrating peptide that mediates cell membrane permeation of PCI. *J Biol Chem* **290**, 3081-3091, doi:10.1074/jbc.M114.581736 (2015).
- 56 Fortenberry, Y. M. *et al.* Essential thrombin residues for inhibition by protein C inhibitor with the cofactors heparin and thrombomodulin. *J Thromb Haemost* **5**, 1486-1492, doi:10.1111/j.1538-7836.2007.02574.x (2007).
- 57 Hayashi, S. & Wakizaka, A. Urinary protein C inhibitor binding region in the A α -chain of human fibrinogen. *Blood coagulation & fibrinolysis : an international journal in haemostasis and thrombosis* **6**, 382-387, doi:10.1097/00001721-199507000-00003 (1995).
- 58 Kise, H., Nishioka, J., Kawamura, J. & Suzuki, K. Characterization of semenogelin II and its molecular interaction with prostate-specific antigen and protein C inhibitor. *European journal of biochemistry* **238**, 88-96, doi:10.1111/j.1432-1033.1996.0088q.x (1996).
- 59 Fortenberry, Y. M., Brandal, S., Bialas, R. C. & Church, F. C. Protein C inhibitor regulates both cathepsin L activity and cell-mediated tumor cell migration. *Biochimica et biophysica acta* **1800**, 580-590, doi:10.1016/j.bbagen.2010.03.003 (2010).
- 60 Zhan, S., Wang, T., Ge, W. & Li, J. Multiple roles of Ring 1 and YY1 binding protein in physiology and disease. *J Cell Mol Med* **22**, 2046-2054, doi:10.1111/jcmm.13503 (2018).
- 61 Gao, Z. *et al.* An AUTS2-Polycomb complex activates gene expression in the CNS. *Nature* **516**, 349-354, doi:10.1038/nature13921 (2014).
- 62 Gonzalez, I., Aparicio, R. & Busturia, A. Functional characterization of the dRYBP gene in Drosophila. *Genetics* **179**, 1373-1388, doi:10.1534/genetics.107.082966 (2008).
- 63 Pirity, M. K., Locker, J. & Schreiber-Agus, N. Rybp/DEDAF is required for early postimplantation and for central nervous system development. *Mol Cell Biol* **25**, 7193-7202, doi:10.1128/MCB.25.16.7193-7202.2005 (2005).
- 64 Schickling, O., Stegh, A. H., Byrd, J. & Peter, M. E. Nuclear localization of DEDD leads to caspase-6 activation through its death effector domain and inhibition of RNA polymerase I dependent transcription. *Cell Death Differ* **8**, 1157-1168, doi:10.1038/sj.cdd.4400928 (2001).
- 65 Zheng, L., Schickling, O., Peter, M. E. & Lenardo, M. J. The death effector domain-associated factor plays distinct regulatory roles in the nucleus and cytoplasm. *J Biol Chem* **276**, 31945-31952, doi:10.1074/jbc.M102799200 (2001).
- 66 Tan, K. *et al.* Tumor suppressor RYBP harbors three nuclear localization signals and its cytoplasm-located mutant exerts more potent anti-cancer activities than corresponding wild type. *Cell Signal* **29**, 127-137, doi:10.1016/j.cellsig.2016.10.011 (2017).
- 67 Takanaga, H., Tokuda, N., Ohtsuki, S., Hosoya, K. & Terasaki, T. ATA2 is predominantly expressed as system A at the blood-brain barrier and acts as brain-to-blood efflux transport for L-proline. *Mol Pharmacol* **61**, 1289-1296, doi:10.1124/mol.61.6.1289 (2002).
- 68 Grewal, S. *et al.* SNAT2 amino acid transporter is regulated by amino acids of the SLC6 gamma-aminobutyric acid transporter subfamily in neocortical neurons and may play no role in delivering glutamine for glutamatergic transmission. *J Biol Chem* **284**, 11224-11236, doi:10.1074/jbc.M806470200 (2009).
- 69 Hoffmann, T. M. *et al.* Effects of Sodium and Amino Acid Substrate Availability upon the Expression and Stability of the SNAT2 (SLC38A2) Amino Acid Transporter. *Front Pharmacol* **9**, 63, doi:10.3389/fphar.2018.00063 (2018).
- 70 Sun, T. P. & Shieh, S. Y. Human FEM1B is required for Rad9 recruitment and CHK1 activation in response to replication stress. *Oncogene* **28**, 1971-1981, doi:10.1038/onc.2009.58 (2009).
- 71 Gilder, A. S., Chen, Y. B., Jackson, R. J., 3rd, Jiang, J. & Maher, J. F. Fem1b promotes ubiquitylation and suppresses transcriptional activity of Gli1. *Biochem Biophys Res Commun* **440**, 431-436, doi:10.1016/j.bbrc.2013.09.090 (2013).
- 72 Subauste, M. C. *et al.* Fem1b, a proapoptotic protein, mediates proteasome inhibitor-induced apoptosis of human colon cancer cells. *Mol Carcinog* **49**, 105-113, doi:10.1002/mc.20594 (2010).
- 73 Stehlik, C. *et al.* The PAAD/PYRIN-only protein POP1/ASC2 is a modulator of ASC-mediated nuclear-factor-kappa B and pro-caspase-1 regulation. *Biochem J* **373**, 101-113, doi:10.1042/BJ20030304 (2003).
- 74 Dobnikar, L. *et al.* Disease-relevant transcriptional signatures identified in individual smooth muscle cells from healthy mouse vessels. *Nat Commun* **9**, 4567, doi:10.1038/s41467-018-06891-x (2018).
- 75 Hurley, M. J., Durrenberger, P. F., Gentleman, S. M., Walls, A. F. & Dexter, D. T. Altered Expression of Brain Proteinase-Activated Receptor-2, Trypsin-2 and Serpin Proteinase Inhibitors in Parkinson's Disease. *J Mol Neurosci* **57**, 48-62, doi:10.1007/s12031-015-0576-8 (2015).

- 76 Morsch, R., Simon, W. & Coleman, P. D. Neurons may live for decades with neurofibrillary tangles. *J Neuropathol Exp Neurol* **58**, 188-197, doi:10.1097/00005072-199902000-00008 (1999).
- 77 Kuchibhotla, K. V. *et al.* Neurofibrillary tangle-bearing neurons are functionally integrated in cortical circuits in vivo. *Proc Natl Acad Sci U S A* **111**, 510-514, doi:10.1073/pnas.1318807111 (2014).
- 78 Neuner, S. M., Tcw, J. & Goate, A. M. Genetic architecture of Alzheimer's disease. *Neurobiol Dis* **143**, 104976, doi:10.1016/j.nbd.2020.104976 (2020).
- 79 Young, A. L. *et al.* Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference. *Nat Commun* **9**, 4273, doi:10.1038/s41467-018-05892-0 (2018).
- 80 Ravits, J. M. & La Spada, A. R. ALS motor phenotype heterogeneity, focal, and spread: deconstructing motor neuron degeneration. *Neurology* **73**, 805-811, doi:10.1212/WNL.0b013e3181b6bbbd (2009).
- 81 Thenganatt, M. A. & Jankovic, J. Parkinson disease subtypes. *JAMA Neurol* **71**, 499-504, doi:10.1001/jamaneurol.2013.6233 (2014).
- 82 Davies, M., Brophy, S., Williams, R. & Taylor, A. The prevalence, severity, and impact of painful diabetic peripheral neuropathy in type 2 diabetes. *Diabetes Care* **29**, 1518-1522, doi:10.2337/dc05-2228 (2006).
- 83 Liesinger, A. M. *et al.* Sex and age interact to determine clinicopathologic differences in Alzheimer's disease. *Acta Neuropathol* **136**, 873-885, doi:10.1007/s00401-018-1908-x (2018).
- 84 Schroeder, A. *et al.* The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol* **7**, 3, doi:10.1186/1471-2199-7-3 (2006).
- 85 Wehmas, L. C. *et al.* Enhanced Quality Metrics for Assessing RNA Derived From Archival Formalin-Fixed Paraffin-Embedded Tissue Samples. *Toxicol Sci* **170**, 357-373, doi:10.1093/toxsci/kfz113 (2019).
- 86 Kalari, K. R. *et al.* MAP-RSeq: Mayo Analysis Pipeline for RNA sequencing. *BMC Bioinformatics* **15**, 224, doi:10.1186/1471-2105-15-224 (2014).
- 87 Hansen, K. D., Irizarry, R. A. & Wu, Z. Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics* **13**, 204-216, doi:10.1093/biostatistics/kxr054 (2012).
- 88 Partek Genomics Suite (Partek Inc., St. Louis, 2020).
- 89 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).
- 90 Scarlet, D., Ertl, R., Aurich, C. & Steinborn, R. The Orthology Clause in the Next Generation Sequencing Era: Novel Reference Genes Identified by RNA-seq in Humans Improve Normalization of Neonatal Equine Ovary RT-qPCR Data. *PLoS One* **10**, e0142122, doi:10.1371/journal.pone.0142122 (2015).
- 91 Yan, L. *et al.* OSAT: a tool for sample-to-batch allocations in genomics experiments. *BMC Genomics* **13**, 689, doi:10.1186/1471-2164-13-689 (2012).
- 92 Waggott, D. *et al.* NanoStringNorm: an extensible R package for the pre-processing of NanoString mRNA and miRNA data. *Bioinformatics (Oxford, England)* **28**, 1546-1548, doi:10.1093/bioinformatics/bts188 (2012).
- 93 Classification and Regression by randomForest (R News, 2002).
- 94 randomForestExplainer: Explaining and Visualizing Random Forests in Terms of Variable Importance (2019).
- 95 Wan, Y.-W. *et al.* Meta-Analysis of the Alzheimer's Disease Human Brain Transcriptome and Functional Dissection in Mouse Models. *Cell Rep* **32**, 107908-107908, doi:10.1016/j.celrep.2020.107908 (2020).
- 96 Mirra, S. S. *et al.* The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurology* **41**, 479-486, doi:10.1212/wnl.41.4.479 (1991).
- 97 Santos, O. A. *et al.* Ethnoracial differences in Alzheimer's disease from the FLorida Autopsied Multi-Ethnic (FLAME) cohort. *Alzheimers Dement* **15**, 635-643, doi:10.1016/j.jalz.2018.12.013 (2019).
- 98 Yen, S. H., Dickson, D. W., Crowe, A., Butler, M. & Shelanski, M. L. Alzheimer's neurofibrillary tangles contain unique epitopes and epitopes in common with the heat-stable microtubule associated proteins tau and MAP2. *Am J Pathol* **126**, 81-91 (1987).
- 99 Ioannidis, J. P. A. The Proposal to Lower P Value Thresholds to .005. *JAMA* **319**, 1429-1430, doi:10.1001/jama.2018.1536 (2018).