A Continuous Extension of Gene Set Enrichment Analysis using the Likelihood Ratio Test Statistics Identifies VEGF as a Candidate Pathway for Alzheimer's disease

Ali Mahzarnia[1*], Michael W. Lutz[2], Alexandra Badea[1,2,3,4]

[1] Department of Radiology, Duke University School of Medicine, Durham, NC, USA

[2] Department of Neurology, Duke University School of Medicine, Durham, NC, USA

[3] Biomedical Engineering, Duke University, Durham, NC, USA

[4] Brain Imaging and Analysis Center, Duke University School of Medicine, Durham, NC, USA

*Corresponding Author:

Ali Mahzarnia, PhD

Rm139, Joseph and Kathleen Bryan Research Building

Duke University Medical Center, Durham, NC, 27710

Ali.mahzarnia@duke.edu

919-684-7654

Abstract

Background

Alzheimer's disease involves brain pathologies such as amyloid plaque depositions and hyperphosphorylated tau tangles and is accompanied by cognitive decline. Identifying the biological mechanisms underlying disease onset and progression based on quantifiable phenotypes will help understand the disease etiology and devise therapies.

Objective

Our objective was to identify molecular pathways associated with AD biomarkers (Amyloid-β and tau) and cognitive status (MMSE) accounting for variables such as age, sex, education, and APOE genotype.

Methods

We introduce a novel pathway-based statistical approach, extending the gene set likelihood ratio test to continuous phenotypes. We first analyzed independently each of the three phenotypes (Amyloid-β, tau, cognition), using continuous gene set likelihood ratio tests to account for covariates, including age, sex, education, and APOE genotype. The analysis involved a large sample size with data available for all three phenotypes, allowing for the identification of common pathways.

Results

We identified 14 pathways significantly associated with Amyloid-β, 5 associated with tau, and 174 associated with MMSE. Surprisingly, the MMSE outcome showed a larger number of significant pathways compared to biomarkers. A single pathway, vascular endothelial growth factor receptor binding (VEGF-RB), exhibited significant associations with all three phenotypes.

Conclusions.

The study's findings highlight the importance of the VEGF signaling pathway in aging in AD. The complex interactions within the VEGF signaling family offer valuable insights for future therapeutic interventions.

Keywords: Continuous GSLRT, Likelihood Ratio-Derived Test Statistics, Alzheimer's Disease, RNA-Seq

Introduction

41 One strategy to better understand and accurately model late-onset Alzheimer's disease, is to
42 incorporate various genetic, clinical, and environmental factors into a cohesive model. This model
43 should establish connections between measurable biomarkers and risk factors. Several key
44 factors play a significant role in shaping the risk of Alzheimer's disease (AD). These factors
45 encompass both genetic influences, such as APOE genotype and sex, as well as environmental
46 elements, including education level, diet, and age. In recent years, the identification and
47 characterization of AD have been facilitated by the use of biomarkers like Amyloid-β (Aβ),
48 phosphorylated tau (tau), and measures of neurodegeneration. Furthermore, AD is distinguished
49 by memory impairment, often evaluated using the Minimental State Evaluation (MMSE).

50 AD has a deleterious impact on American lives, as over 6 million individuals are currently afflicted
51 with Alzheimer's, with a projected twofold increase by 2050. The disease's high mortality rate
52 claims 1 in 3 seniors while imposing significant economic burdens, costing the nation billions [1].
53 However, the biological background conducive to developing AD remains unknown. The aim of
54 this study was to address the existing knowledge gap by pinpointing molecular pathways that play
55 a crucial role in modulating levels of hallmark AD pathologies, as well as memory function.

56 Recent publications have revealed a role for multiple pathways in AD, based on brain proteomics
57 and transcriptomic analyses [2, 3]. These studies have revealed pathways that are relevant not
58 only to neurons but also to cells regulating response to inflammation [4], such as microglia [5].
59 Interestingly endothelial cells, astrocytes and neurons that control neurovascular functions have
60 been shown to play an important role in AD [6]. Other cell types and subcellular components
61 such as mitochondria may be involved [7]. [8] identifies a novel brain-enriched RING finger E3
62 ligase, RNF182, which shows elevated expression in AD brains and may play a role in controlling
63 neurotransmitter release. Pathways involved with filament-based processes, cellular
64 detoxification, and wound healing have also been involved [2, 3]. A decline in sensory function,

including taste has also been reported with aging and AD [9]. Importantly, the vascular endothelial growth factor has been associated with AD (VEGF) [10, 11], and while its role in neurodegeneration is yet to be understood, it presents a druggable target for therapies. However, most studies focused on comparisons of two or three groups of subjects using discrete classification variables, such as case/control, without accounting for the relationships between multiple hallmark biomarkers. Here we propose an approach to detect common gene pathways based on RNA-Seq changes associated with continuous-scale changes in multiple biomarkers and clinical phenotypes.

Our study involves developing a statistical approach centered around gene pathways. This method incorporates estimations of Amyloid-β and tau tangle brain levels, along with memory scores from the MMSE, and integrates them into a comprehensive statistical model. Additionally, we include disease-relevant traits such as age, sex, education, and APOE genotype in this model. The primary goal is to rank the pathways that undergo alterations in AD, considering the influence of each of the biomarkers. By doing so, we can effectively identify the shared pathways across the three domains: amyloid, tau, and MMSE scores.

Moreover, our approach factors in the unique characteristics of human subjects, including age-specific, sex-specific, education-specific, and APOE genotype-specific differences. It also evaluates the significance of the relationship between pathway-level interactions and the presence of AD in relation to each of these factors. We conduct analyses using human transcriptomic data on Alzheimer's disease progression and explore interactions with the APOE genotype.

To identify relevant gene sets [12] introduced Gene Set Enrichment Analysis (GSEA), a robust analytical method for interpreting gene expression data from genome-wide RNA analysis. GSEA focuses on gene sets—groups of genes with shared biological function, chromosomal location, or regulation—and demonstrates effectiveness in identifying common biological pathways e.g. for

90 cancer-related data sets, such as leukemia and lung cancer, where single-gene analysis falls
91 short. This method using the Kolmogorov-Smirnov statistics has a limitation in that it does not
92 account for gene-gene interactions. Still, it has spurred the development of many other methods
93 [13]. Other statistics have been proposed as well, e.g. the gene set likelihood ratio test (gsLRT),
94 which uses a logistic regression model [14]. However, this particular model has its limitations as
95 it is only applicable to binary outcome variables. In contrast, our study deals with continuous-scale
96 variables, specifically amyloid, tau, and MMSE scores. As a result, we require a different model
97 that can accommodate such continuous data. Additionally, we incorporated transcriptomic data,
98 encompassing more than 20 thousand genes, as opposed to the mere 142 genes used in the
99 referenced gsLRT study involving proteomics.

Methodology

100 Our proposition introduces a novel approach utilizing all three phenotypes and biomarkers
101 (amyloid, tau, and MMSE scores) to identify shared and meaningful pathways associated with
102 them. Our methodology harnesses gene-level data, encompassing transcriptomics. By utilizing
103 this data, we generate significance scores at the biological pathway level, serving as indicators of
104 the novelty of observed experimental outcomes.

Data and preprocessing

105 The data sample was taken from a subset of the Religious Orders Study and Rush Memory and
106 Aging Project (ROSMAP) dataset [15-17] that had RNAseq data available from the dorsolateral
107 frontal cortex. ROS has enlisted nuns and brothers since 1994. MAP recruited individuals from
108 the northern Illinois region since 1997. Both studies were run by the same investigators using
109 similar data collection techniques. Thus, the results from both were comparable. For the analyses
110 reported in this paper, the clinical consensus diagnoses of Alzheimer's disease or mild cognitive

111 impairment were used to define a case while the diagnosis of no cognitive impairment/no impaired
112 domains defined controls. Additional covariates for the statistical models were age, sex,
113 education, and APOE genotype. The total sample with both gene expression and clinical data
114 contained 634 subjects, with 433 cases and 201 controls. Demographic information for the
115 sample is summarized in Table 1.

Group	N	age (mean)	age (SD)
All	634	87.93	6.70
Female	406	89.02	6.57
Male	228	85.99	6.51

116 *Table 1- Demographic Information of Participants by Gender. This table presents the*
117 *demographic information for a sample of 634 participants, categorized by gender. It includes the*
118 *mean age and standard deviation (SD) for each group.*

119 In the context of [18], the amyloid and tangles definitions and measurements are as follows.

120 The overall amyloid level is determined as the mean percentage of cortex occupied by Amyloid-
121 β protein in eight specific brain regions. This measurement is obtained through molecular-specific
122 immunohistochemistry, where the Amyloid-β protein is targeted and quantified using image
123 analysis techniques. The Amyloid-β score is calculated in eight brain regions, namely the
124 hippocampus, entorhinal cortex, midfrontal cortex, inferior temporal cortex, angular gyrus,
125 calcarine cortex, anterior cingulate cortex, and superior frontal cortex. At least four of these
126 regions are required to calculate the mean Amyloid-βscore.

127 Tangle density is determined as the mean density of neuronal neurofibrillary tangles in eight
128 specific brain regions. These tangles are identified using molecular-specific
129 immunohistochemistry, employing antibodies specific to abnormally phosphorylated Tau protein,
130 known as AT8. The cortical density of tangles is measured per square millimeter using systematic

sampling. The tangle score is calculated as the mean density in eight brain regions, including the hippocampus, entorhinal cortex, midfrontal cortex, inferior temporal cortex, angular gyrus, calcarine cortex, anterior cingulate cortex, and superior frontal cortex. A minimum of four regions is necessary to compute the mean tangle density.

Clinical phenotypic information for the sample is summarized in Table 2.

Cognitive state/Phenotypes	N Rows	MMSE (mean)	MMSE (SD)	Amyloid (mean)	Amyloid (SD)	log(tangles) (mean)	log(tangles) (SD)
No cognitive impairment	201	28.21	1.6	2.78	3.44	0.32	1.58
MCI: Mild cognitive impairment (one impaired domain) and NO other cause of CI	158	25.89	3.38	4.03	4.63	1.07	1.28
MCI: Mild cognitive impairment (One impaired domain) AND another cause of CI	10	25.19	3.73	1.35	1.33	-0.03	1.3
AD: Alzheimer's dementia and NO other cause of CI (NINCDS PROB AD)	220	13.49	8.22	5.72	4.36	1.83	1.25
AD: Alzheimer's dementia AND another cause of CI (NINCDS POSS AD)	33	15.49	7.94	3.62	3.69	1.41	0.97
Other dementia: Other primary cause of dementia	12	15.17	5.47	2.59	3.6	0.5	1.31

Table 2-Clinical Phenotypic Data for Different Cognitive States. This table presents key clinical phenotypic data for different cognitive states in a diverse sample population. The study includes individuals with no cognitive impairment, mild cognitive impairment (MCI), Alzheimer's dementia

(AD), and other primary causes of dementia. The table provides mean scores and standard deviations for Mini-Mental State Examination (MMSE), amyloid levels, and log(tangles) for each cognitive state/phenotype. The findings highlight distinct cognitive profiles and potential biomarkers associated with various cognitive conditions, contributing to better understanding and targeted interventions for cognitive disorders.

The RNAseq data was obtained from the Accelerating Medicines Partnership Program for Alzheimer's Disease Data Knowledge Portal (https://adknowledgeportal.synapse.org/), specifically, the RNA-seq Harmonization study (https://www.synapse.org/#!Synapse:syn9702085). The ROSMAP data from this study was used to create a combined dataset of RNA-seq data in combination with the three clinical phenotypes of MMSE, amyloid burden, and tangles [19]. The RNA-seq Harmonization study has the goal of creating an RNA-seq database based on a consensus set of analytical tools. The methodological details of the RNA-seq processing are given in Wan et al [20] and at the RNA-Seq reprocessing study website for the ROSMAP project (https://www.synapse.org/#!Synapse:syn8456629). In brief, RNA was extracted from samples consisting of approximately 100 mg of gray matter tissue from the dorsolateral prefrontal cortex. The RNA samples were prepared and sequenced as described in [19]. The reprocessing of the RNAseq data was done using a consensus set of tools with only library type-specific parameters varying between pipelines. Picard (https://broadinstitute.github.io/picard/) was used to generate FASTQs from source BAMs. Generated FASTQ reads were aligned to the GENCODE24 (GRCh38) reference genome using STAR [21] and gene counts were computed for each sample. To evaluate the quality of individual samples and to identify potentially important covariates for expression modeling, two sets of metrics were computed using the CollectAlignmentSummaryMetrics and CollectRnaSeqMetrics functions in Picard. To account for differences between samples, studies, experimental batch effects, and unwanted RNA-Seq specific technical variations library normalization and covariate

adjustments for each study separately using fixed/mixed effects modeling. The workflow consists of the following steps: (i) gene filtering: Genes that are expressed more than 1 CPM (read Counts Per Million total reads) in at least 50% of samples in each tissue and diagnosis category were used for further analysis, (ii) conditional quantile normalization was applied to account for variations in gene length and GC content, (iii) sample outlier detection using principal component analysis and clustering, (iv) Covariates identification and adjustment, where confidence of sampling abundance were estimated using a weighted linear model using voom-limma package in Bioconductor [22]. For the differential expression analysis, fixed/mixed effect linear regression was used with the following models: gene expression ~Diagnosis + Sex + covariates + (1| Donor) or gene expression ~Diagnosis x Sex + covariates + (1|Donor), where each gene is linearly regressed independently with Diagnosis, a variable explaining the AD status of an individual, identified covariates, and donor information as a random effect. Observation weights (if any) were calculated using the voom-limma [22] pipeline such that observations with higher presumed precision will be up-weighted in the linear model fitting process.

The data used for the gsLRT analysis included the phenotypes MMSE, amyloid burden and log(tangles). For gene expression, normalized and filtered counts were used as per the description on the RNA-Seq reprocessing study website (https://www.synapse.org/#!Synapse:syn8456629). The specific datasets used for this study are: RNAseq data is obtained from (https://www.synapse.org/#!Synapse:syn8456638), covariates are recorded in (https://www.synapse.org/#!Synapse:syn11024258), differential expression results are from (https://www.synapse.org/#!Synapse:syn8456721) and clinical phenotypic data was obtained from the ROSMAP data sharing resource (https://www.radc.rush.edu/). These datasets were organized into the specific matrices for gene expression results, covariates, and phenotypes required by the gsLRT program.

Data and Code Availability

ROSMAP resources can be requested at https://www.radc.rush.edu. The Code is available at https://github.com/Ali-Mahzarnia/gslrt2.

Statistical Model

To derive the pathways associated with each AD related phenotypes (amyloid, tau, or MMSE scores), we define the following hypothesis tests. In our models, we utilize a matrix $G \in \mathbb{R}^{n\times m}$ to represent measurement values (e.g., RNA expression) for m genes and n samples. Additionally, we employ a matrix, $X \in \mathbb{R}^{n\times d}$, which contains d=4 covariates such as "sex," years of education", "APOE Genotype" (APOE2,3,4), and "age at death" associated with the samples. Moreover, we utilize a continuous vector $y \in \mathbb{R}^{n\times 1}$ for the n samples that is the phenotype measurements (amyloid, tau, or MMSE scores). To conduct data analysis, we employ nested models for each gene $g_j,\ j = 1,\ldots,m$ and all samples $i = 1,\ldots,n$ as outlined below:

$$H_0 : y_i = x_i\beta^j + g_i^j\alpha^j + e_i$$

$$H_a :\ y_i = x_i\beta^j + g_i^j\alpha^j + Genotype_i * g_i^j\gamma^j + e_i$$

where $e_i \sim N(0\,,\sigma)$. We denote the maximum likelihood estimator of the two above models with $\widehat{\omega}_0 = (\widehat{\beta_0}, \widehat{\alpha_0}\,)$, $\widehat{\omega}_a = (\widehat{\beta_a}, \widehat{\alpha_a}, \widehat{\gamma^j})$.

In order to assess the enhanced explanatory capacity of the interaction term between the gene measurement profile (g_j) and genotype (APOE2,3,4), in contrast to the simpler model that merely includes the covariate matrix X and (g_j), we introduce statistical measures that evaluate the disparity in the models via log-likelihoods at the gene level. These statistics act as our metric for quantifying the supplementary explanatory power.

$$\Lambda^j = 2\,[\,l\,(\widehat{\omega}_a) - l(\widehat{\omega}_0)]$$

where $l(.)$, is the log-likelihood. We define the enrichment score for gene set G_j:

212 $$ES = \sum_{j \in G_j} \Lambda^j$$

213 We calculate p-values for enrichment scores through nperm = 10,000 permutation sampling,
214 which provides more conservative estimates but is computationally intensive. Since gene set tests
215 share overlapping membership and exhibit interdependence, the presented p-values in this
216 context and subsequent tables have not been adjusted for multiple testing. Consequently, they
217 do not possess theoretical guarantees for controlling the False Discovery Rate (FDR). Once the
218 pathways are sorted by their significance or ES, identifying the shared pathways among the three
219 model runs becomes possible.

220 The model utilized in this study closely resembles that developed by Bryan et al [14], based on a
221 logistic model, whereas our model accommodates a continuous outcome variable.

222

Results

223 Using public resources from RNA-seq analyses of prefrontal cortex in ROSMAP participants, we
224 have identified pathways associated for Amyloid-β, tau tangles and MMSE, accounting for
225 covariates i.e. sex, genotype, education, age, and the interactions between RNA expression
226 levels and APOE genotype.

227 We have identified significant pathways when examining the outcome of Amyloid-β. We have
228 chosen a p-value threshold of 0.1 instead of the conventional 0.05 for hypothesis decision-making
229 in order to capture a broader range of potentially significant pathways. This slightly relaxed
230 threshold allows for a more inclusive analysis, potentially uncovering additional pathways that
231 may contribute to our understanding of the relationship between Amyloid-β and the identified
232 pathways. Table 3 and Figure 1 present the 14 significant pathways associated with Amyloid-β,
233 such as TRANSLOCASE ACTIVITY, C3HC4 TYPE RING FINGER DOMAIN BINDING, INSULIN

234 RECEPTOR BINDING, and VASCULAR ENDOTHELIAL GROWTH FACTOR RECEPTOR 2

235 BINDING.

Gene Pathways	p-value	ES
GOMF DNA TRANSLOCASE ACTIVITY	0.032	3.761
GOMF C3HC4 TYPE RING FINGER DOMAIN BINDING	0.040	3.047
GOMF VASCULAR ENDOTHELIAL GROWTH FACTOR RECEPTOR 2 BINDING	0.043	2.445
GOMF SUMO POLYMER BINDING	0.059	2.637
GOMF ATP DEPENDENT PROTEIN DISAGGREGASE ACTIVITY	0.064	2.468
GOMF ACETYLESTERASE ACTIVITY	0.067	2.315
GOMF DOPAMINE NEUROTRANSMITTER RECEPTOR ACTIVITY	0.082	2.383
GOMF BH DOMAIN BINDING	0.084	2.074
GOMF INSULIN RECEPTOR BINDING	0.086	2.217
GOMF SMALL RIBOSOMAL SUBUNIT RRNA BINDING	0.087	2.002
GOMF CLASS I DNA APURINIC OR APYRIMIDINIC SITE ENDONUCLEASE ACTIVITY	0.091	2.298
GOMF SHORT CHAIN CARBOXYLESTERASE ACTIVITY	0.094	2.001
GOMF VASCULAR ENDOTHELIAL GROWTH FACTOR RECEPTOR BINDING	0.098	1.787
GOMF ALPHA N ACETYLGALACTOSAMINIDE ALPHA 2 6 SIALYLTRANSFERASE ACTIVITY	0.099	1.869

236 *Table 3- Significant Pathways Associated with Amyloid-β Using a P-value Threshold of 0.1. This*

237 *table presents the pathways that show significance (p-values smaller than 0.1) in relation to*

238 *Amyloid-β,. The pathways listed provide insights into potential biological mechanisms and*

239 *molecular processes associated with Amyloid-β.*

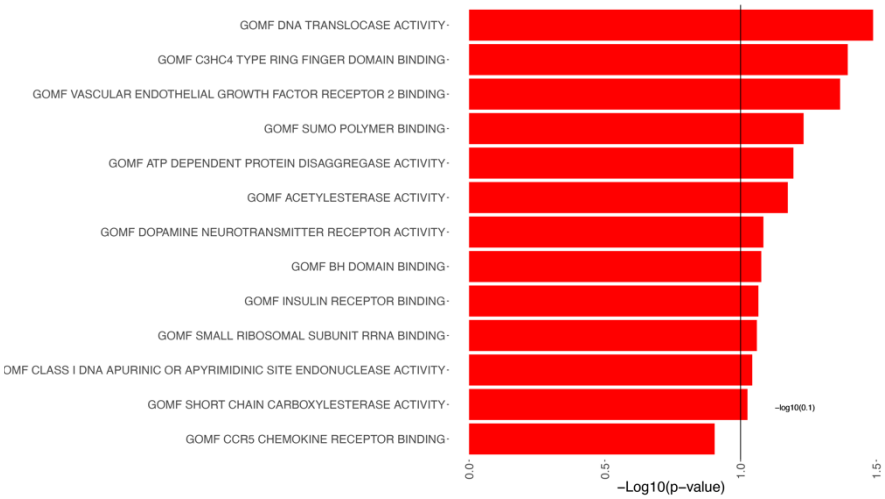

Figure 1-Pathway Analysis of Amyloid-β. Each pathway is represented on the y-axis, while the corresponding -log10(P-value) is represented by bars parallel to the x-axis, positioned in front of each pathway. The length of the bars reflects the statistical significance of the pathways with amyloid burden; longer bars indicating greater significance. This visualization allows for a quick assessment of the significance levels for each pathway, aiding in the identification of key pathways associated with Amyloid-β.

The 5 significant pathways associated with log(tangles) are presented in Table 4 and illustrated in Figure 2 such as BITTER TASTE RECEPTOR ACTIVITY, PROTEIN GLUTAMINE GAMMA GLUTAMYLTRANSFERASE ACTIVITY, TASTE RECEPTOR ACTIVITY, and VASCULAR ENDOTHELIAL GROWTH FACTOR RECEPTOR 2 BINDING.

Gene Pathways	p-value	ES
GOMF BITTER TASTE RECEPTOR ACTIVITY	0.035	3.547
GOMF PROTEIN GLUTAMINE GAMMA GLUTAMYLTRANSFERASE ACTIVITY	0.055	2.658
GOMF TASTE RECEPTOR ACTIVITY	0.066	2.486
GOMF VASCULAR ENDOTHELIAL GROWTH FACTOR RECEPTOR 2 BINDING	0.097	1.910
GOMF HYALURONONGLUCOSAMINIDASE ACTIVITY	0.098	2.074

Table 4- Significant Pathways Associated with log(tangles). This table shows the pathways that exhibited statistical significance (p-values smaller than 0.1) in relation to log(tangles).

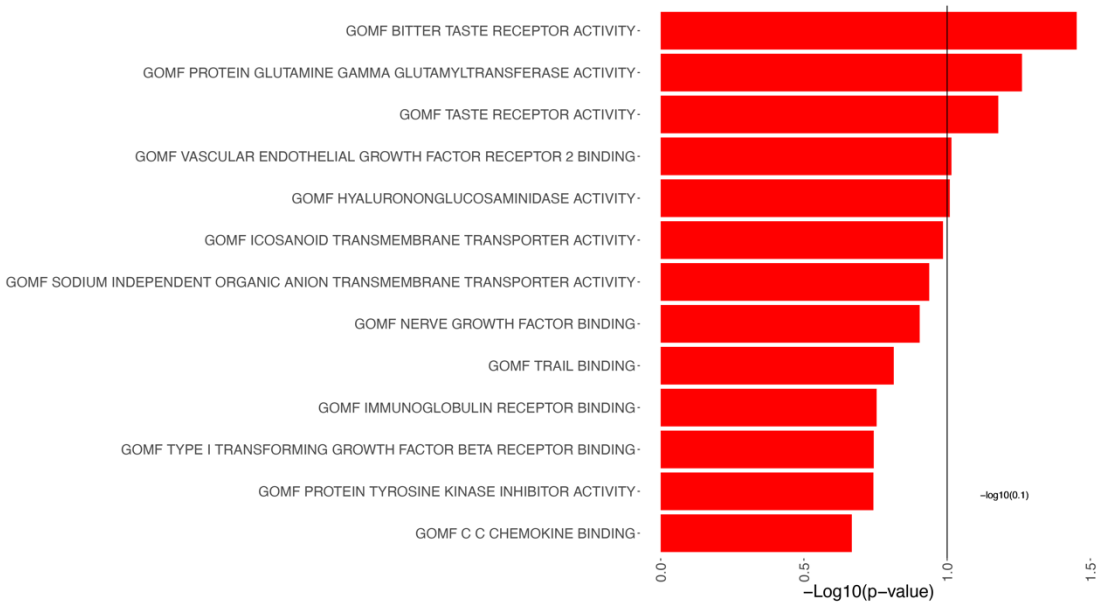

Figure 2- Pathway Analysis of log(tangles). This figure presents the results of the pathway significance analysis for log(tangles). Each pathway is displayed on the y-axis, while bars parallel to the x-axis represent the corresponding -log10(P-value). The length of each bar reflects the statistical significance of the pathway association with log(tangles). Longer bars indicate greater significance/smaller p values.

In addition to the pathways depicted in Figure 3, Table 5 includes the top candidate of such pathways, which represent a subset of significant associations (with p-values smaller than 0.1) with the Mini-Mental State Examination (MMSE) outcome. These results highlight specific pathways that exhibit statistical significance in association with MMSE. Figure 3 provides a partial representation of 174 top pathways, including TRAIL BINDING, VASCULAR ENDOTHELIAL GROWTH FACTOR RECEPTOR BINDING, VASCULAR ENDOTHELIAL GROWTH FACTOR RECEPTOR 2 BINDING.

Gene Pathways	p-value	ES
GOMF TRAIL BINDING	0.0010	7.071
GOMF VASCULAR ENDOTHELIAL GROWTH FACTOR RECEPTOR BINDING	0.0018	4.114
GOMF VASCULAR ENDOTHELIAL GROWTH FACTOR RECEPTOR 2 BINDING	0.0021	4.747
GOMF DEATH RECEPTOR ACTIVITY	0.0023	4.800
GOMF MECHANOSENSITIVE MONOATOMIC ION CHANNEL ACTIVITY	0.0027	4.061
GOMF HEXOKINASE ACTIVITY	0.0031	4.691
GOMF TUMOR NECROSIS FACTOR BINDING	0.0032	5.289
GOMF CC CHEMOKINE BINDING	0.0049	4.886
GOMF NERVE GROWTH FACTOR BINDING	0.0049	4.603
GOMF RAGE RECEPTOR BINDING	0.0060	3.650

Table 5- Significant Gene Pathways Associated with MMSE Outcome. It highlights the top candidate pathways, including TRAIL BINDING and VASCULAR ENDOTHELIAL GROWTH FACTOR RECEPTOR BINDING, among others.

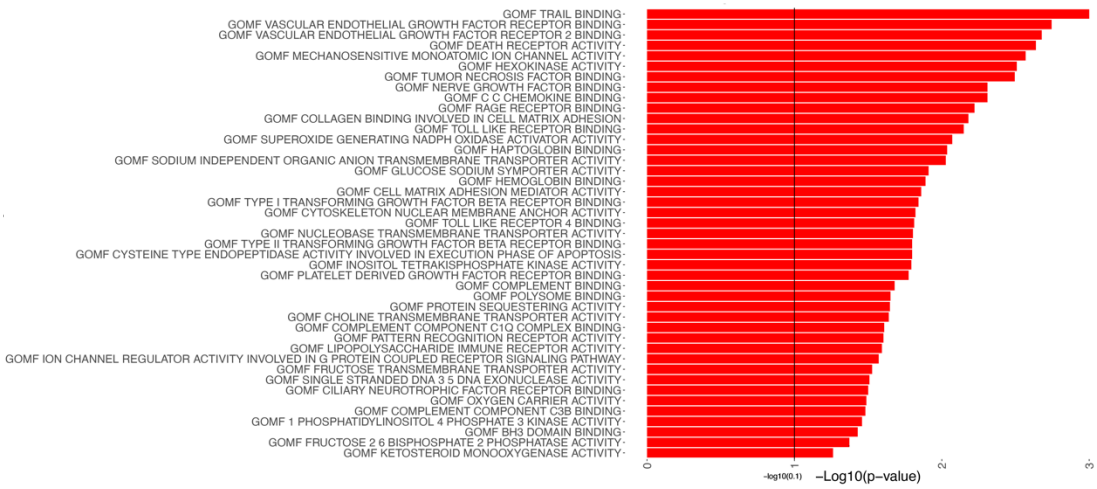

Figure 3- Pathway Analysis of MMSE. This figure shows a portion of the pathway significance analysis results for MMSE. Each pathway is plotted on the y-axis, accompanied by bars parallel to the x-axis that represent the corresponding -log10(P-value). The length of each bar reflects the statistical significance of the pathway, with longer bars indicating higher significance. This visual representation enables a quick evaluation of the significance levels associated with each pathway. It assists in identifying key pathways that are linked to MMSE, contributing to a better understanding of the underlying mechanisms influencing MMSE.

Finally, we investigated the shared pathways among all 3 studies involving Amyloid-β, log(tangles), and MMSE, shown in Figure 4. This analysis allows us to gain insights into the shared biological mechanisms and molecular processes that may contribute to the interplay between Amyloid-β, log(tangles), and MMSE.

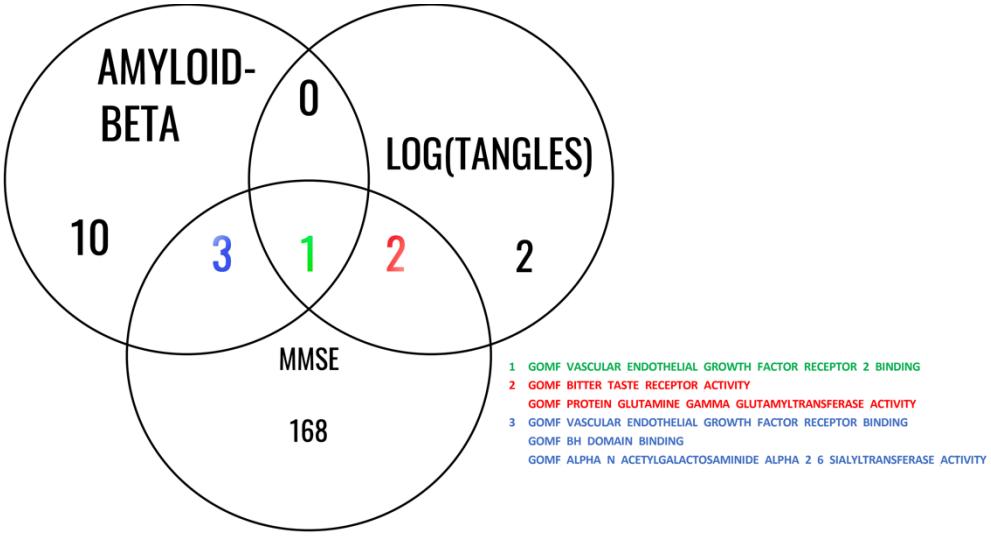

Figure 4- Common Significant Pathways Between for Amyloid-β, log(tangles), and MMSE. This Venn diagram illustrates the common pathways identified among the studies involving Amyloid-β, log(tangles), and MMSE. The diagram consists of overlapping circles that represent each study, with labeled sections indicating the shared pathways among them. The shared pathways are listed within the diagram, providing a concise overview of the biological processes and molecular mechanisms that are consistently implicated across these phenotypes. This analysis highlights the interconnectedness of these factors and underlying pathways that contribute to the associations between Amyloid-β, log(tangles), and MMSE.

Results for the comparisons of expression differences between control and AD samples are shown in Table 6 as reported in the AMP-AD analysis of the dataset (https://www.synapse.org/#!Synapse:syn8456721). Higher expression in AD samples relative to

296 normal cognition is observed for VEGFA, VEGFB, VEGFD, PGF and FLT1 while lower expression
297 in AD samples relative to controls is observed for FLT4, KDR, NRP1 and NRP2. The only
298 significant difference between the two groups after adjusting for multiple comparisons is for
299 VEGFB.

Gene	Comparison	logFC	CI.L	CI.R	AveExpr	t	P.Value	adj.P.Val
VEGFA	AD-CONTROL	0.16	-0.08	0.40	1.86	1.28	0.20	0.43
VEGFB	AD-CONTROL	0.10	0.03	0.16	5.67	2.81	0.01	0.05
VEGFD	AD-CONTROL	0.02	-0.07	0.12	1.10	0.50	0.62	0.79
PGF	AD-CONTROL	0.15	-0.02	0.32	0.17	1.69	0.09	0.26
FLT1	AD-CONTROL	0.14	-0.05	0.32	3.86	1.44	0.15	0.36
FLT4	AD-CONTROL	-0.03	-0.15	0.09	0.15	-0.54	0.59	0.78
KDR	AD-CONTROL	-0.07	-0.20	0.07	1.05	-1.00	0.32	0.55
NRP1	AD-CONTROL	-0.07	-0.18	0.03	2.14	-1.41	0.16	0.37
NRP2	AD-CONTROL	-0.02	-0.10	0.06	1.37	-0.58	0.56	0.75

300 *Table 6- Differential Expression of VEGF family related Genes in Alzheimer's Disease (AD)*
301 *Samples Compared to Controls. This table presents the results of a comparative analysis*
302 *between control and Alzheimer's disease (AD) samples, focusing on the expression differences*
303 *of VEGF-family--related genes.*

Discussion

304 The progression of Alzheimer's disease (AD) manifests through changes in biomarkers that reflect
305 abnormal protein expression, such as Amyloid-β, phosphorylated tau, as well as clinically
306 measurable symptoms including memory decline. In our study, we developed new methods to
307 reveal pathways related to changes in brain RNA-Seq for each of these two neuropathological
308 biomarkers, as well as for MMSE. Surprisingly we observed a larger number of significant
309 pathways for gene expression association with MMSE (174 pathways) than those associated with
310 the two hallmark biomarkers for AD, Amyloid-β (14 pathways), and tau tangles (5 pathways). More
311 importantly, we identified a single pathway, vascular endothelial grown factor receptor binding

312 (VEGF-RB) that was associated with differences reflective of all three phenotypes analyzed by
313 the continuous gene set likelihood ratio test (gsLRT).

314 The methodological advance of this study was the extension of gsLRT from binary (e.g.
315 disease/control status) to continuous value phenotypes. Our study extends prior work by
316 analyzing three continuous scale phenotypes (Amyloid-β , tau tangles, MMSE) in a pathway
317 analysis that accounts for the covariates age, sex and APOE genotype (gsLRT for continuous
318 phenotypes).

319 Among the pathways identified as significant for Amyloid-β we noted several candidate pathways
320 that support changes in DNA repair ability (GOMF DNA TRANSLOCASE ACTIVITY) [23], cell
321 mediated immunity (GOMF C3HC4 TYPE RING FINGER DOMAIN BINDING) [24], apoptosis
322 (GOMF BH DOMAIN BINDING) [25], protein synthesis (GOMF SMALL RIBOSOMAL SUBUNIT
323 RRNA BINDING) [26] and disaggregation (GOMF ATP DEPENDENT PROTEIN
324 DISAGGREGASE ACTIVITY), as well as insulin signaling (GOMF INSULIN RECEPTOR
325 BINDING) [27, 28], which have all been connected to AD.

326 Among the pathways identified as significant for tau tangles we noted two that support alterations
327 in sensory processing, pointing to taste (GOMF BITTER TASTE RECEPTOR ACTIVITY; GOMF
328 TASTE RECEPTOR ACTIVITY). Recent studies have shown alterations in the sour taste [29],
329 while here we identified changes in pathways associated with bitter taste, and taste in general.
330 Since taste and olfaction are closely linked, these results suggest possible changes in olfactory
331 function. Among sensory changes in AD, olfaction has been proposed as one of the more
332 promising biomarkers for early detection [30]. We also noted pathways pointing to differences in
333 the transfer of amino acids across the membrane, cell survival during oxidative stress (GOMF
334 PROTEIN GLUTAMINE GAMMA GLUTAMYLTRANSFERASE ACTIVITY) [31], and glutathione
335 homeostasis, relevant to several neurodegenerative diseases, such as AD, PD and ALS [31, 32].
336 Finally, the role of extracellular matrix was suggested by the presence of the GOMF

337 HYALURONONGLUCOSAMINIDASE ACTIVITY pathway, which has a less understood but
338 complex role in aging and disease [33, 34].

339 Among the large number of pathways related to MMSE, we noted top candidates involved in
340 apoptosis through TRAIL which binds to death receptors, suggesting a relation with immune
341 related mechanisms [35].

342 Amyloid-β and tau biomarkers in combination with measures of neurodegeneration have been
343 associated with the progression of AD neuropathology and memory loss [36-41]. We note that
344 neuropsychological measures such as MMSE can reflect other causes of cognitive impairment
345 including damage to the cerebrovascular system although individuals with moderate to severe
346 Alzheimer's disease tend to have MMSE scores less than 15 [42].

347 The role of the VEGF signaling family in neurodegeneration and Alzheimer's disease has been
348 extensively studied including with multiomic approaches that analyzed bulk RNA sequencing
349 data, single nucleus sequencing data, and mass spectrometry proteomics data [10]. As a
350 consequence of the numerous pathways that intersect with VEGF receptor signaling, it has been
351 difficult to identify the specific receptors and molecules that associate with disease
352 endophenotypes or covariates including age and sex.

353 VEGF includes a family of five ligands (VEGFA, VEGFB, VEGFC, VEGFD, and PGF), three
354 tyrosine kinase receptors (FLT1, FLT4, and KDR), and two modulating receptors (NRP1 and
355 NRP2). The pathway identified as significantly associated with the three phenotypes (GOMF
356 VASCULAR ENDOTHELIAL GROWTH FACTOR RECEPTOR BINDING (GO:0005172) includes
357 the ligands and several other genes. The pathway GOMF VASCULAR ENDOTHELIAL
358 GROWTH FACTOR RECEPTOR ACTIVITY (GO:0005021) includes the tyrosine kinase
359 receptors and modulating receptors).

360 The presence of VEGF alongside Amyloid-β plaques in AD brains and its strong binding to
361 Amyloid-β suggest that VEGF may contribute to neurodegeneration and vascular dysfunction
362 [43]. Additionally, Amyloid-β inhibits VEGF receptor signaling, impairing angiogenesis [44]. VEGF
363 accumulation around amyloid plaques interacts directly with Amyloid-β, rescuing synaptic
364 dysfunction caused by the toxic Amyloid-β oligomers [45].

365 VEGF also interacted with tau and amyloid-β_{42}, predicting hippocampal atrophy and memory
366 decline. Another study [46] revealed that VEGF genes, particularly FLT4 and FLT1, were
367 associated with AD neuropathology and cognition. Higher levels of VEGF were associated with
368 slower hippocampal atrophy and better cognitive function [47]. These findings emphasize the
369 importance of understanding the relationship between VEGF, tau pathology, and AD for potential
370 therapeutic interventions.

371 Several studies have investigated the role of vascular endothelial growth factor (VEGF) in
372 cognitive impairment. For example one study found that VEGF AA genotype is associated with
373 an increased risk of developing AD and MCI, while higher VEGF levels are observed in AD
374 patients [48]. Another study found lower VEGF levels in AD patients and amnestic MCI patients
375 compared to controls, correlating with cognitive decline [49]. Additionally, [50] finds that higher
376 serum VEGF levels in ischemic stroke patients are associated with post-stroke cognitive
377 impairment. Conversely, [51] shows that VEGF signaling is crucial for maintaining cognition and
378 neurogenesis, cautioning against inhibiting VEGF signaling. Interestingly, [52] demonstrates that
379 VEGF levels increase during the early stage of AD but decrease as the disease progresses,
380 suggesting a link between VEGF levels and cognitive decline. VEGF produced by macrophages
381 plays a role in preserving cognitive function in obesity, which can be a risk factor during aging and
382 AD [53] However the literature is still controversial and more work is needed to understand the
383 role of various VEGF isoforms role in modulating cognition [54] .

384 The ensemble of molecules in the VEGF pathways and their interactions have been reported to
385 have varied effects on AD phenotypes. The members of the VEGF measured in the brain and
386 blood have been characterized with respect to cognitive performance, neural and cerebrovascular
387 pathology, and CSF biomarkers [11, 55]. Blood and brain VEGFA has been reported to be
388 protective against memory impairment and brain atrophy in AD [10]. Higher expression levels of
389 VEGFB, PGF, FLT1 and FLT4 are reported to be associated with faster cognitive decline and
390 greater neuropathological lesion development [10, 55] and our data supported this direction of
391 difference in expression levels. The VEGF family is involved in multiple signaling pathways,
392 leading to potentially different effects on AD-related phenotypes; for example, VEGFA can signal
393 through KDR or FLT1 where the receptors can elicit effects in opposing directions [10]. In a study
394 of microglial control of astrocytes in response to microbial metabolites, microglial VEGFB was
395 shown to trigger FLT1 signaling in astrocytes and promote CNS inflammation [56]. Neutropilin
396 expression (NRP1 and NRP2) was decreased AD samples relative to samples from individuals
397 with normal cognition in agreement with prior evidence [10]. NRP1 and NRP2 have well-
398 established roles in angiogenesis[10, 57]. Moreover, interactions between VEGF family proteins
399 with APOE have also been reported. Higher levels of VEGFA were reported to be associated
400 with worse outcomes among APOE E4 carriers and better outcomes among non-E4 carriers [10].

401 Our study has several strengths. First, the extension of the gsLRT to continuous phenotypes
402 allowed a higher statistical level (continuous vs. nominal) of data for the pathway analysis. The
403 relatively large sample size where data was available for all three phenotypes allowed for
404 comparisons at the gene and pathway level. For the pathway/signature analysis, well-established
405 databases including gene ontology were used to enable replication studies and other future work.

406 A limitation of the analysis is that the phenotypes are measured at a single time point.
407 Longitudinal data for the Amyloid-β and tau biomarkers are not available from post-mortem brains,
408 though nuclear imaging or fluid biomarkers present a great promise for the future of such

409 longitudinal studies. MMSE is measured over time, so another possible phenotype to consider
410 would be the decline in MMSE from baseline to death. Studies that examine changes in CSF and
411 blood biomarkers for VEGF over time will provide information on temporal relationships between
412 VEGF family mRNA and protein concentrations and biomarker changes and can open new
413 avenues for exploiting its therapeutic potential [10].

414 Our study proposes a method for pathway identification using continuous phenotypes and public
415 data bases on RNA-Seq, but can potentially be used for proteomic analyses, or extended to
416 multinetwork omic studies and make use of extensive public data base resources or de novo
417 analyses to better understand the mechanistic substrates for neurodegenerative diseases.

418 ***Conflict of Interest***

419 The authors declare that the research was conducted in the absence of any commercial or
420 financial relationships that could be construed as a potential conflict of interest.

421 ***Funding***

422 This work was supported by RF1 AG057895, R01 AG066184, U24 CA220245, RF1 AG070149,
423 P30 AG072958.

Acknowledgment

424 We thank all the participants of ROS and MAP studies. These studies were funded by the National
425 Institute on Aging: P30AG010161 ADCC, P30AG072975 ADRC, R01AG015819 RISK,
426 R01AG017917 MAP, U01AG46152 AMP-AD Pipeline I, U01AG61356 AMP-AD Pipeline II.
427 Furthermore, we are grateful to Dr. Jordan Bryan for helpful discussions.

428

429

References

430 1. Association, A.s., *2019 Alzheimer's disease facts and figures.* Alzheimer's & dementia,
431 2019. **15**(3): p. 321-387.
432 2. Seyfried, N.T., et al., *A Multi-network Approach Identifies Protein-Specific Co-expression*
433 *in Asymptomatic and Symptomatic Alzheimer's Disease.* Cell Systems, 2017. **4**(1): p. 60-
434 72.e4.
435 3. Huang, Z., et al., *Brain proteomic analysis implicates actin filament processes and injury*
436 *response in resilience to Alzheimer's disease.* Nature Communications, 2023. **14**(1).
437 4. McGeer, E.G. and P.L. McGeer, *The importance of inflammatory mechanisms in*
438 *Alzheimer disease.* Experimental Gerontology, 1998. **33**(5): p. 371-378.
439 5. Ofengeim, D., et al., *RIPK1 mediates a disease-associated microglial response in*
440 *Alzheimer's disease.* Proceedings of the National Academy of Sciences of the United
441 States of America, 2017. **114**(41): p. E8788-E8797.
442 6. Sweeney, M.D., S. Ayyadurai, and B.V. Zlokovic, *Pericytes of the neurovascular unit:*
443 *Key functions and signaling pathways.* Nature Neuroscience, 2016. **19**(6): p. 771-783.
444 7. Zhang, Y., et al., *Identification of mitochondrial related signature associated with immune*
445 *microenvironment in Alzheimer's disease.* Journal of Translational Medicine, 2023.
446 **21**(1).
447 8. Liu, Q.Y., et al., *A novel brain-enriched E3 ubiquitin ligase RNF182 is up regulated in the*
448 *brains of Alzheimer's patients and targets ATP6V0C for degradation.* Molecular
449 neurodegeneration, 2008. **3**: p. 1-16.
450 9. Sakai, M., et al., *Decline of gustatory sensitivity with the progression of Alzheimer's*
451 *disease.* International Psychogeriatrics, 2016. **28**(3): p. 511-517.

452 10. Seto, M., et al., *Multi-omic characterization of brain changes in the vascular endothelial*
453 *growth factor family during aging and Alzheimer's disease.* Neurobiol Aging, 2023. **126**:
454 p. 25-33.
455 11. Libby, J.B., et al., *Whole blood transcript and protein abundance of the vascular*
456 *endothelial growth factor family relate to cognitive performance.* Neurobiol Aging, 2023.
457 **124**: p. 11-17.
458 12. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for*
459 *interpreting genome-wide expression profiles.* Proceedings of the National Academy of
460 Sciences, 2005. **102**(43): p. 15545-15550.
461 13. Tamayo, P., et al., *The limitations of simple gene set enrichment analysis assuming*
462 *gene independence.* Stat Methods Med Res, 2016. **25**(1): p. 472-87.
463 14. Bryan, J., et al., *Likelihood ratio statistics for gene set enrichment in Alzheimer's disease*
464 *pathways.* Alzheimer's & Dementia, 2021. **17**(4): p. 561-573.
465 15. Bennett, D.A., et al., *Overview and findings from the Religious Orders Study.* Curr
466 Alzheimer Res, 2012. **9**(6): p. 628-45.
467 16. Bennett, D.A., et al., *Overview and findings from the Rush Memory and Aging Project.*
468 Curr Alzheimer Res, 2012. **9**(6): p. 646-63.
469 17. De Jager, P.L., et al., *Alzheimer's disease: early alterations in brain DNA methylation at*
470 *ANK1, BIN1, RHBDF2 and other loci.* Nat Neurosci, 2014. **17**(9): p. 1156-63.
471 18. Bennett, D.A., et al., *Religious orders study and rush memory and aging project.* Journal
472 of Alzheimer's disease, 2018. **64**(s1): p. S161-S189.
473 19. De Jager, P.L., et al., *A multi-omic atlas of the human frontal cortex for aging and*
474 *Alzheimer's disease research.* Sci Data, 2018. **5**: p. 180142.
475 20. Wan, Y.W., et al., *Meta-Analysis of the Alzheimer's Disease Human Brain Transcriptome*
476 *and Functional Dissection in Mouse Models.* Cell Rep, 2020. **32**(2): p. 107908.

477 21. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1):
478 p. 15-21.
479 22. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing*
480 *and microarray studies.* Nucleic Acids Res, 2015. **43**(7): p. e47.
481 23. Ciccia, A. and S.J. Elledge, *The DNA damage response: making it safe to play with*
482 *knives.* Mol Cell, 2010. **40**(2): p. 179-204.
483 24. Lundegaard, C., et al., *Major histocompatibility complex class I binding predictions as a*
484 *tool in epitope discovery.* Immunology, 2010. **130**(3): p. 309-18.
485 25. Yuan, J. and B.A. Yankner, *Apoptosis in the nervous system.* Nature, 2000. **407**(6805):
486 p. 802-809.
487 26. Hernández-Ortega, K., et al., *Altered Machinery of Protein Synthesis in Alzheimer's:*
488 *From the Nucleolus to the Ribosome.* Brain Pathol, 2016. **26**(5): p. 593-605.
489 27. Xie, L., et al., *Alzheimer's beta-amyloid peptides compete for insulin binding to the*
490 *insulin receptor.* J Neurosci, 2002. **22**(10): p. RC221.
491 28. Matioli, M. and R. Nitrini, *Mechanisms linking brain insulin resistance to Alzheimer's*
492 *disease.* Dement Neuropsychol, 2015. **9**(2): p. 96-102.
493 29. Ferraris, C.A.-O., et al., *Association between Sour Taste SNP KCNJ2-rs236514, Diet*
494 *Quality and Mild Cognitive Impairment in an Elderly Cohort. LID - 10.3390/nu13030719*
495 *[doi] LID - 719.* (2072-6643 (Electronic)).
496 30. Murphy, C., *Olfactory and other sensory impairments in Alzheimer disease.* Nature
497 Reviews Neurology, 2019. **15**(1): p. 11-24.
498 31. Zhang, H., H.J. Forman, and J. Choi, *Gamma-glutamyl transpeptidase in glutathione*
499 *biosynthesis.* Methods Enzymol, 2005. **401**: p. 468-83.
500 32. Fujii, J., et al. *Critical Roles of the Cysteine–Glutathione Axis in the Production of*
501 *γ-Glutamyl Peptides in the Nervous System*. International Journal of Molecular
502 Sciences, 2023. **24**, DOI: 10.3390/ijms24098044.

503 33. Zakusilo, F.T., et al., *Matters of size: Roles of hyaluronan in CNS aging and disease.*
504 Ageing Research Reviews, 2021. **72**: p. 101485.
505 34. Sun, Y., et al., *Role of the Extracellular Matrix in Alzheimer's Disease.* Frontiers in Aging
506 Neuroscience, 2021. **13**.
507 35. Pimentel, J.M., J.Y. Zhou, and G.S. Wu, *The Role of TRAIL in Apoptosis and*
508 *Immunosurveillance in Cancer.* Cancers (Basel), 2023. **15**(10).
509 36. Dubois, B., et al., *Preclinical Alzheimer's disease: Definition, natural history, and*
510 *diagnostic criteria.* Alzheimers Dement, 2016. **12**(3): p. 292-323.
511 37. Jack, C.R., Jr., et al., *NIA-AA Research Framework: Toward a biological definition of*
512 *Alzheimer's disease.* Alzheimers Dement, 2018. **14**(4): p. 535-562.
513 38. Jack, C.R., Jr., et al., *A/T/N: An unbiased descriptive classification scheme for Alzheimer*
514 *disease biomarkers.* Neurology, 2016. **87**(5): p. 539-47.
515 39. Jack, C.R., Jr., et al., *Tracking pathophysiological processes in Alzheimer's disease: an*
516 *updated hypothetical model of dynamic biomarkers.* Lancet Neurol, 2013. **12**(2): p. 207-
517 16.
518 40. Jack, C.R., Jr., et al., *Hypothetical model of dynamic biomarkers of the Alzheimer's*
519 *pathological cascade.* Lancet Neurol, 2010. **9**(1): p. 119-28.
520 41. Knopman, D.S., et al., *The National Institute on Aging and the Alzheimer's Association*
521 *Research Framework for Alzheimer's disease: Perspectives from the Research*
522 *Roundtable.* Alzheimers Dement, 2018. **14**(4): p. 563-575.
523 42. Wimo, A., et al., *The GERAS Study: a prospective observational study of costs and*
524 *resource use in community dwellers with Alzheimer's disease in three European*
525 *countries--study design and baseline findings.* J Alzheimers Dis, 2013. **36**(2): p. 385-99.
526 43. Yang, S., *Bae DG, Kang HJ, Gwag BJ, Gho YS, Chae CB.* Co-accumulation of vascular
527 endothelial growth factor with beta-amyloid in the brain of patients with Alzheimer's
528 disease. Neurobiol Aging, 2004. **25**: p. 283-290.

529 44. Patel, N.S., et al., *Alzheimer's β-amyloid peptide blocks vascular endothelial growth*
530 *factor mediated signaling via direct interaction with VEGFR-2.* Journal of
531 neurochemistry, 2010. **112**(1): p. 66-76.
532 45. Martin, L., et al., *VEGF counteracts amyloid-β-induced synaptic dysfunction.* Cell
533 reports, 2021. **35**(6).
534 46. Mahoney, E.R., et al., *Brain expression of the vascular endothelial growth factor gene*
535 *family in cognitive aging and alzheimer's disease.* Molecular psychiatry, 2021. **26**(3): p.
536 888-896.
537 47. Hohman, T.J., et al., *The role of vascular endothelial growth factor in neurodegeneration*
538 *and cognitive decline: exploring interactions with biomarkers of Alzheimer disease.*
539 JAMA neurology, 2015. **72**(5): p. 520-529.
540 48. Chiappelli, M., et al., *VEGF gene and phenotype relation with Alzheimer's disease and*
541 *mild cognitive impairment.* Rejuvenation research, 2006. **9**(4): p. 485-493.
542 49. Huang, L., J. Jia, and R. Liu, *Decreased serum levels of the angiogenic factors VEGF*
543 *and TGF-β1 in Alzheimer's disease and amnestic mild cognitive impairment.*
544 Neuroscience letters, 2013. **550**: p. 60-63.
545 50. Prodjohardjono, A., et al., *Higher level of acute serum VEGF and larger infarct volume*
546 *are more frequently associated with post-stroke cognitive impairment.* PloS one, 2020.
547 **15**(10): p. e0239370.
548 51. Koester-Hegmann, C., et al., *High-altitude cognitive impairment is prevented by enriched*
549 *environment including exercise via VEGF signaling.* Frontiers in cellular neuroscience,
550 2019. **12**: p. 532.
551 52. Tang, H., et al., *Expression level of vascular endothelial growth factor in hippocampus is*
552 *associated with cognitive impairment in patients with Alzheimer's disease.* Neurobiology
553 of aging, 2013. **34**(5): p. 1412-1415.

554 53. Jais, A., et al., *Myeloid-cell-derived VEGF maintains brain glucose uptake and limits*
555 *cognitive impairment in obesity.* Cell, 2016. **165**(4): p. 882-895.
556 54. Ng, T., et al., *Vascular endothelial growth factor inhibitors and cognitive impairment:*
557 *evidence and controversies.* Expert opinion on drug safety, 2014. **13**(1): p. 83-92.
558 55. Mahoney, E.R., et al., *Brain expression of the vascular endothelial growth factor gene*
559 *family in cognitive aging and alzheimer's disease.* Mol Psychiatry, 2021. **26**(3): p. 888-
560 896.
561 56. Rothhammer, V., et al., *Microglial control of astrocytes in response to microbial*
562 *metabolites.* Nature, 2018. **557**(7707): p. 724-728.
563 57. Lampropoulou, A. and C. Ruhrberg, *Neuropilin regulation of angiogenesis.* Biochem Soc
564 Trans, 2014. **42**(6): p. 1623-8.