

# Detecting Discontinuities in the Topology of Alzheimer's gene Co-expression

A. Samadzelkava<sup>1,2</sup>

**1** CUNY Graduate Center

**2** CCNY

nsamadzelkava@gradcenter.cuny.edu

## Abstract

Alzheimer's disease (AD) emerges from a complex interplay of molecular, cellular, and network-level disturbances that are not easily captured by traditional reductionist frameworks. Conventional analyses of gene expression often rely on thresholded correlation networks or clustering-based module detection, approaches that may obscure nonlinear structure and higher-order organization. Here, we introduce a comparative topological framework that makes use of topological data analysis (TDA) and Mapper algorithm to detect discontinuities - localized disruptions in the topology of gene co-expression space between healthy and AD brain tissue. Using gene expression data from 3 brain regions, we mapped how AD reshapes the global topology of gene–gene relationships. Discontinuity hotspots were identified via variability-based node scoring and subjected to GO Biological Process enrichment analysis. This work illustrates the potential of TDA to uncover disease-relevant structure in high-dimensional transcriptomic data and motivates broader application of shape-based comparative methods in neurodegeneration research and other areas that benefit from comparative analysis.

## Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative disorder characterized by cognitive decline, memory loss, and behavioral changes. It is the most common cause of dementia, accounting for 60–80% of all dementia cases worldwide. The prevalence of AD is rapidly increasing, emphasizing the need for effective diagnostic and therapeutic strategies. Particularly in its late-onset form (LOAD), Alzheimer's disease epitomizes the complexity of human diseases, arising from a non-linear interplay between genetic variants, environmental factors, and molecular networks. Despite decades of intensive research and numerous identified risk factors, the causal mechanisms of LOAD remain elusive, and the quest for effective therapies has been fraught with challenges. Traditional focus on specific pathological features, such as amyloid-beta plaques and tau protein tangles, has not yielded the desired breakthroughs [1, 2, 3]. This highlights the need for a more holistic approach that embraces the multifaceted nature of the disease [4, 5].

Recent efforts to tackle the complexity of Alzheimer's disease include system-wide approaches aiming to pinpoint disruptions in the networks underlying the disease [6]. Gene coexpression networks, also known as covariance or correlation networks, present a successful translation of gene expression profiles into the language of network theory [7, 8]. Comparative analysis of the resultant networks between healthy and diseased conditions has been powerful in determining functional modules—groups of genes with similar expression patterns connected into functional complexes or pathways involved in regulatory or signaling circuits [7, 9, 10, 11, 6]. However, these traditional methods have limitations. Firstly, they often rely on arbitrary thresholding of the coexpression matrix to define network edges, which may exclude potentially important features of the network. Secondly, clustering algorithms used to identify functional modules are essentially point approximations, which may overlook significant features of the data, especially in the presence of complex, non-linear relationships. Furthermore, biological networks, including gene expression networks, exhibit complex structures at various scales, making them notoriously difficult to analyze using traditional clustering methods. The complexity of molecular landscape of LOAD demands analytical tools that can model the high-dimensional and complex relationships inherent in gene expression data [11].

## Topological Data Analysis

Non-linear, data-driven methods are essential for uncovering hidden structures and patterns that linear, assumption-based approaches may miss [12]. They can capture subtle variations and non-linear interactions

critical for understanding disease mechanisms, providing a more accurate representation of the biological processes involved [13].

Topological Data Analysis (TDA) emerges as a promising candidate, offering a novel perspective that robustly addresses the limitations of traditional network-based approaches [14, 15, 16]. Grounded in topology, TDA extracts underlying structure of complex datasets, including high-dimensional gene expression data [17]. Non-linear dimensionality reduction techniques, along with algorithms like Mapper, allow the extraction of complex structures and the overall shape of data without the need for simplification or assumptions about the data’s shape. By focusing on qualitative topological features, TDA provides a framework for exploring the topology of gene coexpression networks, revealing patterns that are missed by traditional statistical methods [18, 12]. This approach has the potential to uncover new insights into the dynamic interplay of genes and pathways involved in AD.

In this project we investigated misaligned genes associated with AD highlighted by leveraging TDA’s Mapper. We then extend the analysis to detect discontinuities that emerge in comparative analysis. Discontinuities that emerge during the analysis are caused by “misaligned” genes — not only those whose expression patterns deviate significantly from those in the healthy state but those who alter networks topology. Identifying such genes could provide valuable information on gene interactions and pinpoint intervention opportunities to restore the network to a healthy state. Ultimately, this topological approach aims to improve our understanding of the molecular mechanisms underlying Alzheimer’s disease and contribute to the development of novel therapeutic strategies.

## Results

We applied the pipeline described below to analyze gene co-expression patterns across three brain regions—the dorsolateral prefrontal cortex (DLPFC), visual cortex (VC), and cerebellum—in healthy and Alzheimer’s disease (AD) conditions. The primary objective was to highlight discontinuities in expression patterns and identify potentially “misplaced” genes whose connectivity diverges from healthy co-expression structure.

Our analysis is grounded in co-expression matrices, which capture the pairwise correlation of gene expression levels. For each condition (healthy and AD) and brain region, we constructed a co-expression matrix.

The healthy co-expression network serves as a reference to define a notion of continuity in gene expression space. Mapper graphs constructed from this network were colored using a continuous color gradient that reflects local similarity in gene co-expression profiles. This approach highlights smooth transitions in expression—genes with similar co-expression are assigned similar colors.

To assess divergence in the disease state, we applied the color scheme derived from the healthy condition to the corresponding AD co-expression graph. Color discontinuities in the Alzheimer’s graph thus point to disrupted co-expression structure and altered gene relationships.

## 1 Pipeline overview

### 1.1 Mapper Algorithm Overview

The Mapper algorithm is a TDA technique that captures the “shape” of complex, high-dimensional datasets in a simplified graph [14]. First, a “lens” (or filter) function is applied to map each data point to a lower-dimensional scalar or vector. This range is then partitioned into overlapping intervals (covers). Within each interval, local subsets of data points are clustered, forming nodes in the Mapper graph. Nodes that share points from overlapping intervals are connected by edges, revealing how local clusters relate globally. This process preserves local groupings while highlighting overall structural features (i.e., topology) of the dataset [15].

### 1.2 Color Continuity

Mapper naturally facilitates smooth coloring of its resulting graph because it segments the data into overlapping bins along the chosen lens function. When assigning colors (for instance, by enumerating nodes and mapping them to a gradient), this adjacency ensures that similar or connected clusters receive closely related hues, leading to gradual shifts in color. As a result, regions of the graph that represent smoothly transitioning structures in the underlying data also display smooth color transitions, reflecting the continuous nature of the data's topology (Figure 1A).

Additionally, we rotated the color gradient by 90 degrees to investigate genes that might be displaced in a direction perpendicular to the initial color gradient, and therefore not highlighted by the original coloring axis (Figure 2A).

### 1.3 Detecting Discontinuities

Once a reference Mapper graph (from healthy controls) was assigned color scheme, we transferred those colors to the Mapper graph constructed from the AD dataset (containing the same genes but with different expression levels). If the AD dataset's structure closely resembled the reference, the transferred colors appeared to flow smoothly across its nodes. However, when the AD dataset diverged from the reference, sharp color jumps—or discontinuities—emerged, indicating clusters or genes that shifted relative to their original placement. These discontinuities revealed where the AD topology deviated most from the baseline, highlighting potentially important structural differences (Figure 1B).

To further quantify regions of altered connectivity, we applied standard deviation as an aggregation function on the AD Mapper graph (Figure 1C). Nodes with high color variability, where constituent genes derive from significantly different or conflicting color groups of the reference, stood out as hotspots of disrupted continuity.

Building upon the identification of the top 20% most variable Mapper nodes we compiled lists of genes from these nodes. Genes within these highly variable clusters were considered prime candidates for functional disruption linked to the disease.

### 1.4 GO Enrichment Analysis

For each region, we performed Gene Ontology (GO) enrichment analysis on the genes from the top-20% variable clusters to determine which biological functions are most impacted under AD conditions. Enrichment was conducted using the Enrichr platform, focusing on GO-Biological Process terms. We applied a significance threshold of  $p < 0.05$  (after adjustment for multiple testing) and ranked GO terms by the combined score (CS) retaining only GO terms with  $CS < 30$ . To benchmark our topological approach against conven-

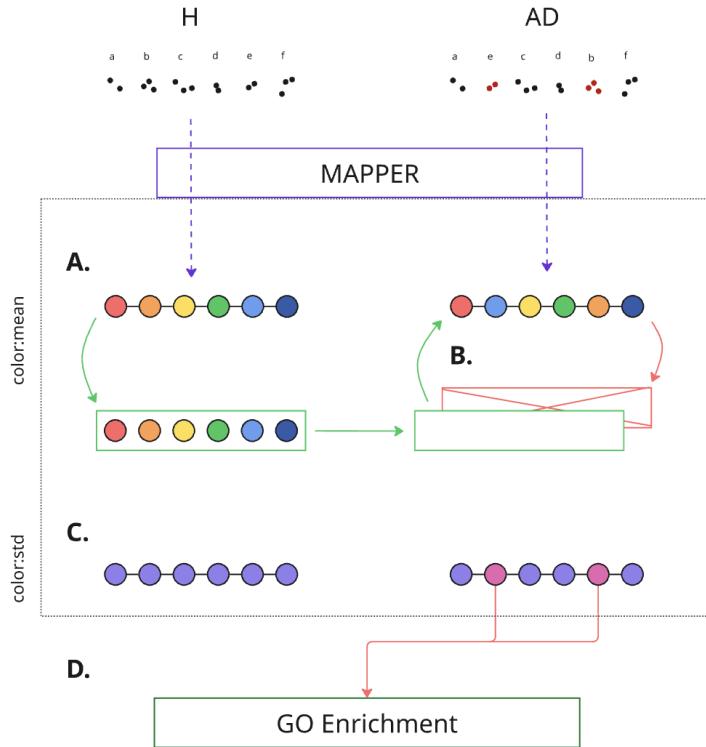


Figure 1: **A.** colors assigned to H. Mapper automatically assigns colors to AD but we discard those taking coloring of H as "ground truth". **B.** We use colors H to color dataset AD highlighting discontinuities of AD **C.** To highlight displaced genes in AD we apply "standard deviation" aggregation function to AD Mapper graph **D.** Genes that fall into highly variable (pink) mapper nodes are extracted for GO enrichment analysis

tional analyses, we also quantified differential gene expression and differential gene co-expression between control and AD samples (Figure5).

## 2 Findings

### 2.1 DLPFC

The DLPFC is known to be one of the most affected brain regions in Alzheimer's disease, and our analysis supports this. Thirteen GO terms showed a combined score higher than 50, with the top terms associated with vascular smooth muscle cell development and differentiation, positive regulation of CD8-positive  $\alpha\beta$  T cell differentiation, and protection from natural killer cell-mediated cytotoxicity. The enrichment of vascular-related pathways (e.g., vascular smooth muscle cell differentiation) aligns with previous research indicating vascular dysfunction as a critical factor in Alzheimer's pathology. Cerebrovascular abnormalities, including impaired vascular smooth muscle cell function, have been consistently associated with AD progression, emphasizing the role of compromised cerebral blood flow and vascular integrity in neurodegeneration [19, 20].

Additionally, immune response pathways, such as the regulation of T cell differentiation and protection from natural killer cell activity, underscore the emerging concept of immune dysregulation and inflammation in Alzheimer's pathology. Prior literature demonstrates significant T cell involvement and immune-mediated damage contributing to AD progression [21, 22]. Interestingly, several novel findings emerged from this analysis, including GO terms like positive regulation of isomerase activity and regulation of glial cell proliferation. These processes have less established roles in AD and may represent potential novel targets for therapeutic exploration or further mechanistic studies.

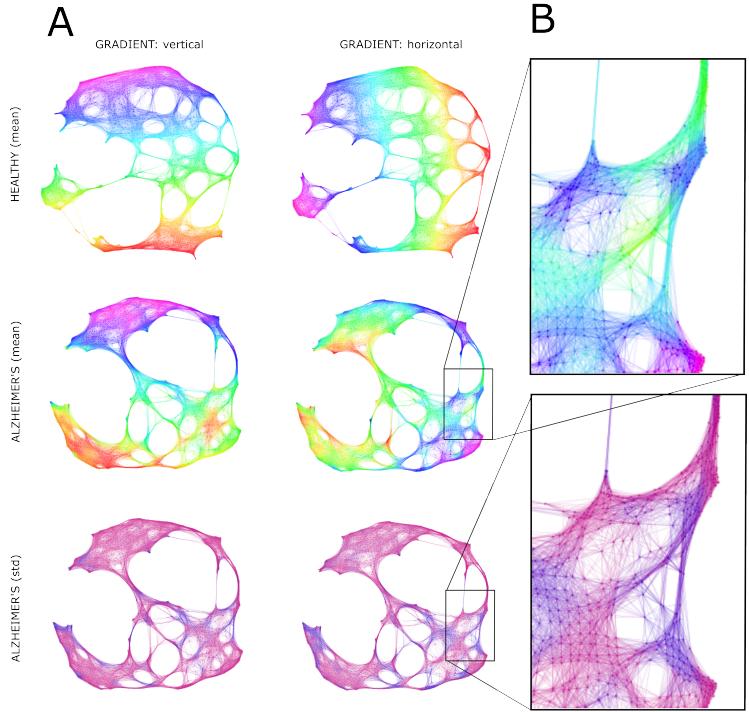


Figure 2: Mapper graph of DLPFC gene co-expression network (healthy reference coloring). axes and Nodes representing gene clusters are colored according to the reference (healthy) smooth expression gradient. B. Discontinuities in the AD graph's colors indicate network perturbations in AD.

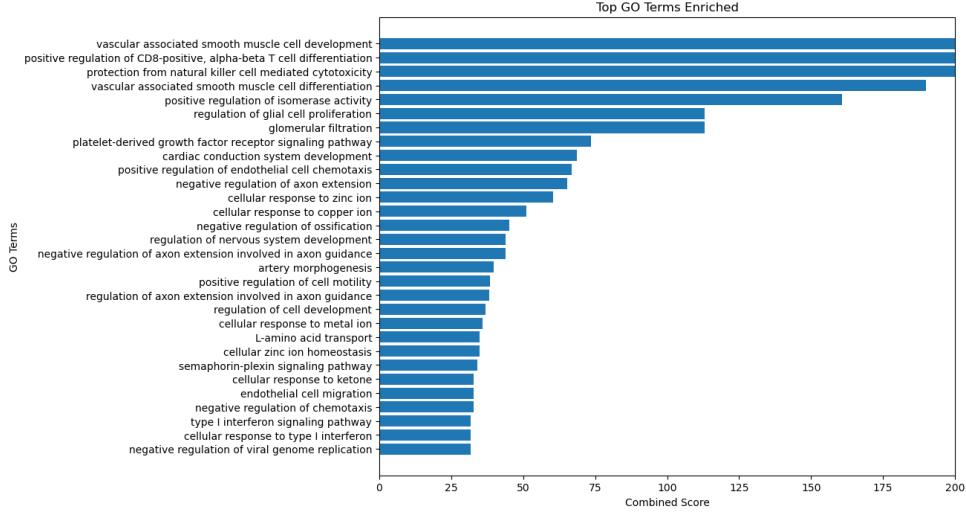


Figure 3: Top enriched GO terms in DLPFC sorted by combined score

Figure 4 (DLPFC dendrogram) illustrates hierarchical clustering of enriched GO terms based on gene overlap, revealing four major clusters of biological processes in DLPFC. The largest cluster (red) predominantly includes immune and cellular regulation pathways (e.g., positive regulation of CD8-positive T cell differentiation, platelet-derived growth factor receptor signaling), consistent with literature emphasizing immune dysregulation in AD [21, 22]. Another cluster (green) groups vascular-associated processes such as vascular smooth muscle cell development and differentiation, aligning with known cerebrovascular dysfunction in AD [19, 20]. An orange cluster highlights novel functional relationships, including regulation of axon extension, glomerular filtration, and cellular responses to interferon—pathways less well-established in AD and thus intriguing new avenues for investigation. The clear delineation between clusters underscores the complexity of AD mechanisms, suggesting both established and novel gene networks are at play in the DLPFC.

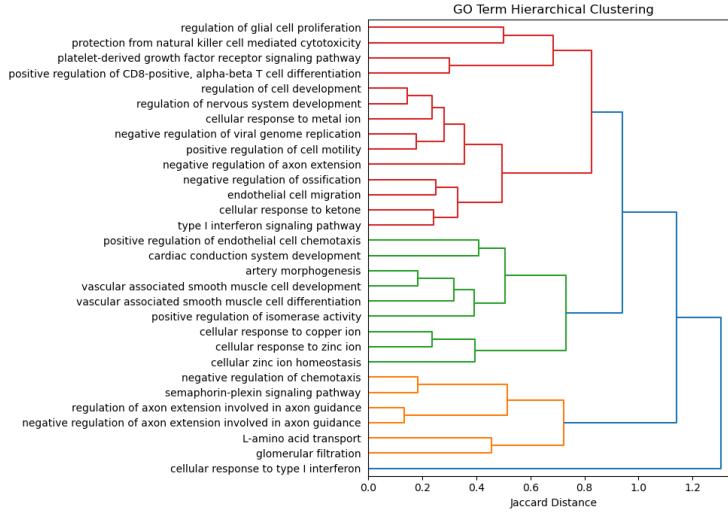


Figure 4: Hierarchical clustering of enriched GO terms in DLPFC.

The DLPFC cnetplot (Figure 5) maps connections between enriched GO terms and their associated genes, highlighting the diverse involvement of certain genes across multiple biological processes. Some genes (e.g., *MET*, *PRKD1*) appear only in single processes, indicating specialized roles. For example, *MET* has been linked to neuronal survival pathways relevant to AD [23], whereas *PRKD1*'s role in AD-related pathways

is less documented, representing a novel area for exploration. In contrast, genes like *HES1* span multiple GO terms, connecting vascular smooth muscle cell differentiation with glial cell proliferation—processes that intersect via Notch signaling [24, 25]. Such genes with broad involvement across pathways underscore the complexity of AD molecular networks and suggest that both specialized and multifunctional genes could be key therapeutic targets.

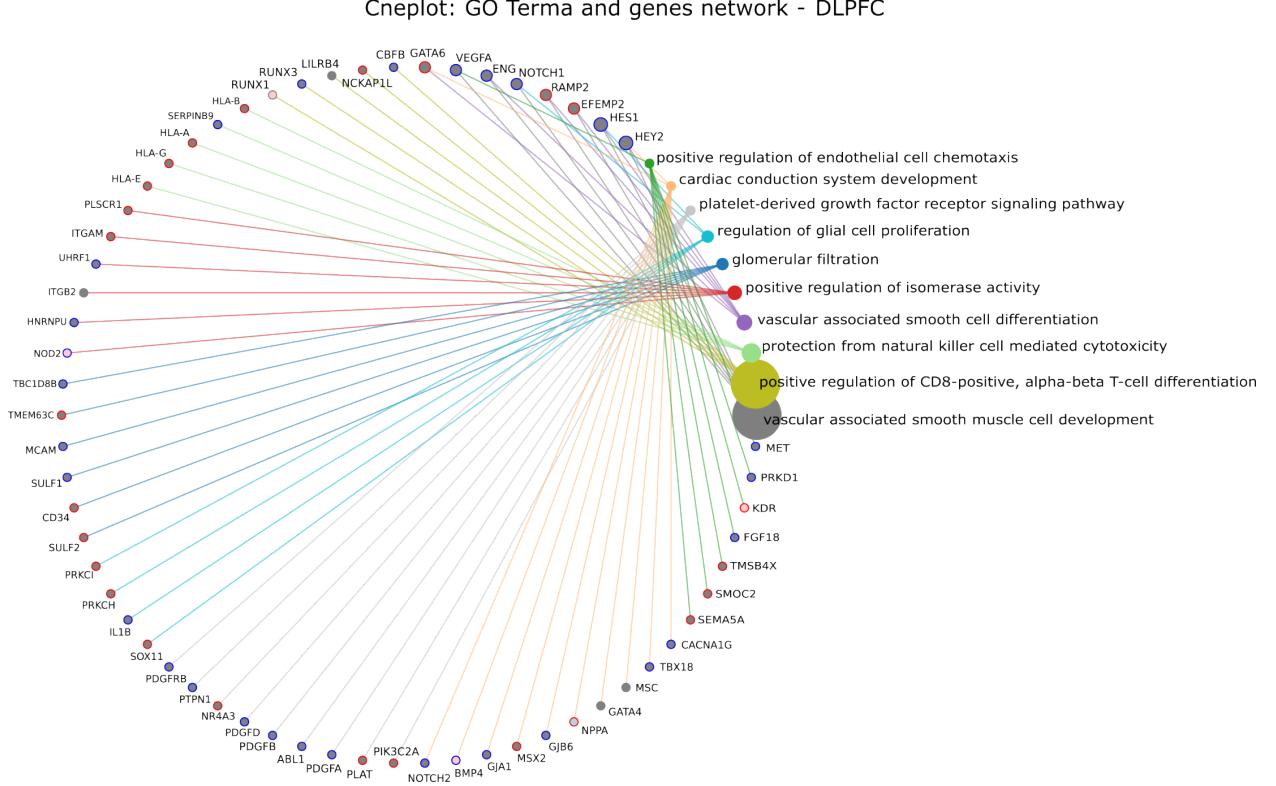


Figure 5: Top 10 GO term–gene network with enriched biological processes in DLPFC. GO terms scaled according to their combined enrichment scores. Genes are scaled by their in-degree (number of GO term associations). The interior color of each gene node reflects its differential expression status (up-regulated, down-regulated, or unchanged), whereas border color denotes the direction of its differential coexpression change between conditions.

## 2.2 Visual Cortex

The visual cortex (VC) is traditionally considered less prominently affected in AD compared to regions like the DLPFC; neuropathological changes in VC often appear later and are less severe [26, 27]. Nonetheless, our analysis identified several significantly enriched GO terms in VC, suggesting subtle yet important biological disturbances. Eleven GO terms showed notable enrichment in AD, with “skeletal myofibril assembly,” “positive regulation of endothelial cell chemotaxis,” and “positive regulation of glial cell differentiation” among the most pronounced.

The enrichment of pathways related to endothelial chemotaxis and smooth muscle contraction supports evidence linking vascular dysfunction and cerebral hypoperfusion to AD pathology, even in less severely affected regions [19, 20]. Additionally, GO terms like “cellular zinc ion homeostasis” and “response to unfolded protein” align with established molecular abnormalities in AD, particularly regarding protein misfolding stress and metal ion dysregulation [28, 29]. Interestingly, the significant enrichment of “skeletal myofibril assembly” and “platelet degranulation” in VC represents novel findings, as these processes have not been traditionally associated with AD pathology in this region, offering intriguing new targets for research.

The dendrogram for enriched GO terms in VC (Figure 6B) reveals distinct clusters: the largest clus-

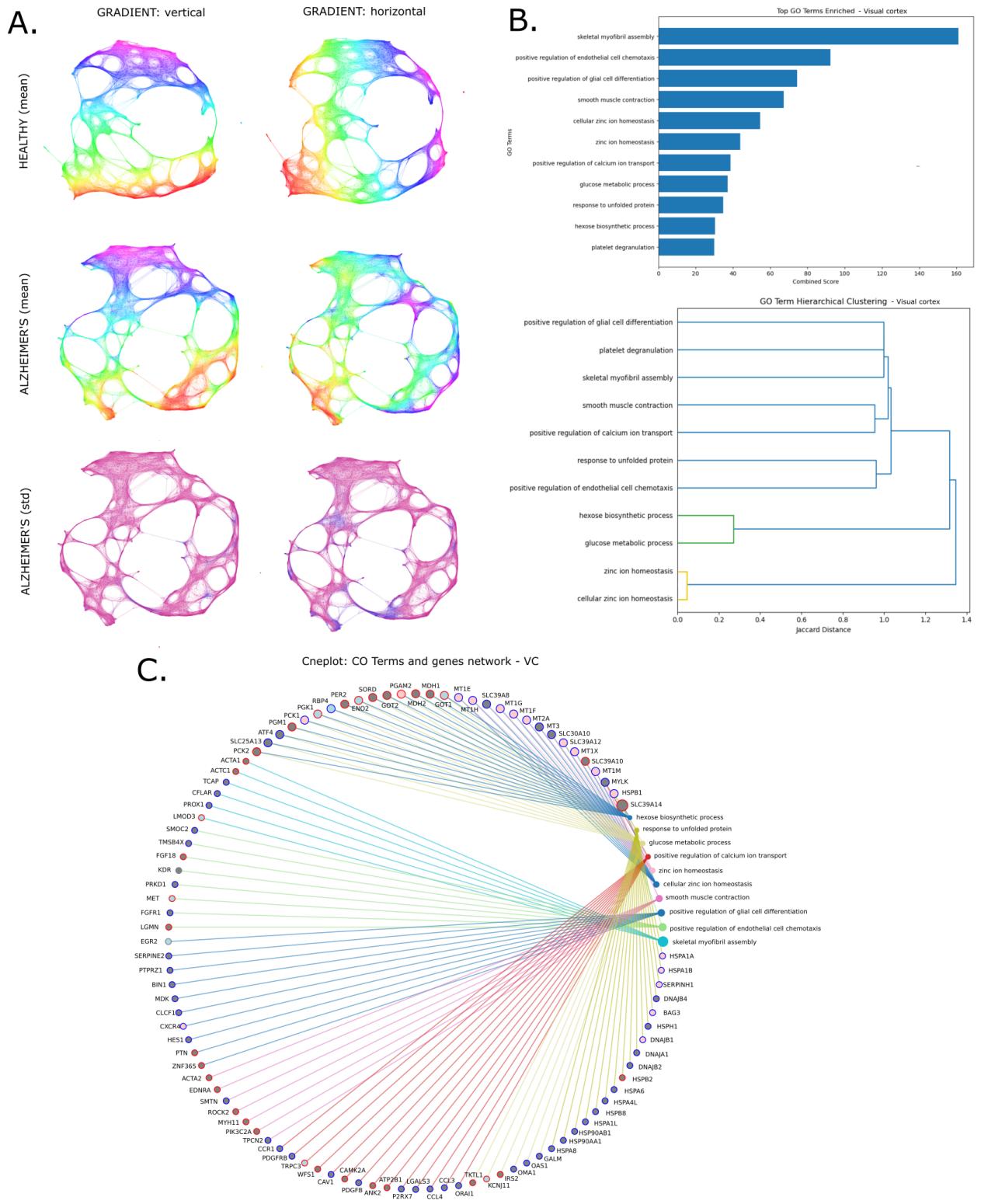


Figure 6: **A.** Mapper graph of VC gene network (healthy coloring). Compared to DLPFC, the VC shows fewer disrupted regions (color discontinuities) in AD. **B.** Bar graph and hierarchical clustering of enriched GO terms in VC. **C.** Cneplot of VC (same as DLPFC)

ter (blue) groups terms related to vascular and stress-response pathways (e.g., endothelial cell chemotaxis, response to unfolded protein, smooth muscle contraction), consistent with vascular and proteostasis disturbances in AD [19, 29]. A smaller green cluster groups metabolic processes (hexose biosynthesis, glucose metabolism), aligning with known alterations in cerebral glucose metabolism—an AD hallmark [30]. The orange cluster, marked by essentially zero Jaccard distance between terms, tightly pairs “zinc ion homeostasis” with “cellular zinc ion homeostasis,” highlighting metal ion dysregulation as a critical aspect in VC as well [28].

The VC cnetplot reveals that the gene *SLC39A14* (ZIP14, a zinc transporter) is implicated in three key enriched processes: hexose biosynthetic process, cellular zinc ion homeostasis, and response to unfolded protein. This underscores *SLC39A14*’s multifaceted role in the VC’s molecular landscape. As a member of the ZIP family of metal ion transporters (primarily Zn, but also Mn and Fe), *SLC39A14* has established roles in cellular zinc regulation and metabolic signaling [31, 32]. Notably, recent preclinical AD models have demonstrated reduced ZIP14 levels in cortex under sporadic AD conditions—a deficit reversible with dietary zinc supplementation [33]. Zinc dyshomeostasis is a recognized feature of AD pathology, correlated with A $\beta$  aggregation and impaired synaptic function [34, 35]. *SLC39A14*’s association with hexose biosynthesis also suggests a metabolic dimension, resonating with its known involvement in glucose metabolism and insulin regulation in non-neuronal contexts [36]. Meanwhile, its linkage to the unfolded protein response hints at a potential role in proteostasis under neurodegenerative stress [29]. Taken together, these connections bridge metal ion regulation, metabolic control, and protein-folding mechanisms in AD. *SLC39A14* may serve as a nexus of vulnerability in cortical AD pathology, meriting deeper functional characterization.

### 2.3 Cerebellum

In contrast to DLPFC and VC, the cerebellum (CB) in our analysis did not show any GO Biological Process terms meeting our significance criteria after correction. This aligns with clinical and neuropathological observations that cerebellar involvement in AD is minimal compared to cortical regions. The lack of significant enrichment suggests that, within the sensitivity of our approach, the cerebellum’s gene co-expression network remains relatively intact or unchanged in AD, or that any changes do not converge on common GO-defined pathways. While individual gene disruptions in CB may occur in AD, they might be too subtle or heterogeneous to yield a significant GO term enrichment.

## 3 Tau- and Amyloid- $\beta$ -Relevant Processes Across DLPFC and Visual Cortex

Across both DLPFC and VC, the enriched GO terms point to upstream pathways that modulate tau and amyloid- $\beta$  (A $\beta$ ) pathology. In DLPFC, several broad mechanistic themes emerge. Semaphorin–plexin signaling, PDGF $\rightarrow$ c-Abl activity, metal-ion homeostasis, and type-I interferon responses all appear as upstream modulators of pathological tau, consistent with models in which kinase activation, Zn $^{2+}$ /Cu $^{2+}$  imbalance, and inflammatory tone influence tau phosphorylation and aggregation [37, 3, 38, 39]. Notably, these pathways act without the need for increased *MAPT* expression itself, aligning with contemporary views that tau dysfunction in AD arises from network-level stressors rather than direct *MAPT* dysregulation.

A $\beta$ -related processes in DLPFC map onto known modulators of A $\beta$  production, clearance, and toxicity. Lipid-handling and ApoE-associated pathways influence A $\beta$  deposition and removal, while immune and endolysosomal modules regulate plaque clearance and secretase activity [40, 41, 42]. Current evidence suggests a convergence between the two hallmark pathologies: A $\beta$  can amplify tau phosphorylation, and tau burden can exacerbate A $\beta$ -driven synaptic dysfunction [2]. The DLPFC GO-term profile reflects this interplay, highlighting pathways relevant to both arms of AD pathology.

In the visual cortex, enrichment patterns shift toward vascular, metabolic, and proteostasis pathways. Metal-ion regulation again appears, relevant to both tau phase behavior and A $\beta$  aggregation [43, 44]. Calcium-handling terms relate to Ca $^{2+}$  dysregulation frameworks in which A $\beta$  exposure can trigger kinase cascades leading to tau hyperphosphorylation [45]. Carbohydrate-metabolism terms implicate the O-GlcNAc axis, whose impairment promotes tau phosphorylation [46, 47]. ER stress and UPR-related processes correspond to PERK-eIF2 $\alpha$  signaling changes that affect both BACE1 activity and neuronal tau burden [48]. Vascular and platelet-associated pathways—tied to APP/secretase release from platelets and to endothelial A $\beta$  transport via receptors such as LRP1—connect VC enrichment to A $\beta$  handling and cerebrovascular

contribution to AD [49, 50, 51, 52, 53].

Taken together, the GO-term signatures from DLPFC and VC converge on broad biological themes—signaling stress, ion regulation, immune tone, vascular function, and proteostasis—that interface with both tau and A $\beta$  pathways, albeit with region-specific emphases.

## Discussion

This topological approach offers a new way to examine the structure of gene co-expression changes. By using Mapper to identify comparative discontinuities in data - that is, breaks or shifts in the topology of gene co-expression space relative to controls - we detect network changes that standard methods may miss. In the DLPFC, these shifts point to coordinated immune, vascular, and other cellular processes, consistent with the view that AD involves multiple interacting pathways.

There are, however, important limitations. The Mapper algorithm depends on several parameter choices (lens function, cover resolution, clustering threshold), and different settings can highlight different aspects of the dataset. While these parameters are more general and less restrictive than rigid assumptions about the shape of data, they might present challenges for data interpretation. It remains unclear which combinations of said parameters are most informative for biological interpretation, and this requires systematic evaluation. Our approach may also benefit from a percolation-like procedure to assess how stable these comparative discontinuities are across parameter ranges.

In summary, using a topological lens to track comparative discontinuities in TDA shows how localized genetic changes accumulate into broader network disturbances. This pipeline is extremely generalizable and can be applied to many forms of complex data - proteomic, epigenomic, imaging, longitudinal datasets or any data that requires comparison. Extending this work to other data types or time-series samples could clarify how these network disruptions arise and evolve over disease progression. Ultimately, topological tools in conjunction with discontinuity detection may help identify disruptions in complex biological systems.

## Acknowledgments

The authors declare no competing interests.

## Data

Data is publicly available at <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE44772>

## Code Availability

Code can be found at [https://github.com/ayasamadzelkava/genes\\_Alzh\\_TDA](https://github.com/ayasamadzelkava/genes_Alzh_TDA)

## Additional Resources

Interactive map of genes and GO terms can be found at  
<https://embed.kumu.io/ed76a53d20b9ad7cd4aac7df81103f0>

## Materials and Methods

### Data Collection and Preprocessing

**Data Collection** Our study utilized gene expression data from GSE44772 dataset, obtained from the Gene Expression Omnibus (GEO) repository. The GSE44772 SuperSeries is composed of three SubSeries, each focusing on a different brain region: cerebellum, dorsolateral prefrontal cortex, and visual cortex. gene expression matrix contained  $G = 39,302$  probe-level measurements. Each individual contributed tissue from

3 brain regions. This SuperSeries provides multi-tissue gene expression profiles of the human brain ( $N = 230$  individuals: 129 AD and 101 controls)), with a total of 690 samples across the three SubSeries. Gene expression profiling for these samples was conducted using the Rosetta/Merck Human 44K v1.1 microarray platform and contains 39,302 probe-level gene measurements..

**Data Preprocessing** Raw microarray expression values were examined for missing or infinite entries, which were filled using linear interpolation across samples. To mitigate confounding, expression values were adjusted for multiple sample-level covariates known to influence transcriptomic profiles: pH, age, RNA integrity number (RIN), gender, batch, post-mortem interval (PMI), tissue, and preservation method.

**Filtering genes.** Microarray probes lacking a valid gene symbol (i.e., probes annotated as predicted or uncharacterized genes) were excluded. Only probes with an assigned *Gene Symbol* were retained.

Some genes were represented by multiple probes. To obtain a single expression value per gene, expression values from all probes mapping to the same gene symbol were averaged.

**Selection of the most variable genes.** To focus on genes exhibiting the strongest biological signal and reduce noise, we extracted the most variable subset of genes. Gene-wise variability was quantified using the interquartile range (IQR) across all samples. Genes were ranked by IQR, and the top one-third were retained. This yielded a final high-variability gene set comprising  $G = 6,507$  genes, which served as the input for downstream network-based analyses.

**Co-expression Matrix Construction** Using the preprocessed matrices, we computed Pearson correlation coefficients for each pair of genes, separately for AD and control groups in each brain region. This yielded a gene–gene co-expression matrix for every condition and region (dimensions  $N \times N$  for  $N = 6,507$  genes). These matrices form the input for subsequent topological analyses.

## Topological Data Analysis

### 3.1 Application of the Mapper Algorithm

All Mapper computations were performed using the `python KeplerMapper` library.

**Dimensionality Reduction** Dimensionality reduction was performed by utilizing a combination of Isomap and Uniform Manifold Approximation and Projection (UMAP).

- **Isomap:** (`n_components=100`) Isomap is a non-linear dimensionality reduction method that preserves the geodesic distances between all pairs of data points, effectively capturing the global manifold structure of the data. We set the number of components to 100 to retain sufficient variance and structural information from the original high-dimensional space.
- **UMAP:** (`n_components=2`) After reducing the dimensions with Isomap, we further projected the data onto a two-dimensional space using UMAP.

The combination of Isomap and UMAP allows us to capture both the global and local structures of the gene expression data. Isomap preserves the overall manifold geometry, while UMAP maintains the local relationships between genes.

**Parameter selection** We manually selected the covering parameters to maximize the topological complexity of the resulting Mapper graph—specifically aiming to reduce the number of disconnected components while increasing the number of loops or holes, which are indicative of nontrivial topological features. Parameters were chosen to ensure sufficient overlap between intervals, thereby promoting continuity in the graph, and to provide a sufficient number of intervals to capture fine-grained structural variations within the data by method of manual percolation and visual inspection.

Covering parameters were set as follows:

Brain Region	Parameter	Healthy	Alzheimer's
DLPFC	Percentage overlap	0.72	0.72
	Number of intervals	80	80
Cerebellum	Percentage overlap	0.68	0.70
	Number of intervals	80	80
Visual Cortex	Percentage overlap	0.72	0.72
	Number of intervals	80	80

Table 1: Mapper parameters used for each brain region under healthy and Alzheimer's conditions.

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm was applied to cluster genes based on their expression profiles. DBSCAN identifies clusters as areas of high density separated by areas of low density, making it effective for detecting clusters of arbitrary shape and handling noise in the data.

The clustering step internally converts each pairwise correlation value  $c$  into a distance using the standard transformation

$$d = 1 - c,$$

such that genes with higher correlation correspond to smaller distances.

**Gene Coloring** Gene coloring is performed by KepplerMapper. Each gene within the Mapper graph was assigned a color based on its DBSCAN cluster label derived from the healthy dataset. A distinct color was mapped to each unique cluster label, creating a visual representation of the clustering structure in the gene expression data. This color-coding allowed us to track how genes grouped in the healthy state are distributed in the Mapper graphs of both healthy and Alzheimer's datasets.

**Node Coloring** In the Mapper graph, nodes represent groups of genes. Since each node contains multiple genes, we employed two *aggregation functions* to assign a single representative color to each node:

1. Mean Color Value: (associated with DBSCAN cluster labels) of all genes within a node. This approach provides an average representation of the gene clusters present in the node.
2. Standard Deviation of Colors: to assess the variability of genes in the node.

### 3.2 Discontinuity Detection

Discontinuity of AD dataset was highlighted by using Healthy dataset as a "continuous ground truth".

Let  $G$  be the set of genes and let  $PC$  be the corresponding point cloud. Applying Mapper to  $PC$  produces a Mapper graph  $MG$  consisting of nodes

$$MG\_node \subseteq G.$$

Each node is assigned a numerical color value in the process of Mapper application that results on a smooth color gradient on the graph. Each gene  $g \in G$  inherits the color of the Mapper node it belongs to:

$$g \in MG\_node \implies \text{color}(g) = MG\_node\_color(MG\_node).$$

To compare healthy and Alzheimer's data, we first compute  $MG$  and  $MG\_node\_color$  for the healthy point cloud. We then take the colors assigned to genes in the healthy case and transfer these gene-level colors to the Alzheimer's Mapper graph  $MG^{AD}$ . Thus, each gene in  $MG^{AD}$  receives the color associated with its label from the healthy graph.

By examining color transitions between adjacent nodes in  $MG^{AD}$ , we identify discontinuities: abrupt changes in the transferred colors indicate regions where the Alzheimer's structure diverges from the healthy organization.

For each node  $n \in MG^{\text{AD}}$ , we compute the standard deviation of the numerical color values of the genes it contains,

$$\sigma(n) = \text{StdDev}\{\text{color}(g) \mid g \in n\},$$

to quantify heterogeneity within the node. Nodes with the highest  $\sigma(n)$  (top 20% in each brain region: DLPFC, cerebellum, visual cortex) are selected as the most variable. Genes contained in these nodes correspond to regions exhibiting topological misalignment relative to the healthy state.

**Rotation of the Coloring Gradient** To ensure that topological discontinuities were not missed due to the orientation of the coloring gradient, we applied an additional relabeling step that effectively rotates the coloring axis. Mapper coloring based on a one-dimensional lens is sensitive only to variation along that specific axis; discontinuities aligned orthogonally will remain undetected. To address this, we reordered the cluster colors according to the vertical coordinate of each cluster's centroid in the two-dimensional lens space. For each cluster  $c$ , we computed its centroid position

$$\bar{y}_c = \frac{1}{|S_c|} \sum_{i \in S_c} y_i,$$

where  $S_c$  denotes the set of points assigned to cluster  $c$  and  $y_i$  is the second coordinate of point  $i$  in the projected space. Clusters were then sorted by  $\bar{y}_c$ , and new color labels were assigned in this top-to-bottom order. This procedure produces a coloring aligned with the orthogonal direction of the original gradient, allowing the Mapper analysis to detect structural discontinuities regardless of their orientation in the lens.

Extracted genes were used for consequent GO enrichment analysis.

### 3.3 Mapper Graph Visualization

Mapper graphs for figures (Figure 2 and Figure 6A) were generated using Gephi. Sigmoid function was applied to STD color scheme to highlight regions with high standard deviation.

## GO (Gene Ontology) Enrichment Analysis

### 3.4 Enrichr

For the present study, we focused on Gene Ontology (GO) Biological Process (BP) annotations. GO enrichment was performed with Enrichr (<https://maayanlab.cloud/Enrichr/>) via its Python API (`gp.enrichr`), using the “GO\_Biological\_Process\_2021” gene set library and *Homo sapiens* as the organism. Enrichr combines results from multiple gene set libraries; in our analysis we restricted attention to this GO BP 2021 library.

For each brain region, we thus obtained a table of enriched GO Biological Process terms with associated statistics (including raw  $p$ -value, adjusted  $p$ -value, odds ratio, and combined score). Adjusted  $p$ -value is computed using the Benjamini-Hochberg (BH) procedure to correct for multiple hypothesis testing and control the false discovery rate (FDR). Given a list of  $m$  hypotheses with corresponding  $p$ -values  $p_1, p_2, \dots, p_m$ , sorted in ascending order, the BH-adjusted  $p$ -value for the  $i$ -th ranked term is:

$$p_i^{\text{adj}} = \frac{p_i \cdot m}{i}$$

where  $i$  is the rank of the  $p$ -value in the sorted list and  $m$  is the total number of terms tested. Terms with  $p < 0.05$  (Fisher's exact test, Benjamini-Hochberg correction for multiple testing) were considered significant. To quantify the prominence of each process, we ranked GO terms by Enrichr's combined score. The combined score is a composite metric that integrates the significance from the Fisher exact test (via the  $p$ -value) and the deviation from the expected rank (via a z-score). It is calculated as:

$$\text{Combined Score} = \log(p) \cdot z$$

where  $p$  is the raw  $p$ -value from the enrichment test and  $z$  is the z-score reflecting the deviation of the observed rank from the expected rank. A high combined score indicates that a GO term is both statistically significant (low  $p$ -value) and strongly overrepresented (high odds ratio) in the gene list. In addition, we applied a combined-score cutoff and retained only terms with combined score  $\geq 30$ .

### 3.5 Hierarchical clustering/dendrogram

To visualize functional similarity among enriched Gene Ontology (GO) terms, we constructed a hierarchical dendrogram based on shared gene content. First, we assembled a binary matrix  $M \in \{0, 1\}^{n \times g}$ , where each row represents a GO term and each column represents a unique gene. Each entry  $M_{ij} = 1$  if gene  $j$  is annotated to GO term  $i$ , and 0 otherwise. We computed pairwise Jaccard distances between GO terms using:

$$D_J(A, B) = 1 - \frac{|A \cap B|}{|A \cup B|}$$

where  $A$  and  $B$  are the sets of genes associated with two GO terms to quantifies dissimilarity based on shared gene content. The resulting distance matrix was subjected to agglomerative hierarchical clustering using Ward's linkage method, which minimizes the total within-cluster variance at each merging step.

### 3.6 Cnetplot Construction: GO Term–Gene Bipartite Network

To visualize the functional landscape of enriched biological processes, we constructed bipartite networks connecting GO terms and their associated genes using a custom circular network plot (cnetplot). From the list of enriched GO terms, we selected the top  $k$  terms by highest combined score (see Section *GO Enrichment Analysis*). Each GO term  $t_i \in T$  and gene  $g_j \in G$  was added as a node in a graph  $\mathcal{G} = (V, E)$ , where  $V = T \cup G$ , and an undirected edge  $(t_i, g_j) \in E$  was created if gene  $g_j$  was annotated to term  $t_i$ .

**Differential Expression Analysis** Differential expression between control and Alzheimer's disease (AD) samples was assessed at the gene level using a robust linear modeling framework. For each gene, expression values from all individuals were modeled using a robust linear regression of the form

$$y_i = \beta_0 + \beta_1 \text{Condition}_i + \varepsilon_i,$$

where  $\text{Condition}_i$  indicates group membership (0 = control, 1 = AD),  $\beta_1$  represents the estimated  $\log_2$  fold change, and  $\varepsilon_i$  captures residual noise. A p-value for differential expression was obtained for the coefficient  $\beta_1$ , and multiple hypothesis correction was performed using the Benjamini–Hochberg false discovery rate (FDR). Genes exceeding both a  $\log_2$  fold-change threshold ( $|\beta_1| > 1$ ) and an FDR-adjusted significance threshold ( $q < 0.05$ ) were classified as differentially expressed and assigned a color label reflecting up-regulation (red), down-regulation (blue), or no significant change (gray) for visualization.

**Differential Coexpression Analysis** To characterize changes in gene–gene relationships between conditions, we computed separate Pearson correlation matrices for control and AD samples. A differential coexpression matrix was then obtained by subtracting control correlations from AD correlations,

$$\Delta r_{ij} = r_{ij}^{(\text{AD})} - r_{ij}^{(\text{Ctrl})},$$

for each gene pair  $(i, j)$ . For each gene, differential connectivity was summarized by computing (i) the average absolute change in correlation magnitude and (ii) the average signed change in correlation direction across all its gene–gene interactions. Genes with an average magnitude change exceeding a predefined threshold (0.5) were considered differentially coexpressed; positive shifts were assigned red labels, negative shifts blue labels, and non-significant changes gray. These labels were used to highlight genes showing substantial rewiring of coexpression structure between conditions.

## References

- [1] Giulia Paroni et al. “Understanding the Amyloid Hypothesis in Alzheimer’s Disease”. In: *Journal of Alzheimer’s Disease* 68.2 (Jan. 1, 2019), pp. 493–510. DOI: 10.3233/jad-180802. PMID: 30883346.
- [2] X. Zhuang et al. “The Interplay Between Accumulation of Amyloid-Beta and Tau Proteins, PANoptosis, and Inflammation in Alzheimer’s Disease.” In: *Neuromolecular medicine* (2024). DOI: 10.1007/s12017-024-08815-z. PMID: 39751702.
- [3] Jiyoen Kim et al. “TYK2 Regulates Tau Levels, Phosphorylation and Aggregation in a Tauopathy Mouse Model”. In: *Nature Neuroscience* (2024). DOI: 10.1038/s41593-024-01777-2. PMID: 39528671.
- [4] Gabrielle Zuniga et al. “A Century of Questions: Retrospective Study of the Controversy and Efficacy of Alzheimer’s Disease Models”. In: (May 1, 2015). DOI: 10.15781/t2jd4pp3c.
- [5] Rudy J. Castellani et al. “Molecular Pathogenesis of Alzheimer’s Disease: Reductionist versus Expansionist Approaches.” In: *International Journal of Molecular Sciences* 10.3 (Mar. 26, 2009), pp. 1386–1406. DOI: 10.3390/ijms10031386. PMID: 19399255.
- [6] Bin Zhang et al. “Integrated Systems Approach Identifies Genetic Nodes and Networks in Late-Onset Alzheimer’s Disease”. In: *Cell* 153.3 (Apr. 25, 2013), pp. 707–720. DOI: 10.1016/j.cell.2013.03.030. PMID: 23622250.
- [7] Eric E. Schadt et al. “An Integrative Genomics Approach to Infer Causal Associations between Gene Expression and Disease”. In: *Nature Genetics* 37.7 (June 29, 2005), pp. 710–717. DOI: 10.1038/ng1589. PMID: 15965475.
- [8] Steve Horvath et al. “Geometric Interpretation of Gene Coexpression Network Analysis”. In: *PLoS Computational Biology* 4.8 (Aug. 15, 2008). Ed. by Satoru Miyano, e1000117. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1000117. URL: <https://dx.plos.org/10.1371/journal.pcbi.1000117> (visited on 07/12/2024).
- [9] Julie P Merchant et al. “Predictive Network Analysis Identifies *JMJD6* and Other Novel Key Drivers in Alzheimer’s Disease”. In: *Cold Spring Harbor Laboratory - bioRxiv* (Oct. 22, 2022). DOI: 10.1101/2022.10.19.512949.
- [10] Lingfei Wang et al. “Detection of Regulator Genes and eQTLs in Gene Networks”. In: *Springer International Publishing* (Jan. 1, 2016), pp. 1–23. DOI: 10.1007/978-3-319-43335-6\_1.
- [11] Mete Civelek et al. “Systems Genetics Approaches to Understand Complex Traits”. In: *Nature Reviews Genetics* 15.1 (Jan. 1, 2014), pp. 34–48. DOI: 10.1038/nrg3575. PMID: 24296534.
- [12] Yara Skaf et al. “Topological Data Analysis in Biomedicine: A Review”. In: *Journal of Biomedical Informatics* (May 1, 2022), pp. 104082–104082. DOI: 10.1016/j.jbi.2022.104082. PMID: 35508272.
- [13] Sipko van Dam et al. “Gene Co-Expression Analysis for Functional Classification and Gene–Disease Predictions”. In: *Briefings in Bioinformatics* 19.4 (July 20, 2018), pp. 575–592. ISSN: 1477-4054. DOI: 10.1093/bib/bbw139. URL: <https://doi.org/10.1093/bib/bbw139> (visited on 05/28/2024).
- [14] Frédéric Chazal et al. “An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists”. In: *Frontiers in Artificial Intelligence* 4 (Sept. 29, 2021), p. 667963. ISSN: 2624-8212. DOI: 10.3389/frai.2021.667963. URL: <https://www.frontiersin.org/articles/10.3389/frai.2021.667963/full> (visited on 08/07/2025).
- [15] Gunnar Carlsson. “Topology and Data”. In: *Bulletin of the American Mathematical Society* 46.2 (Jan. 29, 2009), pp. 255–308. ISSN: 0273-0979. DOI: 10.1090/S0273-0979-09-01249-X. URL: <http://www.ams.org/journal-getitem?pii=S0273-0979-09-01249-X> (visited on 05/28/2023).
- [16] Zhengqian Lyu. “Application of Topological Data Analysis in Complex Network Structure Identification”. In: 136 (2025).
- [17] Alice Patania et al. “Topological Gene Expression Networks Recapitulate Brain Anatomy and Function”. In: *Network Neuroscience* 3.3 (Jan. 2019), pp. 744–762. ISSN: 2472-1751. DOI: 10.1162/netn\_a\_00094. URL: <https://direct.mit.edu/netn/article/3/3/744-762/2179> (visited on 08/27/2024).

- [18] Tamal K. Dey et al. "Gene expression data classification using topology and machine learning models." In: *BMC Bioinformatics* (2022). DOI: 10.1186/s12859-022-04704-z.
- [19] Berislav V. Zlokovic. "Neurovascular pathways to neurodegeneration in Alzheimer's disease and other disorders". In: *Nature Reviews Neuroscience* 12.12 (2011), pp. 723–738. DOI: 10.1038/nrn3114.
- [20] Costantino Iadecola. "The pathobiology of vascular dementia". In: *Neuron* 80.4 (2013), pp. 844–866. DOI: 10.1016/j.neuron.2013.10.008.
- [21] David Gate et al. "Clonally expanded CD8 T cells patrol the cerebrospinal fluid in Alzheimer's disease". In: *Nature* 577 (2020), pp. 399–404. DOI: 10.1038/s41586-019-1895-7.
- [22] Brianne M. Bettcher et al. "Peripheral and central immune system crosstalk in Alzheimer disease – a research prospectus". In: *Nature Reviews Neurology* 17.11 (2021), pp. 689–701. DOI: 10.1038/s41582-021-00549-x.
- [23] Aimee L. Wright et al. "HGF/c-Met Signaling: A Critical Regulator of Neuronal Survival and Function in Alzheimer's Disease". In: *Neurodegenerative Diseases* 15 (2015), pp. 133–137.
- [24] Thomas Gridley. "Notch signaling in the vasculature". In: *Current Topics in Developmental Biology* 92 (2010), pp. 277–309. DOI: 10.1016/S0070-2153(10)92009-7.
- [25] Justin D. Lathia et al. "Notch: from neural development to neurological disorders". In: *Journal of Neurochemistry* 107.6 (2008), pp. 1471–1481. DOI: 10.1111/j.1471-4159.2008.05715.x.
- [26] Heiko Braak et al. "Neuropathological staging of Alzheimer-related changes". In: *Acta Neuropathologica* 82.4 (1991), pp. 239–259. DOI: 10.1007/BF00308809.
- [27] Ramasamy Thangavel et al. "Modular and laminar pathology of Brodmann's area 37 in Alzheimer's disease". In: *Neuroscience* 152.1 (2008), pp. 50–55. DOI: 10.1016/j.neuroscience.2007.12.025.
- [28] Mark A. Lovell et al. "Copper, iron and zinc in Alzheimer's disease senile plaques". In: *Journal of the Neurological Sciences* 158.1 (1998), pp. 47–52. DOI: 10.1016/S0022-510X(98)00092-6.
- [29] Claudio Hetz et al. "ER stress and the unfolded protein response in neurodegeneration". In: *Nature Reviews Neurology* 13 (2017), pp. 477–491. DOI: 10.1038/nrneurol.2017.99.
- [30] Lisa Mosconi. "Brain glucose metabolism in the early and specific diagnosis of Alzheimer's disease". In: *European Journal of Nuclear Medicine and Molecular Imaging* 32.4 (2005), pp. 486–510. DOI: 10.1007/s00259-005-1762-7.
- [31] Toshiyuki Fukada et al. "Zinc signaling in health and disease". In: *Biological Chemistry* 399.10 (2018), pp. 1285–1297. DOI: 10.1515/hsz-2018-0242.
- [32] Supak Jenkitkasemwong et al. "SLC39A14 is required for the development of hepatocellular iron overload in murine models of hereditary hemochromatosis". In: *Cell Metabolism* 22.1 (2015), pp. 138–150. DOI: 10.1016/j.cmet.2015.05.002.
- [33] Katarzyna A. Rychlik et al. "Zinc supplementation restores ZIP14 levels in a sporadic Alzheimer's disease rat model". In: *Neuroscience Letters* (2023).
- [34] Paul A. Adlard et al. "Metals and Alzheimer's disease: A decade of progress". In: *Journal of Alzheimer's Disease* 62.3 (2018), pp. 1369–1379. DOI: 10.3233/JAD-170875.
- [35] Stefano L. Sensi et al. "Copper and Zinc Dysregulation in Alzheimer's Disease". In: *Trends in Pharmacological Sciences* 39.12 (2018), pp. 1049–1063. DOI: 10.1016/j.tips.2018.10.001.
- [36] Catalina Troche et al. "Hepatic ZIP14-mediated zinc transport contributes to endosomal insulin receptor trafficking and glucose metabolism". In: *Journal of Biological Chemistry* 291.47 (2016), pp. 23939–23951. DOI: 10.1074/jbc.M116.747709.
- [37] Eun Hee Ahn et al. "Molecular Mechanisms of Alzheimer's Disease Induced by Amyloid- $\beta$  and Tau Phosphorylation Along with RhoA Activity: Perspective of RhoA/Rho-Associated Protein Kinase Inhibitors for Neuronal Therapy". In: *Cells* (2025). DOI: 10.3390/cells14020089. pmid: 39851517.
- [38] Virender Singh et al. "Zinc Promotes Liquid–Liquid Phase Separation of Tau Protein". In: *Journal of Biological Chemistry* 295.18 (May 1, 2020), pp. 5850–5856. DOI: 10.1074/jbc.ac120.013166. pmid: 32229582.

- [39] Sophie Sanford et al. “The type-I Interferon Response Potentiates Seeded Tau Aggregation and Exacerbates Tau Pathology”. In: *Alzheimer’s & Dementia* (Oct. 17, 2023). doi: 10.1002/alz.13493. pmid: 37849026.
- [40] Christian Behl. “In 2024, the Amyloid-Cascade-Hypothesis Still Remains a Working Hypothesis, No Less but Certainly No More”. In: *Frontiers in Aging Neuroscience* (2024). doi: 10.3389/fnagi.2024.1459224. pmid: 39295642.
- [41] Madia Lozupone et al. “Impact of Apolipoprotein E Isoforms on Sporadic Alzheimer’s Disease”. In: *Neural Regeneration Research* 19.1 (Jan. 1, 2024), pp. 80–83. doi: 10.4103/1673-5374.375316. pmid: 37488848.
- [42] Zezhong Lv et al. “Clearance of  $\beta$ -Amyloid and Synapses by the Optogenetic Depolarization of Microglia Is Complement Selective”. In: *Neuron* (2024). doi: 10.1016/j.neuron.2023.12.003. pmid: 38295790.
- [43] Anna Abelein et al. “Metal Binding of Alzheimer’s Amyloid- $\beta$  and Its Effect on Aggregation and Toxicity”. In: *Accounts of Chemical Research* (2023).
- [44] Philippe Faller et al. “Copper and zinc binding to amyloid- $\beta$ : roles in aggregation and toxicity”. In: *Coordination Chemistry Reviews* (2009).
- [45] Marina Calvo-Rodríguez et al. “Therapeutic Strategies to Target Calcium Dysregulation in Alzheimer’s Disease”. In: *Frontiers in Neuroscience* 14 (2020).
- [46] Feng Liu et al. “O-GlcNAcylation regulates phosphorylation of tau: evidence for a reciprocal relationship”. In: *Proceedings of the National Academy of Sciences* 101.29 (2004), pp. 10804–10809. doi: 10.1073/pnas.0400348101.
- [47] Feng Liu et al. “Reduced O-GlcNAcylation links lower brain glucose metabolism and tau pathology in Alzheimer’s disease”. In: *Proceedings of the National Academy of Sciences* 106.13 (2009), pp. 4912–4917. doi: 10.1073/pnas.0900358106.
- [48] Ali Ajoolabady et al. “ER stress and UPR in Alzheimer’s disease: mechanisms, pathogenesis, treatments”. In: *Cell Death & Disease* (2022).
- [49] Guojun Evin et al. “Platelets and Alzheimer’s disease: Potential of APP as a biomarker”. In: *Experimental Gerontology* 47.2 (2012), pp. 85–89.
- [50] J. A. Johnston et al. “Platelet  $\beta$ -secretase activity is increased in Alzheimer’s disease”. In: *Neurobiology of Aging* 29.6 (2008), pp. 861–868.
- [51] Mikhail Inyushin et al. “On the Role of Platelet-Generated Amyloid Beta Peptides in Blood–Brain Barrier Permeability and Brain Amyloidosis”. In: *Frontiers in Immunology* 11 (2020), p. 571083. doi: 10.3389/fimmu.2020.571083.
- [52] Stefania Petralla et al. “Low-Density Lipoprotein Receptor-Related Protein 1 as a Therapeutic Target for Alzheimer’s Disease”. In: *International Journal of Molecular Sciences* (2024).
- [53] Dong Wang et al. “Relationship Between Amyloid- $\beta$  Deposition and Blood–Brain Barrier Damage in Alzheimer’s Disease”. In: *Frontiers in Cellular Neuroscience* 15 (2021), p. 695479.