Integrated Metabolomics and Transcriptomics Analysis Identifies Molecular Subtypes within the Early and Late Mild Cognitive Impairment Stages of Alzheimer's Disease

Shashank Yadev[1], Shu Zhou[1], Zachary Karas[1], Yueyang Zhang[1], Lana X Garmire[1]*, Alzheimer's Disease Neuroimaging Initiative

[1]Department of Computational Medicine and Bioinformatics, the University of Michigan. Ann Arbor, USA. 48105

*: corresponding author email address: lgarmire@med.umich.edu

Abstract

Alzheimer's disease (AD) is a highly heterogeneous neurodegenerative condition. The current study identified clinically relevant molecular subtypes of the early and late mild cognitive impairment (EMCI and LMCI) stages of AD using 401 patients' data from the ADNI consortium. We integrated patients' metabolomics data with the PBMC transcriptomics data using an unsupervised clustering method called Similarity Network Fusion (SNF), and identified two subtypes in early and late MCI patients, respectively. The differences between these subtypes' metabolite concentrations and gene expression well correlate with physio-pathogenesis for AD, based on cognitive measurements, pseudo-trajectory analysis, and longitudinal analysis of dementia diagnosis. We detected many dysregulated processes between subtypes, such as aminoacyl-tRNA biosynthesis, immune system activity, zinc imbalances. While immune-related pathways are commonly dysregulated pathways in EMCI and LMCI stages, oxidative stress is prevalent in EMCI, whereas metabolic abnormality is enriched in LMCI. Refined subtypes within EMCI and LMCI are a step-forward toward more personalized treatment strategies for progressing patients before AD diagnosis.

Introduction

Alzheimer's Disease (AD) is a terminal neurodegenerative disease with irreversible cognitive impairment (Smoller, 2010). Patients with AD experience memory loss, lose motor function, have trouble eating and swallowing, and struggle to communicate with their family members (Frank, 1994; Ballardsy, 2011). In 2010, roughly 35.6 million people worldwide lived with AD, and this number is projected to increase to 115.4 million by the year 2050 due to the increasing average lifespan (Prince, 2013). Effective therapeutics would improve the lives of millions of patients and their families and lighten the economic burden on societies (Jia, 2018). However, there is still no cure for AD (Cisbani, 2021) despite the significant effort in developing treatments (Yiannopoulou, 2013).

There are three general stages along the AD continuum: first preclinical, then mild cognitive impairment (MCI), and finally dementia. Since identifying the MCI stage of AD in the late 1990s (Petersen, 1998), researchers have recognized this intermediate stage as having more potential for therapeutic interventions than the dementia stage (Edmonds, 2019). Many lifestyle and pharmaceutical interventions have been tested at the MCI stage, with some degrees of success (Ströhle, 2015; Cummings, 2007; Olazarán, 2004). However, further improvement in the MCI population by therapeutic interventions is still elusive, largely because of the characteristic heterogeneity of AD (Horr, 2015). Thus, developing a more granular understanding of the MCI stage of AD could provide new directions for further research and elucidate better targets for clinical applications (Mitelpunkt, 2020).

Metabolomics is an emerging omics field and provides a temporally specific and global view of an individual's ongoing biological processes associated with disease phenotypes (Dettmer, 2007). Metabolomics is sensitive to transient biological processes, transcriptomics data have uncovered more targeted biomarkers in cancer studies as compared to other omics platforms (Gao, 2018). While there is a

growing body of metabolomics research for AD (Wilkins, 2018), the multi-omics AD study to integrate metabolomics to other omics data types is currently lacking (Badhwar, 2020). Therefore, in this study we aim to use the multi-omics approach to better understand the heterogeneity of AD and to identify coherent subtypes (Badhwar, 2020; Liu, 2014).

In particular, we hypothesize that integrating transcriptomics and metabolomics data will yield more subtypes in MCI, an important pre-clinical stage of the AD continuum. Using data from MCI patients in the ADNI cohort, we applied an unsupervised multi-omics integration method called Similarity Network Fusion (SNF) (Wang, 2014) and identified two subtypes in EMCI (EMCI-1, EMCI-2) and LMCI (LMCI-1 and LMCI-2) groups, respectively. We confirmed these subtypes by various approaches: (1) we compared them with cognitively normal (CN) and AD patients to illuminate the characteristics of MCI in AD progression through different stages; (2) we performed pseudotime construction and trajectory analysis using the clinical information to validate the gradual change of subtypes' severities; (3) we conducted survival analysis to demonstrate the trends of earlier onset time of AD in the order of LMCI-2, LMCI-1, EMCI-2, EMCI-1 groups.

Materials & Methods

The ADNI Dataset

The ADNI consortium was launched in 2004 (Alzheimer's Disease Neuroimaging Initiative, 2017). This consortium began as a public-private collaborative effort across many countries. There have been multiple phases of the ADNI project, starting with ADNI-1, followed by ADNI-GO, then ADNI-2, and ADNI-3. The ADNI database includes clinical data, cognitive assessments, genomics and metabolomics data, MRI images, PET images, and biospecimens (Mueller, 2005). In our analyses, we used transcriptomics and metabolomics data. There are 681 patients with both metabolomics and transcriptomics data from blood samples. Among them, 401 patients are classified as either EMCI (201) or LMCI (200), as shown in **Table 1.**

Transcriptomics and metabolomics data preprocessing.

The raw transcriptomics data from ADNI contain gene expression data from over 48,000 transcripts. We used the average expression values of multiple transcripts to represent each gene, resulting in 20,032 genes. We used median values among duplicated assays, followed by quantile normalization before downstream analysis.

For the metabolomics data from ADNI, we first combined P180 Kit data from ADNI-1 and ADNI-2/GO, obtained from previous reports (St John-Williams, 2019). To correct the batch effects, the correction factor for each analyte was calculated by dividing its average on a given plate by its global average. Metabolites were then removed if their coefficient of variation across plates was greater than 20% or more than 40% of its measurements across all subjects were not available. We excluded 107 patients from our analyses who reportedly did not fast before the blood draw, to minimize the confounding effect due to the diet. We also consolidated data from the 36 blinded replicates by calculating the average of the blinded duplicates for each of these subjects. Lastly, we log-2 transformed the values, centered and scaled the data, and replaced any values that were >3.0 standard deviation away from the center in either direction with 3.0 and -3.0, following the original reports. Such procedures yield 172 metabolites in 1,572 patients across every stage of AD, among which 681 are classified as MCI.

Adjustment of metabolomics and transcriptomics data by cell type

We adjusted the transcriptomics data from PBMC by proportions of different cell types in leukocytes using the GEDIT package (Nadel, 2021). Specifically, we adjusted for CD8 T-Cell, CD4 T-Cell, CD56 Natural Killer Cell, CD14 Monocytes, and B-Cell proportions. We used a reference dataset from one of GEDIT's reference matrices, the Human Body Atlas (Su, 2004), with the abovementioned five cell types.

Dimension reduction of transcriptomics and metabolomics data using autoencoders

Given the considerable differences in the total numbers of metabolomics and genes (172 vs. ~20,000), we first used an autoencoder (Yasenko, 2020) to reduce the dimensionality of gene expression to 172, to match the metabolomics data. We set the bottleneck layer of the autoencoder for gene expression as 172. We then integrate the metabolomics data with a higher-order representation of the transcriptomics data from the autoencoder's bottleneck layer. This integration of transcriptomics data and metabolomics data forms the basis for further analyses.

Subtype identification within EMCI and LMCI

This used SNF for integration between gene expression and metabolomics data (Wang, 2014). SNF first constructs sample similarity matrices for each omics data, and then aggregates these matrices in a non-linear fashion. We first calculated the Silhouette Coefficients (SC) and Calinski Harabasz (CH) Scores for cluster sizes ranging from 2 to 15 to find the optimal number of clusters to fit the data from EMCI and LMCI patients, separately. An SC ranges between -1 to 1 to determine how distinguishable the clusters are from one another, with values closer to 1 being more distinguishable.

Differential Expression analysis

With these newly defined subgroups, we first performed differential expression (DE) analysis to find differentially expressed genes, between the EMCI and LMCI subtypes. We added covariates to the R *limma* package, to adjust for age, gender, education, race, marital status, and ApoE-ε4 carrier status. We also adjusted the DE by leukocyte cell types. We then performed gene set enrichment analysis (GSEA) on the DE genes ($p < 0.05$) to uncover the upregulated and downregulated gene sets. For metabolomics data, we conducted DE analysis similarly using the Limma package, followed by enrichment analysis by MetaboAnalyst 5.0 web tool (Pang et. al. 2021).

Contextualizing Patient Subtypes with Clinical Data

We clustered patients within the EMCI and LMCI group using metabolomics and transcriptomics data and then checked whether these subsamples showed any significant differences in cognitive measures, which are relied upon to diagnose patients as healthy, MCI, or AD (Edmonds, 2019). The ADNI consortium includes data from a battery of cognitive tests, with key measures being the Mini-Mental State Exam (MMSE) and the Clinical Dementia Rating (CDR). We analyzed differences between the subtypes for these metrics and additionally investigated patients' scores on the Functional Activities Questionnaire (FAQ), the Alzheimer's Disease Assessment Scale (ADAS-11, ADAS-13, ADAS-Q4), the Trail Making Test (TRABSCOR), the Delayed Recall Total score (LDELTOTAL), and subtests from the Rey Auditory Verbal Learning Test (RAVLT-Immediate, RAVLT-Learning, RAVLT-Forgetting, RAVLT-Percent Forgetting). For each group, we first calculated its average scores on each of the cognitive measures, and since the tests use different metrics, we then scaled and centered the values across all the groups. We used the Complex Heatmap package's default clustering method (distance metric = ‘euclidean’, linkage method = ‘complete’) (Gu, 2016) to measure the proximity between subtypes along with the progression from CN to AD, based on their cognitive measures.

Trajectory analysis for patient subgroups

Trajectory analysis represents an ordering of patient subtypes based on consecutive states of disease progression, from a completely normal state (root node) to a final state (end node). For any clinical trajectory, pseudotime reconstruction quantifies the progression between states or nodes based on the data used in the analysis. In the current study, to evaluate the severity and order of dementia progression, we compared the cognitive scores of EMCI and LMCI subtypes with those of CN and AD patients. We used the ‘ClinTrajan’ python package (Golovenkin et. al, 2020) to quantify the pseudotime of dementia progression states based on the cognitive measures. We set the parameters for principal tree calculation as a number of nodes = 40, α=0.01, μ=0.1, λ=0.05. We chose the trajectory that goes from CN patients (with the least dementia risk) as root nodes to AD patients (highest level dementia and cognitive impairment).

Time-to-event (dementia diagnosis) analysis

Time to event survival analysis shows the differences in the probabilities of dementia between each state during the progression of disease. We used the longitudinal clinical diagnosis data from ADNI with three labels: 'Control', 'MCI' and 'Dementia' and checked the latest diagnosis of each patient after their first visit every six months. We used their latest diagnosis as the event at each checkpoint and calculated the percentage of the label 'Dementia' among all of the three labels. And at last we plotted the dementia percentage against each checkpoint time for 96 months.

Code availability

The code for this project is written in R and Python, and available at https://github.com/xinformatics/adnisubtyping

Results

Mutliomics integration through similarity network fusion

First, considering there are many more transcriptomics features than metabolomics features, we reduced the dimensionality of the transcriptomics dataset from 20,032 to 172 using a stacked autoencoder and setting the bottleneck layer size to 172 (**Fig. 1a**). Then transcriptomics data were split into train and test sets in a ratio of 80:20. We trained the autoencoder on the training set and evaluated the Mean Squared Error (MSE) metric on the test set (Test Set - MSE: 0.1175, suggesting a high fidelity representation of the transcriptomics data. Our model training procedure minimized the MSE metric at epoch 10 (**Fig. 1b)**, Hence we saved the model weights at epoch 10 for further use. For data integration, we extracted the output of the autoencoder bottleneck layers, which represented a low-dimensional representation of the transcriptomics data. We calculated the patient affinity matrix for both transcriptomics and metabolomics as the input to the similarity network fusion method (**Fig 1c**) (Wang, Bo, et al. 2014)

Next, we performed spectral clustering on the SNF-integrated patient affinity matrix in the EMCI cohort. We varied the number of clusters from 2 to 10 and calculated the SC and CH scores, two metrics measuring the fitness of the subclusters, respectively. We found that in the multi omics case, the optimum number of clusters is 2 for both metrics (**Fig. 2a and b**). We performed differential expression (DE) analysis for both the cell-type adjusted transcriptomics data and metabolomics data using the limma package and identified a high number of DE genes and DE metabolites illustrating a clear difference between the clusters. In the multiomics approach for subpopulations within the EMCI groups, there were 235 DE genes (**Fig. 2c**). The top ten differentially expressed genes between the EMCI subpopulations are LUC7L3, KDM6A, ZMAT1, HK3, XIST, PNISR, CHORDC1, FBXW7, ZNF140, and DDX3Y. For the metabolomics data, there were 103 differentially expressed metabolites for the EMCI group (**Fig. 2d**). For the EMCI group, the top ten differentially expressed metabolites are primarily Sphingolipids and Glycerophospholipids: Glycerophospholipid (PC ae C42:2), Sphingolipid SM C16:0, Glycerophospholipid (PC aa C36:1), Glycerophospholipid (PC ae C44:3), Glycerophospholipid (PC ae C42:3), Glycerophospholipid (PC ae C36:3), Glycerophospholipid (PC ae C34:1), Glycerophospholipid (PC ae C36:1), Glycerophospholipid (PC ae C34:2), Sphingolipid SM C24:0.

Similarly, we performed spectral clustering on the SNF-integrated patient similarity matrix in the LMCI group. We varied the number of clusters from 2 to 10, and found that the optimum number of clusters was 2, using SC and CH scores (**Fig. 2e and f**). Similar to the EMCI patients, we also found a large number of DE genes and DE metabolites between the LMCI subtypes: 329 DE genes and 121 DE metabolites (**Fig. 2g and h**). The top ten differentially expressed genes between the LMCI subpopulations were ABCB10, ABCB4, ADAP2, ADCY4, ADD3, ADO, AGGF1, AKAP11, AKAP13, AKIRIN1 (**Fig. 2g**). The top ten differentially expressed metabolites were Sphingolipid SM (OH) C22:2, Glycerophospholipid (PC ae C34:1), Sphingolipid SM C16:1, Glycerophospholipid (PC ae C32:2), Sphingolipid SM C20:2,

Sphingolipid SM (OH) C14:1, Glycerophospholipid (PC ae C34:0), Glycerophospholipid (PC aa C32:0), Sphingolipid SM C18:0, Glycerophospholipid (PC aa C36:1) (**Fig. 2h**).

Gene Set Enrichment Analysis and Metabolomic enrichment for the subtypes in EMCI group

We next conducted systems biology analysis using the DE genes. In the EMCI group, with the GSEA analysis, we found 5 significantly ($P<0.05$) enriched Gene Ontology (GO) terms and 4 significant suppressed GO terms in the EMCI-2 subtype compared to the EMCI-1 subtype (**Fig. 3a**). The activated GO terms are: defense response to other organisms, myeloid cell activation involved in immune response, neutrophil activation involved in immune response, myeloid leukocyte mediated immunity, and multi-organism process. The suppressed GO functions in the EMCI-2 subtype compared to the EMCI-1 subtype are nuclear body, zinc ion binding, neuron part, and transition metal ion binding. The network of enriched or suppressed pathways and their leading edge genes for the EMCI subtypes is shown in **Fig. 3b**. The zinc ion binding pathway is distinct and disconnected with the other pathways, and its leading-edge genes are CHORDC1, CPA3, MIB1, NEXL1, SEC23A, UBR3, UBR5, ZMAT1, ZNF117, ZNF84, and ZNF92. For the remaining pathways, the leading edge genes are BPI, CAMP, DDX3Y, GBP3, HK3, IF127, IRF7, LCN2, LTF, PGLYRP1, RETN, RPS4Y1, and SLPI.

For the metabolomic pathways, EMCI-2 shows two activated and five suppressed pathways compared to EMCI-1 (**Table 2; Fig. 3c**). The activated pathways are aminoacyl-tRNA biosynthesis and histidine metabolism, with leading-edge metabolites Glycine, L-Proline, and L-Glutamic acid for aminoacyl-tRNA biosynthesis, and L-Glutamic acid, histamine, and histidine decarboxylase for histidine metabolism. The suppressed pathways are Linoleic acid metabolism (Linoleic acid, PC(16:0/16:0)), Glutathione metabolism (Glycine, L-Glutamic acid), Porphyrin and chlorophyll metabolism (Glycine, L-Glutamic acid), Glyoxylate and dicarboxylate metabolism (Glycine, L-Glutamic acid), and Glycerophospholipid metabolism (PC(16:0/16:0),
LysoPC(18:1(9Z))

Gene Set Enrichment Analysis and Metabolomic enrichment for the subtypes in LMCI group

For the LMCI subtypes, we found five significantly (P<0.05) activated and five suppressed GO pathways in the LMCI-2 group compared to the LMCI-1 group (**Fig. 3d**). The five activated pathways are myeloid leukocyte-mediated immunity, leukocyte mediated immunity, exocytosis, cell activation involved in immune response, and leukocyte activation involved in the immune response. The five suppressed GO terms in LMCI-2 are lymphocyte differentiation, adaptive immune response, leukocyte differentiation, zinc ion binding, and DNA repair. The network of activated pathways and their leading edge genes for the LMCI subtypes can be seen in **Fig 3e**. The leading edge genes are BPI, CTSG, DEFA1, DEFA4, ELANE, HK3, KIR3DL1, LCN2, MMP9, MPO, ORM1, PGLYRP1, PLEK, RIF1, and SPON2.

For the significantly regulated metabolomic pathways in the LMCI-2 vs LMCI-1 subtypes, there are eight activated pathways, all related to amino acid metabolism (**Table 3; Fig 3f**). These pathways are:: aminoacyl-tRNA biosynthesis, valine, leucine and isoleucine biosynthesis, alanine, aspartate and glutamate metabolism, arginine biosynthesis, arginine and proline metabolism, nitrogen metabolism, D-glutamine and D-glutamate metabolism, and histidine metabolism. The leading edge metabolites are shown in Fig 3f. There is no suppressed metabolic pathway.

Validation of molecular subtypes by clinical phenotypes

We first compared the cognitive measures among the subtypes and found a clear progression from healthy controls to AD (**Fig. 6a**). For some cognitive measures, a lower score means cognitively healthier such as ADAS-11, ADAS-13, ADAS-Q4, RAVLT-Forgetting, RAVLT-Percent Forgetting, CDRSB, FAQ, TRABSCOR). For other measures, a higher score means cognitively healthier, such as RAVLT-Learning, RAVLT-Immediate, LDELTOTAL, and MMSE. As shown in the heatmap clustering result, there is a clear transition from control, through EMCI and LMCI to AD patients. Control patients show the lowest scores for ADAS-11, ADAS-13, ADAS-Q4, RAVLT-Forgetting, RAVLT-Percent Forgetting, CDRSB, FAQ, and

TRABSCORE, and the highest scores for RAVLT-Learning, RAVLT-Immediate, LDELTOTAL, and MMSE. Progressing from healthy controls to EMCI-1, EMCI-2, LMCI-1, LMCI-2, and AD, this pattern gradually reverses (**Table 4**). Interestingly, the EMCI-2 and LMCI-1 subtypes are clustered next to one another, further demonstrating the fine-grained subtle difference between EMCI and LMCI subtypes, in the continuum of disease progression.

In order to validate the progression of severity among the subtypes, we concatenated the cognitive measures from the EMCI and LMCI subtypes with those from dementia patients and healthy controls, and computed the clinical trajectory using the elastic principal tree (EPT) algorithm. A principal tree comprises an assembly of principal curves representing the topology between the samples (Albergante et. al 2020). We set the trajectory root node as the healthy control patients with the least risk of dementia (**Fig 4b**). EPT allows us to visualize the location of patients in each category on the tree, from healthy control to MCI and AD. There is a gradual shift in patients' risk categories while traversing along the trajectory, illustrating the consecutive states of dementia progression among the subtypes (**Fig 4c**). There is also a high correlation (r2 value = 0.86) between the dementia subtypes along the categories and the reconstructed pseudotime (**Fig 4d**). These results again validated the gradual shift in dementia risk in the order of the patients from CN, through EMCI-1, EMCI-2, LMCI-1, LMCI-2 to AD, as reflected by the trajectory of dementia progression. Lastly, we also examined the percentage of the patients in each subtype who were later clinically diagnosed as having “dementia” in the following check-ups every six months since their first screening (**Fig 4e**). Again it shows higher percentages of dementia in each check-up time, when the states change from the earliest stage of EMCI-1 to the latest stage of LMCI-2. Thus, using different analyses, the clinical relevance of the four subtypes within MCI patients are confirmed.

Discussion

The aim of the current study is to identify clinically relevant molecular subtypes in the MCI cohort of AD patients, by integrating metabolomics and gene expression data.. We focused on patients in both the EMCI and the LMCI groups in the ADNI dataset, as the MCI stage is a sensitive window important for therapeutic intervention. We found that both EMCI and LMCI can be further split into two relatively even subtypes, which we call EMCI-1, EMCI-2, LMCI-1, and LMCI-2. Based on the relationship with the cognitive measures, trajectory analysis and the longitudinal dementia outcome of these patients, the progression of severity in the order of EMCI-1, EMCI-2, LMCI-1, and LMCI-2 is validated. At molecular level, these subtypes differ significantly based on gene expression, metabolite intensity, and biological pathways. To the best of our knowledge, this is the first time these biomarkers have been reported between subtypes in the EMCI group and LMCI group, and have been linked to cognitive measures at such a granular level.

Some dysregulated pathways are shared between the EMCI and LMCI subtypes. These include aminoacyl-tRNA biosynthesis, leukocyte-mediated immunity, and zinc ion binding. Within the EMCI and LMCI subtypes, both aminoacyl-tRNA biosynthesis and leukocyte-mediated immunity are more activated in the more severe subtype. However, zinc ion binding pathway is suppressed in the more severe subtype, for both EMCI and LMCI. The consistent trends of changes between EMCI and LMCI subpopulations suggests these pathway changes may exist on a continuum as the disease progresses, similar to report in other studies (Pietronigro, 2016; Yuan, 2014).

The importance of the immune system and inflammation in the progression of AD has been well known(McGeer, 1989; Akiyama, 2000). During the course of the disease, there is a breakdown in the blood-brain barrier (BBB; Song, 2015), allowing leukocytes to enter the brain (Pietronigro, 2016). Neutrophils are also active in chronic inflammatory diseases, which can contribute to tissue damage in diseases such as AD (Wu, 2020). AD patients have higher levels of reactive oxygen species in the peripheral blood neutrophils (Dong, 2018). Our results suggest that the immune response and neuroinflammation characteristic of AD can further be distinguished between patient subtypes as early as the EMCI stage. Both

EMCI-2 and LMCI-2 also show higher activation of the aminoacyl-tRNA biosynthesis. However, the activity of Aminoacyl-tRNA as it relates to AD disease progression is complex and not well understood (Kron, 2013), although numerous studies have mentioned this pathway in their results (Trushina, 2013; Wang, 2020; Hou, 2020). A suppressed zinc ion binding pathway, on the other hand, has been linked to oxidative stress by means of lipid peroxidation, an increase in osmotic fragility of erythrocyte membranes, and a decrease in protein turnover (Jurowski, 2014; Roohani, 2013). Suppressed zinc ion binding pathway may suggest accumulation of zinc, which can worsen neurodegenerative diseases, leading to an increase in Amyloid-beta proteins and neurofibrillary tangles (Wojtunik-Kulesza, 2019).

There also exist differences in enriched pathways, comparing EMCI subtypes to LMCI subtypes. For the EMCI subtypes, oxidative stress is activated while linoleic acid metabolism is repressed in the more severe subtypes. However, for LMCI subtypes, amino-acid metabolism related pathways are significantly dys-regulated in the more severe subtype. For example, pathways involved in arginine do not reach significance during the EMCI stages, but significantly differentiate the LMCI-2 subpopulation from LMCI-1 patients. These results suggest that LMCI have bigger variations compared to EMCI subtypes.

Noticeably, in the EMCI-2 subtype, lipid pathways, including linoleic acid metabolism and glutathione metabolism are suppressed. Linoleic acid has antioxidant, anti-inflammatory, and anti-apoptotic effects (Zhang, 2020), and has been shown that levels of linoleic acid decrease progressively over the course of Alzheimer’s disease (Iuliano, 2013). Glutathione is considered the most important component of the brain's antioxidant mechanisms (Liu, 2005). Suppression of glutathione metabolism in the EMCI-2 subpopulation compared to EMCI-1 appears to correspond well to the increased oxidative stress activities.

For the LMCI group, the activated pathways include valine, leucine, and isoleucine biosynthesis, alanine, aspartate, and glutamate metabolism, and arginine biosynthesis. Dysregulation in the glutamatergic system is well-documented for AD etiology and can affect memory, cognition, and behavior (Findley, 2019). In

addition, an increase in glutamate induces excitotoxicity and cell death in AD patients (Wang, 2017). A dysregulation of arginine also has been well-documented in AD etiology (Altiné-Samey, 2021), and studies have reported that there is an abnormally high amount of arginine used in AD brains, but a decreased bioavailability of this metabolite (Kan, 2015). By contrast, the roles of both alanine and aspartate are poorly understood (Graham, 2014). For the valine, leucine, and isoleucine biosynthesis pathway, there are conflicting findings showing that MCI patients exhibit an increase in these amino acids compared to AD (Di Costanzo, 2020), while other studies have shown the opposite, that there are lower levels of valine specifically in MCI, and higher levels of valine in AD (Ibáñez, 2012).

There are some limitations worth mentioning in our current work. Most notably, even though the ADNI consortium represents a large-scale collaborative effort and contains a comprehensive database of elderly patients, the population is overwhelmingly white (92% in EMCI, 94.5% in LMCI). Additionally, our sample sizes were relatively small. Confirmation of the observations to other ethnicities or racial groups or cohorts should be followed up.. Nevertheless, our study revealed subtypes in the EMCI and LMCI groups, and identified critical role oxidative stress and metabolic pathways associated with progression in pre-AD continuum. Such knowledge will help guide more effective drug development tailored towards subpopulations of patients in the MCI stage.

References:

Akiyama, Haruhiko, et al. "Inflammation and Alzheimer's disease." Neurobiology of aging 21.3 (2000): 383-421.

Albergante, L.; Mirkes, E.; Bac, J.; Chen, H.; Martin, A.; Faure, L.; Barillot, E.; Pinello, L.; Gorban, A.; Zinovyev, A. Robust and Scalable Learning of Complex Intrinsic Dataset Geometry via ElPiGraph. Entropy 2020, 22, 296.

Alzheimer's Disease Neuroimaging Initiative. (2017). Adni : about. ADNI. Retrieved March 29, 2022, from https://adni.loni.usc.edu/about/

Alzheimer's Association, Thies, W., & Bleiler, L. (2013). 2013 Alzheimer's disease facts and figures. Alzheimer's & dementia, 9(2), 208-245.

Amoroso, N., Diacono, D., Fanizzi, A., La Rocca, M., Monaco, A., Lombardi, A., ... & Alzheimer's Disease Neuroimaging Initiative. (2018). Deep learning reveals Alzheimer's disease onset in MCI subjects: results from an international challenge. Journal of neuroscience methods, 302, 3-9.

Arevalo-Rodriguez, I., Smailagic, N., i Figuls, M. R., Ciapponi, A., Sanchez-Perez, E., Giannakou, A., ... & Cullum, S. (2015). Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI). Cochrane Database of Systematic Reviews, (3).

Badhwar, A., McFall, G. P., Sapkota, S., Black, S. E., Chertkow, H., Duchesne, S., ... & Bellec, P. (2020). Brain, 143(5), 1315-1331.

Ball, D., Adames, N., Reischmann, N., Barik, D., Franck, C., Tyson, J. J., & Peccoud, J. (2013). Measurement and modeling of transcriptional noise in the cell cycle regulatory network. Cell Cycle, 12(19), 3392-3407.

Ballardsych, C., Gauthier, S., Corbett, A., Brayne, C., Aarsland, D., & Jones, E. (2011). Alzheimer's disease. Lancet, 377(9770), 1019-1031.

Beales, A., Bates, K., Cartwright, J., & Whitworth, A. (2019). Lost for words: Perspectives and experiences of people with primary progressive aphasia and Alzheimer's disease and their families of participation in a lexical retrieval intervention. International Journal of Speech-Language Pathology, 21(5), 483-492.

Breijyeh, Z., & Karaman, R. (2020). Comprehensive review on Alzheimer's disease: Causes and treatment. Molecules, 25(24), 5789.

Chattopadhyay, A., & Lu, T. P. (2019). Gene-gene interaction: the curse of dimensionality. Annals of translational medicine, 7(24).

Chipi, E., Salvadori, N., Farotti, L., & Parnetti, L. (2019). Biomarker-based signature of Alzheimer's disease in pre-MCI individuals. Brain Sciences, 9(9), 213.

Cisbani, G., & Rivest, S. (2021). Targeting innate immunity to protect and cure Alzheimer's disease: Opportunities and pitfalls. Molecular Psychiatry, 1-12.

Connor, S. C., Hansen, M. K., Corner, A., Smith, R. F., & Ryan, T. E. (2010). Integration of metabolomics and transcriptomics data to aid biomarker discovery in type 2 diabetes. Molecular BioSystems, 6(5), 909-921.

Cummings, J. L., Doody, R., & Clark, C. (2007). Disease-modifying therapies for Alzheimer disease: challenges to early intervention. Neurology, 69(16), 1622-1634.

Dettmer, K., Aronov, P. A., & Hammock, B. D. (2007). Mass spectrometry-based metabolomics. Mass spectrometry reviews, 26(1), 51-78.

Di Costanzo, Alfonso, et al. "Blood biomarkers indicate that the preclinical stages of Alzheimer's disease present overlapping molecular features." Scientific reports 10.1 (2020): 1-19.

Dong, Yuan, et al. "Neutrophil hyperactivation correlates with Alzheimer's disease progression." Annals of neurology 83.2 (2018): 387-405.

Edmonds, E. C., McDonald, C. R., Marshall, A., Thomas, K. R., Eppig, J., Weigand, A. J., ... & Alzheimer's Disease Neuroimaging Initiative. (2019). Early versus late MCI: Improved MCI staging using a neuropsychological approach. Alzheimer's & Dementia, 15(5), 699-708.

Farooqui, Akhlaq A., Wei-Yi Ong, and Tahira Farooqui. "Lipid mediators in the nucleus: Their potential contribution to Alzheimer's disease." Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids 1801.8 (2010): 906-916.

Findley, Caleigh A., et al. "Amyloid beta-related alterations to glutamate signaling dynamics during Alzheimer's disease progression." ASN neuro 11 (2019): 1759091419855541.

Frank, E. M. (1994). Effect of Alzheimer's disease on communication function. Journal of the South Carolina Medical Association (1975), 90(9), 417-423.
Frisardi, Vincenza, et al. "Glycerophospholipids and glycerophospholipid-derived lipid mediators: a complex meshwork in Alzheimer's disease pathology." Progress in lipid research 50.4 (2011): 313-330.
Fujisaka, S., Avila-Pacheco, J., Soto, M., Kostic, A., Dreyfuss, J. M., Pan, H., ... & Kahn, C. R. (2018). Diet, genetics, and the gut microbiome drive dynamic changes in plasma metabolites. Cell reports, 22(11), 3072-3086.
Ganeshan, K., & Chawla, A. (2014). Metabolic regulation of immune responses. Annual review of immunology, 32, 609-634.
Gao, Galen F., et al. "Before and after: comparison of legacy and harmonized TCGA genomic data commons' data." Cell systems 9.1 (2019): 24-34.
Gatz, M., Pedersen, N. L., Berg, S., Johansson, B., Johansson, K., Mortimer, J. A., ... & Ahlbom, A. tool(1997). Heritability for Alzheimer's disease: the study of dementia in Swedish twins. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 52(2), M117-M125.
Gelman, C. R., & Rhames, K. (2018). In their own words: the experience and needs of children in younger-onset Alzheimer's disease and other dementias families. Dementia, 17(3), 337-358.
Graham, Stewart F., Christian Holscher, and Brian D. Green. "Metabolic signatures of human Alzheimer's disease (AD): 1H NMR analysis of the polar metabolome of post-mortem brain tissue." Metabolomics 10.4 (2014): 744-753.
Gu, Zuguang, Roland Eils, and Matthias Schlesner. "Complex heatmaps reveal patterns and correlations in multidimensional genomic data." Bioinformatics 32.18 (2016): 2847-2849.
Habes, M., Grothe, M. J., Tunc, B., McMillan, C., Wolk, D. A., & Davatzikos, C. (2020). Disentangling heterogeneity in Alzheimer's disease and related dementias using data-driven methods. Biological psychiatry, 88(1), 70-82.
Horgusluoglu, E., Neff, R., Song, W. M., Wang, M., Wang, Q., Arnold, M., ... & Alzheimer Disease Metabolomics Consortium. (2021). Integrative metabolomics-genomics approach reveals key metabolic pathways and regulators of Alzheimer's disease. Alzheimer's & Dementia.
Horr, T., Messinger-Rapport, B., & Pillai, J. A. (2015). Systematic review of strengths and limitations of randomized controlled trials for non-pharmacological interventions in mild cognitive impairment: focus on Alzheimer's disease. The journal of nutrition, health & aging, 19(2), 141-153.
Hou, Zhitao, et al. "Metabolomics Study on Improvement of Mild Cognitive Impairment by SCPE: a UPLC-Q/TOF-MS Mass Spectrometric Study." (2020).
Ibáñez, Clara, et al. "Toward a predictive model of Alzheimer's disease progression using capillary electrophoresis–mass spectrometry metabolomics." Analytical chemistry 84.20 (2012): 8532-8540.
Iuliano, Luigi, et al. "Plasma fatty acid lipidomics in amnestic mild cognitive impairment and Alzheimer's disease." Journal of Alzheimer's Disease 36.3 (2013): 545-553.
Jack Jr, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., ... & Silverberg, N. (2018). NIA-AA research framework: toward a biological definition of Alzheimer's disease. Alzheimer's & Dementia, 14(4), 535-562.
Jia, J., Wei, C., Chen, S., Li, F., Tang, Y., Qin, W., ... & Gauthier, S. (2018). The cost of Alzheimer's disease in China and re-estimation of costs worldwide. Alzheimer's & Dementia, 14(4), 483-491.
John-Williams St, Mahmoudiandehkordi L, Arnold S, et al. (2019). Bile acids targeted metabolomics and medication classification data in the ADNI1 and ADNIGO/2 cohorts. Sci Data 6, 212.

Jurowski, Kamil, et al. "Biological consequences of zinc deficiency in the pathomechanisms of selected diseases." JBIC Journal of Biological Inorganic Chemistry 19.7 (2014): 1069-1079.
Kan, Matthew J., et al. "Arginine deprivation and immune suppression in a mouse model of Alzheimer's disease." Journal of Neuroscience 35.15 (2015): 5969-5982.
Khoury, R., & Ghossoub, E. (2019). Diagnostic biomarkers of Alzheimer's disease: a state-of-the-art review. Biomarkers in Neuropsychiatry, 1, 100005.
Korczyn, A. D. (2012). Why have we failed to cure Alzheimer's disease?. Journal of Alzheimer's Disease, 29(2), 275-282.
Kron, Michael, and Michael Härtlein. "Aminoacyl-tRNA synthetases and disease." Madame Curie Bioscience Database [Internet] (2013).
Lee, J. C., Kim, S. J., Hong, S., & Kim, Y. (2019). Diagnosis of Alzheimer's disease utilizing amyloid and tau as fluid biomarkers. Experimental & molecular medicine, 51(5), 1-10.
Li, Shuzhao, Andrei Todor, and Ruiyan Luo. "Blood transcriptomics and metabolomics for personalized medicine." Computational and structural biotechnology journal 14 (2016): 1-7.
Liu, F., Wee, C. Y., Chen, H., & Shen, D. (2014). Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer's Disease and mild cognitive impairment identification. NeuroImage, 84, 466-475.
Liu, Honglei, et al. "Gender differences in glutathione metabolism in Alzheimer's disease." Journal of Neuroscience Research 79.6 (2005): 861-867.
Locascio, J. J., Growdon, J. H., & Corkin, S. (1995). Cognitive test performance in detecting, staging, and tracking Alzheimer's disease. Archives of neurology, 52(11), 1087-1099.
M Bakulski, K., S Rozek, L., C Dolinoy, D., L Paulson, H., & Hu, H. (2012). Alzheimer's disease and environmental exposure to lead: the epidemiologic evidence and potential role of epigenetics. Current Alzheimer Research, 9(5), 563-573.
McGeer, P. L., et al. "Immune system response in Alzheimer's disease." Canadian Journal of Neurological Sciences 16.S4 (1989): 516-527.
Mitelpunkt, A., Galili, T., Kozlovski, T., Bregman, N., Shachar, N., Markus-Kalish, M., & Benjamini, Y. (2020). Novel Alzheimer's disease subtypes identified using a data and knowledge driven strategy. Scientific reports, 10(1), 1-13.
Morris, J. C., & Cummings, J. (2005). Mild cognitive impairment (MCI) represents early-stage Alzheimer's disease. Journal of Alzheimer's Disease, 7(3), 235-239.
Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C., Jagust, W., ... & Beckett, L. (2005). The Alzheimer's disease neuroimaging initiative. Neuroimaging Clinics, 15(4), 869-877.
Mullane, K., & Williams, M. (2020). Alzheimer's disease beyond amyloid: Can the repetitive failures of amyloid-targeted therapeutics inform future approaches to dementia drug discovery?. Biochemical pharmacology, 177, 113945.
Nadel, B. B., Lopez, D., Montoya, D. J., Ma, F., Waddel, H., Khan, M. M., ... & Pellegrini, M. (2021). The Gene Expression Deconvolution Interactive Tool (GEDIT): accurate cell type quantification from gene expression data. GigaScience, 10(2), giab002.
Olazaran, J., Muñiz, R., Reisberg, B., Peña-Casanova, J., Del Ser, T., Cruz-Jentoft, A. J., ... & Sevilla, C. (2004). Benefits of cognitive-motor intervention in MCI and mild to moderate Alzheimer disease. Neurology, 63(12), 2348-2353.
Pang, Zhiqiang, et al. "MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights." Nucleic acids research 49.W1 (2021): W388-W396.

Parker, T. D., Slattery, C. F., Zhang, J., Nicholas, J. M., Paterson, R. W., Foulkes, A. J., ... & Schott, J. M. (2018). Cortical microstructure in young onset Alzheimer's disease using neurite orientation dispersion and density imaging. Human brain mapping, 39(7), 3005-3017.

Petersen, R. C. (1998). Clinical subtypes of Alzheimer's disease. Dementia and geriatric cognitive disorders, 9(Suppl. 3), 16-24.

Petersen, R. C., Aisen, P. S., Beckett, L. A., Donohue, M. C., Gamst, A. C., Harvey, D. J., ... & Weiner, M. W. (2010). Alzheimer's disease neuroimaging initiative (ADNI): clinical characterization. Neurology, 74(3), 201-209.

Petry, S., Cummings, J. L., Hill, M. A., & Shapira, J. (1988). Personality alterations in dementia of the Alzheimer type. Archives of Neurology, 45(11), 1187-1190.

Pietronigro, Enrica, Elena Zenaro, and Gabriela Constantin. "Imaging of leukocyte trafficking in Alzheimer's disease." Frontiers in Immunology 7 (2016): 33.

Portet, F., Ousset, P. J., Visser, P. J., Frisoni, G. B., Nobili, F., Scheltens, P., ... & MCI Working Group of the European Consortium on Alzheimer's Disease (EADC. (2006). Mild cognitive impairment (MCI) in medical practice: a critical review of the concept and new diagnostic procedure. Report of the MCI Working Group of the European Consortium on Alzheimer's Disease. Journal of Neurology, Neurosurgery & Psychiatry, 77(6), 714-718.

Prince, M., Bryce, R., Albanese, E., Wimo, A., Ribeiro, W., & Ferri, C. P. (2013). The global prevalence of dementia: a systematic review and metaanalysis. Alzheimer's & dementia, 9(1), 63-75.

Qian, X., Hamad, B., & Dias-Lalcaca, G. (2015). The Alzheimer disease market. Nature reviews. Drug discovery, 14(10), 675.

Ritchie, M. E., Phipson, B., Wu, D. I., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research, 43(7), e47-e47.

Roohani, Nazanin, et al. "Zinc and its importance for human health: An integrative review." Journal of research in medical sciences: the official journal of Isfahan University of Medical Sciences 18.2 (2013): 144.

Sancesario, G. M., & Bernardini, S. (2018). Alzheimer's disease in the omics era. Clinical biochemistry, 59, 9-16.

Schneider, L. S., & Sano, M. (2009). Current Alzheimer's disease clinical trials: methods and placebo outcomes. Alzheimer's & dementia, 5(5), 388-397.

Shao, Y., Ye, G., Ren, S., Piao, H. L., Zhao, X., Lu, X., ... & Xu, G. (2018). Metabolomics and transcriptomics profiles reveal the dysregulation of the tricarboxylic acid cycle and related mechanisms in prostate cancer. International journal of cancer, 143(2), 396-407.

Smoller, E. S. (2010). I can't remember: family stories of Alzheimer's disease. Temple University Press.

Song, Juhyun, and Jong Eun Lee. "miR-155 is involved in Alzheimer's disease by regulating T lymphocyte function." Frontiers in aging neuroscience 7 (2015): 61.

Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., ... & Phelps, C. H. (2011). Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & dementia, 7(3), 280-292.

Ströhle, A., Schmidt, D. K., Schultz, F., Fricke, N., Staden, T., Hellweg, R., ... & Rieckmann, N. (2015). Drug and exercise treatment of Alzheimer disease and mild cognitive impairment: a systematic review and

meta-analysis of effects on cognition in randomized controlled trials. The American Journal of Geriatric Psychiatry, 23(12), 1234-1249.
Su, A. I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K. A., Block, D., ... & Hogenesch, J. B. (2004). A gene atlas of the mouse and human protein-encoding transcriptomes. Proceedings of the National Academy of Sciences, 101(16), 6062-6067.
Trushina, Eugenia, et al. "Identification of altered metabolic pathways in plasma and CSF in mild cognitive impairment and Alzheimer's disease using metabolomics." PloS one 8.5 (2013): e63644.
Tynkkynen, Juho, et al. "Association of branched-chain amino acids and other circulating metabolites with risk of incident dementia and Alzheimer's disease: a prospective study in eight cohorts." Alzheimer's & Dementia 14.6 (2018): 723-733.
Wainaina, M. N., Chen, Z., & Zhong, C. (2014). Environmental factors in the development and progression of late-onset Alzheimer's disease. Neuroscience bulletin, 30(2), 253-270.
Wang, B., Mezlini, A. M., Demir, F., Fiume, M., Tu, Z., Brudno, M., ... & Goldenberg, A. (2014). Similarity network fusion for aggregating data types on a genomic scale. Nature methods, 11(3), 333-337.
Wang, Bo, et al. "Similarity network fusion for aggregating data types on a genomic scale." Nature methods 11.3 (2014): 333-337.
Wang, Rui, and P. Hemachandra Reddy. "Role of glutamate and NMDA receptors in Alzheimer's disease." Journal of Alzheimer's Disease 57.4 (2017): 1041-1048.
Wang, Ruiqi, et al. "Circulating metabolites associated with brain MRI markers of Alzheimer's disease: Neuroimaging/imaging and genetics." Alzheimer's & Dementia 16 (2020): e044283.
Wilkins, J. M., & Trushina, E. (2018). Application of metabolomics in Alzheimer's disease. Frontiers in neurology, 8, 719.
Williams, M. M., Storandt, M., Roe, C. M., & Morris, J. C. (2013). Progression of Alzheimer's disease as measured by Clinical Dementia Rating Sum of Boxes scores. Alzheimer's & Dementia, 9(1), S39-S44.
Wojtunik-Kulesza, Karolina, Anna Oniszczuk, and Monika Waksmundzka-Hajnos. "An attempt to elucidate the role of iron and zinc ions in development of Alzheimer's and Parkinson's diseases." Biomedicine & Pharmacotherapy 111 (2019): 1277-1289.
Wu, Che-Yuan, et al. "Neutrophil activation in Alzheimer's disease and mild cognitive impairment: A systematic review and meta-analysis of protein markers in blood and cerebrospinal fluid." Ageing Research Reviews 62 (2020): 101130.
Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., ... & Yu, G. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The Innovation, 2(3), 100141.
Yasenko, L., Klyatchenko, Y., & Tarasenko-Klyatchenko, O. (2020, May). Image noise reduction by denoising autoencoder. In 2020 IEEE 11th International Conference on Dependable Systems, Services and Technologies (DESSERT) (pp. 351-355). IEEE.
Yiannopoulou, Konstantina G., and Sokratis G. Papageorgiou. "Current and future treatments for Alzheimer's disease." Therapeutic advances in neurological disorders 6.1 (2013): 19-33.
Yuan, Ye, et al. "Zinc and its effects on oxidative stress in Alzheimer's disease." Neurological Sciences 35.6 (2014): 923-928.
Zhao, X., Zhang, Z., Hu, B., Huang, W., Yuan, C., & Zou, L. (2018). Response of gut microbiota to metabolite changes induced by endurance exercise. Frontiers in microbiology, 9, 765.

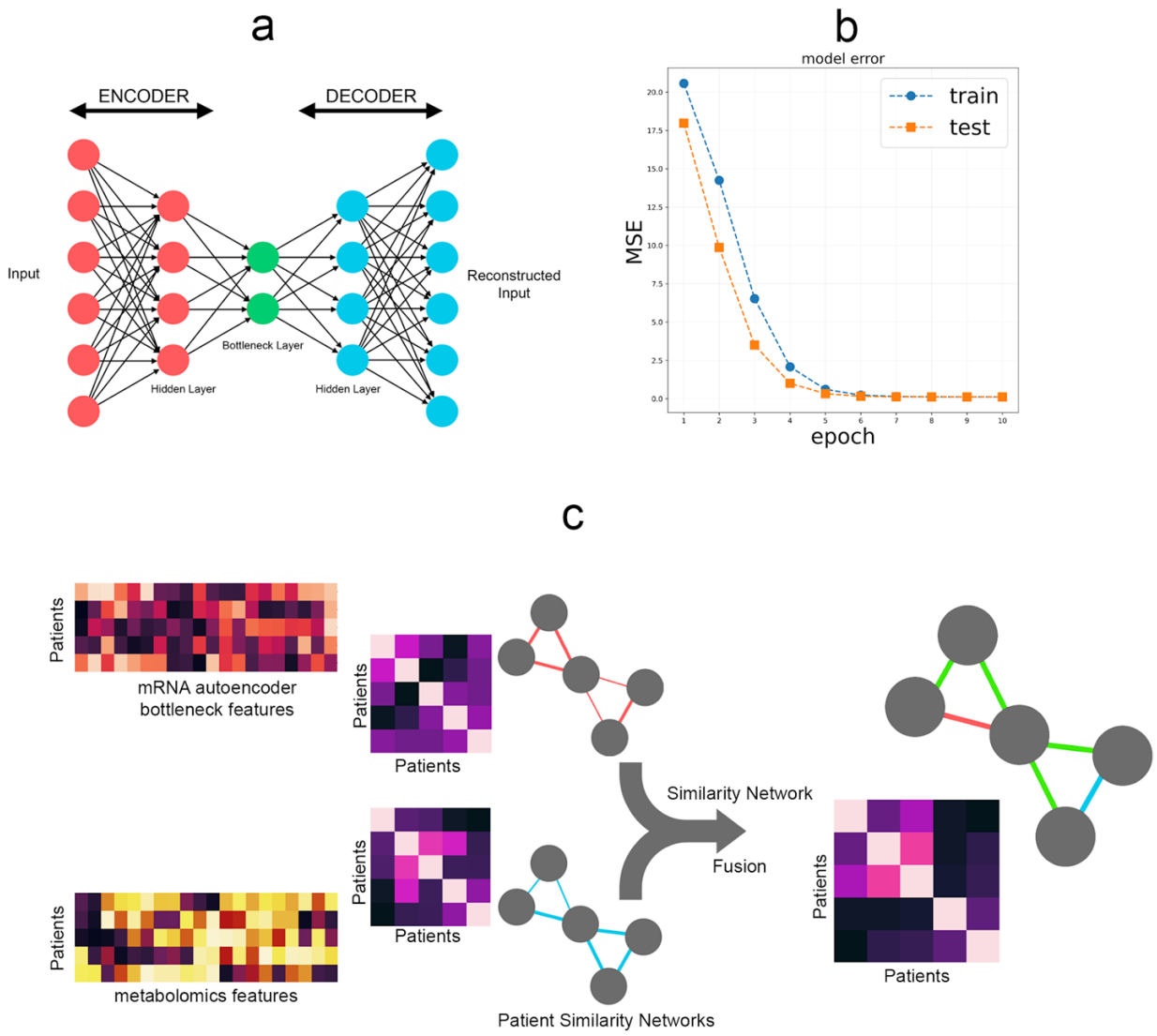

Figure 1 Procedures for integrating metabolomics data with transcriptomics data. (a) Dimensionality reduction for transcriptomics data, using stacked autoencoders. The number of bottleneck layer of features are designed to be the same as the metabolomics features. (b) Mean squared error for both training and test set vs training epoch for the stacked autoencoder showing that the model has converged. (c) Integrating transcriptomics and metabolomics with similarity network fusion (SNF), which are then used to identify subtypes in EMCI and LMCI, respectively.

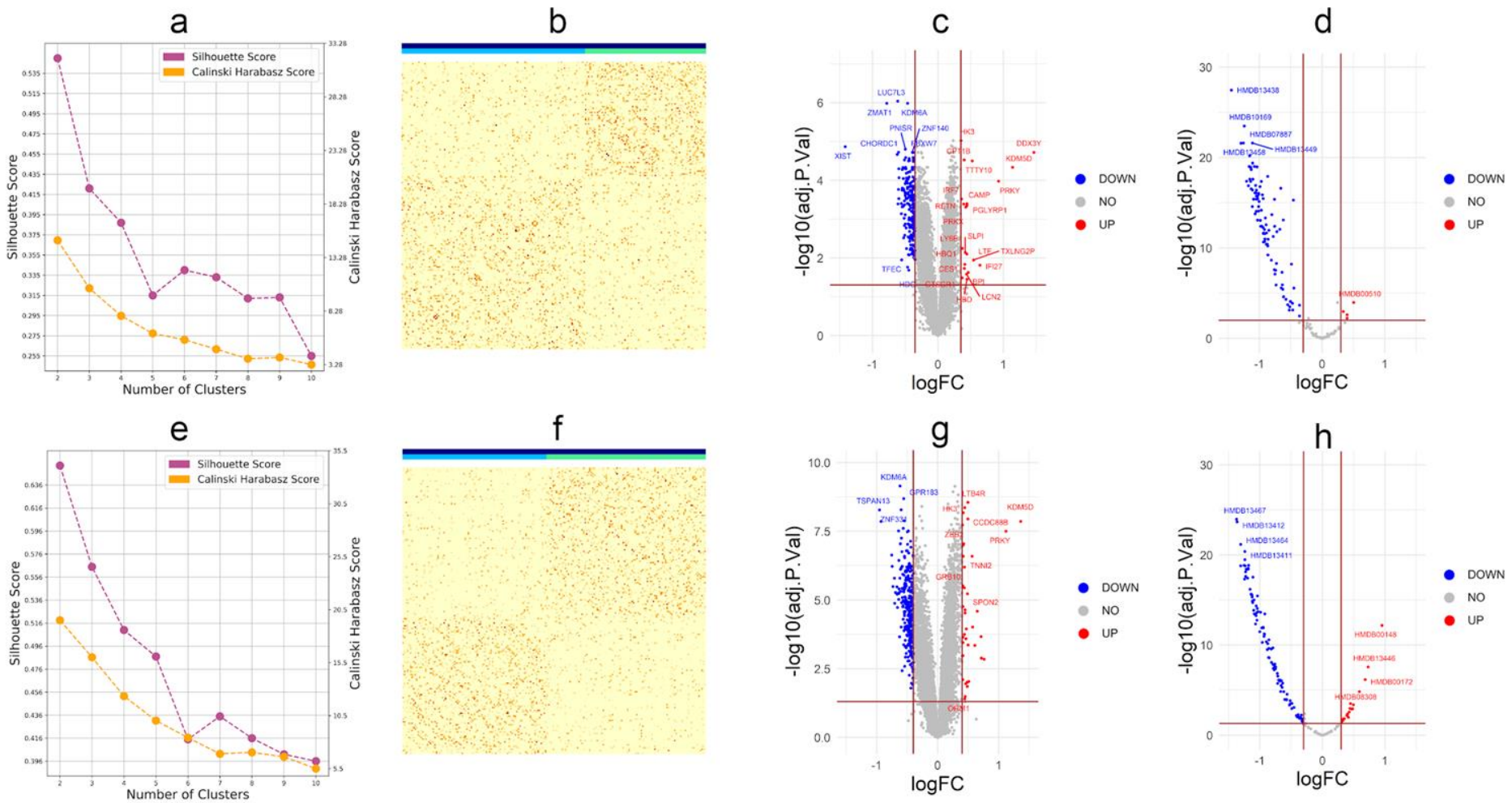

*Figure 2. Multi-omics analyses of the **EMCI** group (top) and the **LMCI** group (bottom) (a) Estimation of optimum number of clusters (from spectral clustering) in SNF integrated data in EMCI samples. Silhouette score and Calinski Harabasz score are used as metrics. (b) Visualization of patient-level spectral clustering results for the two clusters based on SNF integrated data. (c) Differentially expressed genes for the two clusters. (d) Differentially expressed metabolites for the two clusters. (e-h) plots are done the same way as in (a-d), but for the LMCI group.*

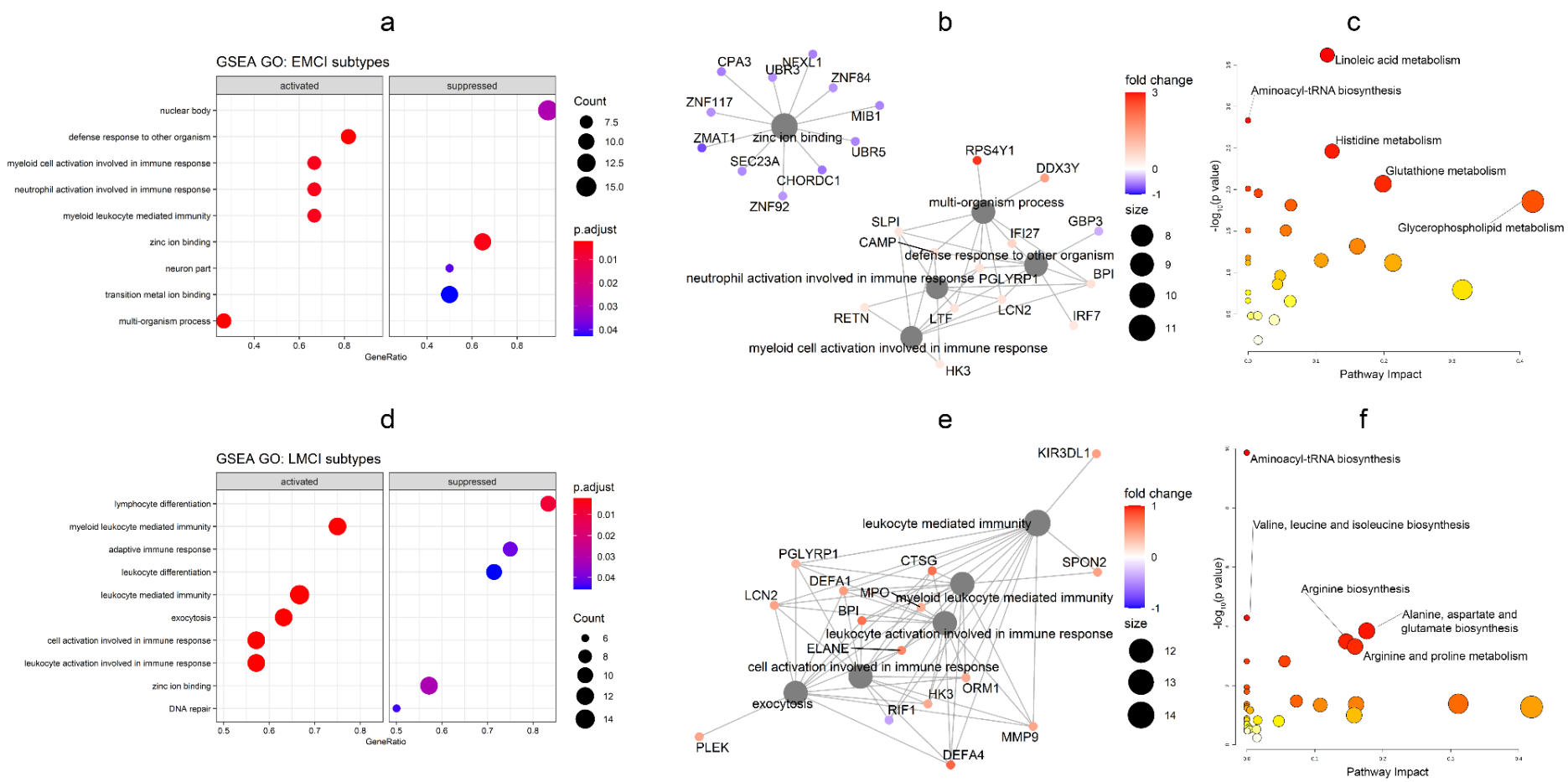

*Figure 5. Gene set enrichment analysis (GSEA) of activated and suppressed pathways between **EMCI-1** vs. **EMCI-2** (top) and **LMCI-1** vs. **LMCI-2** (bottom) using the differentially expressed (DE) genes between the two subtypes. (a) Dysregulated pathways based on DE genes in the EMCI subtypes. (b) Network result showing the DE genes associated with the dysregulated pathways in the EMCI subtypes. (c) Enriched metabolic pathways between the EMCI subtypes. (d-f) Same analyses as done in (a-c), but in LMCI subtypes.*

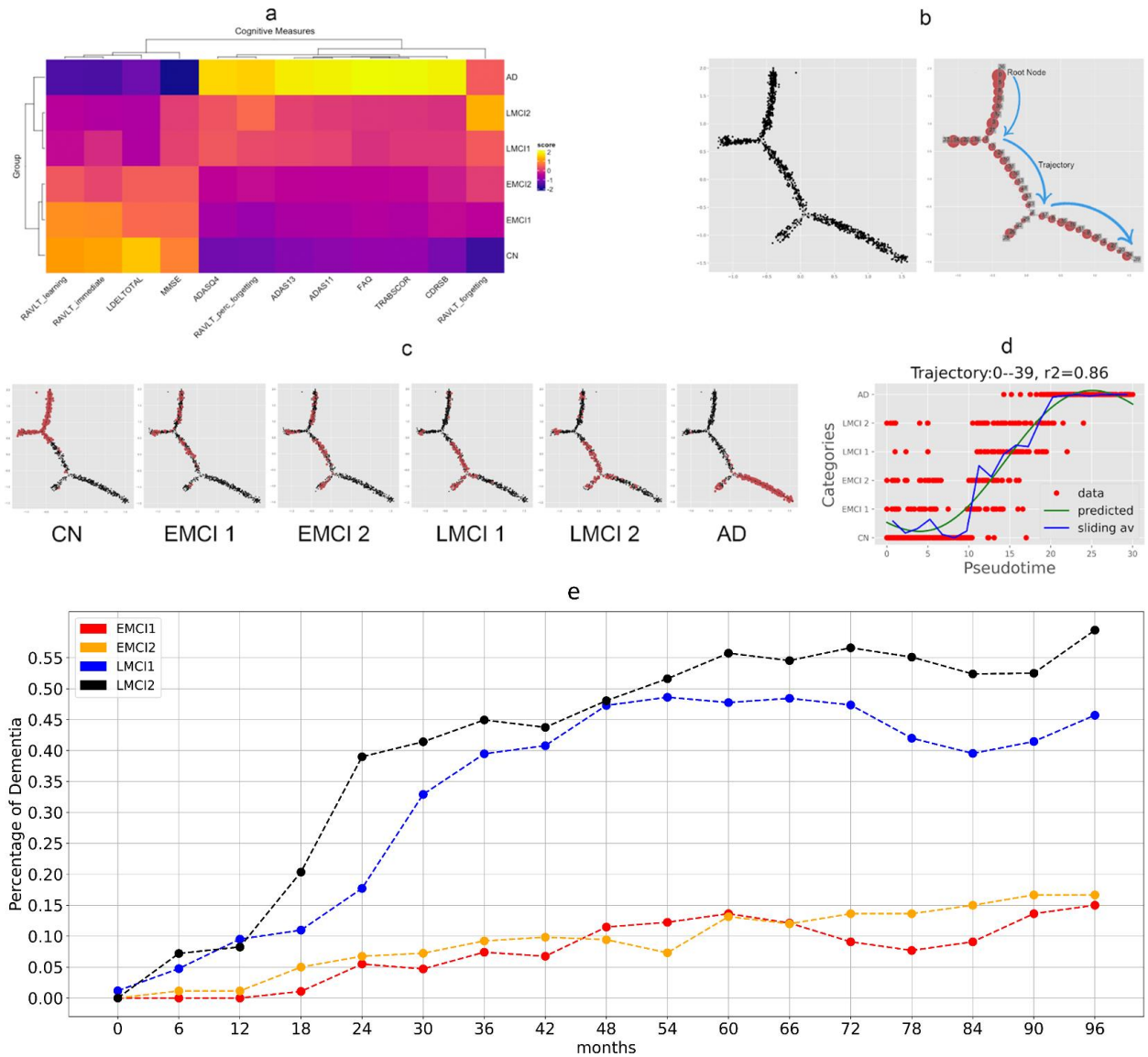

Figure 3. Validation of the molecular subtypes by clinical and phenotypic information. (a) Hierarchical clustering based comparison of the EMCI and LMCI subtypes with 'Control' and 'Dementia' patients based on cognitive measures. (b) Principal tree illustrating the multidimensional structure in the Alzheimer's dataset. Visualization of disease progression trajectory.. Controls are shown as the root node and the patients with highest risk dementia are shown as the final state. (c) Ordered patient phenotypes from 'Control', EMCI-1, EMCI-2, LMCI-1, LMCI-2 and AD(dementia) on the trajectory tree. The consecutive states of progression of Alzheimer disease are evident. (d) Clinical trajectory regression analysis for Trajectory 0-39 (Control-MCI-Dementia) with pseudotime representing the degree of disease progression along the trajectory. (e) Time-to-diagnosis plot for the percentage of dementia in every six-month check up, among the EMCI and LMCI subtypes.

Tables:

Table 1. Descriptive statistics for the subsample of ADNI data used in the current study. This includes patients diagnosed as either EMCI or LMCI at baseline who have both transcriptomics data and CSF metabolomics data available (* - significant differences between the two groups (p < 0.001)).

	No. of cases (n)	**Age (mean)**	**Gender**	**Race (%)**	**ApoE-ε4 carrier**
Early Mild Cognitive Impairment	201	70.85	94 Females	White (92.0) Black (3.5) Asian (1.5) Hawaiian (0.5) More than one (2.5)	64 carriers
Late Mild Cognitive Impairment	200	73.54*	76 Females	White (94.5) Black (2.5) Asian (2) Hawaiian (0.5) More than one (0.5)	74 carriers

Table 3. Significantly activated (+) or suppressed (-) metabolomic pathways in EMCI-2 subtype, compared to EMCI-1, based on metabolite enrichment analysis.

Significant Pathways	Matched Features	Total Features	Matched Compounds	unadjusted p-values	FDR-corrected p-values
Linoleic acid metabolism (-)	Linoleic acid, PC(16:0/16:0)	17	2	0.0002	0.0062
Aminoacyl-tRNA biosynthesis (+)	Glycine, L-Proline, L-Glutamic acid	74	3	0.0015	0.0191
Histidine metabolism (+)	L-Glutamic acid, Histamine, Histidine decarboxylase	32	2	0.0034	0.0300
Glutathione metabolism (-)	Glycine, L-Glutamic acid	56	2	0.0085	0.0479
Porphyrin and chlorophyll metabolism (-)	Glycine, L-Glutamic acid	53	2	0.0097	0.0479
Glyoxylate and dicarboxylate metabolism (-)	Glycine, L-Glutamic acid	56	2	0.0110	0.0479
Glycerophospholipid metabolism (-)	PC(16:0/16:0), LysoPC(18:1(9Z))	86	2	0.0139	0.0500

Table 3. Significantly activated (+) metabolomic pathways in LMCI-2 subtype, compared to LMCI-1, based on metabolite enrichment analysis.

Significant Pathways	Matched Features	Total Features	Matched Compounds	Raw p-values	FDR-corrected p-values
Aminoacyl-tRNA biosynthesis (+)	L-Phenylalanine, L-Glutamine, L-Aspartic acid, L-Valine, L-Alanine, L-Isoleucine, L-Threonine, L-Proline, L-Glutamic acid	74	9	1.36E-10	4.50E-09
Valine, leucine and isoleucine biosynthesis (+)	L-Threonine, L-Isoleucine, L-Valine	12	3	5.16E-05	0.0008
Alanine, aspartate and glutamate metabolism (+)	L-Aspartic acid, L-Alanine, L-Glutamic acid, L-Glutamine	61	4	0.0001	0.0016
Arginine biosynthesis (+)	L-Glutamic acid, L-Aspartic acid, L-Glutamine	27	3	0.0003	0.0027
Arginine and proline metabolism (+)	Putrescine, 4-Hydroxyproline, L-Proline, L-Glutamic acid	78	4	0.0004	0.0032
Nitrogen metabolism (+)	L-Glutamic acid, L-Glutamine	10	2	0.0015	0.0071
D-Glutamine and D-glutamate metabolism (+)	L-Glutamic acid, L-Glutamine	10	2	0.0015	0.0071
Histidine metabolism (+)	L-Glutamic acid, L-Aspartic acid	32	2	0.0113	0.0469

Table 4. Scores for each group on cognitive measures. Cognitive measures include the Clinical Dementia Rating (CDR), the Alzheimer's Disease Assessment Scale (ADAS-11, ADAS-13, ADAS-Q4), the Mini Mental State Exam (MMSE), subtests from the Rey Auditory Verbal Learning Test (RAVLT-Immediate,

RAVLT-Learning, RAVLT-Forgetting, RAVLT-Percent Forgetting), the Delayed Recall Total score (LDELTOTAL), the Trail Making Test (TRABSCOR), and the Functional Activities Questionnaire (FAQ).

	CDRSB	**ADAS11**	**ADAS13**	**ADASQ4**	**MMSE**	**RAVLT Immediate**	**RAVLT Learning**	**RAVLT Forgetting**	**RAVLT Percent Forgetting**	**LDELTOTAL**	**TRABSCOR**	**FAQ**
CN	0.0318	5.8769	9.0548	2.7640	29.0581	45.3590	5.9436	3.5789	33.7744	13.2828	82.3616	0.1248
EMCI-1	1.1505	7.4782	11.7826	3.8387	28.4516	43.3548	5.7634	4.1183	41.0667	9.1613	92.75	1.5435
EMCI-2	1.3101	8.3611	13.3704	4.3704	28.1944	36.9167	4.7870	4.3426	49.8148	8.6204	101.2	2.1574
LMCI-1	1.7	11.1247	17.7595	5.9739	27.5391	34.4783	3.6348	4.4870	61.3610	4.1913	117.7544	3.6957
LMCI-2	1.5706	10.6551	17.3454	5.8706	27.4353	30.8471	3.4824	4.8824	68.7140	4.1294	116.2143	4.0235
AD	4.4248	19.6361	29.9617	8.6283	23.1947	22.8244	1.8422	4.4940	89.1390	1.3392	197.4188	13.2248