Can Small Molecules Provide Clues on Disease Progression in Cerebrospinal Fluid from Mild Cognitive Impairment and Alzheimer's Disease Patients?

Begoña Talavera Andújar[1], Arnaud Mary[1], Carmen Venegas[1], Tiejun Cheng[2], Leonid Zaslavsky[2], Evan E. Bolton[2], Michael T. Heneka[1], Emma L. Schymanski[1*]*

1. Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Avenue du Swing 6, L-4367 Belvaux, Luxembourg

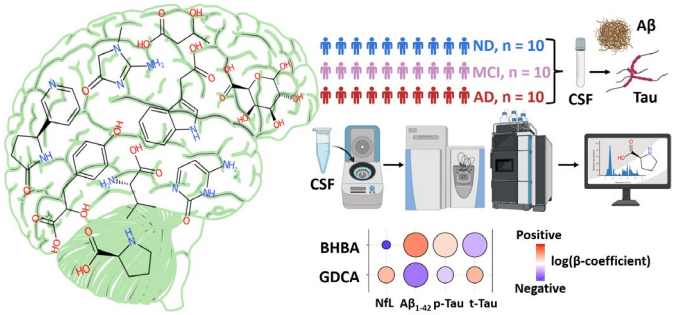
2. National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

ORCIDs: BTA: 0000-0002-3430-9255, AM: 0000-0002-6597-766X, CV: 0000-0002-8902-4986, TC: 0000-0002-4486-3356, LZ: 0000-0001-5873-4873, EEB: 0000-0002-5959-6190, MTH: 0000-0003-4996-1630, ELS: 0000-0001-6868-8145

ABSTRACT

Alzheimer's Disease (AD) is a complex and multifactorial neurodegenerative disease, which is currently diagnosed via clinical symptoms and non-specific biomarkers (such as $A\beta_{1\text{-}42}$, t-Tau, and p-Tau) measured in cerebrospinal fluid (CSF), which alone do not provide sufficient insights into disease progression. In this pilot study, these biomarkers were complemented with small molecule analysis using non-target high resolution mass spectrometry (NT-HRMS) coupled to liquid chromatography (LC) on the CSF of three groups; AD, Mild Cognitive Impairment (MCI) due to AD, and a non-demented control group (ND). An open source cheminformatics pipeline based on MS-DIAL and patRoon was enhanced using CSF- and AD-specific suspect lists to assist in data interpretation. ChemRICH analysis revealed a significant increase of hydroxybutyrates in AD, including 3-hydroxy butanoic acid (BHBA), which was found at higher levels in AD compared to MCI and ND. Furthermore, a highly sensitive target LC-MS method was used to quantify 35 bile acids (BAs) in the CSF, revealing several statistically significant differences including higher dehydrolithocholic acid levels and decreased conjugated BAs levels in AD. This work provides several promising small molecule hypotheses that could be used to help track the progression of AD in CSF samples.

GRAPHICAL ABSTRACT



KEYWORDS: High-resolution mass spectrometry, liquid chromatography, exposomics, metabolomics, cheminformatics, bile acids

SYNOPSIS: The combination of non-target LC-HRMS and target LC-MS in cerebrospinal fluid reveals potential molecular markers that could indicate Alzheimer's Disease progression

1. INTRODUCTION

Alzheimer's Disease (AD) is a complex and multifactorial neurodegenerative disease influenced by genetics, lifestyle, and environmental factors. AD is the most common form of dementia, and its prevalence is expected to increase from 50 million people in 2010 to 113 million by 2050 worldwide[1,2]. AD is often divided into three stages: (1) preclinical stage characterized by normal cognitive ability, (2) prodromal stage characterized by mild cognitive impairment (MCI) and (3) dementia stage[1,3]. Current diagnosis relies on clinical symptoms and pathological alterations indicated by biomarkers such as reduced Amyloid-$\beta_{1\text{-}42}$ ($A\beta_{1\text{-}42}$) or increased p-Tau and t-Tau concentrations in cerebrospinal fluid (CSF). Neurofilament light (NfL), a neuronal cytoplasmatic protein highly expressed in large caliber myelinated axons, has also recently emerged as a nonspecific biomarker of neurodegeneration. CSF and blood NfL levels are elevated in multiple neurodegenerative diseases, including AD, in response to axonal damage[4,5]. AD pathology starts decades before the clinical symptoms appear. Moreover Aβ and Tau protein are quite stable in clinical AD, and may not always differentiate AD from other types of dementia, leading to a high rate of misdiagnosis in the early stages[6,7].

Since CSF is already collected for AD diagnosis, further investigation into the small molecule signatures (e.g., via metabolomics and exposomics) could provide new insights to better understand disease progression and identify individuals at risk. CSF is the closest biological fluid to the brain, such that abnormalities in this matrix are directly related to pathological changes in the brain[8]. Despite its biological significance, the number of metabolomics/exposomics studies in CSF samples remains low. This due to the invasive and thus precious nature of the sample (requiring lumbar puncture) combined with methodological challenges, including the lack of standard material[9], and the relatively low chemical concentrations in CSF compared to other

matrices like blood[10]. Previous studies have revealed that alterations in various metabolomics pathways are associated with AD and MCI, including the energy metabolism, fatty acid oxidation, amino acids, and lipid biosynthesis[11–15]. Recently, bile acids (BAs) were proposed to be involved in the AD pathogenesis[16–18], but have not yet been explored in CSF in the context of MCI and AD.

High-resolution mass spectrometry (HRMS) coupled to liquid chromatography (LC) is a well-suited platform to study the chemical composition of CSF, due to the polar nature of the matrix. The current work explores the CSF of three groups of subjects: non-demented (ND) control group, MCI due to AD (which offers the opportunity to study the disease progression), and AD. Non-target LC-HRMS was performed coupled to two different analytical columns and using various software and cheminformatics approaches for data analysis to detect small molecules potentially associated with disease progression, complemented by a highly sensitive target LC-MS method to quantify extremely low concentrations of BAs in CSF. Finally, the potential associations between clinical AD biomarkers ($A\beta_{1-42}$, t-Tau, p-Tau and NfL) and chemicals identified in the CSF were investigated to determine which small molecule signatures could serve as potential biomarkers of disease progression for future investigations in a larger cohort of patients.

2. MATERIALS AND METHODS

2.1. Sample collection and biomarker assessment

30 CSF human samples (**Table 1**) were extracted by lumbar puncture and stored at –80 °C until analysis. Informed consent for research purposes was obtained by the ethics committee approval of the University Hospital of Bonn Ethics Commission (#279/10). Further details are provided in the Supporting Information (SI), **Section S1.1, Figure S1** and **Table S1.**

Table 1. Clinical characteristics of the cohort. (a) Chi-square p-value was computed for the categorical variable (sex). ANOVA p-values were calculated for the rest of characteristics. (b) NfL concentrations were measured in n =9 for ND group. **Table S1** for further details.

Clinical characteristics	**ND** (n = 10)	**MCI (due to AD)** (n = 10)	**AD** (n = 10)	**p-value[(a)]**
Sex (female/male)	6/4	1/9	8/2	0.0055
Age (years), mean ± SD	53.2 ± 16.71	66.0 ± 10.24	69.9 ± 12.96	0.0269
t-Tau (pg/mL), mean ± SD	236.9 ± 67.8	295.2 ± 68.4	549.2 ± 208.36	6.32E-05
p-Tau (pg/mL), mean ± SD	24.5 ± 7.9	37.1 ± 10.9	95.9 ± 41.1	3.30E-06
$A\beta_{1-40}$ (pg/mL), mean ± SD	4043.4 ± 1467.7	5980.2 ± 1598.1	5639.7 ± 2178.5	0.0637
$A\beta_{1-42}$ (pg/mL), mean ± SD	293.4 ± 128.8	385.8 ± 130.1	241.8 ± 109.8	0.0588
NfL (pg/mL) [(b)], mean ± SD	290.2 ± 213.2	551.2 ± 240.1	1320 ± 1932.86	0.1731

2.2. Non-target and suspect screening

Sample preparation

CSF samples were mixed with ethanol, vortexed, incubated (-20°C) and centrifuged, as described by Song et al.[19]. The supernatant was evaporated to dryness and reconstituted using Milli-Q water:MeOH:MeCN (2:1:1, v/v/v). Ten internal standards (IS) were added (**Table S2**) and pooled Quality Control (QC) samples were prepared according to recent recommendations (SI, **Section 1.2, Figure S2**). The sample preparation method was first tested on artificial CSF samples

(HelloBio Ltd, UK) using the same protocol as above, with the addition of 10 µL of a mixture containing 121 polar chemical standards (50 µM) to serve as reference standards later. Further details are given the SI **(S1.3**, **Table S3** and **Figure S3)**.

Instrumental analysis

Analytical measurements were performed on an Accela LC system coupled to a Q Exactive™ HF mass spectrometer (both Thermo Scientific) using electrospray ionization (ESI) in both positive (+) and negative (-) modes. BEH C18 reversed phase (RPLC) (1.7 µm, 2.1×150 mm) and SeQuant® ZIC-pHILIC polymer (HILIC) (5 µm, 150×2.1 mm) columns were used, in separate runs, to detect a broader range of chemicals. The HRMS was operated in full scan profile mode with scan range 60-900 *m/z* using the methods described in Talavera-Andújar et al[20]. QC samples were analyzed prior the first sample and every three or four sample injections.

Disease-specific chemical lists

New disease-specific database (*AD-database*) and suspect lists (*TOP1*, *SC20* and *AD-CTD*) were created to explore the CSF metabolome and exposome of MCI and AD subjects (**Figure 1**). First, the *AD-database* was created through literature mining[21], integrating chemicals co-occurring with 27 Medical Subject Headings (MeSH) related to AD or symptoms, given in **Table S4**. This database was filtered to create smaller lists based on reverse neighboring relations (*TOP1*) and co-occurrence scores (*SC20*), as detailed in **S1.4** and **Figures S4&5**. A list of chemicals (*AD-CTD*) specifically related to AD in the Comparative Toxicogenomic Database (CTD), was also extracted from PubChem[22]. These lists were complemented with the publicly available CSF Human Metabolome database (*HMDB-CSF*)[23,24] and *PubChemLite* for Exposomics (*PCL*)[25,26]. The associated code and lists are available on GitLab[27] and Zenodo[28] respectively.

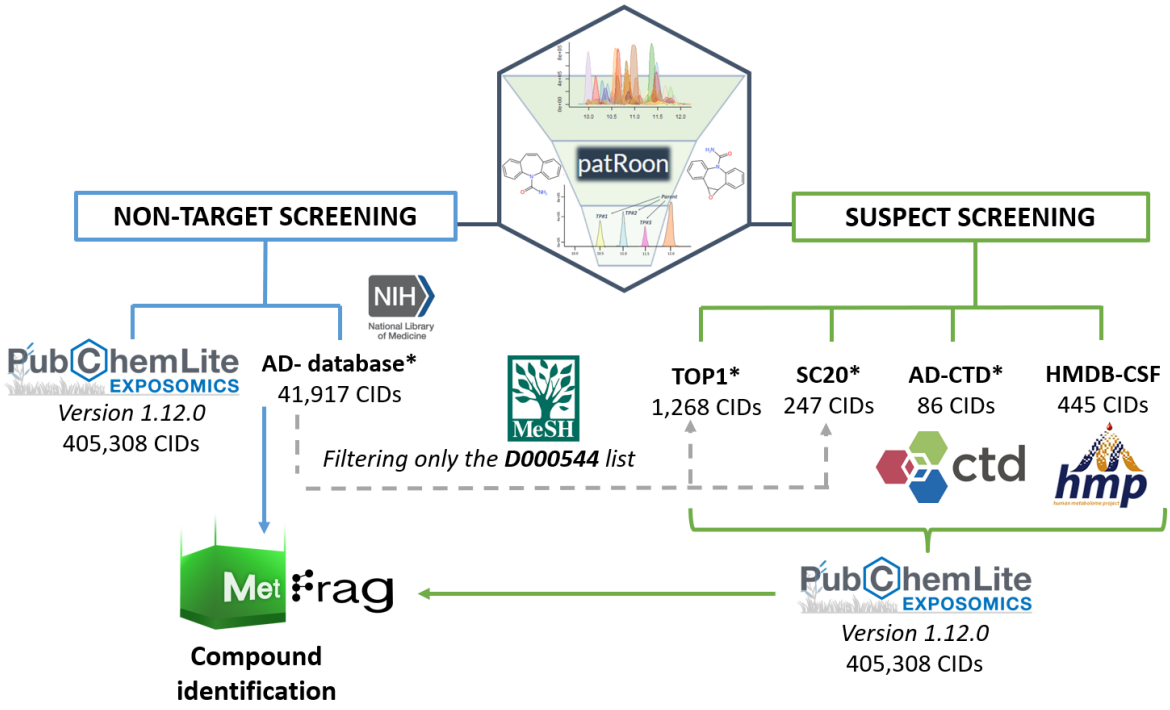


Figure 1. Databases and suspect lists employed for the non-target screening (left) and the suspect screening (right) analysis with patRoon. *Indicates databases/suspect lists created for the purpose of this study[27–29]. See main text, **S1.4** and **Table S4** for more details. CID: PubChem **C**ompound **ID**entifier.

Data processing

Raw LC-HRMS files were converted to .mzML using ProteoWizard MSConvert (version 3.0.20331.3768aa6e9 64-bit) and analyzed with MS-DIAL (version 4.90)[30], MS-FINDER (version 3.52)[31,32] and patRoon (version 2.1.0)[33,34] (see **S1.5** for details). MS-DIAL was used to perform non-target analysis via MSP-formatted libraries (MSMS-Public-Pos-VS17 and MSMS-Public-Neg-VS17 for (+) and (-) mode, respectively) using the parameters in **Table S5**. Features without a tentative candidate via MS-DIAL were uploaded to MS-FINDER to annotate them via *in silico* fragmentation (**Figure S6**). patRoon was employed for both suspect and non-target screening (**Figure 1**); all scripts including parameters and settings are available in GitLab[27].

After the analysis with MS-DIAL and patRoon, peak intensity tables were used to filter features based on the QC samples (see **S1.5**). The remaining features were manually checked and annotated

using three different sets of criteria tailored to the three different data analysis approaches. Briefly, MS-DIAL features were annotated based on the library spectral match using the *Dot product* (0-100). Level 2a was assigned with *Dot product* ≥70, and ≥3 ion fragments matching with a known structure in the library, while Level 2b was assigned to features with the same requirements but unknown structure in library (these are spectra that are commonly detected in samples belonging to unknown structures). Level 3a was assigned to features with 50 ≤ *Dot product* ≤*70* and ≥3 ion fragments matching, while Level 3b or Level 3c were assigned when <3 ion fragments were matching with known and unknown structures, respectively. Level 3d and 3e corresponded to features annotated via MS-FINDER, detailed in **Table S6**.

For features identified through patRoon non-target screening, the *individualMoNAscore* (0-1) was employed for the annotation. Level 2a was assigned when *individualMoNAscore*> 0.9. Level 3a was considered when the score was in the range of 0.7-0.9, and Level 3b when 0.4-0.7, as previously described[20,35]. Chemicals identified by patRoon suspect screening were automatically annotated following pre-defined rules specified in the handbook[36].

Identifications were considered Level 1 when the match between the standard and tentative candidate (in the CSF) yielded a *SpectrumSimilarity* score ≥ 0.7 and the retention time (RT) shift was <1 min. *OrgMassSpecR*[37–39] was used to calculate spectral similarity. Xcalibur Qual Browser (version 4.1.31.9) was used to check the RT and to extract the MS/MS information.

Peak intensity tables of the annotated features were pre-processed with MetaboAnalyst 5.0[40,41] by filtering (interquartile range option), normalization by sum, log transformation (base 10), and pareto scaling. Finally, Level 1-3 compounds were classified using the HMDB disposition ontology[42,43], pathways information[44] in the PubChem Classification Browser, and/or literature associated with PubChem records via co-occurrence scores[21]. See **S1.5** and GitLab[27] for details.

2.3. Target screening of BAs

The target study of BAs used an Agilent 1290 LC system (Waters C18 column, 1.7 μm, 2.1×150 mm) coupled to a Sciex 7500 QQQ MS in multiple-reaction monitoring (MRM) mode with (-) detection, as described by Han et al.[45]. A 10 μM standard mixture (94 BAs in total, **Table S7**) was prepared in an IS solution of UDCA-D_4 in MeCN. Next, 60 μL of each CSF sample was mixed with 140 μL of the IS solution, vortexed, sonicated and centrifuged, then dried and dissolved in 40 μL of 50% MeCN. 15 μL per sample were injected in the LC-MS system. BA concentrations were calculated by interpolating the constructed IS calibrated linear-regression curves of individual BAs, with the peak area ratios measured from injections of the sample solutions.

2.4.Statistical analysis

One-way analysis of variance (ANOVA) with post-hoc Tukey's Honestly Significant Difference (HSD) test for multiple comparisons was computed via R (*aov* and *TukeyHSD* functions). Compounds with post-hoc test p-values < 0.05 were considered as statistically significant. Chemical Similarity Enrichment Analysis (ChemRICH)[46,47] was performed to explore differentially regulated clusters of metabolites between ND-AD, MCI-AD and ND-MCI. Linear multiple regression analysis (via *lm* function in R) was used to analyze the relationship between the biomarker concentrations ($A\beta_{1-40}$, $A\beta_{1-42}$, p-Tau, t-Tau and NfL) and the relevant compounds found in CSF. Plots were created with R, Excel, and GraphPad Prism (version 10.1.0).

3. RESULTS AND DISCUSSION

3.1. Non-target characterization of CSF in MCI and AD

Compound annotation and classification

CSF samples were analyzed by RPLC and HILIC and annotated with patRoon (suspect and non-target screening) and MS-DIAL. The total number of Level 1-3 annotations can be found in the SI: **Table S8** for patRoon suspect screening, **Table S9** patRoon non-target screening, and **Table S10** for MS-DIAL. **Figure 2** summarizes the patRoon annotations (the equivalent UpSet plot for RPLC is given in **Figure S8**), while **Figure 3** shows the MS-DIAL annotations.

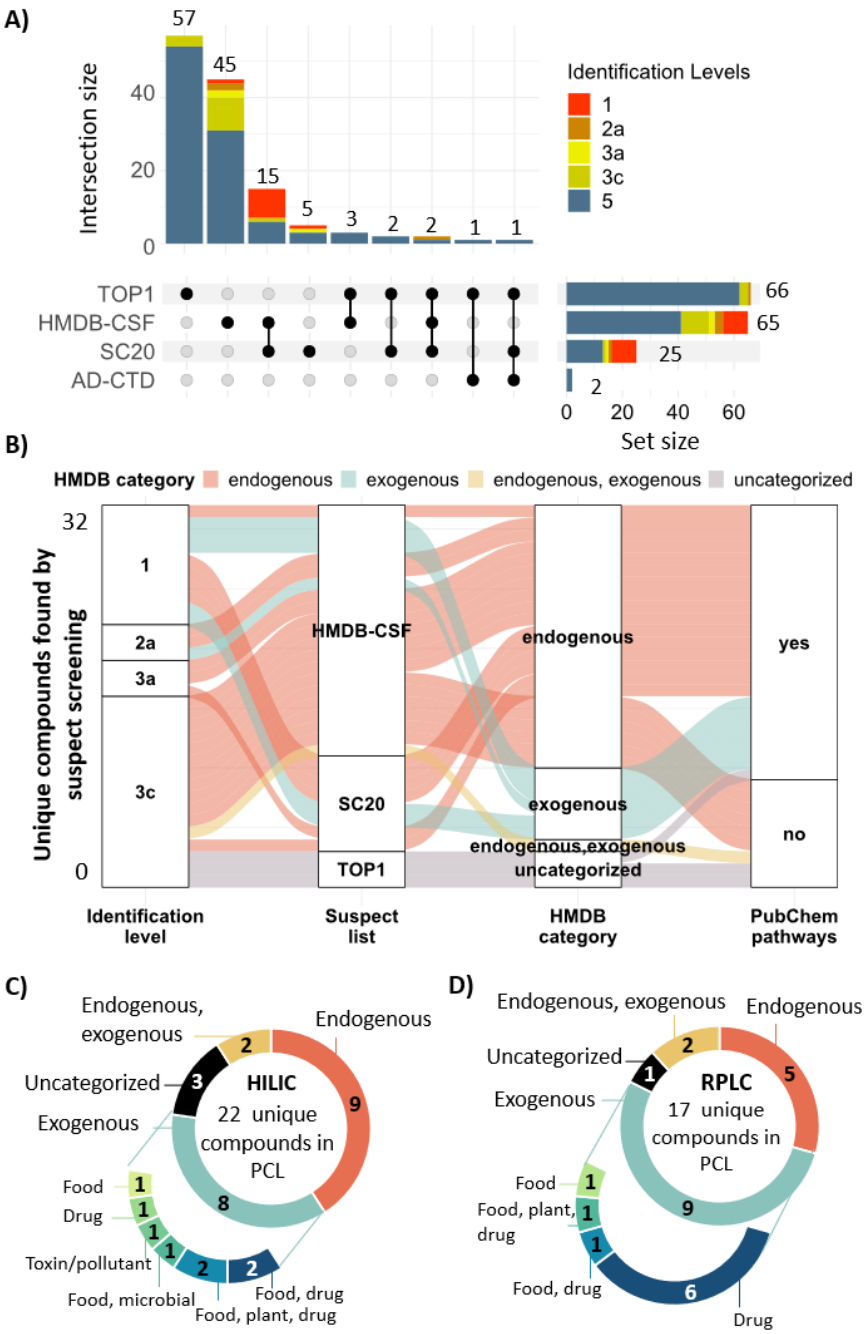


Figure 2. (**A**) UpSet plot representing the number of annotated features in each suspect lists plus overlap across lists using HILIC. See **Figure S8** for RPLC results. (**B**) Alluvial plot showing the HMDB categories of the features annotated by each suspect screening approach. RPLC and HILIC annotations were combined, and duplicates were removed prior plotting (32 unique compounds in total). The presence (or not) of PubChem pathways information is indicated in the last column. (**C,D**) Pie charts showing the classification of the compounds identified by patRoon non-target screening with PCL by HILIC (C) and RPLC (D). See Table S8.1 for detailed information about Level 1-3 compounds, and Table S8.2 for the Level 5 compounds.

Overall, the overlap between the different suspect lists was low (**Figure 2A)** confirming the need for the complementary suspect screening approaches applied here. While most of the unique features were found in the largest lists (*TOP1*, *HMDB-CSF*), the highest confidence features (Level 1) were exclusively found in *HMDB-CSF* (metabolites previously identified in CSF) and *SC20* (chemicals associated with AD through literature mining). Thus, the filtering method utilizing co-occurrence scores to generate the *SC20* list was more effective than the reverse neighboring relations approach for generating the *TOP1* list. In stark contrast to the previous work with PD[20], the *AD-CTD* list did not reveal any confident annotations, and only two in total. Interestingly, a considerably higher number of unique features was observed in the *HMDB-CSF* suspect list using HILIC (45 features, **Figure 2A)**, compared to RPLC (15 features, **Figure S8**), suggesting that HILIC is a more effective chromatographic approach for CSF analysis, likely due to the matrix's polarity.

The origin of the annotated chemicals is explored in **Figure 2B,** revealing that most are endogenous. Since exogenous species are typically present at trace levels compared to endogenous metabolites, it is challenging to capture both concurrently; furthermore, detection in CSF requires exogenous species to cross the blood brain barrier (BBB), which regulates the passage of substances to the CNS and CSF[48]. **Figure 2B** also shows why it can be challenging to distinguish between endogenous and exogenous compounds when interpreting exposomics results, as this is often difficult to disambiguate in various resources. Some exogenous compounds according to HMDB are associated with PubChem Pathways information, suggesting a potential endogenous nature. Examples include amino acids such as histidine, tryptophan and phenylalanine which can be synthesized endogenously by humans or obtained exogenously from the diet. Interestingly, only

compounds annotated at the lowest confidence level shown (Level 3c) from the *TOP1* list could not be verified with information available in either HMDB or PubChem Pathways.

Annotation with *PubChemLite* revealed 22 (HILIC, **Figure 2C**) and 17 (RPLC, **Figure 2D**) unique features between Level 1 and 3 (**Table S9**). The same compounds were identified using the *AD-database* except for metoprolol acid (Level 2a) and L-beta-homolysine (Level 3a). Despite only 18,677 chemicals overlapping between *PubChemLite* and the *AD-database* (**Figure S9**), most of the features annotated in the CSF samples were within this overlap. Although both databases (*PubChemLite* and the *AD-database*) focus on the exposome and include primarily exogenous compounds (**Figure S10**), most of the annotations by the HILIC method were categorized as endogenous (**Figure 2C**), revealing that the annotation results were not biased by the nature of the database. The RPLC method captured a lower number of endogenous chemicals (**Figure 2D**). Drugs constituted the primary subcategory among the exogenous compounds identified by RPLC, whereas HILIC revealed a more diverse array of exogenous substances (**Figure 2C**).

Using MS-DIAL and the public MSPs, 611 unique compounds were annotated between Level 1-3 (excluding Level 2b and 3c, as their structure is unknown in the libraries), including 271 (RPLC) and 340 (HILIC), **Table S10.** Overall, HILIC was the preferred LC approach for the CSF analysis, with better chromatographic separation on average for compounds detected in both modes. In general, Level 1-2a compounds tended to be endogenous while Level 3e were mainly uncategorized with no PubChem pathways information, as noted above (**Figure 3A**). Interestingly, some of the uncategorized compounds in HMDB had PubChem information (**Figure 3B**). Drugs and food (**Figure 3C**) constituted the main exogenous subcategories of the MS-DIAL annotations.

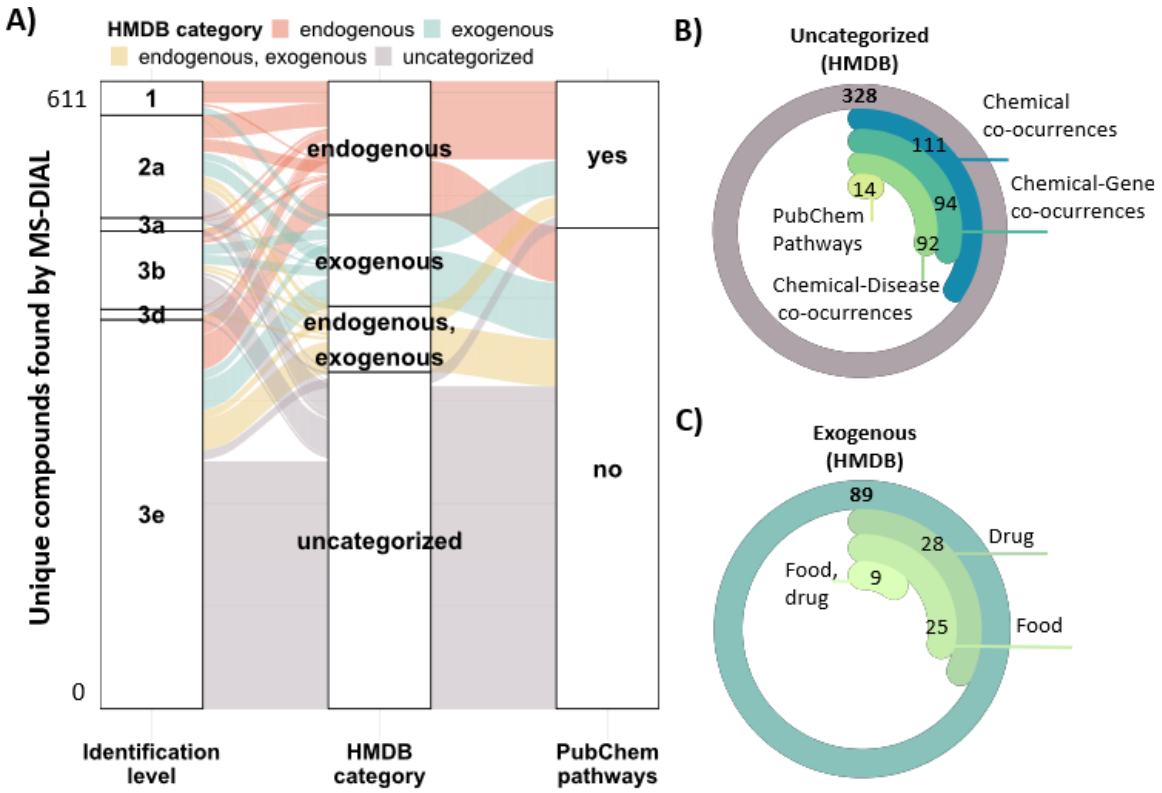


Figure 3. (A) Alluvial plot showing the HMDB categories of the Level 1-3 features annotated using MS-DIAL MSPs (+) and (-) libraries. The presence (or not) of PubChem pathways information is indicated in the last column. This plot represents the 611 unique identifications found by RPLC and HILIC. **(B)** Pie Chart representing how many of the uncategorized compounds by HMDB (grey bar) have literature knowledge via PubChem Classification Browser. **(C)** Pie chart showing the exogenous subcategories (by HMDB) of the unique MS-DIAL identifications found using RPLC and HILIC.

Statistically significant chemicals in MCI and AD

Cheminformatics and statistical approaches were used to identify significant chemicals potentially associated with disease progression. Twelve Level 1-2a features were identified as statistically significant (Tukey's HSD post-hoc p-value < 0.05), summarized in **Table 2** and **Figure S11A-L**. Full results, including Level 3 features, are available in **Table S8-S10.**

Table 2. Statistically relevant compounds found by MS-DIAL and patRoon. Only Level 1 and Level 2a annotations are included. *Indicates p-value <0.05. **Adducts were $[M+H]^+$ for (+) and $[M-H]^-$ for (-) mode. IL: Identification Level. See **Tables S8-S10** and **Figure S11** for detailed information.

								Post-hoc p-values			ANOVA
Chemical name	**rt (min)**	***m/z*** **	**LC mode**	**IL**	**HMDB category**	**PubChem pathways**	**Library/database/ suspect list**	**MCI-AD**	**ND-AD**	**ND-MCI**	**p-value**
Valine	6.99	118.0862	HILIC (+)	1	exogenous	yes	PCL/AD-database	0.9991	0.0320*	0.0293*	0.0156*
Proline	6.93	116.0706	HILIC (+)	1	endogenous	yes	PCL/AD-database	0.1432	0.7350	0.0303*	0.0325*
N-Acetylhistidine	8.14	198.0872	HILIC (+)	2a	endogenous, exogenous	no	MSDIAL-MSPs	0.4226	0.3859	0.0374*	0.0477*
3-hydroxybutanoic acid (BHBA)	5.92	103.0396	HILIC (-)	1	endogenous	yes	MSDIAL-MSPs	0.0042*	0.0150*	0.8637	0.0030*
Indole-3-acetic acid (IAA)	13.66	176.0706	RPLC (+)	1	endogenous	yes	PCL/AD-database/SC20/ HMDB-CSF	0.7248	0.0390*	0.0064*	0.0061*
4-Hydroxyphenyl lactic acid (4-HPLA)	6.46	181.0495	HILIC (-)	2a	endogenous, exogenous	no	MSDIAL-MSPs	0.9935	0.0574	0.0455*	0.0285*
Adenine	3.42	136.0617	HILIC (+)	1	endogenous	yes	MSDIAL-MSPs	0.9958	0.0213*	0.0174*	0.0091*
Cytosine	5.90	112.0505	HILIC (+)	2a	endogenous	yes	MSDIAL-MSPs	0.0216*	0.6122	0.1575	0.0253*
Galacturonic acid	12.24	193.0342	HILIC (-)	1	endogenous	yes	MSDIAL-MSPs	0.9534	0.0753	0.0403*	0.0305*
Threonic acid	10.48	135.0291	HILIC (-)	1	endogenous	no	MSDIAL-MSPs	0.4707	0.3345	0.0361*	0.0457*
Cotinine	1.76	177.1021	HILIC (+)	1	endogenous	no	MSDIAL-MSPs	0.0915	0.8744	0.0320*	0.0284*
Diazepam	17.11	285.0787	RPLC (+)	2a	exogenous	no	MSDIAL-MSPs	0.0373*	0.6052	0.2429	0.0447*

289 Significantly altered levels of valine and proline were observed, consistent with prior research
290 indicating disrupted amino acids pathways in AD[11,49–51], potentially due to the alteration of
291 different neurotransmitters. Valine (**Figure S11A**) was significantly decreased in MCI and AD
292 compared to the ND group, in line with previous studies reporting decreased valine levels in AD,
293 associated with impaired neurotransmission and cognitive function[11,13]. Furthermore, adenine
294 (**Figure S11G**) was also significantly reduced in both MCI and AD, compared to ND. These results
295 are consistent with previous studies performed in mice[52], indicating that the purine metabolism
296 pathway may be altered in AD and potentially play an important role in the pathogenesis.
297 3-hydroxybutanoic acid (BHBA, **Figure S11D**), was found with statistically higher levels in AD
298 compared to the other two groups. BHBA is the most abundant ketone in the human circulation
299 and may be involved in many brain functions including neurotransmission, neuroinflammation
300 and myelination[53]. The higher levels in the AD group may be due to increased fat degradation and
301 thus ketone formation as a physiological response to energy shortage in the brain[53–55]. Ketogenic
302 diets, which elevate BHBA concentrations, may also contribute to increased levels, but dietary
303 information was not available for these samples.
304 Statistically higher levels of indole-3-acetic acid (IAA, **Figure S11E**) were found in the MCI
305 and AD groups compared to ND, aligning with a recent study reporting significantly higher levels
306 of IAA in the plasma of AD patients compared to control subjects[56]. IAA was previously found
307 upregulated in the CSF of MCI[57]. IAA has shown proinflammatory and prooxidant effects[58],
308 potentially contributing to neurodegeneration, such that higher levels in MCI might serve as an
309 inflammatory indicator, as previously suggested[56].
310 Galacturonic acid (**Figure S11I**) was found with statistically higher levels in the MCI compared
311 to the ND group, along with a similar but non-significant trend in the AD group compared to ND.

Galacturonic acid is the major component of pectin, found in fruits and vegetables, where the higher levels found in both AD and MCI groups could be due to increased BBB permeability, associated to the dementia status. Hence, this could be a biomarker of BBB dysfunction[59].

Significantly higher levels of cotinine, the main metabolite of nicotine, were found in the MCI group compared to ND (**Figure S11K**). Interestingly, tobacco smoking has been correlated with a lower incidence of AD. Cotinine has shown to prevent memory loss and inhibit Aβ aggregation without the toxicity and addictive properties of its precursor (nicotine)[60,61].

Statistically relevant chemical clusters

ChemRICH analysis was performed to facilitate the biological interpretation of the non-target results, as it accounts for both endogenous and exogenous chemicals. Unlike other common approaches such as pathway enrichment analysis, ChemRICH's p-values do not rely on the size of a background database (e.g., KEGG)[46], avoiding overrepresentation issues. Instead, ChemRICH is study-specific, with a self-contained size. The analysis here considered Level 1-3 annotations; the results are given in **Figure 4** (including key annotations in each cluster) and **Table S11**.

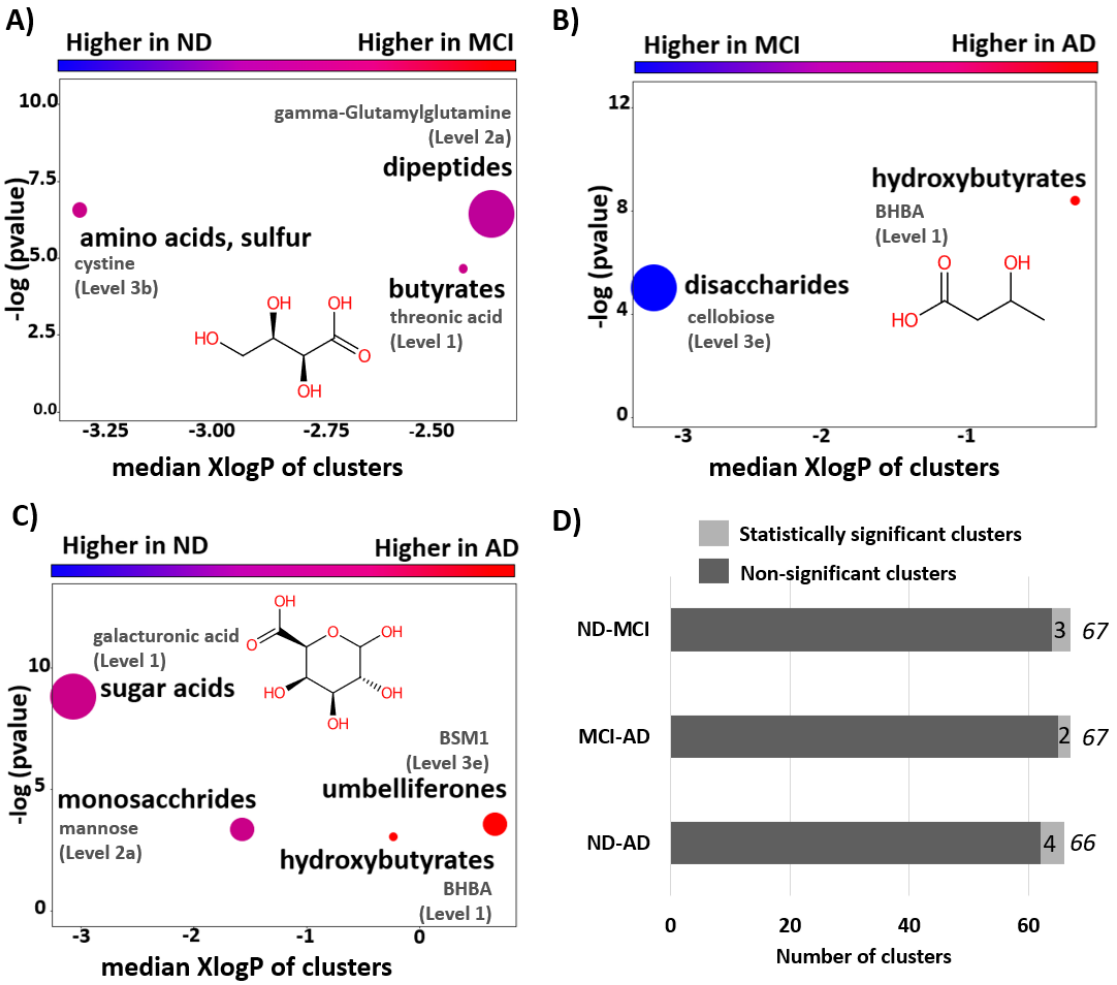


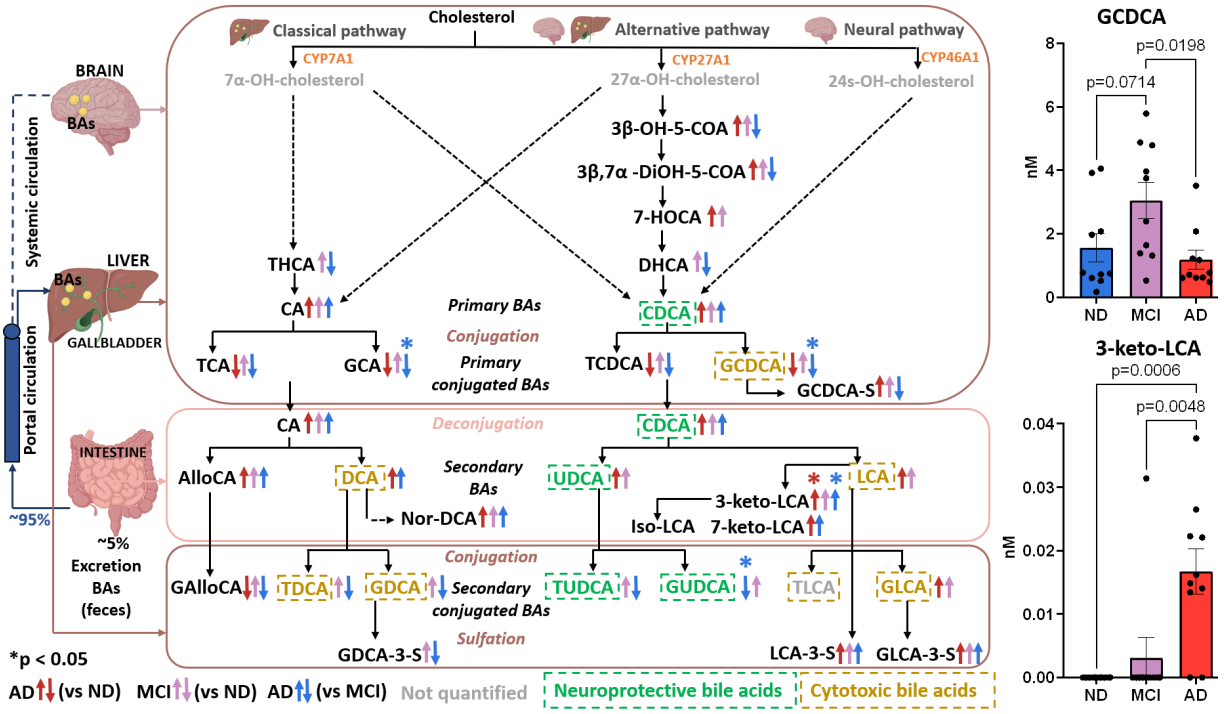
Figure 4. ChemRICH analysis between (A) ND-MCI, (B) MC-AD and (C) ND-AD. Enrichment p-values are given by the Kolmogorov–Smirnov-test. Each dot represents a significantly altered cluster of chemicals (p-value<0.05). Dot size is proportional to the number of metabolites in the cluster. The node color scale shows the proportion of increased (red) or decreased (blue) metabolite levels in MCI (A) or AD (B, C). Purple-color nodes have both increased and decreased metabolites. Names of key metabolites in each cluster are displayed in grey; structures are shown only for Level 1 key metabolites. (D) Bar plot representing the total number of clusters identified in each of the comparisons. BSM1=2-(2-methyl-4-oxochromen-5-yl)acetic acid. See Table S11 for further details. Chemical structures were drawn with CDK Depict[62].

Different significant chemical clusters were found across the three comparisons. The ND-MCI comparison (**Figure 4A)** revealed significant alterations in three chemical clusters: sulfur amino acids, dipeptides and butyrates, with threonic acid (**Figure S11J)** the key metabolite of the latter. The MCI-AD comparison (**Figure 4B**) revealed two significant clusters - disaccharides (decreased

in AD) and hydroxybutyrates (increased in AD, with key metabolite BHBA, discussed above). This last cluster was also statistically altered comparing ND-AD (**Figure 4C**). Umbelliferones (increased in AD), monosaccharides and sugar acids were also significant in ND-AD. While the total number of chemical clusters identified was almost the same in the three cases (**Figure 4D),** only one significant cluster (hydroxybutyrates) overlapped between MCI-AD and ND-AD.

3.2.Target study of BAs in CSF of MCI and AD

Of the 94 bile acids included in the targeted method, 35 were quantified in the CSF samples (**Table S12**). An overview of these results is given in **Figure 5**, significant results are marked with an asterisk. The identification of BAs in CSF implies a possible source either through systemic circulation uptake or local synthesis within the brain[63]. While a previous study quantified BAs precursors in the CSF of AD patients[64], to our knowledge this is the first time that BAs are quantified in CSF in the context of MCI and AD.



357

358 **Figure 5.** Schematic representation of the BAs (and precursors) quantified in this study. Primary
359 BAs are synthetized from cholesterol through different pathways. The classical pathway in the
360 liver (top left) is responsible for most BA synthesis. The alternative pathway (top middle) occurs
361 in other tissues besides the liver, such as the brain. The neural pathway (top right) takes place in
362 the brain and is responsible for the majority of cholesterol turnover in the CNS. Primary BAs, after
363 conjugation with taurine or glycine in the liver, are secreted into the bile and transported to the gut
364 where the gut bacteria deconjugate the conjugated BAs, generating secondary BAs (middle box).
365 Most of the BAs (95%) are reabsorbed in the ileum via portal circulation to the liver. Only a small
366 portion escapes the enterohepatic circulation and reaches the systemic circulation. Arrows indicate
367 the higher (↑) or lower (↓) concentrations. No arrow means no differences between groups. Right
368 box plots show the mean concentration with standard error of the mean of GCDCA (top) and 3-
369 keto-LCA (bottom). p= Tukey's HSD post-hoc p-value. Note that $p<0.1$ are displayed although
370 only $p<0.05$ are considered statistically significant in this work, marked with an asterisk (*) in the
371 scheme. Abbreviations: 3-keto-LCA, dehydrolithocholic acid; 3β,7α-DiOH-5-COA, 3β-7α-
372 DiOH-5 cholestenoic acid; 3β-OH-5-COA, 3β-OH-5-cholestenoic acid; 7-HOCA, 7α-hydroxy-3-
373 oxo-4-cholestenoic acid; 7-Keto-LCA, 7 ketolithocholic acid; AlloCA, allocholic acid; CA, cholic
374 acid; CDCA, chenodeoxycholic acid; DCA, deoxycholic acid; DHCA, 3α,7α-
375 dihydroxycholestanoic acid; GAlloCA, glycoallocholic acid; GCA, glycocholic acid; GCDCA,
376 glycochenodeoxycholic acid; GDCA, glycochenodeoxycholic acid; GLCA, glycolithocholic acid;
377 GUDCA; glycoursodeoxycholic acid; HCA, hyocholic acid; LCA, lithocholic acid; Nor-DCA,
378 nordeoxycholic acid; TCA, taurocholic acid; TCDCA, taurochenodeoxycholic acid; TDCA,
379 tauoursodeoxycholic acid; THCA, 3α,7α,12α-trihydroxycholestanoic acid; TUDCA,
380 tauroursodexycholic acid; UDCA, ursodeoxycholic acid. Adapted from [16,65–68].

Primary BAs are synthetized from cholesterol through different pathways (top of **Figure 5**). While the classical pathway in the liver is responsible for most BA synthesis, the brain uses the alternative and neural pathways to clear cholesterol, leading to the production of BAs[18,63]. The intermediates of the alternative pathway are explained in **S2.2** and **Figure S12.** The primary BAs, cholic acid (CA) and chenodeoxycholic acid (CDCA) presented a non-significant higher trend in MCI and AD compared to the ND group (**Figure S13**). In contrast, the glycine conjugated primary BAs glycocholic acid (GCA) and glycochenodeoxycholic acid (GCDCA), top right of **Figure 5**, showed significantly lower concentrations in the AD compared to the MCI group. Elevated concentrations of these two BAs were previously reported in AD plasma samples compared with control subjects[17,69]. Therefore, CSF concentrations may not always reflect circulating BAs levels. Furthermore, the conjugates with taurine, taurocholic acid (TCA) and taurochenodeoxycholic acid (TCDA) exhibited the same low trend in the AD group compared to the others, without statistical significance (**Figure S13**).

The secondary and cytotoxic BAs, lithocholic acid (LCA), and deoxycholic acid (DCA) increased non-significantly in MCI and AD compared with ND subjects (**Figure S14**). Higher levels of LCA and DCA were previously noted in AD blood samples[17,65,69], suggesting LCA as putative biomarker for AD[69]. Interestingly, 3-keto-LCA, the major metabolite of LCA, was found with statistically higher concentrations in AD compared to MCI and ND (bottom right of **Figure 5**). This is a microbial metabolite, not previously reported in CSF, which could reflect the importance of the microbiota-gut-brain axis (MGBA) in neurodegeneration.

As observed with the primary conjugated BAs, the secondary conjugated BAs showed lower trends in the AD compared to the MCI group (**Figure S15**). Significantly lower concentrations of GUDCA were found in AD compared to the MCI group, in line with a previous study identifying

GUDCA as a potential blood marker for early diagnosis that could predict the onset of AD or MCI with 2-3 years and 90% of accuracy[70,71]. This highlights the importance of the MCI group to study early disease biomarkers. While the average GCDCA concentration across the 30 samples was 1.92 nM, the average of GCDCA-S was 871.56 nM. This trend was also observed for GDCA (0.60) and GDCA-S (866.73), see **Figure S16**. Sulfation, catalyzed by SULT2A1 in humans, is an important detoxification pathway of BAs. The resulting sulfated BAs are less toxic and more soluble, leading to reduced intestinal absorption and enhanced fecal and urinary excretion[72,73]. This suggests that the brain may utilize sulfation to mitigate the BAs toxicity, as SULT2A1 is expressed not only in the liver but also in the brain[72].

To explore whether the observed dysregulation of conjugated BAs in AD is linked to enzymatic differences in taurine and glycine conjugation, the ratios GDCA:DCA, TDCA:DCA, GCA:CA and TCA:CA were calculated (**Figure S17**) as previously described[65]. A significant decrease in GDCA:DCA was found in AD compared to MCI, suggesting a change in the processes involving glycine conjugation in the liver. This is in contrast with a previous study in blood where no significant changes were observed[65].

The presence of secondary cytotoxic BAs with higher concentrations in the MCI and AD groups supports a previously published hypothesis[65] that gut microbiome dysregulation leads to increased production of cytotoxic secondary BAs and derivatives such as 3-keto-LCA. Moreover, elevated hydrophobic BAs in blood, such as DCA and CDCA can alter the BBB permeability[17,65], which might explain some of these results.

3.3.Correlation between altered metabolites in CSF and classical biomarkers

Finally, the correlation between the altered metabolites and the concentrations of the diagnostic biomarkers $A\beta_{1-42}$, p-Tau, t-Tau, and NfL in CSF was explored (**Figure 6** and **Table S13-15**). Age, sex, and $A\beta_{1-40}$ concentrations were considered as covariates to compute the different linear models. Although a positive β-coefficient indicates a positive association between two variables (e.g., t-Tau and valine levels), an association does not necessarily imply causation.

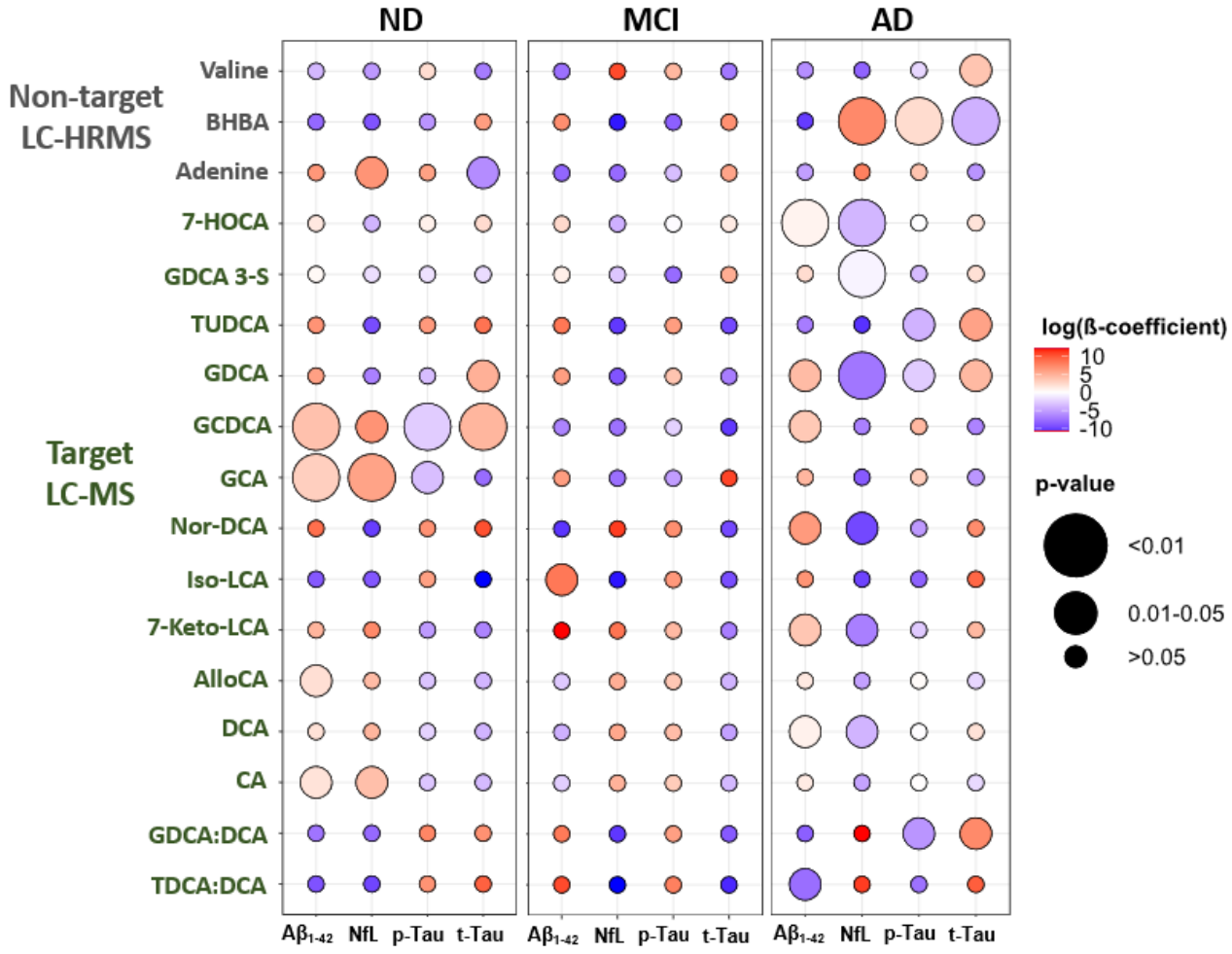


Figure 6. Associations between the statistically relevant compounds found in CSF, by non-target (first three rows) and target screening, and $A\beta_{1-42}$, NfL, t-Tau and p-Tau concentrations. Color represents the log transformed β-coefficients. Positive and negative associations are indicated by the red and blue colors, respectively. Note that only compounds with a statistically significant association are illustrated. **Table S13-15** for further details. BHBA: 3-hydroxybutanoic acid, see **Figure 5** for the rest of abbreviations.

Since previous studies have shown that metabolic changes in AD blood and CSF were associated with the disease status and pathological alterations (e.g., brain atrophy)[11,74–76], correlating the identified metabolites in CSF with the classical AD biomarkers might reveal additional insights[75]. The results here show multiple significant associations between the altered compounds found in CSF from AD (e.g., BHBA, GDCA, Nor-DCA) and Aβ, Tau, and NfL levels (right panel of **Figure 6**). In contrast, the associations in the MCI group were weaker (middle panel of **Figure 6**), with only one significant association (Iso-LCA and $A\beta_{1-42}$). The ND group (left panel of **Figure 6**) presented various significant associations with the classical biomarkers, most strikingly for GCDCA and GCA. Some of them correlate (positively or negatively) as in the AD group (e.g., GDCA, GCDCA and $A\beta_{1-42}$ were positively correlated in both groups), potentially due to a disease-independent relationship, as explained by Jacobs et al[76].

Briefly, significant positive associations were found for BHBA with NfL and p-Tau in the AD group, with a significant negative association for t-Tau. The ND group exhibited the opposite but non-significant correlations. This disparity in associations may suggest disease-specific patterns in AD. Multiple significant associations were found between the quantified BAs and the classical biomarkers in AD. In short, the neurotoxic GCDCA, GDCA, and the ratio GDCA:DCA were significantly and positively associated with t-Tau in the AD group, the first association in line with a previous work performed in serum from AD[74]. Additionally, a statistically significant negative association was observed between the TDCA:DCA ratio and $A\beta_{1-42}$, in the AD group, which might be associated with higher cerebral amyloid burden[11].

4. Future Perspectives

The identification of chemicals with statistically higher levels in the MCI compared to the ND (e.g., galacturonic acid, IAA, 4-HPLA) shows the importance of this group for the early identification of individuals at risk. However external factors, including diet, medication, and exercise, may account for some of the observed chemical differences across groups, such that more information about these factors would enhance the interpretation of findings. Notably, one AD patient exhibited high outlier levels for NfL and BHBA compared with the other patients in the group (**Figure S18**), and it would be interesting to investigate whether this is due to AD pathology or to environmental factors such as a ketogenic diet, as previously discussed.

The HILIC LC method (**Table 2**) appears to be the most suitable method for future non-target experiments. Since most of the enriched clusters (**Figure 4**) are highly hydrophilic (logP < -1), an expansion of the non-target methods to explore the hydrophobic part (lipidome) of the CSF could reveal extra information in future efforts. Overall, while MS-DIAL provided a higher number of annotated chemicals, the combination of different software and suspect lists enhanced the annotation of a variety of chemicals, increasing the general understanding of the CSF metabolome/exposome. Furthermore, the identification of some chemicals in this study (e.g. galacturonic acid, threonic acid, N-acetylhistidine, and some of the BAs) could help expand the current HMDB-CSF database, as they are not yet included in this resource.

This study highlights the possible role of the microbiota-gut-brain-axis (MGBA) in the disease progression, as some metabolites found altered in MCI and/or AD, such as 3-keto-LCA, are produced by the human microbiome. However, the role of BAs in CSF needs to be further investigated, as the link between peripheral and central BAs is poorly understood. Matching

samples of CSF, plasma, and feces would be needed to study the influence of microbiota composition.

Although the low sample size can be considered as a limitation of this study, the development of novel non-target cheminformatics approaches together with the highly sensitive target study of BAs in CSF provides valuables insights into this complex matrix (CSF) and disease progression. Multiple significant molecules were found in both MCI and AD compared to ND. Moreover, some significant associations between the altered metabolites and the CSF biomarkers in AD were observed. Further studies in larger cohort of samples will be necessary to validate the promising hypothesis and results presented here to determine which of these small molecules may reveal insights into disease progression.

ASSOCIATED CONTENT

Supporting Information.

The following files are available free of charge on the ACS Publication website and via DOI.

A Word file contains figures and additional details regarding material and methods (**S1**), results and discussion (**S2**) plus **Figures S1-S18**. An Excel file contains supplementary tables **Table S1-S15**.

Data availability

The code functions, and files associated with this manuscript are provided in the ECI GitLab repository (https://gitlab.lcsb.uni.lu/eci/AD-CSF). The PubChemLite database (https://doi.org/10.5281/zenodo.6936117) and database/suspect lists created here (https://doi.org/10.5281/zenodo.8014420) are available for download on Zenodo[28,29]

AUTHOR INFORMATION

Corresponding Author

*Email: begona.talavera@uni.lu and emma.schymanski@uni.lu

Author Contributions

BTA: conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, writing – original draft (lead), reviewing and editing; AM: formal analysis, investigation, writing- review and editing; CV: investigation, writing- review and editing; TC: methodology, software, writing - review and editing; LZ: methodology, software, writing – review and editing; EEB: conceptualization, resources, software, supervision, writing – reviewing and editing; MTH: conceptualization, funding acquisition, resources, supervision, writing – reviewing and editing; ELS: conceptualization, data curation, resources, software, supervision, writing – original draft (supporting), writing – review and editing.

Ethics declarations

Informed consent for use of samples and data for research purposes was given with the local ethics committee approval (University Hospital of Bonn Ethics Commission #279/10). This work does not contain identifiable data of the subjects or any other specific individual person's data.

Funding Sources

BTA is part of the "Microbiomes in One Health" PhD training program, which is supported by the PRIDE doctoral research funding scheme (PRIDE/11823097) of the Luxembourg National Research Fund (FNR). CV was funded by an FNR CORE Junior Grant ("NeuroFlame", C20/BM/14548100). The work of EEB, TC, and LZ was supported by the National Center for Biotechnology Information of the National Library of Medicine (NLM), National Institutes of Health. ELS acknowledges funding support from the Luxembourg National Research Fund (FNR)

for project A18/BM/12341006, MTH acknowledges funding support from the FNR within the PEARL programme (FNR/16745220).

ACKNOWLEDGMENT

BTA acknowledges support from Gianfranco Frigerio during sample preparation and advice from Corey Griffith and Lorenzo Favilli during data processing/interpretation. Katyeny Manuela Da Silva is acknowledged for her inputs during the manuscript revision. We thank the Metabolomics Platform of the LCSB for their support with the LC-HRMS analysis and other Environmental Cheminformatics and PubChem team members who contributed to this work indirectly via other collaborative and scientific activities. The target LC-MS BA analysis was performed by the Genome BC Proteomics Centre of the University of Victoria (Canada).

REFERENCES

(1) Knopman, D. S.; Amieva, H.; Petersen, R. C.; Chételat, G.; Holtzman, D. M.; Hyman, B. T.; Nixon, R. A.; Jones, D. T. Alzheimer Disease. *Nat Rev Dis Primers* **2021**, *7* (1), 1–21. https://doi.org/10.1038/s41572-021-00269-y.

(2) Reveglia, P.; Paolillo, C.; Ferretti, G.; De Carlo, A.; Angiolillo, A.; Nasso, R.; Caputo, M.; Matrone, C.; Di Costanzo, A.; Corso, G. Challenges in LC–MS-Based Metabolomics for Alzheimer's Disease Early Detection: Targeted Approaches versus Untargeted Approaches. *Metabolomics* **2021**, *17* (9), 78. https://doi.org/10.1007/s11306-021-01828-w.

(3) *Duration of preclinical, prodromal, and dementia stages of Alzheimer's disease in relation to age, sex, and APOE genotype*. https://doi.org/10.1016/j.jalz.2019.04.001.

(4) Mielke, M. M.; Syrjanen, J. A.; Blennow, K.; Zetterberg, H.; Vemuri, P.; Skoog, I.; Machulda, M. M.; Kremers, W. K.; Knopman, D. S.; Jack, C.; Petersen, R. C.; Kern, S.

Plasma and CSF Neurofilament Light. *Neurology* **2019**, *93* (3), e252–e260. https://doi.org/10.1212/WNL.0000000000007767.

(5) Gaetani, L.; Blennow, K.; Calabresi, P.; Di Filippo, M.; Parnetti, L.; Zetterberg, H. Neurofilament Light Chain as a Biomarker in Neurological Disorders. *J Neurol Neurosurg Psychiatry* **2019**, *90* (8), 870–881. https://doi.org/10.1136/jnnp-2018-320106.

(6) Dhiman, K.; Blennow, K.; Zetterberg, H.; Martins, R. N.; Gupta, V. B. Cerebrospinal Fluid Biomarkers for Understanding Multiple Aspects of Alzheimer's Disease Pathogenesis. *Cell Mol Life Sci* **2019**, *76* (10), 1833–1863. https://doi.org/10.1007/s00018-019-03040-5.

(7) Microbiota-Gut-Brain Axis in the Alzheimer's Disease Pathology - an Overview. *Neuroscience Research* **2022**, *181*, 17–21. https://doi.org/10.1016/j.neures.2022.05.003.

(8) Yan, J.; Kuzhiumparambil, U.; Bandodkar, S.; Dale, R. C.; Fu, S. Cerebrospinal Fluid Metabolomics: Detection of Neuroinflammation in Human Central Nervous System Disease. *Clin Transl Immunology* **2021**, *10* (8), e1318. https://doi.org/10.1002/cti2.1318.

(9) Lippa, K. A.; Aristizabal-Henao, J. J.; Beger, R. D.; Bowden, J. A.; Broeckling, C.; Beecher, C.; Clay Davis, W.; Dunn, W. B.; Flores, R.; Goodacre, R.; Gouveia, G. J.; Harms, A. C.; Hartung, T.; Jones, C. M.; Lewis, M. R.; Ntai, I.; Percy, A. J.; Raftery, D.; Schock, T. B.; Sun, J.; Theodoridis, G.; Tayyari, F.; Torta, F.; Ulmer, C. Z.; Wilson, I.; Ubhi, B. K. Reference Materials for MS-Based Untargeted Metabolomics and Lipidomics: A Review by the Metabolomics Quality Assurance and Quality Control Consortium (mQACC). *Metabolomics* **2022**, *18* (4), 24. https://doi.org/10.1007/s11306-021-01848-6.

(10) Stoop, M. P.; Coulier, L.; Rosenling, T.; Shi, S.; Smolinska, A. M.; Buydens, L.; Ampt, K.; Stingl, C.; Dane, A.; Muilwijk, B.; Luitwieler, R. L.; Smitt, P. A. E. S.; Hintzen, R. Q.; Bischoff, R.; Wijmenga, S. S.; Hankemeier, T.; Gool, A. J. van; Luider, T. M. Quantitative

Proteomics and Metabolomics Analysis of Normal Human Cerebrospinal Fluid Samples*. *Molecular & Cellular Proteomics* **2010**, *9* (9), 2063–2075. https://doi.org/10.1074/mcp.M110.000877.

(11) van der Velpen, V.; Teav, T.; Gallart-Ayala, H.; Mehl, F.; Konz, I.; Clark, C.; Oikonomidi, A.; Peyratout, G.; Henry, H.; Delorenzi, M.; Ivanisevic, J.; Popp, J. Systemic and Central Nervous System Metabolic Alterations in Alzheimer's Disease. *Alz Res Therapy* **2019**, *11* (1), 93. https://doi.org/10.1186/s13195-019-0551-7.

(12) Trushina, E.; Dutta, T.; Persson, X.-M. T.; Mielke, M. M.; Petersen, R. C. Identification of Altered Metabolic Pathways in Plasma and CSF in Mild Cognitive Impairment and Alzheimer's Disease Using Metabolomics. *PLoS One* **2013**, *8* (5), e63644. https://doi.org/10.1371/journal.pone.0063644.

(13) Nielsen, J. E.; Maltesen, R. G.; Havelund, J. F.; Færgeman, N. J.; Gotfredsen, C. H.; Vestergård, K.; Kristensen, S. R.; Pedersen, S. Characterising Alzheimer's Disease through Integrative NMR- and LC-MS-Based Metabolomics. *Metabolism Open* **2021**, *12*, 100125. https://doi.org/10.1016/j.metop.2021.100125.

(14) Dong, R.; Denier-Fields, D. N.; Lu, Q.; Suridjan, I.; Kollmorgen, G.; Wild, N.; Betthauser, T. J.; Carlsson, C. M.; Asthana, S.; Johnson, S. C.; Zetterberg, H.; Blennow, K.; Engelman, C. D. Principal Components from Untargeted Cerebrospinal Fluid Metabolomics Associated with Alzheimer's Disease Biomarkers. *Neurobiology of Aging* **2022**, *117*, 12–23. https://doi.org/10.1016/j.neurobiolaging.2022.04.009.

(15) Tynkkynen, J.; Chouraki, V.; van der Lee, S. J.; Hernesniemi, J.; Yang, Q.; Li, S.; Beiser, A.; Larson, M. G.; Sääksjärvi, K.; Shipley, M. J.; Singh-Manoux, A.; Gerszten, R. E.; Wang, T. J.; Havulinna, A. S.; Würtz, P.; Fischer, K.; Demirkan, A.; Ikram, M. A.; Amin, N.;

Lehtimäki, T.; Kähönen, M.; Perola, M.; Metspalu, A.; Kangas, A. J.; Soininen, P.; Ala-Korpela, M.; Vasan, R. S.; Kivimäki, M.; van Duijn, C. M.; Seshadri, S.; Salomaa, V. 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. *Alzheimers Dement* **2018**, *14* (6), 723–733. https://doi.org/10.1016/j.jalz.2018.01.003.

(16) Mulak, A. Bile Acids as Key Modulators of the Brain-Gut-Microbiota Axis in Alzheimer's Disease. *J Alzheimers Dis* **2021**, *84* (2), 461–477. https://doi.org/10.3233/JAD-210608.

(17) Ehtezazi, T.; Rahman, K.; Davies, R.; Leach, A. G. The Pathological Effects of Circulating Hydrophobic Bile Acids in Alzheimer's Disease. *J Alzheimers Dis Rep* **2023**, *7* (1), 173–211. https://doi.org/10.3233/ADR-220071.

(18) Baloni, P.; Funk, C. C.; Yan, J.; Yurkovich, J. T.; Kueider-Paisley, A.; Nho, K.; Heinken, A.; Jia, W.; Mahmoudiandehkordi, S.; Louie, G.; Saykin, A. J.; Arnold, M.; Kastenmüller, G.; Griffiths, W. J.; Thiele, I.; Kaddurah-Daouk, R.; Price, N. D.; Kaddurah-Daouk, R.; Kueider-Paisley, A.; Louie, G.; Doraiswamy, P. M.; Blach, C.; Moseley, A.; Thompson, J. W.; Mahmoudiandehkhordi, S.; Welsh-Balmer, K.; Plassman, B.; Saykin, A.; Nho, K.; Kastenmüller, G.; Arnold, M.; Bhattacharyya, S.; Han, X.; Baillie, R.; Fiehn, O.; Barupal, D.; Meikle, P.; Mazmanian, S.; Kling, M.; Shaw, L.; Trojanowski, J.; Toledo, J.; Van Duijin, C.; Hankemier, T.; Thiele, I.; Heinken, A.; Price, N.; Funk, C.; Baloni, P.; Jia, W.; Wishart, D.; Brinton, R.; Chang, R.; Farrer, L.; Au, R.; Qiu, W.; Würtz, P.; Mangravite, L.; Krumsiek, J.; Newman, J.; Zhang, B.; Moreno, H. Metabolic Network Analysis Reveals Altered Bile Acid Synthesis and Metabolism in Alzheimer's Disease. *Cell Reports Medicine* **2020**, *1* (8), 100138. https://doi.org/10.1016/j.xcrm.2020.100138.

(19) Song, Z.; Wang, M.; Zhu, Z.; Tang, G.; Liu, Y.; Chai, Y. Optimization of Pretreatment Methods for Cerebrospinal Fluid Metabolomics Based on Ultrahigh Performance Liquid Chromatography/Mass Spectrometry. *Journal of Pharmaceutical and Biomedical Analysis* **2021**, *197*, 113938. https://doi.org/10.1016/j.jpba.2021.113938.

(20) Talavera Andújar, B.; Aurich, D.; Aho, V. T. E.; Singh, R. R.; Cheng, T.; Zaslavsky, L.; Bolton, E. E.; Mollenhauer, B.; Wilmes, P.; Schymanski, E. L. Studying the Parkinson's Disease Metabolome and Exposome in Biological Samples through Different Analytical and Cheminformatics Approaches: A Pilot Study. *Anal Bioanal Chem* **2022**. https://doi.org/validanti.

(21) Zaslavsky, L.; Cheng, T.; Gindulyte, A.; He, S.; Kim, S.; Li, Q.; Thiessen, P.; Yu, B.; Bolton, E. E. Discovering and Summarizing Relationships Between Chemicals, Genes, Proteins, and Diseases in PubChem. *Front. Res. Metr. Anal.* **2021**, *6*, 689059. https://doi.org/10.3389/frma.2021.689059.

(22) Kim, S.; Cheng, T.; He, S.; Thiessen, P. A.; Li, Q.; Gindulyte, A.; Bolton, E. E. PubChem Protein, Gene, Pathway, and Taxonomy Data Collections: Bridging Biology and Chemistry through Target-Centric Views of PubChem Data. *J Mol Biol* **2022**, *434* (11), 167514. https://doi.org/10.1016/j.jmb.2022.167514.

(23) Wishart, D. S.; Lewis, M. J.; Morrissey, J. A.; Flegel, M. D.; Jeroncic, K.; Xiong, Y.; Cheng, D.; Eisner, R.; Gautam, B.; Tzur, D.; Sawhney, S.; Bamforth, F.; Greiner, R.; Li, L. The Human Cerebrospinal Fluid Metabolome. *Journal of Chromatography B* **2008**, *871* (2), 164–173. https://doi.org/10.1016/j.jchromb.2008.05.001.

(24) *CSF Metabolome*. https://csfmetabolome.ca/ (accessed 2023-06-06).

(25) PubChemLite for Exposomics. https://doi.org/10.5281/zenodo.6936117.

(26) Schymanski, E. L.; Kondić, T.; Neumann, S.; Thiessen, P. A.; Zhang, J.; Bolton, E. E. Empowering Large Chemical Knowledge Bases for Exposomics: PubChemLite Meets MetFrag. *J Cheminform* **2021**, *13* (1), 19. https://doi.org/10.1186/s13321-021-00489-0.

(27) *Environmental Cheminformatics / AD-CSF · GitLab*. GitLab. https://gitlab.lcsb.uni.lu/eci/AD-CSF (accessed 2023-06-12).

(28) Andújar, B. T.; Mary, A.; Cheng, T.; Zaslavsky, L.; Bolton, E. E.; Heneka, M. T.; Schymanski, E. L. PubChem Lists to Study the Exposome of Mild Cognitive Impairment and Alzheimer's Disease on Cerebrospinal Fluid, 2023. https://doi.org/10.5281/zenodo.8014420.

(29) Bolton, E.; Schymanski, E.; Kondic, T.; Thiessen, P.; Zhang, J. PubChemLite for Exposomics, 2020. https://doi.org/10.5281/zenodo.4183801.

(30) Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; VanderGheynst, J.; Fiehn, O.; Arita, M. MS-DIAL: Data-Independent MS/MS Deconvolution for Comprehensive Metabolome Analysis. *Nat Methods* **2015**, *12* (6), 523–526. https://doi.org/10.1038/nmeth.3393.

(31) Tsugawa, H.; Kind, T.; Nakabayashi, R.; Yukihira, D.; Tanaka, W.; Cajka, T.; Saito, K.; Fiehn, O.; Arita, M. Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software. *Anal. Chem.* **2016**, *88* (16), 7946–7958. https://doi.org/10.1021/acs.analchem.6b00770.

(32) Lai, Z.; Tsugawa, H.; Wohlgemuth, G.; Mehta, S.; Mueller, M.; Zheng, Y.; Ogiwara, A.; Meissen, J.; Showalter, M.; Takeuchi, K.; Kind, T.; Beal, P.; Arita, M.; Fiehn, O. Identifying Metabolites by Integrating Metabolome Databases with Mass Spectrometry Cheminformatics. *Nat Methods* **2018**, *15* (1), 53–56. https://doi.org/10.1038/nmeth.4512.

(33) Helmus, R.; ter Laak, T. L.; van Wezel, A. P.; de Voogt, P.; Schymanski, E. L. patRoon: Open Source Software Platform for Environmental Mass Spectrometry Based Non-Target Screening. *Journal of Cheminformatics* **2021**, *13* (1), 1. https://doi.org/10.1186/s13321-020-00477-w.

(34) Helmus, R.; Velde, B. van de; Brunner, A. M.; Laak, T. L. ter; Wezel, A. P. van; Schymanski, E. L. patRoon 2.0: Improved Non-Target Analysis Workflows Including Automated Transformation Product Screening. *Journal of Open Source Software* **2022**, *7* (71), 4029. https://doi.org/10.21105/joss.04029.

(35) Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J. Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environ. Sci. Technol.* **2014**, *48* (4), 2097–2098. https://doi.org/10.1021/es5002105.

(36) Helmus, R. *patRoon Handbook*.

(37) Dodder, N.; Mullen, K. OrgMassSpecR: Organic Mass Spectrometry, 2017. https://CRAN.R-project.org/package=OrgMassSpecR (accessed 2022-08-01).

(38) Liao, W.-L.; Heo, G.-Y.; Dodder, N. G.; Pikuleva, I. A.; Turko, I. V. Optimizing the Conditions of a Multiple Reaction Monitoring Assay for Membrane Proteins: Quantification of Cytochrome P450 11A1 and Adrenodoxin Reductase in Bovine Adrenal Cortex and Retina. *Anal. Chem.* **2010**, *82* (13), 5760–5767. https://doi.org/10.1021/ac100811x.

(39) Hoh, E.; Dodder, N. G.; Lehotay, S. J.; Pangallo, K. C.; Reddy, C. M.; Maruya, K. A. Nontargeted Comprehensive Two-Dimensional Gas Chromatography/Time-of-Flight Mass Spectrometry Method and Software for Inventorying Persistent and Bioaccumulative

Contaminants in Marine Environments. *Environ. Sci. Technol.* **2012**, *46* (15), 8001–8008. https://doi.org/10.1021/es301139q.

(40) Chong, J.; Xia, J. Using MetaboAnalyst 4.0 for Metabolomics Data Analysis, Interpretation, and Integration with Other Omics Data. In *Computational Methods and Data Analysis for Metabolomics*; Li, S., Ed.; Methods in Molecular Biology; Springer US: New York, NY, 2020; pp 337–360. https://doi.org/10.1007/978-1-0716-0239-3_17.

(41) *MetaboAnalyst*. https://www.metaboanalyst.ca/ (accessed 2022-03-10).

(42) Wishart, D. S.; Feunang, Y. D.; Marcu, A.; Guo, A. C.; Liang, K.; Vázquez-Fresno, R.; Sajed, T.; Johnson, D.; Li, C.; Karu, N.; Sayeeda, Z.; Lo, E.; Assempour, N.; Berjanskii, M.; Singhal, S.; Arndt, D.; Liang, Y.; Badran, H.; Grant, J.; Serra-Cayuela, A.; Liu, Y.; Mandal, R.; Neveu, V.; Pon, A.; Knox, C.; Wilson, M.; Manach, C.; Scalbert, A. HMDB 4.0: The Human Metabolome Database for 2018. *Nucleic Acids Res* **2018**, *46* (Database issue), D608–D617. https://doi.org/10.1093/nar/gkx1089.

(43) *Human Metabolome Database*. https://hmdb.ca/ (accessed 2023-12-05).

(44) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E. PubChem 2023 Update. *Nucleic Acids Research* **2023**, *51* (D1), D1373–D1380. https://doi.org/10.1093/nar/gkac956.

(45) Han, J.; Liu, Y.; Wang, R.; Yang, J.; Ling, V.; Borchers, C. H. Metabolic Profiling of Bile Acids in Human and Mouse Blood by LC-MS/MS in Combination with Phospholipid-Depletion Solid-Phase Extraction. *Anal Chem* **2015**, *87* (2), 1127–1136. https://doi.org/10.1021/ac503816u.

(46) Barupal, D. K.; Fiehn, O. Chemical Similarity Enrichment Analysis (ChemRICH) as Alternative to Biochemical Pathway Mapping for Metabolomic Datasets. *Sci Rep* **2017**, *7* (1), 14567. https://doi.org/10.1038/s41598-017-15231-w.

(47) *ChemRICH*. https://chemrich.fiehnlab.ucdavis.edu/ (accessed 2023-11-08).

(48) Obermeier, B.; Verma, A.; Ransohoff, R. M. Chapter 3 - The Blood–Brain Barrier. In *Handbook of Clinical Neurology*; Pittock, S. J., Vincent, A., Eds.; Autoimmune Neurology; Elsevier, 2016; Vol. 133, pp 39–59. https://doi.org/10.1016/B978-0-444-63432-0.00003-7.

(49) Ibáñez, C.; Simó, C.; Martín-Álvarez, P. J.; Kivipelto, M.; Winblad, B.; Cedazo-Mínguez, A.; Cifuentes, A. *Toward a Predictive Model of Alzheimer's Disease Progression Using Capillary Electrophoresis–Mass Spectrometry Metabolomics*. ACS Publications. https://doi.org/10.1021/ac301243k.

(50) Xie, K.; Qin, Q.; Long, Z.; Yang, Y.; Peng, C.; Xi, C.; Li, L.; Wu, Z.; Daria, V.; Zhao, Y.; Wang, F.; Wang, M. High-Throughput Metabolomics for Discovering Potential Biomarkers and Identifying Metabolic Mechanisms in Aging and Alzheimer's Disease. *Front Cell Dev Biol* **2021**, *9*, 602887. https://doi.org/10.3389/fcell.2021.602887.

(51) Ibáñez, C.; Simó, C.; Martín-Álvarez, P. J.; Kivipelto, M.; Winblad, B.; Cedazo-Mínguez, A.; Cifuentes, A. Toward a Predictive Model of Alzheimer's Disease Progression Using Capillary Electrophoresis–Mass Spectrometry Metabolomics. *Anal. Chem.* **2012**, *84* (20), 8532–8540. https://doi.org/10.1021/ac301243k.

(52) Esteve, C.; Jones, E. A.; Kell, D. B.; Boutin, H.; McDonnell, L. A. Mass Spectrometry Imaging Shows Major Derangements in Neurogranin and in Purine Metabolism in the Triple-Knockout 3×Tg Alzheimer Mouse Model. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2017**, *1865* (7), 747–754. https://doi.org/10.1016/j.bbapap.2017.04.002.

(53) Jensen, N. J.; Wodschow, H. Z.; Nilsson, M.; Rungby, J. Effects of Ketone Bodies on Brain Metabolism and Function in Neurodegenerative Diseases. *International Journal of Molecular Sciences* **2020**, *21* (22), 8767. https://doi.org/10.3390/ijms21228767.

(54) Reger, M. A.; Henderson, S. T.; Hale, C.; Cholerton, B.; Baker, L. D.; Watson, G. S.; Hyde, K.; Chapman, D.; Craft, S. Effects of β-Hydroxybutyrate on Cognition in Memory-Impaired Adults. *Neurobiology of Aging* **2004**, *25* (3), 311–314. https://doi.org/10.1016/S0197-4580(03)00087-3.

(55) Shippy, D. C.; Wilhelm, C.; Viharkumar, P. A.; Raife, T. J.; Ulland, T. K. β-Hydroxybutyrate Inhibits Inflammasome Activation to Attenuate Alzheimer's Disease Pathology. *Journal of Neuroinflammation* **2020**, *17* (1), 280. https://doi.org/10.1186/s12974-020-01948-5.

(56) Lin, C.-H.; Lin, Y.-N.; Lane, H.-Y.; Chen, C.-J. The Identification of a Potential Plasma Metabolite Marker for Alzheimer's Disease by LC-MS Untargeted Metabolomics. *Journal of Chromatography B* **2023**, *1222*, 123686. https://doi.org/10.1016/j.jchromb.2023.123686.

(57) Kaddurah-Daouk, R.; Zhu, H.; Sharma, S.; Bogdanov, M.; Rozen, S. G.; Matson, W.; Oki, N. O.; Motsinger-Reif, A. A.; Churchill, E.; Lei, Z.; Appleby, D.; Kling, M. A.; Trojanowski, J. Q.; Doraiswamy, P. M.; Arnold, S. E.; Pharmacometabolomics Research Network. Alterations in Metabolic Pathways and Networks in Alzheimer's Disease. *Transl Psychiatry* **2013**, *3* (4), e244. https://doi.org/10.1038/tp.2013.18.

(58) Dou, L.; Sallée, M.; Cerini, C.; Poitevin, S.; Gondouin, B.; Jourde-Chiche, N.; Fallague, K.; Brunet, P.; Calaf, R.; Dussol, B.; Mallet, B.; Dignat-George, F.; Burtey, S. The Cardiovascular Effect of the Uremic Solute Indole-3 Acetic Acid. *Journal of the American Society of Nephrology* **2015**, *26* (4), 876–887. https://doi.org/10.1681/ASN.2013121283.

(59) Beukema, M.; Faas, M. M.; de Vos, P. The Effects of Different Dietary Fiber Pectin Structures on the Gastrointestinal Immune Barrier: Impact via Gut Microbiota and Direct Effects on Immune Cells. *Exp Mol Med* **2020**, *52* (9), 1364–1376. https://doi.org/10.1038/s12276-020-0449-2.

(60) Echeverria, V.; Zeitlin, R. Cotinine: A Potential New Therapeutic Agent against Alzheimer's Disease. *CNS Neurosci Ther* **2012**, *18* (7), 517–523. https://doi.org/10.1111/j.1755-5949.2012.00317.x.

(61) Patel, S.; Grizzell, J. A.; Holmes, R.; Zeitlin, R.; Solomon, R.; Sutton, T. L.; Rohani, A.; Charry, L. C.; Iarkov, A.; Mori, T.; Echeverria Moran, V. Cotinine Halts the Advance of Alzheimer's Disease-like Pathology and Associated Depressive-like Behavior in Tg6799 Mice. *Front Aging Neurosci* **2014**, *6*, 162. https://doi.org/10.3389/fnagi.2014.00162.

(62) Mayfield, J. CDK Depict Web Interface, 2023. https://www.simolecule.com/cdkdepict/depict.html (accessed 2023-03-09).

(63) Monteiro-Cardoso, V. F.; Corlianò, M.; Singaraja, R. R. Bile Acids: A Communication Channel in the Gut-Brain Axis. *Neuromol Med* **2021**, *23* (1), 99–117. https://doi.org/10.1007/s12017-020-08625-z.

(64) Griffiths, W. J.; Abdel-Khalik, J.; Yutuc, E.; Roman, G.; Warner, M.; Gustafsson, J.-Å.; Wang, Y. Concentrations of Bile Acid Precursors in Cerebrospinal Fluid of Alzheimer's Disease Patients. *Free Radic Biol Med* **2019**, *134*, 42–52. https://doi.org/10.1016/j.freeradbiomed.2018.12.020.

(65) MahmoudianDehkordi, S.; Arnold, M.; Nho, K.; Ahmad, S.; Jia, W.; Xie, G.; Louie, G.; Kueider-Paisley, A.; Moseley, M. A.; Thompson, J. W.; St John Williams, L.; Tenenbaum, J. D.; Blach, C.; Baillie, R.; Han, X.; Bhattacharyya, S.; Toledo, J. B.; Schafferer, S.; Klein,

S.; Koal, T.; Risacher, S. L.; Allan Kling, M.; Motsinger-Reif, A.; Rotroff, D. M.; Jack, J.; Hankemeier, T.; Bennett, D. A.; De Jager, P. L.; Trojanowski, J. Q.; Shaw, L. M.; Weiner, M. W.; Doraiswamy, P. M.; Van Duijn, C. M.; Saykin, A. J.; Kastenmüller, G.; Kaddurah-Daouk, R.; for the Alzheimer's Disease Neuroimaging Initiative and the Alzheimer Disease Metabolomics Consortium. Altered Bile Acid Profile Associates with Cognitive Impairment in Alzheimer's Disease—An Emerging Role for Gut Microbiome. *Alzheimer's & Dementia* **2019**, *15* (1), 76–92. https://doi.org/10.1016/j.jalz.2018.07.217.

(66) *Frontiers | Gut Microbiome-Linked Metabolites in the Pathobiology of Major Depression With or Without Anxiety—A Role for Bile Acids*. https://www.frontiersin.org/articles/10.3389/fnins.2022.937906/full (accessed 2023-11-01).

(67) Mertens, K. L.; Kalsbeek, A.; Soeters, M. R.; Eggink, H. M. Bile Acid Signaling Pathways from the Enterohepatic Circulation to the Central Nervous System. *Frontiers in Neuroscience* **2017**, *11*.

(68) Chiang, J. Y. L.; Ferrell, J. M. Bile Acid Metabolism in Liver Pathobiology. *gene expr* **2018**, *18* (2), 71–87. https://doi.org/10.3727/105221618X15156018385515.

(69) Marksteiner, J.; Blasko, I.; Kemmler, G.; Koal, T.; Humpel, C. Bile Acid Quantification of 20 Plasma Metabolites Identifies Lithocholic Acid as a Putative Biomarker in Alzheimer's Disease. *Metabolomics* **2017**, *14* (1), 1. https://doi.org/10.1007/s11306-017-1297-5.

(70) Mapstone, M.; Cheema, A. K.; Fiandaca, M. S.; Zhong, X.; Mhyre, T. R.; MacArthur, L. H.; Hall, W. J.; Fisher, S. G.; Peterson, D. R.; Haley, J. M.; Nazar, M. D.; Rich, S. A.; Berlau, D. J.; Peltz, C. B.; Tan, M. T.; Kawas, C. H.; Federoff, H. J. Plasma Phospholipids Identify Antecedent Memory Impairment in Older Adults. *Nat Med* **2014**, *20* (4), 415–418. https://doi.org/10.1038/nm.3466.

(71) Connell, E.; Le Gall, G.; Pontifex, M. G.; Sami, S.; Cryan, J. F.; Clarke, G.; Müller, M.; Vauzour, D. Microbial-Derived Metabolites as a Risk Factor of Age-Related Cognitive Decline and Dementia. *Molecular Neurodegeneration* **2022**, *17* (1), 43. https://doi.org/10.1186/s13024-022-00548-6.

(72) *Bile Acid Sulfation: A Pathway of Bile Acid Elimination and Detoxification | Toxicological Sciences | Oxford Academic*. https://academic.oup.com/toxsci/article/108/2/225/1664383 (accessed 2023-11-17).

(73) Novel Insights into Bile Acid Detoxification via CYP, UGT and SULT Enzymes. *Toxicology in Vitro* **2023**, *87*, 105533. https://doi.org/10.1016/j.tiv.2022.105533.

(74) Nho, K.; Kueider-Paisley, A.; MahmoudianDehkordi, S.; Arnold, M.; Risacher, S. L.; Louie, G.; Blach, C.; Baillie, R.; Han, X.; Kastenmüller, G.; Jia, W.; Xie, G.; Ahmad, S.; Hankemeier, T.; van Duijn, C. M.; Trojanowski, J. Q.; Shaw, L. M.; Weiner, M. W.; Doraiswamy, P. M.; Saykin, A. J.; Kaddurah-Daouk, R.; Alzheimer's Disease Neuroimaging Initiative and the Alzheimer Disease Metabolomics Consortium. Altered Bile Acid Profile in Mild Cognitive Impairment and Alzheimer's Disease: Relationship to Neuroimaging and CSF Biomarkers. *Alzheimers Dement* **2019**, *15* (2), 232–244. https://doi.org/10.1016/j.jalz.2018.08.012.

(75) Dong, R.; Lu, Q.; Kang, H.; Suridjan, I.; Kollmorgen, G.; Wild, N.; Deming, Y.; Van Hulle, C. A.; Anderson, R. M.; Zetterberg, H.; Blennow, K.; Carlsson, C. M.; Asthana, S.; Johnson, S. C.; Engelman, C. D. CSF Metabolites Associated with Biomarkers of Alzheimer's Disease Pathology. *Frontiers in Aging Neuroscience* **2023**, *15*.

(76) Jacobs, K. R.; Lim, C. K.; Blennow, K.; Zetterberg, H.; Chatterjee, P.; Martins, R. N.; Brew, B. J.; Guillemin, G. J.; Lovejoy, D. B. Correlation between Plasma and CSF Concentrations

of Kynurenine Pathway Metabolites in Alzheimer's Disease and Relationship to Amyloid-β and Tau. *Neurobiology of Aging* **2019**, *80*, 11–20. https://doi.org/10.1016/j.neurobiolaging.2019.03.015.