

# SCIENTIFIC DATA

OPEN

DATA DESCRIPTOR

## Bile acids targeted metabolomics and medication classification data in the ADNI1 and ADNIGO/2 cohorts

Received: 19 March 2019

Accepted: 10 July 2019

Published online: 17 October 2019

Lisa St. John-Williams<sup>1</sup>, Siamak Mahmoudiandehkordi<sup>2</sup>, Matthias Arnold<sup>1,2,3</sup>, Tyler Massaro<sup>4</sup>, Colette Blach<sup>5</sup>, Gabi Kastenmüller<sup>3,6</sup>, Gregory Louie<sup>2</sup>, Alexandra Kueider-Paisley<sup>2</sup>, Xianlin Han<sup>7</sup>, Rebecca Baillie<sup>8</sup>, Alison A. Motsinger-Reif<sup>9</sup>, Daniel Rotroff<sup>10</sup>, Kwangsik Nho<sup>11</sup>, Andrew J. Saykin<sup>11</sup>, Shannon L. Risacher<sup>11</sup>, Therese Koal<sup>12</sup>, M. Arthur Moseley<sup>1</sup>, Jessica D. Tenenbaum<sup>13</sup>, J. Will Thompson<sup>1</sup>, Rima Kaddurah-Daouk<sup>2,14</sup>, Alzheimer's Disease Neuroimaging Initiative\* & Alzheimer's Disease Metabolomics Consortium\*

Alzheimer's disease (AD) is the most common cause of dementia. The mechanism of disease development and progression is not well understood, but increasing evidence suggests multifactorial etiology, with a number of genetic, environmental, and aging-related factors. There is a growing body of evidence that metabolic defects may contribute to this complex disease. To interrogate the relationship between system level metabolites and disease susceptibility and progression, the AD Metabolomics Consortium (ADMC) in partnership with AD Neuroimaging Initiative (ADNI) is creating a comprehensive biochemical database for patients in the ADNI1 cohort. We used the Biocrates Bile Acids platform to evaluate the association of metabolic levels with disease risk and progression. We detail the quantitative metabolomics data generated on the baseline samples from ADNI1 and ADNIGO/2 (370 cognitively normal, 887 mild cognitive impairment, and 305 AD). Similar to our previous reports on ADNI1, we present the tools for data quality control and initial analysis. This data descriptor represents the third in a series of comprehensive metabolomics datasets from the ADMC on the ADNI.

### Background and Summary

With the dramatic increase of older adults around the world, Alzheimer's disease (AD) has become a major public health challenge<sup>1</sup>. AD is the leading cause of dementia, and is clinically defined by an insidious onset and a progressive loss of memory and other cognitive functions that effects a person's ability to function in daily activities<sup>2</sup>.

<sup>1</sup>Proteomics and Metabolomics Shared Resource, Center for Genomics and Computational Biology, Duke University, Durham, NC, USA. <sup>2</sup>Department of Psychiatry & Behavioral Sciences, Duke University, Durham, NC, USA. <sup>3</sup>Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. <sup>4</sup>Duke Clinical Research Institute, Duke University, Durham, NC, USA. <sup>5</sup>Duke Molecular Physiology Institute, Duke University, Durham, NC, USA. <sup>6</sup>German Center for Diabetes Research, Neuherberg, Germany. <sup>7</sup>University of Texas Health Science Center at San Antonio, San Antonio, TX, USA. <sup>8</sup>Rosa & Co LLC, San Carlos, CA, USA. <sup>9</sup>Bioinformatics Research Center, Department of Statistics, North Carolina State University, Raleigh, NC, USA. <sup>10</sup>Department of Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH, USA. <sup>11</sup>Department of Radiology and Imaging Sciences and the Indiana Alzheimer Disease Center, Indiana University School of Medicine, Indianapolis, IN, USA. <sup>12</sup>BIOCRATES Life Sciences AG, Innsbruck, Austria. <sup>13</sup>Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, USA. <sup>14</sup>Duke Institute for Brain Sciences, Duke University, Durham, NC, USA. These authors contributed equally: Lisa St. John-Williams and Siamak Mahmoudiandehkordi. \*A full list of Alzheimer's Disease Neuroimaging Initiative Consortium members is available at: [http://adni.loni.usc.edu/wp-content/uploads/how\\_to\\_apply/ADNI\\_Acknowledgement\\_List.pdf](http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf). A full list of Alzheimer's Disease Metabolomics Consortium members is available at: <https://sites.duke.edu/adnimetab/team/>. Correspondence and requests for materials should be addressed to R.K.-D. (email: [kaddu001@mc.duke.edu](mailto:kaddu001@mc.duke.edu))

AD is a complex and progressive disorder: the cognitive and functional decline is preceded by a pre-clinical phase known as mild cognitive impairment (MCI). MCI is a complex syndrome characterized by memory failures that may be considered as an intermediate stage in the development of AD and is distinct from normal aging<sup>3</sup>.

The etiology of AD is highly complex and multifactorial. Accumulating evidence highlights numerous biochemical perturbations that are suggested to play a role in AD. These include the characteristic deposition of  $\beta$ -amyloid plaques (A $\beta$ ), hyperphosphorylation of tau protein, oxidative stress, inflammation, abnormal metal homeostasis, as well as disruption in energetic and neurotransmitter pathways, among others<sup>4–9</sup>. AD has traditionally been considered primarily a neurodegenerative disorder of the central nervous system, but there is increasing evidence that the pathological processes associated with disease may also manifest in the peripheral system<sup>10–12</sup>.

Our limited understanding of the etiology is reflected in the limited options for therapeutic treatments<sup>13</sup>. There are currently a limited number of treatments available, and they have only modest effects<sup>14</sup>. A recent review of drug development in AD discusses these issues in detail<sup>15</sup>. Improved mechanistic understanding of disease onset and progression is central to more efficient AD drug development and will lead to improved therapeutic approaches and targets.

To better understand this complex etiology, the application of metabolomics for AD research has the potential to monitor molecular alterations associated with disease pathogenesis and progression, as well as to discover candidate diagnostic biomarkers. The rapidly emerging field of metabolomics combines strategies to identify and quantify cellular metabolites using sophisticated analytical technologies with the application of statistical and multi-variant methods for information extraction. Current technologies allow for the high throughput collection of a large number of metabolites<sup>16,17</sup>. Initial studies have demonstrated the potential of metabolomics in AD, demonstrating that metabolomics markers may serve as biomarkers for the disease, and help unravel the complex biochemical pathways involved in the disease<sup>4,5,18–23</sup>.

These successes motivate the collection of metabolomics data in large, well-powered cohorts. The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a public-private partnership that has established a landmark longitudinal cohort to increase the rate of scientific discovery in AD. Data collection for the ADNI cohort is comprehensive. Across thousands of subjects, ADNI researchers collect MRI and PET images, genetics, cognitive tests, CSF and blood biomarkers, etc.<sup>24,25</sup>. Details of the ADNI efforts can be found at [www.adni-info.org](http://www.adni-info.org). The Alzheimer's Disease Metabolomics Consortium (ADMC) is working with ADNI to add metabolomics data to the vast collection of data collected for this cohort. The data collected through the ADMC will provide a resource to interrogate global metabolomics changes within the ADNI cohort, to enhance the systems level data available. A total of eight targeted and non-targeted metabolomics platforms are being used by the ADMC, with the second of these described in the current manuscript.

Herein, we describe the use of the Biocrates Bile Acids kit to profile baseline serum samples from the ADNI1 and ADNIGO/2 cohorts. We also apply the medication mapping approach performed previously on ADNI1 to the ADNIGO2 cohort<sup>23</sup>. These data are intended to aid in the discovery of metabolic features associated with disease risk, progression, or other clinically and biologically relevant outcomes. We describe both the data collection and tools and resources for data processing, quality control, and analysis.

## Methods

**Alzheimer's disease neuroimaging initiative (ADNI) cohort.** ADNI data was obtained through the University of Southern California's Laboratory of Neuroimaging (LONI) data repository (<http://adni.loni.ucla.edu/>), where data and results have been made accessible through the AMP-AD Knowledge Portal (<https://ampad-portal.org>). The AMP-AD Knowledge Portal is the distribution site for data, analysis results, analytical methodology and research tools generated by the AMP-AD Target Discovery and Preclinical Validation Consortium and multiple Consortia and research programs supported by the National Institute on Aging. Institutional Review Board approval and written informed consent was obtained at each of the participating institutions. Data obtained included: demographic information, clinical assessment data, clinical diagnosis, etc. Detailed information on the ADNI cohort is described in Petersen *et al.* 2010 and at <http://www.adni-info.org/><sup>24</sup>. ADNI data collection is ongoing, and variables are continually updated through their LONI resource. Data presented here are from January 2016, and represent a snapshot of the clinical data for analytical reproducibility. A summary of key demographic and clinical variables are summarized in Table 1 for ADNI 1 and Table 2 for ADNIGO/2.

**Serum collection and sample management.** Samples were collected at the baseline study visit. Blood samples were collected in the morning, after overnight fasting (except where explicitly annotated). Standard operating procedures are detailed at ([www.adni-info.org](http://www.adni-info.org)). Briefly, duplicate blood samples were collected in two bar-coded 10 mL red-top plastic Vacutainer blood tubes, allowed to clot for 30 minutes, and then centrifuged at 3000 rpm (1500 rcf) for 15 minutes. The serum was then transferred into a bar-coded 13 mL polypropylene transfer tube, capped and frozen in dry ice. Frozen samples were overnighted to the ADNI biomarker core laboratory at the University of Pennsylvania Medical Center. Samples were thawed and aliquoted to 0.5 mL samples, then once more for individual laboratory analyses. A 20  $\mu$ L sample aliquot was delivered to the Duke Proteomics and Metabolomics Shared Resource (Durham, NC) for analysis with the Bile Acids platform, as detailed below.

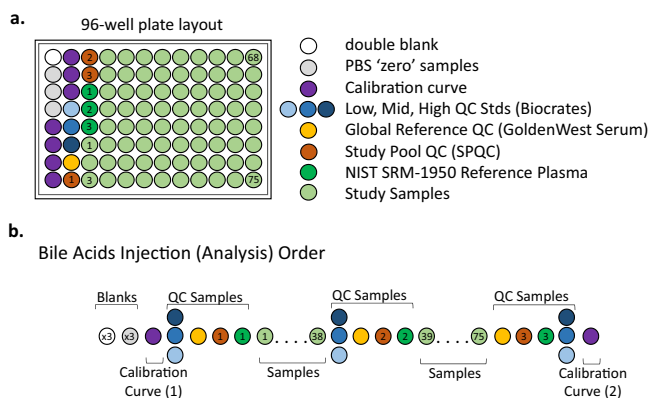
**Metabolomics analysis using the biocrates bile acids kit.** *Sample preparation.* Samples were prepared and analyzed in the Duke Proteomics and Metabolomics Shared Resource using the Biocrates<sup>®</sup> Bile Acids Kit (Biocrates Life Sciences AG, Innsbruck, Austria) in accordance with the user manual. In brief, 10  $\mu$ L of the supplied internal standard solution were added to each well (except for the zero sample) on a filterspot of the 96-well extraction plate. After drying under a gentle stream of nitrogen 10  $\mu$ L of each serum sample, quality control (QC) samples, blank, zero sample, or calibration standard were added to the appropriate wells (Fig. 1).

	CN (n = 193)	LMCI (n = 357)	AD (n = 172)	p-value
Age, year M (SD)	75.60 (4.90)	74.71 (7.44)	75.09 (7.35)	0.84
Sex, Female % (n)	50% (96)	35% (126)	49% (84)	<0.001
APOE ε4, % (n)	28% (54)	53% (189)	67% (115)	<0.001
MMSE, M (SD)	29.17 (0.96)	26.96 (1.81)	23.44 (2.00)	<0.0001
ADAS-Cog13, M (SD)	9.33 (4.16)	18.61 (6.32)	28.57 (7.66)	<0.0001

**Table 1.** Demographics and clinical data of studied ADNI subjects at baseline. Abbreviations: AD-Alzheimer's Disease; ADAS-Cog13: Alzheimer's Disease Assessment Scale-Cognitive subscale; CN-cognitively normal, MMSE-Mini-Mental State Exam, LMCI-late mild cognitive impairment.

	CN (n = 177)	SMC (n = 98)	EMC (n = 284)	LMCI (n = 148)	AD (n = 133)	p-value
Age (years)	73.46 (6.29)	72.18 (5.63)	71.12 (7.51)	72.12 (7.64)	74.21 (8.33)	<0.001
Sex, Female % (n)	53% (94)	57% (56)	46% (130)	48% (71)	41% (55)	0.09
APOE ε4, % (n)	28% (50)	33% (32)	43% (121)	57% (84)	65% (87)	<0.001
MMSE, M (SD)	29.03 (1.27)	29.00 (1.20)	28.32 (1.60)	27.64 (1.79)	23.06 (2.06)	<0.001
ADAS-Cog13, M (SD)	9.05 (4.48)	8.78 (4.12)	12.64 (5.40)	18.80 (7.31)	31.10 (8.67)	<0.001

**Table 2.** Demographics and clinical data of studied ADNIGO/2 subjects at baseline. Abbreviations: AD-Alzheimer's disease; ADAS-Cog13: Alzheimer's Disease Assessment Scale-Cognitive subscale; CN-cognitively normal; EMC: early mild cognitive impairment; MMSE-Mini-Mental State Exam; LMCI-late MCI; SMC-subjective memory complaints.



**Fig. 1** Plate layout for quantitative bile acids analysis in ADNI cohorts. **(a)** 96-well plate layout used for sample preparation and data collection for the bile acids metabolomics analysis by LC-MS/MS. Each of the plates analyzed in the study used the same lot of calibrators, Biocrates QCs, study pool QC (SPQC), GoldenWest Serum and NIST SRM-1950 plasma. **(b)** Analysis order for each plate showing how the calibration curve and QC samples bracket the actual sample analyses, following FDA guidance for regulated bioanalysis, in order to decrease the likelihood of intraplate bias.

The plate was then dried under a gentle stream of nitrogen. Sample extract elution was performed with methanol. Sample extracts were diluted with water for UPLC-MS/MS.

**Quality control samples.** The analysis of the samples using the Biocrates® Bile Acids Kit was performed using four specific sets of quality controls. First, low/mid/high level QC samples provided by Biocrates Life Sciences AG were prepared and analyzed on each plate as recommended by the manufacturer. These QC samples were used for a technical validation of each kit plate. Second, to allow appropriate inter-plate abundance scaling based specifically on this cohort of samples, we generated a Study Pool QC (SPQC) by combining approximately 10 µL from the first 75 samples for analysis. This sample was frozen in aliquots of 45 µL then prepared and analyzed three times on each plate. Third, the study utilized 18 blinded analytical duplicates and 15 blinded analytical triplicates for ADNI 1 and ADNIGO/2, respectively. These replicates, obtained from the same serum draw, were scattered throughout the study in a manner blinded to the investigators until data was sent to the ADNI informatics core for unblinding. The commonly used reference materials NIST SRM-1950 plasma (n = 3 per plate) and GoldenWest serum pool (n = 1 per plate) were also analyzed on each plate to allow cross-comparison against other sample cohorts in the future.

ADNI 1			ADNI GO/2			Processing Level	<i>R</i> Script Summary <sup>a</sup>
Samples	Subjects	Analytes	Samples	Subjects	Analytes		
819	801	20	904	874	20	Matrix Level 1	
815	797	20	904	874	20	Matrix Level 2	Scale to SPQC, remove thawed samples
815	797	15	904	874	15	Matrix Level 3	Remove highly variable analytes
744	726	15	878	848	15	Matrix Level 4	Impute missing values, filter clinical outliers (metadata)
722	722	15	840	840	15	Matrix Level 5	Combine replicate values per subject, filter statistical outliers, log transform data

**Fig. 2** Workflow description for data curation and scaling of the bile acids data. The use of Levels breaks the workflow into discrete steps which can be applied to multiple metabolomics data types, and will be consistent across the eight metabolite datasets collected for ADNI. Filtering for ADNI 1 is shown on left, and ADNI GO/2 is shown on the right. The workflow executed in *R* is described on the right. \*Subjects flagged for exclusion in Level 4 are not physically excluded from the table until Level 5.

Figure 1a shows the preparation layout for the 96-well plates as utilized in this study. In total, eleven plates were prepared in order to analyze serum samples for ADNI1 (Oct–Nov 2015) along with the blinded replicates. The same approach was used in the analysis of ADNI GO/2 cohort (Oct–Nov 2016) plus blinded replicates. The blank, zero sample, calibration standards, and Low/Mid/High QC samples provided with the kit were arranged as recommended by Biocrates. In order to improve the ability to compare results with other metabolomics studies and reduce plate-to-plate batch effects, seven additional wells were used for the additional QC samples as described above: three wells for the study pool QC (SPQC), one well for the GoldenWest Pooled Serum Standard, and three wells for the NIST SRM-1950 Standard Reference Plasma. The remaining 75 wells were used for cohort samples. The analysis order of each plate is summarized in Fig. 1b. The order was arranged to maximize quantitative accuracy and precision within a plate, and limit the potential for batch effects. The analysis order included running the standard curve twice, once at the beginning and end of the samples. The Biocrates QCs and GoldenWest Serum QC were prepared once but injected in technical triplicate, once before, in the middle (after 38 samples), and at the end of the sample set. The SPQC samples ( $n = 3$ ) were each analyzed once, with one analysis before, in the middle, and one after all samples on the plate. The NIST SRM-1950 plasma ( $n = 3$ ) were also analyzed once each at the beginning, middle, and end of the cohort samples. Bracketing the standard curves and nesting the analytical samples between the QCs offers the best chance of observing any system drift and assuring optimal instrument performance across the sample set.

**Quantitative UPLC-MS/MS analysis.** Mass spectrometry analysis was performed based on Standard Operating Procedure (SOP #8111) provided by Biocrates for the Bile Acids kit. Chromatographic separation of the analytes was performed using an ACQUITY UPLC System (Waters Corporation) using a proprietary reverse-phase UPLC and guard column provided by Biocrates then quantified by calibration curve using a linear regression with  $1/x^2$  weighting. Samples were introduced directly into a Xevo TQ-S mass spectrometer (Waters Corporation) using negative electrospray ionization operating in the Multiple Reaction Monitoring (MRM) mode. MRM and pseudo-MRM transitions (compound-specific ions) for each analyte and internal standard were collected over a scheduled retention time window using tune files and acquisition methods provided in the Biocrates® Bile Acids kit. The UPLC data were imported into TargetLynx (Waters Corporation) for peak integration, calibration, and concentration calculations. The UPLC data from TargetLynx were analyzed using Biocrates' MetIDQ v5.4.8 software. The kit data are reported in detail in the Supplemental Information on LONI, along with a color-coded key denoting samples that were below the limit of detection (<LOD) or below the lowest calibration standard (<LLOQ). The data generated for the study samples and SPQC samples can be downloaded using the appropriate links contained in Online-only Table 1.

**Data processing.** *R* v3.2.4 ([www.r-project.org](http://www.r-project.org)) statistical software was used for data analysis. *R* scripts for data processing are available at links provided in Online-only Table 1. The workflow for processing data from the raw set of concentration values for each subject in each cohort to something that is prepared for statistical analysis includes a series of important steps as previously described for the AbsoluteIDQ p180 platform dataset in ADNI 1<sup>23</sup>. Figure 2 provides the graphical representation of the important steps in this process for the Bile Acids datasets for ADNI 1 and ADNI 2/GO. We briefly describe the overall data processing with different levels of the data, as different stages of quality control and processing below, but detailed descriptions of each step can be found in prior publications<sup>23</sup> or in the supplemental material described in Online-only Table 1. The input to this pipeline, formerly called "Level 0", is the \*.xlsx or \*.csv format measured concentrations exported from the Biocrates software, MetIDQ.

The first step of QC was to exclude four samples that were inadvertently included in the ADNI 1 cohort, and to scale the quantitative value of each metabolite across plates and cohorts using the pooled NIST samples that were analyzed three times in ADNI1 and twice in ADNIGO/2 on each plate. Scaling was done by dividing the global average of each metabolite level by its average within the plate. These batch effect adjusted values are included in the Intermediate Data Level 2. Overall, this scaling factor was small (typically less than 10%), as can be seen by the raw reported NIST SRM-1950 values reported in Online-only Table 2.

The next step of QC involved filtering based on quality metrics. We routinely applied filter criteria to each of the metabolites (based on the blinded ADNI 1 duplicates or ADNI 2/GO triplicates) to allow only the most robust analytes to be included in downstream analysis. Separately for each cohort, we used a coefficient of variation (CV) <30% across plates to filter out metabolites with limited variation and therefore statistical power for analysis. Next, we used an intraclass correlation coefficient (ICC) between the values for the blinded duplicate (or triplicate) analyses >0.6. Finally, analytes with >40% of measurements below the lower limit of detection (<LOD) were filtered out. This filtered data represents the Level 3 data matrix. This filtering reduced the total number of analytes reported from 20 analytes (Level 2) to 15 analytes (Level 3). The filter QC results are presented in detail in the Supplementary Table 1.

The next step in data processing performs missing value replacement, by imputing any values reported as '<LOD' were using LOD/2 value for each specific analyte. Additionally, we screened for outliers for removal prior to analysis. In ADNI 1, there were a total of 71 samples identified as outliers based on the following criteria: 69 samples identified as non-fasting, 2 samples lacking corresponding body mass index (BMI) values, and 1 for which no baseline medication record was reported. The resulting Intermediate Data – Level 4 contained  $n = 744$  samples (726 subjects) and  $n = 15$  analytes. In ADNI GO/2, there were a total of 26 non-fasting samples that resulted in the Intermediate Data – Level 4 contained  $n = 878$  samples (848 subjects) and  $n = 15$  analytes.

The final step of data processing (Level 5) prior to analysis achieved the following goals. First, duplicate/triplicates measures for the blinded duplicates were averaged to give singular values for each analyte for each sample. An additional screen for outliers was performed from a statistical perspective using Principal Components Analysis (PCA). Principal components that explained >90% cumulative variance were selected, and any subjects located more than 7 SD from the mean were filtered out as outliers. This identified 4 and 6 additional samples, respectively, that were excluded from the final data matrices for ADNI 1 and ADNI 2/GO. Finally, all metabolites were log<sub>2</sub> transformed. The final, analysis-ready data matrix (Fig. 2, Matrix Level 5) contained 722 and 840 subjects in ADNI 1 and ADNI GO/2, respectively, and 15 analytes.

Because this processing represents an initial analysis, relatively conservative quality control was performed. A major motivation of the different levels of data within the workflow was to provide data at different levels of processing for further evaluation, including analyses targeting different hypotheses and using different assumptions than undertaken in the initial analysis. Full transparency is emphasized, as data utilized or generated in each step of the pipeline is available in Online Table 1<sup>26–40</sup>.

**Collection and curation of medication data.** As with any other observational cohort, there are a number of challenges with confounding in the study design. We previously reported an informatics framework which utilizes the free text medication information available on LONI as well as the National Library of Medicine's (NLM) RxNorm API (application programming interface) to match drug names to standardized drug concept identifiers, thus creating a common set of Boolean flags for each patient, annotating whether or not they were taking a drug in a specific class<sup>23</sup>. These binary flags make it much more straightforward to include medications in statistical analysis to assess potential confounding in subsequent association analyses. The same approach was utilized in this work, applied now to both ADNI1 and ADNIGO/2 cohorts. The code for this processing pipeline, along with documentation and API configuration files is available in Synapse: <https://doi.org/10.7303/syn7477310><sup>41</sup>. The final table of Boolean variables for each drug class at the baseline visit for ADNI1 is available at <https://doi.org/10.7303/syn7440367.1><sup>42</sup>, and ADNIGO/2 is available <https://doi.org/10.7303/syn12179110.1><sup>43</sup>.

**Data Records.** Online-only Table 1 details the location of each of the files generated in this study (called Data in the "Type" column) or utilized in this study, which are hosted on the AMP-AD Knowledge Portal (<https://ampadportal.org>). The Online-only Table 1 includes links to R scripts for data processing, processing pipeline for medication data along with documentation and API configuration files, as well as links to data from both ADNI 1 and ADNI GO/2 for Bile Acids in various stages<sup>26–45</sup>.

Data use restrictions prohibit the distribution of any ADNI clinical or demographic data outside of LONI, so the data files to be input into the processing pipeline are hosted there. Researchers can apply for access to the ADNI data at <https://ida.loni.usc.edu/collaboration/access/appLicense.jsp>. Online-only Table 1 details the location of each of the files needed to reproduce the data presented here. The data must be downloaded from LONI (clinical data and medication data), and placed in the same directory as the scripts provided. There are points in some of the scripts where manual intervention is necessary as detailed in readme files that accompany the scripts.

## Technical Validation

The Bile Acids kit from Biocrates was validated according to European Medicine Agency Guideline on bio-analytical method validation. Additionally, the methodologies utilized in this study performed within the Duke Proteomics and Metabolomics Shared Resource include bracketing calibration curves and quality control samples throughout the run list of each plate, consistent with the FDA Guidance on Bioanalytical Method Validation. Additionally, the measurement of 17 human and 20 total bile acids in plasma using this kit was recently shown to have bias <30% and an average %CV <15% across 12 laboratories in an international ring trial<sup>46</sup>. Each kit includes an automated technical validation based on Quality Control samples provided by the vendor (Biocrates AG), which are plasma samples spiked at three different levels. The low, mid, and high QC samples serve to verify

the analytical performance of each kit once the data is imported into the MetIDQ software package (Biocrates AG). A specific standard (calibrator) was considered valid, and included in sample quantification, when the back-calculated quantity (bias) was within 20% of the expected value at the lower limit of quantification (LLOQ) and within 15% at all points above the LLOQ. As a technical note, specific compounds within this kit can have high MS background (for example LCA), and it is important to utilize the test mix provided by Biocrates as a system suitability test. The most common reason for high background prior to sample analysis was found to be contaminants in the formic acid or methanol used to make the mobile phase, which was alleviated by utilizing formic acid only from glass ampules and fresh bottles of LC-MS grade methanol.

## Usage Notes

The general ADNI data use agreements and access policies apply to investigators using the metabolomics data. For details on how to apply for access and rules of usage, please see: <http://adni.loni.usc.edu/data-samples/access-data/>.

It is also important to note that users must apply for access and accounts for both ADNI and Synapse separately.

## Code Availability

We are highly committed to sharing all resources used to produce this data and analysis. Primary distribution of the scripts used in analysis is available through Sage Bionetworks' Synapse platform through the AMP-AD Knowledge Portal (<https://ampadportal.org>), with links to the data from both ADNI 1 and ADNI GO/2 for Bile Acids in various stages, as well as the R scripts used to process the data, available at the links shown in Online-only Table 1.

## References

- Brookmeyer, R., Johnson, E., Ziegler-Graham, K. & Arrighi, H. M. Forecasting the global burden of Alzheimer's disease. *Alzheimer's & dementia: the journal of the Alzheimer's Association* **3**, 186–191, <https://doi.org/10.1016/j.jalz.2007.04.381> (2007).
- McKhann, G. M. *et al.* The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* **7**, 263–269, <https://doi.org/10.1016/j.jalz.2011.03.005> (2011).
- Petersen, R. C. *et al.* Mild cognitive impairment: Ten years later. *Archives of Neurology* **66**, 1447–1455, <https://doi.org/10.1001/archneuro.2009.266> (2009).
- Han, X., Holtzman, D. M. & McKeel, D. W. Jr. Plasmalogen deficiency in early Alzheimer's disease subjects and in animal models: molecular characterization using electrospray ionization mass spectrometry. *J Neurochem* **77**, 1168–1180 (2001).
- Han, X. *et al.* Metabolomics in early Alzheimer's disease: identification of altered plasma sphingolipidome using shotgun lipidomics. *PLoS One* **6**, e21643, <https://doi.org/10.1371/journal.pone.0021643> (2011).
- Toledo, J. B. *et al.* Metabolic network failures in Alzheimer's disease: A biochemical road map. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* **13**, 965–984, <https://doi.org/10.1016/j.jalz.2017.01.020> (2017).
- Kaddurah-Daouk, R. *et al.* Metabolomic changes in autopsy-confirmed Alzheimer's disease. *Alzheimer's & dementia: the journal of the Alzheimer's Association* **7**, 309–317, <https://doi.org/10.1016/j.jalz.2010.06.001> (2011).
- Motsinger-Reif, A. A. *et al.* Comparing metabolomic and pathologic biomarkers alone and in combination for discriminating Alzheimer's disease from normal cognitive aging. *Acta neuropathologica communications* **1**, 28, <https://doi.org/10.1186/2051-5960-1-28> (2013).
- Kaddurah-Daouk, R. *et al.* Alterations in metabolic pathways and networks in mild cognitive impairment and early Alzheimer's disease. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* **9**, P571, <https://doi.org/10.1016/j.jalz.2013.05.1126> (2013).
- Mahmoudian Dehkordi, S. *et al.* Altered bile acid profile associates with cognitive impairment in Alzheimer's disease—An emerging role for gut microbiome. *Alzheimer's & dementia: the journal of the Alzheimer's Association*, <https://doi.org/10.1016/j.jalz.2018.07.217> (2018).
- Nho, K. *et al.* Altered bile acid profile in mild cognitive impairment and Alzheimer's disease: Relationship to neuroimaging and CSF biomarkers. *Alzheimer's & dementia: the journal of the Alzheimer's Association*, <https://doi.org/10.1016/j.jalz.2018.08.012> (2018).
- Pistollato, F. *et al.* Role of gut microbiota and nutrients in amyloid formation and pathogenesis of Alzheimer disease. *Nutr Rev* **74**, 624–634, <https://doi.org/10.1093/nutrit/nuw023> (2016).
- Haas, C. Strategies, development, and pitfalls of therapeutic options for Alzheimer's disease. *Journal of Alzheimer's disease: JAD* **28**, 241–281, <https://doi.org/10.3233/jad-2011-110986> (2012).
- Szeto, J. Y. Y. & Lewis, S. J. G. Current Treatment Options for Alzheimer's Disease and Parkinson's Disease Dementia. *Current neuropharmacology* **14**, 326–338, <https://doi.org/10.2174/1570159X14666151208112754> (2016).
- Cummings, J. *et al.* Drug development in Alzheimer's disease: the path to 2025. *Alzheimer's Research & Therapy* **8**, 39, <https://doi.org/10.1186/s13195-016-0207-9> (2016).
- Kaddurah-Daouk, R. & Krishnan, K. R. Metabolomics: a global biochemical approach to the study of central nervous system diseases. *Neuropsychopharmacology: official publication of the American College of Neuropsychopharmacology* **34**, 173–186, <https://doi.org/10.1038/npp.2008.174> (2009).
- Kaddurah-Daouk, R., Kristal, B. S. & Weinshilboum, R. M. Metabolomics: A global biochemical approach to drug response and disease. *Annu Rev Pharmacol* **48**, 653–683, <https://doi.org/10.1146/annurev.pharmtox.48.113006.094715> (2008).
- Fiandaca, M. S. *et al.* Plasma 24-metabolite Panel Predicts Preclinical Transition to Clinical Stages of Alzheimer's Disease. *Front Neurol* **6**, 237, <https://doi.org/10.3389/fneur.2015.00237> (2015).
- Klavins, K. *et al.* The ratio of phosphatidylcholines to lysophosphatidylcholines in plasma differentiates healthy controls from patients with Alzheimer's disease and mild cognitive impairment. *Alzheimer's & dementia: the journal of the Alzheimer's Association* **1**, 295–302 (2015).
- Mapstone, M. *et al.* Plasma phospholipids identify antecedent memory impairment in older adults. *Nat Med* **20**, 415–418, <https://doi.org/10.1038/nm.3466> (2014).
- Simpson, B. N. *et al.* Blood metabolite markers of cognitive performance and brain function in aging. *J Cereb Blood Flow Metab* **36**, 1212–1223, <https://doi.org/10.1177/0271678X15611678> (2016).
- Wood, P. L. *et al.* Circulating plasmalogen levels and Alzheimer Disease Assessment Scale-Cognitive scores in Alzheimer patients. *J Psychiatry Neurosci* **35**, 59–62 (2010).
- St John-Williams, L. *et al.* Targeted metabolomics and medication classification data from participants in the ADNI1 cohort. *Sci Data* **4**, 170140, <https://doi.org/10.1038/sdata.2017.140> (2017).

24. Petersen, R. C. *et al.* Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. *Neurology* **74**, 201–209, <https://doi.org/10.1212/WNL.0b013e3181cb3e25> (2010).
25. Weiner, M. W. *et al.* Impact of the Alzheimer's Disease Neuroimaging Initiative, 2004 to 2014. *Alzheimer's & dementia: the journal of the Alzheimer's Association* **11**, 865–884, <https://doi.org/10.1016/j.jalz.2015.04.005> (2015).
26. ADMC ADNI1 study. *Synapse*. <https://doi.org/10.7303/syn5592519> (2016).
27. ADMC ADNI2 GO study. *Synapse*. <https://doi.org/10.7303/syn9705278> (2017).
28. ADMC ADNI1 Bile Acids Data. *Synapse*. <https://doi.org/10.7303/syn12036817.1> (2018).
29. ADMC ADNI1 Bile Acids Dictionary. *Synapse*. <https://doi.org/10.7303/syn12036821.1> (2018).
30. ADMC ADNI1 Bile Acids Methods. *Synapse*. <https://doi.org/10.7303/syn12036820.1> (2018).
31. ADNI2-GO Bile Acids Data. *Synapse*. <https://doi.org/10.7303/syn9779093.1> (2017).
32. ADNI2-GO Bile Acids Data Dictionary. *Synapse*. <https://doi.org/10.7303/syn9779094.1> (2017).
33. ADMC ADNI2-GO Bile Acids Method Description. *Synapse*. <https://doi.org/10.7303/syn9779078.1> (2017).
34. Pipeline Structure. *Synapse*. <https://doi.org/10.7303/syn12036815> (2018).
35. Bile Acids QC and NIST Data. *Synapse*. <https://doi.org/10.7303/syn9779088.1> (2017).
36. ADNI1 BA LOD values. *Synapse*. <https://doi.org/10.7303/syn12046012.1> (2018).
37. ADNI GO 2 Bile Acids QC and NIST Data. *Synapse*. <https://doi.org/10.7303/syn9779088.1> (2017).
38. ADNI GO 2 Bile Acids LOD values. *Synapse*. <https://doi.org/10.7303/syn9779079.1> (2017).
39. ADNI 1 Bile Acid Ratios. *Synapse*. <https://doi.org/10.7303/syn12046208.1> (2018).
40. ADNI Fasting Status. *Synapse*. <https://doi.org/10.7303/syn12046023.1> (2018).
41. ADNI 1 participant baseline medications mapped to drug classes. *Synapse*. <https://doi.org/10.7303/syn7440367.1> (2016).
42. ADMC Duke ADNI2-GO Drug Classes. *Synapse*. <https://doi.org/10.7303/syn12179110.1> (2017).
43. Medication mapping pipeline. *Synapse*. <https://doi.org/10.7303/syn7477310> (2018).
44. ADNI Clinical Variables. *Synapse*. <https://doi.org/10.7303/syn7477271.1> (2016).
45. ADNI Medication Data. *Synapse*. <https://doi.org/10.7303/syn7829508.1> (2016).
46. Pham, H. T. *et al.* Inter-Laboratory Robustness of Next-Generation Bile Acid Study in Mice and Humans: International Ring Trial Involving 12 Laboratories. *J. Appl. Lab Med.* **01**, 129–142 (2016).

## Acknowledgements

We thank Lisa Howerton for her enthusiastic coordination and administrative support of this work. The results published here are in whole or in part based on data obtained from the AMP-AD Knowledge Portal (<https://doi.org/10.7303/syn2580853>). Funding for ADMC (Alzheimer's Disease Metabolomics Consortium, led by Dr. Kaddurah-Daouk at Duke University) was provided by National Institute of Aging under its AMP-AD (Accelerated Medicines Partnership for Alzheimer Disease, NIA #1R01AG046171) and MOVE-AD (Molecular Mechanisms of Vascular Etiology, NIA #1R01AG0151550) Programs. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health ([www.fnih.org](http://www.fnih.org)). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

## Author Contributions

St. John-Williams, Thompson, MahmoudianDehkordi, and Arnold had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Data management: Blach. Concept and design: Kaddurah-Daouk led concept and design team that included all co-authors. Sample Preparation and Data Collection: St. John-Williams. Drafting of the manuscript: Thompson, Blach, Motsinger-Reif, Tenenbaum, Kueider-Paisley, Kaddurah-Daouk. Biochemical, genomics and medications integration: Kastenmüller, Baillie, Han, Risacher, Koal. Data deposition: Alzheimer's Disease Neuroimaging Initiative (see note). Harmonization of methods: Alzheimer's Disease Metabolomics Consortium (see note). Technical, bibliographic research and/or material support: Louie. Critical revision of the manuscript for important intellectual content: Saykin, Moseley, Kaddurah-Daouk. Obtained funding: Kaddurah-Daouk. Supervision: Saykin, Moseley, Thompson, Kaddurah-Daouk. Data used in preparation of this descriptor were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database ([adni.loni.usc.edu](http://adni.loni.usc.edu)). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report.

## Additional Information

**Supplementary Information** is available for this paper at <https://doi.org/10.1038/s41597-019-0181-8>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2019