Title
The Metabolomic Signatures of HAND

Permalink
https://escholarship.org/uc/item/36b8k5xc

Author
Husband, Makhai Omarei

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

The Metabolomic Signatures of HAND

A thesis submitted in partial satisfaction of the

requirements for the degree Master of Science

in

Biology

by

Makhai O. Husband

Committee in charge:

       Professor Pieter C. Dorrestein , Chair
       Professor Stacey Glasgow, Co-Chair
       Professor Michelli Faria De Oliveira
       Professor Keefe Reuther

2021

The thesis of Makhai O. Husband is approved,

and it is acceptable in quality and form for

publication on microfilm and electronically.

University of California San Diego

2021

Dedication

I dedicate this work to my family

who believed in me in times when I did not believe in myself,

and to my friends that pushed me to being my best self

Thank you to everyone,

For being in my life

Table of Contents

List of Abbreviations

PLWH   People Living With Alzheimers

HIV   Human Immunodeficiency Virus

HAND   HIV Associated Neurocognitive Disorder

AD   Alzheimer's Disease

CSF   Cerebrospinal Fluid

RNA   Ribonucleic Acid

AIDS   Acquired immunodeficiency syndrome

MRI   Magnetic Resonance Imaging

CT   Computerized Tomography

ADRC   Alzheimer's Disease Research Center

HNRP   HIV Neurobehavioral Research Program

MCI   Mild Cognitive Impairment

GNPS   Global Natural Products Social Molecular Networking

FBMN   Feature Base Molecular Networking

PLS-DA   Principal Least Square – Discriminant Analysis

IQR   Interquartile Range

VIP   Variance of importance

List of Figures

List of Tables

Acknowledgements

I would like to acknowledge Professor Pieter Dorrestein for his support as the chair in my committee. His role as a mentor and his overall guidance throughout the endeavor was invaluable.

I would also like to acknowledge the other members of the Dorrestein lab, in particular that of Emmanuel Elijah and Kelly Weldon to whom this project could not have been completed without.

Vita

2016 Bachelor of Science, University of California San Diego

2021 Master of Science, University of California San Diego

Publications

"ReDU: a framework to find and reanalyze public mass spectrometry data." Nat Methods. 2020 Sep;17(9):901-904. doi: 10.1038/s41592-020-0916-7. Epub 2020 Aug 17. PMID: 32807955; PMCID: PMC7968862.

Fields of Study

Major Field: Human Biology

Studies in Metabolome

Professor Pieter C. Dorrestein

Abstract of the Thesis

The Metabolomic Signatures of HAND

by

Makhai O. Husband

Master of Science in Biology

University of California San Diego, 2021

Professor Pieter Dorrestein, Chair
Professor Stacey Glasgow, Co Chair

Throughout the span of nearly 30 years, the treatment for HIV has progressed from being considered a death sentence to now allowing people living with HIV (PLWH) to live longer and fulfilling lives[1]. While current treatments for HIV can stop the progression of HIV, they are not able to cure the disease altogether and the consequence of a longer lifespan among the PLWH population is the emergence of chronic pathogens due to the long term exposure of HIV[1]. In regards to HIV in the nervous system, one of the most prevalent pathogens affecting PLWH is that of HIV Associated Neurocognitive Disorder (HAND). Around 30-50% of people with HIV experience symptoms of HAND to varying degrees. Symptoms include impairments in cognitive, behavioral, and motor functions[2]. While HIV treatments can reduce the severity of these symptoms, there are still the prevalent challenges for distinguishing the disease from other neurodegenerative diseases such as Alzheimer's disease (AD)[3]. Furthermore, there is a disparity

regarding our knowledge on the underlying metabolites for HAND that drives the progression of the disease[3]. Thus differentiating the metabolomic profiles between HIV and HAND is the focus of this study. The profiles were compared for similarities and differences through the cohort groups of plasma and cerebrospinal fluid (CSF) from patients of varying levels of HIV and cognitive statuses.

Introduction

A. HIV Pathogenesis

Human immunodeficiency virus, otherwise known as HIV, is an RNA retrovirus that targets the immune system. The internal composition of the HIV virus is that of a virus spike made from two enveloping proteins (gp120 and gp41), a lipid bilayer, various enzymes, and a viral capsid containing an rna genome. HIV targets the immune system through the infection of helper T cells (CD4$^+$ T cells), macrophages, and dendritic cells[4]. Through connections with glycoproteins on the surface of these cells, HIV can identify the immune cells, fuse with their cell membrane, and then release the contents necessary for replication, including, reverse transcriptase, integrase, ribonuclease, and protease. After entering the cell, reverse transcription occurs and the viral reverse transcriptase transcribes the single-stranded RNA genome into double-stranded DNA, which is finally integrated into a host chromosome where it proceeds to replicate the virus[4].

Pathogenesis occurs as a consequence of the depletion of the cells in the immune system. In particular the number of CD4$^+$ T cells serves as an indicator of the progression of HIV in the body[6]. The CD4 count in a healthy person ranges from 500 to 1,600 cells per cubic millimeter of blood (cells/mm3), which a person in the first and second stages of HIV can maintain with proper treatment[6]. However, in the third stage of HIV, the CD4 count is lower than 200 cell/mm which is below the critical level necessary to maintain cell-mediated immunity leading to the development of AIDS [6]. AIDS (acquired immunodeficiency syndrome) is the late stage of HIV infection that occurs when the body's immune system is badly damaged because of the virus. The most prevalent symptom of AIDS is the emergence of opportunistic infections and cancers that would not be expected to occur in someone with a healthy immune system. While the

average life expectancy of people with AIDS is 3 years after being diagnosed, typically most people with HIV do not progress to AIDS due to proper medication use.

B. Pathogenesis of HAND

The infection of HIV in the body results in a myriad of changes that are separate from the immune system. HIV Associated Neurocognitive Disorders (HAND) is a direct consequence of the infection of HIV reaching the brain. HAND denotes a classification of neurological disorders of varying severities that are all associated with HIV infections as well as its progression to AIDS[8]. Infection of the brain occurs as a result of HIV exploiting infected cells traveling to the brain and thus circumventing the blood brain barrier. As HIV infects the glial and macrophage cells in the brain, the cells release toxins that promote the apoptosis of neurons as well as triggering other glial cells (most notably astrocytes) to harm neurons[9]. It has also been found that gp120 (an HIV protein) is involved in the process of inhibiting neurogenesis which results in the blocking of stem cells in the brain from producing new nerve cells.

Symptoms vary in severity from patient to patient but generally include a progressive decline in cognition, memory, behavior, and motor functions. The level of severity of HAND depends on a variety of factors from the stage of HIV, duration of time with HIV, and HIV treatment[10]. It should be noted that viral load does not necessarily correlate with the progression of cognitive decline. The diagnostic process of HAND includes longitudinal examinations of a patient's cognition in addition to MRI and CT scans[10]. Regarding brain anatomy, HAND has been associated with reduced white matter volume as well as an overall decline of volume in the cortical and subcortical regions of the brain[11]. Despite these indicators, there are still a plethora

of unknowns regarding reliable biomarkers of HAND which is what this study seeks out to explore.

   C. CSF and Plasma

The purpose of utilizing both cerebrospinal fluid (CSF) and plasma samples for our data is twofold. HIV presents itself via a measurable viral load in both CSF and plasma which we are able to collect from a single patient. Not only does this increase the sample size of our data, we are additionally able to form comparisons between the metabolomic changes that occur as a result of HIV and cognitive impairment in different regions of the body.

For the study, we evaluated approximately 500 samples of paired CSF and plasma samples from participants enrolled in the UCSD HIV Neurobehavioral Research Program (HNRP) as well as the Shiley-Marcos Alzheimer's Disease Research Center (ADRC). Regarding the samples from the HNRP cohort, they were derived from participants enrolled in three ongoing longitudinal research studies managed by the department between the years 1999 and the present. The total study follow-up periods in longitudinal studies ranges from 1 to 20 years. All procedures related to samples collection, laboratory procedures, neurobehavioral assessments and clinical data collection at enrollment and longitudinal visits under the HNRP were standardized across all the studies and all participants enrolled provided written informed consent. In regards to the classification of HAND, the Frascati criteria was implemented. The Frascati criteria consists of the participants completing a comprehensive neurocognitive test battery that takes several forms, including, an oral reading-based premorbid verbal intelligence estimate, as well as multiple measures of speed of information processing, learning, memory (delayed recall), executive functions, verbal fluency, attention/working memory and complex

motor skills. This is necessary to ensure that the study covers the 7 cognitive domains that are attributed to affected HIV-associated CNS dysfunction.

Furthermore, we have retrospectively selected 200 HIV-infected participants from the parent HNRP studies. Each participant must have had available pairs of CSF and blood samples, a minimum of 2 longitudinal visits coupled with neurocognitive assessments, and lastly a suppressed level of blood and CSF HIV viral load. Our grouping criteria will be based on HAND classification: (i) participants who had normal cognition at baseline and had incident HAND at a subsequent time-points (HAND group, n=100) and, (ii) age-matched participants who had normal cognition at baseline and did not develop HAND in any of the subsequent time-points (no-HAND group, n=100). For our study assays, we will include baseline paired blood and CSF samples for each participant in each group, if possible, the immediately-preceding visit to the one where participants (within the HAND group) first met diagnostic criteria for HAND. For participants within the HAND group, we will also include paired CSF and blood at the visit when the participant first met the diagnostic for HAND (HAND time-point). First, because we cannot anticipate how far early molecular changes underlying incident HAND will appear; second, by evaluating both time-points, we evaluate what HAND-related molecular signatures will start early (at baseline) and will be persistent at the HAND time-point, and which ones will manifest only during the HAND event.

Regarding samples from ADRC the method of sample selection was similar to that of the HNRP cohorts. Our selection criteria includes participants with available stored CSF and blood samples with at least 2 longitudinal visits with neurocognitive assessments. Additionally we retrospectively selected 100 participants. The study groups were defined based on neurocognitive classification: participants (i) who had Mild Neurocognitive Impairment (MCI) at baseline and

developed AD in the subsequent time-point (AD group, n=50) and, (ii) age-matched participants who had normal cognition during all observational time-points (control group, n=50). In contrast to the HNRP, we evaluated paired stored blood and CSF samples only at baseline for both ADRC groups. This decision is based on the evidence that AD pathophysiology is likely to begin several years before the emergence of clinical symptoms. Therefore, we are confident that we will be able to identify molecular signatures during the pre-AD stage (i.e. MCI), when compared to cognitively normal controls.

Research objective

My research objective is to further expand our understanding of the metabolomic signatures of HAND through the identification of metabolomic profiles of patients living with HAND and analyzing the similarities and differences of HAND with other patient groups such as HIV (with no cognitive impairment), Alzheimer's, and people with no ailments (control group).

Sample Method Optimization

Sample Collection and Analysis

The initial Plasma and CSF sample collections were done by the ADRC and HNRP organizations using the previously described criteria. The samples were later acquired by the Dorrestein lab for further processing by my associate Emmanuel Elijah. The methodology began with thawing the samples on ice for 30 minutes (after being previously stored in a -80 C container). Once thawed, the samples were extracted using a Phree Phospholipid Removal Kit with 80% methanol. The samples were then vortexed at a maximum speed of 1300 rpm for two minutes. Afterwards, the samples were dried using a centrifugal low pressure system and resuspended in 50% methanol. Finally, the samples were analyzed using an ultra-high pressure LC-MS system (Vanquish, Thermo Scientific) coupled to an Orbitrap mass spectrometer (Q Exactive, Thermo Scientific).

Metabolomic analysis

The primary tool for our metabolomic analysis was the Global Natural Products Social Molecular Networking (GNPS) ecosystem. As described on its homepage "GNPS is a web-based mass spectrometry ecosystem that aims to be an open-access knowledge base for community-wide organization and sharing of raw, processed or identified tandem mass (MS/MS) spectrometry data"[12]. By inputting the necessary metadata and MZmine feature table into the GNPS interface, it is possible to assess an array of information such as Molecular Networks, Feature-Base Molecular Networks, as well as performing various statistical analysis.

Molecular networks can be best described as the visual representations of the chemical space derived from our mass spectrometry (MS/MS) experiments [12]. The networks are generated by grouping sets of spectra from related molecules (utilizing both known and unknown spectra).

9

Additionally, Cytoscape is used in conjunction with GNPS. Cytoscape is an open source software platform utilized in a similar fashion to GNPS for the visualization and annotation of molecular networks [13]. What separates cytoscape's visualization to that of GNPS is that while GNPS is limited to observations of a single component at a time, cytoscape is able to visualize the entire molecular network at once. In regards to the actual visualization, spectra are represented in these molecular networks as nodes that are connected by edges which indicate similar fragmentation. Similar fragmentation indicates similarity in the structure among the nodes. Edges also discern mass shifts between the nodes that provide further information that can be discerned given the appropriate metadata.

Feature Based Molecular Networking (FBMN) is another form of analysis that relies on feature detection to determine a molecule's abundance and aligns these abundances across a cohort of samples into consensus features. Corresponding MS/MS spectra are assigned to each consensus feature and analyzed with spectral library search and molecular networking. The advantages that FBMN provides are ... 1. More accurate quantification of molecules 2. Resolution of isomeric compounds 3. Reduction of redundancy of MS/MS molecules [14]

Furthermore, from the GNPS data generated, we were able to utilize metaboanalyst.ca, an internet platform that allows us to generate PLS-DA and Random Forest tables[15], for the purpose of determining key metabolites in our data. This was accomplished by utilizing data tables derived from GNPS that were then uploaded to the statistical analysis category of the metaboanlyst.ca site. Before analysis could be conducted, the data was filtered using the interquartile range (IQR) in addition to quantile normalization as well as being auto-scaled.

PLS-DA or PLS-Discriminant Analysis is a linear classification model that serves as a feature selector and classifier. PLS methods generate linear combinations of the original

metabolomics variables to maximize the separation among groups[16]. Through the VIP score derived from the analysis we can assess the level of importance for certain metabolites in determining the classification among our groups of interest. VIP score stands for "Variable Importance in Projection" and serves as a measure of a variable's importance in the PLS-DA model. By calculating the weighted sum of the squared correlations between the original variable that we are observing and that of the PLS-DA components, the score summarizes the contribution that variable makes to the model [17].

Random forest is a form of analysis that utilizes a machine learning algorithm to determine the importance of certain metabolites in their contribution for predicting groups[18]. The forest aspect of the name denotes how the algorithm creates several decision trees that are then merged together in order to create an accurate and stable prediction model. The purpose of random forest in this experiment is for the model to determine metabolites of interest through an assessment of the accuracy of the model for discerning the experimental groups using the Meandecreaseaccuracy values. Meandecreaseaccuracy is a measure of how much of the accuracy the prediction model decreases due to the removal of the variable, thus the higher the decrease, the higher the importance of the role of that metabolite in the separation of the groups.

In both the PLS-DA and Random forest data generated, metaboanalyst also provided the relative abundances of the identified metabolites in respect to the groups we are investigating. Furthermore, to better visualize and evaluate the significance of these abundances, box plots and P-values were generated from the feature tables used for metaboanalyst.

Results

Figure 1. Plasma Separation in all 4 groups a) Partial Least Square Discriminant Analysis (PLS-DA) of the Normal, Alzheimer's, HIV, and HAND groups. The top metabolites (features) driving separation between the different statuses are listed in their order of importance (VIP score) b) Random Forest results expressing the top 15 contributing metabolites with the Mean Decrease Accuracy (MDA) representing their level of accuracy.

a) PLS-DA
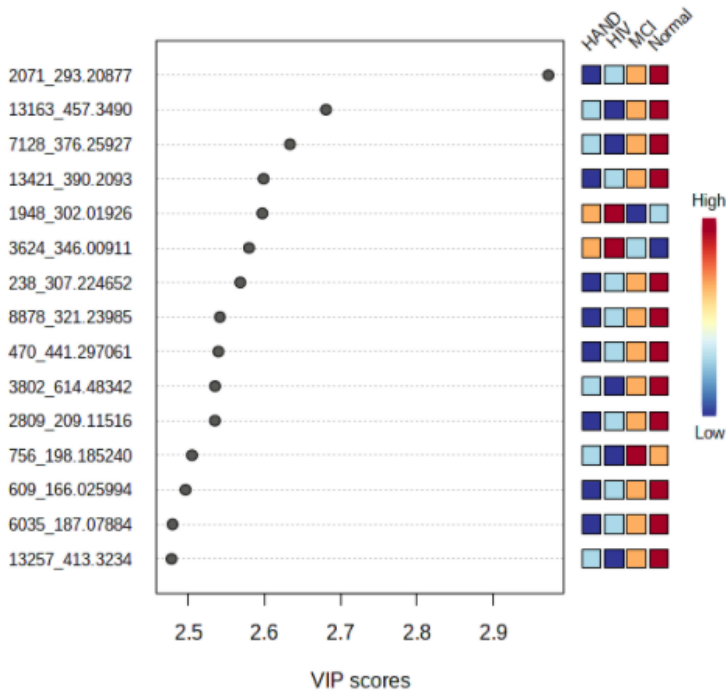
b) Random F
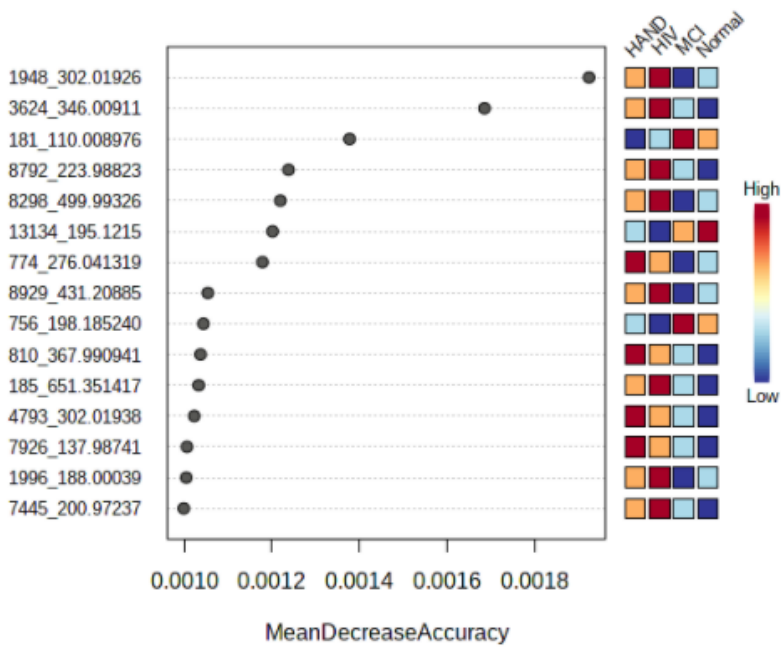
The legend for Figure 1 is described below:

HIV represents the HIV positive patients with no cognitive impairment, HAND represents HIV positive patients with cognitive impairment, Normal represents HIV negative patients with no cognitive impairment, and MCI represents HIV negative patients with cognitive impairment.

The top features driving the separation in our patient groups can be found with the VIP and MDA (mean decrease accuracy scores) values in the figure above along the x axis. The higher VIP score signifies the increased importance of that metabolite for the separation in the patient groups for the PLS-DA plot and the lower the MDA score denotes the importance of that metabolite for the Random Forest model. The rightmost side of both PLS-DA and Random Forest figures denote the relative abundance of the metabolite in relation to the stated groups via the color gradient chart. The leftmost side on the y axis of both the PLS-DA and Random Forest figures denotes the Cluster ID (which can be discerned as the collection of numbers prior to the "_") found in the GNPS job which is attached below.

https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=4bd384e8a048435ca7444927f2de5e47

Table 1. Identified PLS-DA and Random forest features of Figure 1 a) The identified features of Figure 1 are displayed in relation to their annotated name, molecular network, and P-Value. b) Represents the identified metabolite in conjunction to a boxplot representing the abundances of the metabolite amongst the groups of interest.

a)

clusterid	Metabolite	Networked
2071	Unknown	Ethylenediaminetetraacetic acid EDTA
7128	Unknown	Diphenoxylate

Table 1. Identified PLS-DA and Random forest features of Figure 1 a) The identified features of Figure 1 are displayed in relation to their annotated name, molecular network, and P-Value. b) Represents the identified metabolite in conjunction to a boxplot representing the abundances of the metabolite amongst the groups of interest, Continued

clusterid	Metabolite	Networked
238	cis-5,8,11-Eicosatrienoic acid	Monolinolenin (9c,12c,15c)
609	Unknown	PhenylAlanine Histidine Naproxen

Table 1. Identified PLS-DA and Random forest features of Figure 1 a) The identified features of Figure 1 are displayed in relation to their annotated name, molecular network, and P-Value. b) Represents the identified metabolite in conjunction to a boxplot representing the abundances of the metabolite amongst the groups of interest, Continued

b)

clusterid	P-value	Box Plot
2071	HAND-HIV 1.729117e-01 MCI-HIV 2.119522e-05 Normal-HIV 3.373356e-08 MCI-HAND 4.842241e-07 Normal-HAND 5.487521e-10 Normal-MCI 5.477714e-01	
7128	HAND-HIV 7.147595e-01 MCI-HIV 1.075883e-15 Normal-HIV 3.809063e-22 MCI-HAND 3.930367e-13 Normal-HAND 5.825625e-18 Normal-MCI 7.147595e-01	

Table 1. Identified PLS-DA and Random forest features of Figure 1 a) The identified features of Figure 1 are displayed in relation to their annotated name, molecular network, and P-Value. b) Represents the identified metabolite in conjunction to a boxplot representing the abundances of the metabolite amongst the groups of interest. Continued.

clusterid	P-value	Box Plot
238	HAND-HIV 1.891434e-01 MCI-HIV 2.439170e-03 Normal-HIV 2.151621e-05 MCI-HAND 1.199040e-04 Normal-HAND 6.360281e-07 Normal-MCI 4.811102e-01	
609	HAND-HIV 5.411904e-02 MCI-HIV 9.476600e-04 Normal-HIV 2.505267e-06 MCI-HAND 5.042195e-06 Normal-HAND 3.066249e-09 Normal-MCI 4.178753e-01	

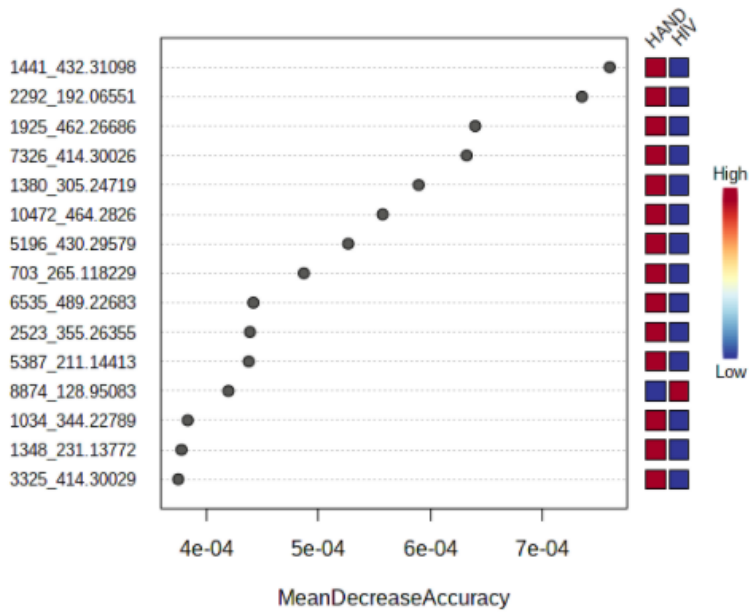The legend of Table 1 is described below.

Cluster ID signifies the ID # of the metabolite denoted by the GNPS job stated in Figure 1. All Cluster IDs displayed in the table were derived from the PLS-DA and Random Forest plots denoted in Figure 1. While other Cluster IDs were found in the previously stated plots, some were excluded if the Cluster ID failed to generate both a direct and indirect annotation. The term Metabolite in relation to the table describes the direct annotation that the Cluster ID is associated with in the GNPS job. Additionally, the term "Networked" refers to the annotated metabolites found in the molecular network of the stated metabolite. Furthermore, the term P-Value on the table is the value derived from the Dunn's test performed on the stated groups of interest to determine whether or not separation could be considered significant (to which significance was determined as a value below .05). In a similar vein, the boxplots on the rightmost side of the table serve as visual displays of the separation between groups in relation to their abundance. It should be noted that grouping nomenclature remains the same as it was in Figure 1

Figure 2. Plasma HIV vs HAND a) Partial Least Square Discriminant Analysis (PLS-DA) of HIV and HAND groups. The top metabolites (features) driving separation between the different statuses are listed in their order of importance (VIP score) b) Random Forest results expressing the top 15 contributing metabolites with the Mean Decrease Accuracy (MDA) representing their level of accuracy.

a) PLS-DA
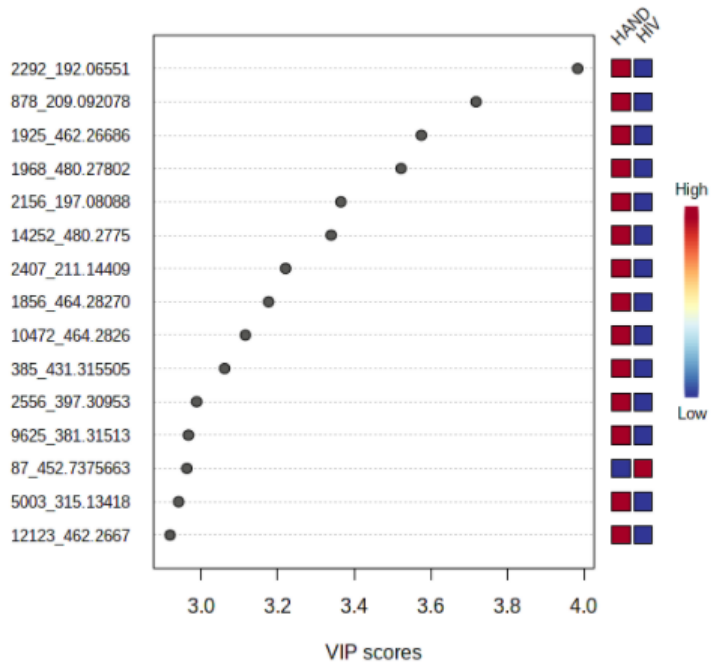
b) Random F

The legend for Figure 2 is described below:

        HIV represents the HIV positive patients with no cognitive impairment and HAND represents HIV positive patients with cognitive impairment.

        The top features driving the separation in our patient groups can be found with the VIP and MDA (mean decrease accuracy scores) values in the figure above on the x axis. The higher VIP score signifies the increased importance of that metabolite for the separation in the patient groups for the PLS-DA plot and the lower the MDA score denotes the importance of that metabolite for the Random Forest model. The rightmost side of both PLS-DA and Random Forest figures denote the relative abundance of the metabolite in relation to the stated groups via the color gradient chart. The leftmost side along the y axis of both the PLS-DA and Random Forest figures denotes the Cluster ID (which can be discerned as the collection of numbers prior to the "_") found in the GNPS job which is attached below.

https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=4bd384e8a048435ca7444927f2de5e47

Table 2. Identified PLS-DA and Random forest features of Figure 2 a) The identified features of Figure 2a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 2a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 2b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 2b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups.

a) PLS-DA

clusterid	Metabolite	Networked
2292	L-Kynurenine	KYNURENINE
878	KYNURENINE	L-Kynurenine
1925	taurocholic acid	GLYCOCHENODEOXYCHOLATE GLYCOCHOLATE glycoursodeoxycholic acid glycocholic acid taurohyodeoxycholic acid taurodeoxycholic acid

Table 2. Identified PLS-DA and Random forest features of Figure 2 a) The identified features of Figure 2a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 2a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 2b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 2b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued

clusterid	Metabolite	Networked
1968	Taurocholic acid	GLYCOCHENODEOXYCHOLATE GLYCOCHOLATE glycoursodeoxycholic acid glycocholic acid taurohyodeoxycholic acid taurodeoxycholic acid
14252	Taurocholic acid	GLYCOCHENODEOXYCHOLATE GLYCOCHOLATE glycoursodeoxycholic acid glycocholic acid taurohyodeoxycholic acid taurodeoxycholic acid
2407	cyclo(L-Leu-L-Pro)	Phe-Pro Val-Leu

Table 2. Identified PLS-DA and Random forest features of Figure 2 a) The identified features of Figure 2a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 2a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 2b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 2b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued

clusterid	Metabolite	Networked
1856	Taurohyodeoxycholic acid	GLYCOCHENODEOXYCHOLATE GLYCOCHOLATE glycoursodeoxycholic acid glycocholic acid taurohyodeoxycholic acid taurodeoxycholic acid
10472	Taurohyodeoxycholic acid	GLYCOCHENODEOXYCHOLATE GLYCOCHOLATE glycoursodeoxycholic acid glycocholic acid taurohyodeoxycholic acid taurodeoxycholic acid

Table 2. Identified PLS-DA and Random forest features of Figure 2 a) The identified features of Figure 2a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 2a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 2b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 2b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued

clusterid	Metabolite	Networked
385	Unknown	7alpha-Hydroxy-4-cholesten-3-one
2556	Unknown	Cholic acid
9625	Unknown	Cholic acid
12123	Taurocholic acid	GLYCOCHENODEOXYCHOLATE GLYCOCHOLATE glycoursodeoxycholic acid glycocholic acid taurohyodeoxycholic acid taurodeoxycholic acid
2831	Urobilin	

Table 2. Identified PLS-DA and Random forest features of Figure 2 a) The identified features of Figure 2a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 2a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 2b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 2b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued

clusterid	Metabolite	Networked
1380	15-OxoEDE	Stearidonic acid Stearidonic acid ethyl ester .omega.-3 Arachidonic acid ethyl ester from 13-Keto-9Z,11E-octadecadienoic acid 13S-Hydroxy-9Z,11E,15Z-octadecatrienoic acid cis-8,11,14-Eicosatrienoic acid
2916	Unknown	Cholic acid
14662	Unknown	Ursodeoxycholic acid

Table 2. Identified PLS-DA and Random forest features of Figure 2 a) The identified features of Figure 2a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 2a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The id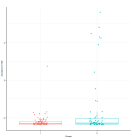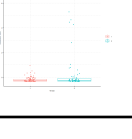entified features of Figure 2b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 2b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued
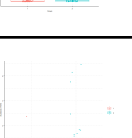
clusterid	P-value	Box Plot
3607	cyclo(L-Leu-L-Pro)	Phe-Pro Val-Leu cyclo(L-Val-L-Pro)
4146	Unknown	Cholesterol DEOXYCHOLATE 3.beta.-Hydroxy-5-cholenoic acid

b)
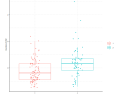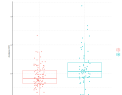
clusterid	P-value	Box Plot
2292	HAND-HIV 0.0002292024	
878	HAND-HIV 0.0005951189	

Table 2. Identified PLS-DA and Random forest features of Figure 2 a) The identified features of Figure 2a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 2a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 2b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 2b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued
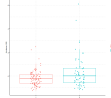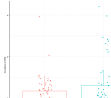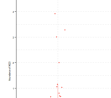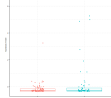
clusterid	P-value	Box Plot
1925	HAND-HIV 0.09479958	
1968	HAND-HIV 0.09091621	
14252	HAND-HIV 0.2004768	
2407	HAND-HIV 0.006632724	
1856	HAND-HIV 0.0505009	
10472	HAND-HIV 0.1849772	

30

Table 2. Identified PLS-DA and Random forest features of Figure 2 a) The identified features of Figure 2a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 2a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 2b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 2b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued

clusterid	P-value	Box Plot
385	HAND-HIV 0.05847079	
2556	HAND-HIV 0.01972246	
9625	HAND-HIV 0.002009635	
12123	HAND-HIV 0.3591336	

Table 2. Identified PLS-DA and Random forest features of Figure 2 a) The identified features of Figure 2a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 2a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The id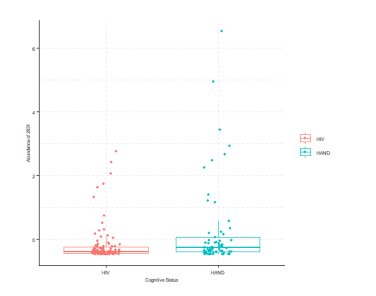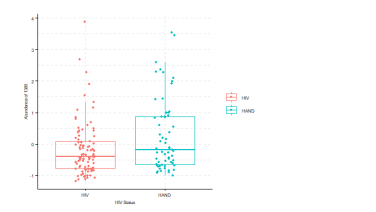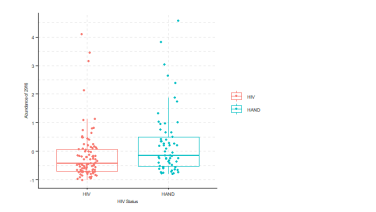entified features of Figure 2b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 2b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued

clusterid	P-value	Box Plot
2831	0.00354184891389575	
1380	0.0316995088645984	
2916	0.00375388391244947	

Table 2. Identified PLS-DA and Random forest features of Figure 2 a) The identified features of Figure 2a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 2a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 2b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 2b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued

clusterid	P-value	Box Plot
14662	0.0621952126742744	
3607	0.0210000481041515	
4146	0.123150837823242	

Table 2. Identified PLS-DA and Random forest features of Figure 2 a) The identified features of Figure 2a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 2a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 2b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 2b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued

c) Random forest

clusterid	Metabolite	Networked
1441	GLYCOCHENODEOXYCHOLATE	1925 Network
2292	L-Kynurenine	KYNURENINE
1925	Taurocholic acid	GLYCOCHENODEOXYCHOLATE GLYCOCHOLATE glycoursodeoxycholic acid glycocholic acid taurohyodeoxycholic acid taurodeoxycholic acid
1380	15-OxoEDE	.omega.-3 Arachidonic acid ethyl ester 13-Keto-9Z,11E-octadecadienoic acid Stearidonic acid ethyl ester

Table 2. Identified PLS-DA and Random forest features of Figure 2 a) The identified features of Figure 2a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 2a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 2b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 2b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued

clusterid	Metabolite	Networked
10472	Taurohyodeoxycholic acid	GLYCOCHENODEOXYCHOLATE GLYCOCHOLATE glycoursodeoxycholic acid glycocholic acid taurohyodeoxycholic acid taurodeoxycholic acid
5196	GLYCOCHOLATE	GLYCOCHENODEOXYCHOLATE GLYCOCHOLATE glycoursodeoxycholic acid glycocholic acid taurohyodeoxycholic acid taurodeoxycholic acid
703	phenylacetylglutamine	Unknown

Table 2. Identified PLS-DA and Random forest features of Figure 2 a) The identified features of Figure 2a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 2a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 2b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 2b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued
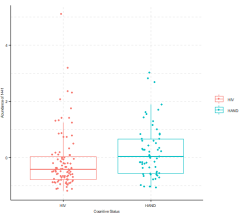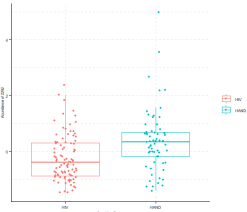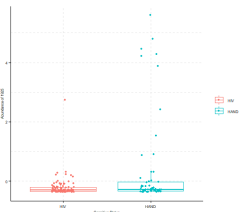
clusterid	Metabolite	Networked
2523	Cholic acid	Unknown
5387	Unknown	L-Arginine
3325	glycoursodeoxycholic acid	GLYCOCHENODEOXYCHOLATE GLYCOCHOLATE glycoursodeoxycholic acid glycocholic acid taurohyodeoxycholic acid taurodeoxycholic acid

Table 2. Identified PLS-DA and Random forest features of Figure 2 a) The identified features of Figure 2a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 2a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 2b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 2b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued

d)

clusterid	P-value	Box Plot
1441	0.00211009459695082	
2292	0.000229202389459476	
1925	0.0947995792612951	

Table 2. Identified PLS-DA and Random forest features of Figure 2 a) The identified features of Figure 2a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 2a in conjunction to a boxplot representing their relative abundance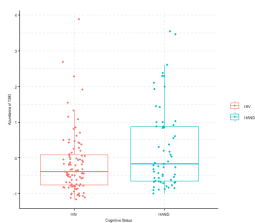s in addition to a p-value test between the stated groups. c) The id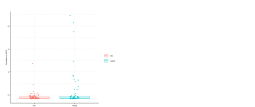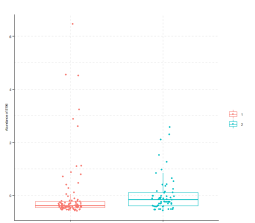entified features of Figure 2b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 2b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued
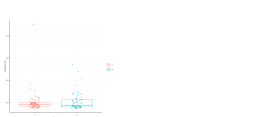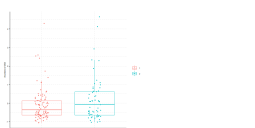
clusterid	P-value	Box Plot
1380	HAND-HIV 0.0316995088645984	
10472	HAND-HIV 0.184977150432481	
5196	HAND-HIV 0.0006755309	
703	HAND-HIV 0.8959781	
2523	HAND-HIV 0.1743778	

Table 2. Identified PLS-DA and Random forest features of Figure 2 a) The identified features of Figure 2a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 2a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 2b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 2b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued

| clusterid	P-value	Box Plot
5387	HAND-HIV 0.3100676	
3325	HAND-HIV 0.002499242	

The legend of Table 2 is described below.

      Cluster ID signifies the ID # of the metabolite denoted by the GNPS job stated in Figure 2. All Cluster IDs displayed in the table were derived from the PLS-DA and Random Forest plots denoted in Figure 2. While other Cluster IDs were found in the previously stated plots, some were excluded if the Cluster ID failed to generate both a direct and indirect annotation. The term Metabolite in relation to the table describes the direct annotation that the Cluster ID is associated with in the GNPS job. Additionally, the term "Networked" refers to the annotated metabolites found in the molecular network of the stated metabolite. Furthermore, the term P-Value on the table is the value derived from the Dunn's test performed on the stated groups of interest to determine whether or not separation could be considered significant (to which significance was determined as a value below .05). In a similar vein, the boxplots on the rightmost side of the table serve as visual displays of the separation between groups in relation to their abundance. It should be noted that grouping nomenclature remains the same as it was in Figure 2.

Figure 3. Plasma Alzheiers vs Normal a) Partial Least Square Discriminant Analysis (PLS-DA) of Normal and Alzheimer's groups. The top metabolites (features) driving separation between the different statuses are listed in their order of importance (VIP score) b) Random Forest results expressing the top 15 contributing metabolites with the Mean Decrease Accuracy (MDA) representing their level of accuracy.

a) PLS-DA
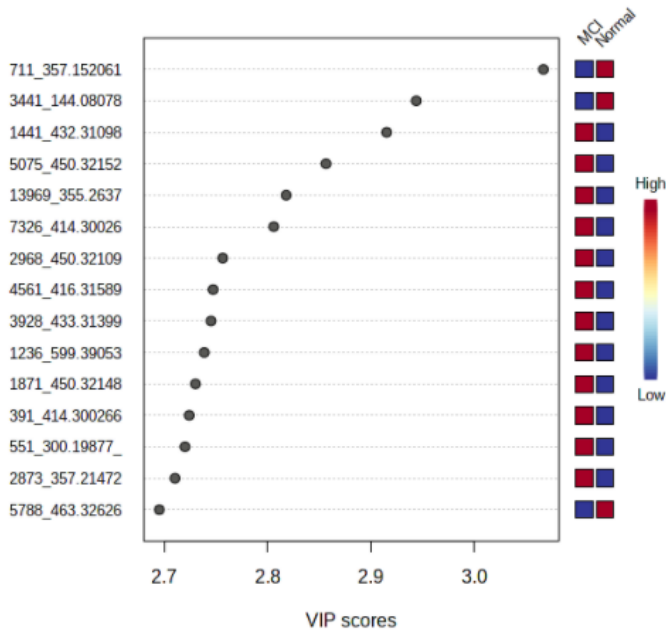
b) Random F
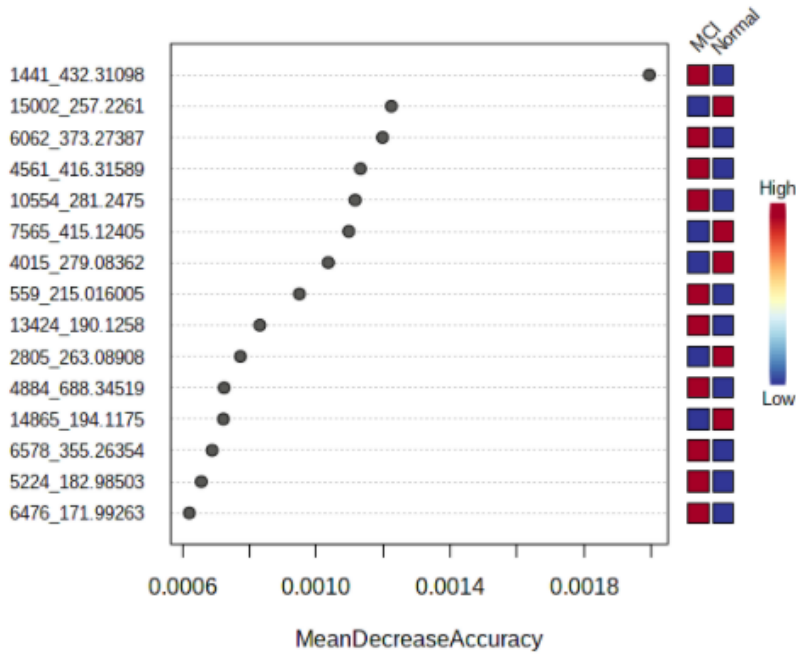
The legend for Figure 3 is described below:

Normal represents HIV negative patients with no cognitive impairment and MCI represents HIV negative patients with cognitive impairment.

The top features driving the separation in our patient groups can be found with the VIP and MDA (mean decrease accuracy scores) values in the figure above on the x axis. The higher VIP score signifies the increased importance of that metabolite for the separation in the patient groups for the PLS-DA plot and the lower the MDA score denotes the importance of that metabolite for the Random Forest model. The rightmost side of both PLS-DA and Random Forest figures denote the relative abundance of the metabolite in relation to the stated groups via the color gradient chart. The leftmost side along the y axis of both the PLS-DA and Random Forest figures denotes the Cluster ID (which can be discerned as the collection of numbers prior to the "_") found in the GNPS job which is attached below.
https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=4bd384e8a048435ca7444927f2de5e47

Table 3. Identified PLS-DA and Random forest features of Figure 3 a) The identified features of Figure 3a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 3a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 3b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 3b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups.

a) PLS-DA

clusterid	Metabolite	Networked
1441	GLYCOCHENODEOXYCHOLATE	GLYCOCHENODEOXYCHOLATE GLYCOCHOLATE glycoursodeoxycholic acid glycocholic acid taurohyodeoxycholic acid taurodeoxycholic acid
5075	Glycodeoxycholic acid	GLYCOCHENODEOXYCHOLATE GLYCOCHOLATE glycoursodeoxycholic acid glycocholic acid taurohyodeoxycholic acid taurodeoxycholic acid

Table 3. Identified PLS-DA and Random forest features of Figure 3 a) The identified features of Figure 3a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 3a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 3b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 3b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued

clusterid	Metabolite	Networked
13969	Cholic acid	Cholic acid
2968	glycoursodeoxycholic acid	GLYCOCHENODEOXYCHOLATE GLYCOCHOLATE glycoursodeoxycholic acid glycocholic acid taurohyodeoxycholic acid taurodeoxycholic acid
4561	Unknown	LITHOCHOLYLTAURINE
1871	Glycodeoxycholic acid	GLYCOCHENODEOXYCHOLATE GLYCOCHOLATE glycoursodeoxycholic acid glycocholic acid taurohyodeoxycholic acid taurodeoxycholic acid

Table 3. Identified PLS-DA and Random forest features of Figure 3 a) The identified features of Figure 3a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 3a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 3b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 3b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued

clusterid	Metabolite	Networked
391	glycoursodeoxycholic acid	GLYCOCHENODEOXYCHOLATE GLYCOCHOLATE glycoursodeoxycholic acid glycocholic acid taurohyodeoxycholic acid taurodeoxycholic acid
2894	Unknown	Coniferyl aldehyde

Table 3. Identified PLS-DA and Random forest features of Figure 3 a) The identified features of Figure 3a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 3a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 3b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 3b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued

clusterid	Metabolite	Networked
2086	Unknown	Bilirubin
8845	Unknown	Bilirubin

b)
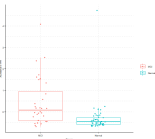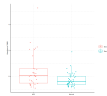
clusterid	P-value	Box Plot
1441	Normal-MCI 4.837549e-05	
5075	Normal-MCI 0.0003700517	
13969	Normal-MCI 0.01031663	
2968	Normal-MCI 0.0006300713	

Table 3. Identified PLS-DA and Random forest features of Figure 3 a) The identified features of Figure 3a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 3a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 3b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 3b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued
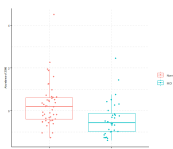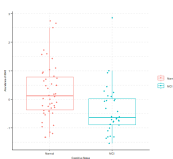
clusterid	P-value	Box Plot
4561	Normal-MCI 0.003166888	
1871	Normal-MCI 0.0005845865	
391	Normal-MCI 0.002552646	
2894	Normal-MCI 0.00366872125890628	
2086	Normal-MCI 0.000744522599027003	
8845	Normal-MCI 0.00223055633927521	

48

Table 3. Identified PLS-DA and Random forest features of Figure 3 a) The identified features of Figure 3a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 3a in conjunction to a boxplot representing their relative abundance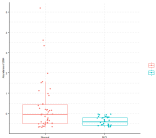s in addition to a p-value test between the stated groups. c) The identified features of Figure 3b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 3b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued

c) Random F

clusterid	Metabolite	Networked
1441	GLYCOCHENODEOXYCHOLATE	GLYCOCHENODEOXYCHOLATE GLYCOCHOLATE glycoursodeoxycholic acid glycocholic acid taurohyodeoxycholic acid taurodeoxycholic acid
15002	Unknown	12(13)-Epoxy-9Z-octadecenoic acid
6062	Cholic acid	Cholic acid
4561	Unknown	LITHOCHOLYLTAURINE
10554	Linoleic acid	Ricinoleic acid methyl ester Conjugated linoleic Acid (10E,12Z)

Table 3. Identified PLS-DA and Random forest features of Figure 3 a) The identified features of Figure 3a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 3a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 3b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 3b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued

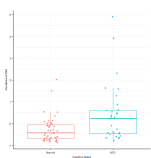| clusterid	Metabolite	Networked
14865	Unknown	Carbamazepine|benzo[b][1]benzazepine-11-carboxamide Carbmazepine
6578	Cholic acid	Cholic Acid

d)
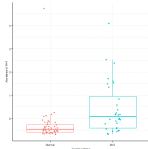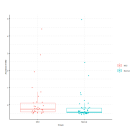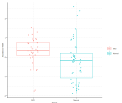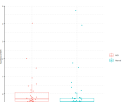
clusterid	P-value	Box Plot
1441	Normal-MCI 4.83754852231829e-05	
15002	Normal-MCI 0.08979971	
6062	Normal-MCI 0.01540023	

Table 3. Identified PLS-DA and Random forest features of Figure 3 a) The identified features of Figure 3a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 3a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 3b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 3b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued

clusterid	P-value	Box Plot
4561	Normal-MCI 0.00316688823260113	
10554	Normal-MCI 0.004594092	
14865	Normal-MCI 0.1987757	
6578	Normal-MCI 0.00445013	

The legend of Table 3 is described below.

Cluster ID signifies the ID # of the metabolite denoted by the GNPS job stated in Figure 3. All Cluster IDs displayed in the table were derived from the PLS-DA and Random Forest plots denoted in Figure 3. While other Cluster IDs were found in the previously stated plots, some were excluded if the Cluster ID failed to generate both a direct and indirect annotation. The term Metabolite in relation to the table describes the direct annotation that the Cluster ID is associated with in the GNPS job. Additionally, the term "Networked" refers to the annotated metabolites found in the molecular network of the stated metabolite. Furthermore, the term P-Value on the table is the value derived from the Dunn's test performed on the stated groups of interest to determine whether or not separation could be considered significant (to which significance was determined as a value below .05). In a similar vein, the boxplots on the rightmost side of the table serve as visual displays of the separation between groups in relation to their abundance. It should be noted that grouping nomenclature remains the same as it was in Figure 3.

Figure 4. CSF Separation in all 4 groups a) Partial Least Square Discriminant Analysis (PLS-DA) of the Normal, Alzheimer's, HIV, and HAND groups. The top metabolites (features) driving separation between the different statuses are listed in their order of importance (VIP score) b) Random Forest results expressing the top 15 contributing metabolites with the Mean Decrease Accuracy (MDA) representing their level of accuracy.

a) PLS-DA
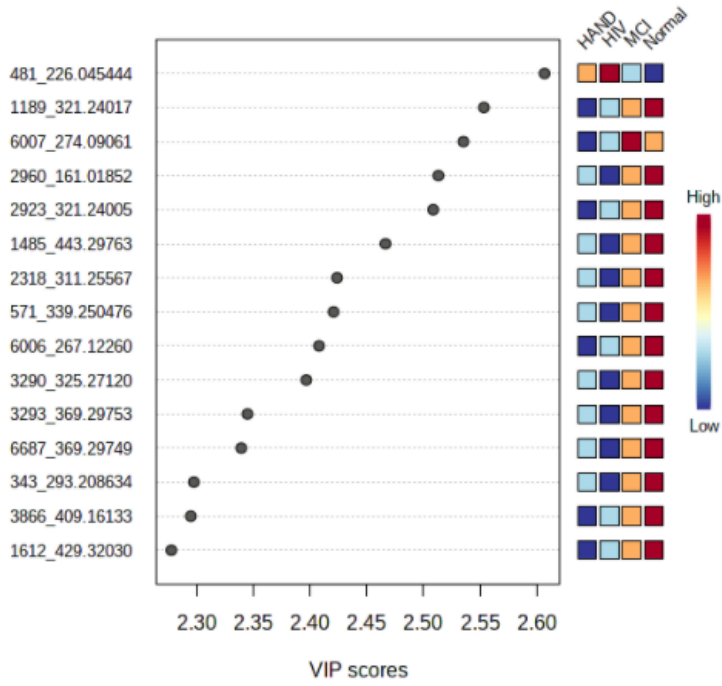
b) Random F
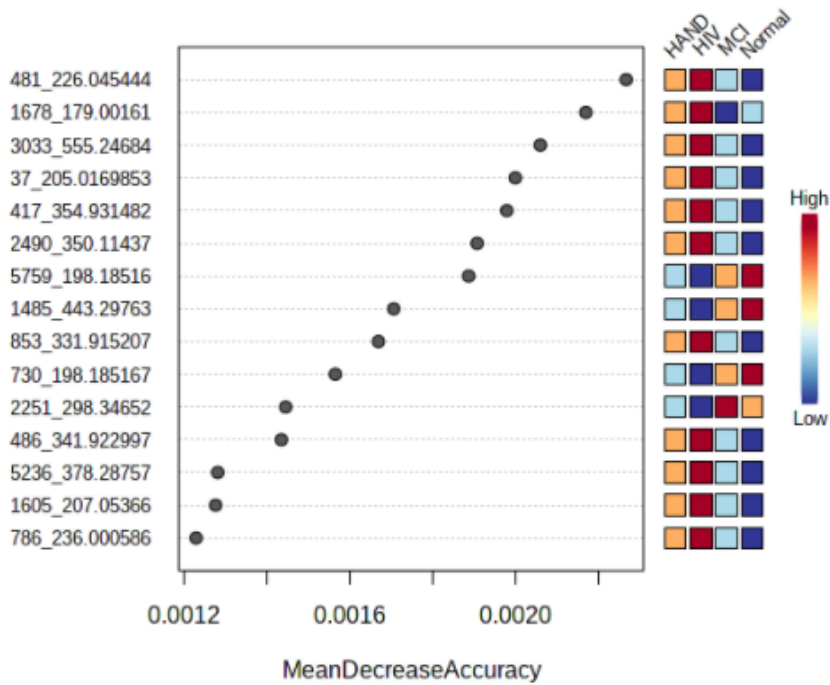
The legend for Figure 4 is described below:

       HIV represents the HIV positive patients with no cognitive impairment, HAND represents HIV positive patients with cognitive impairment, Normal represents HIV negative patients with no cognitive impairment, and MCI represents HIV negative patients with cognitive impairment.

       The top features driving the separation in our patient groups can be found with the VIP and MDA (mean decrease accuracy scores) values in the figure above. The higher VIP score signifies the increased importance of that metabolite for the separation in the patient groups for the PLS-DA plot and the lower the MDA score denotes the importance of that metabolite for the Random Forest model. The rightmost side of both PLS-DA and Random Forest figures denote the relative abundance of the metabolite in relation to the stated groups via the color gradient chart. The leftmost side of both the PLS-DA and Random Forest figures denotes the Cluster ID (which can be discerned as the collection of numbers prior to the "_") found in the GNPS job which is attached below.
https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=76a37187625a4a1c9841c6c269c915f6

Table 4. Identified PLS-DA and Random forest features of Figure 4 a) The identified features of Figure 4a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 4a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 4b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 4b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups.

a) PLS-DA

clusterid	Metabolite	Networked
1189	Unknown	Monolinolenin (9c,12c,15c) cis-8,11,14-Eicosatrienoic acid 9Z,11E,13E-Octadecatrienoic acid methyl ester
2923	Unknown	Monolinolenin (9c,12c,15c) cis-8,11,14-Eicosatrienoic acid 9Z,11E,13E-Octadecatrienoic acid methyl ester
3290	Unknown	Monolinolenin (9c,12c,15c) cis-8,11,14-Eicosatrienoic acid 9Z,11E,13E-Octadecatrienoic acid methyl ester

Table 4. Identified PLS-DA and Random forest features of Figure 4 a) The identified features of Figure 4a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 4a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 4b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 4b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued
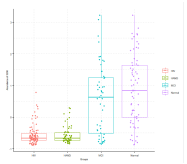
b)

clusterid	P - Value	Box Plot
1189	HAND-HIV 9.767603e-01 MCI-HIV 1.729833e-13 Normal-HIV 1.217057e-17 MCI-HAND 1.191694e-11 Normal-HAND 8.727770e-15 Normal-MCI 8.025919e-01	
2923	HAND-HIV 5.305822e-01 MCI-HIV 4.459662e-13 Normal-HIV 3.792701e-16 MCI-HAND 3.328186e-13 Normal-HAND 3.792701e-16 Normal-MCI 8.096011e-01	

Table 4. Identified PLS-DA and Random forest features of Figure 4 a) The identified features of Figure 4a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 4a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 4b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 4b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued

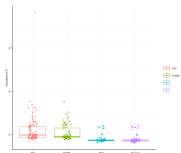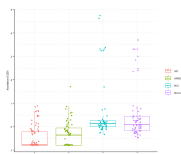| clusterid	Metabolite	Networked
3290	HAND-HIV 9.115886e-01 MCI-HIV 7.549058e-09 Normal-HIV 1.248290e-16 MCI-HAND 1.489837e-07 Normal-HAND 1.060073e-13 Normal-MCI 1.124122e-01	

c) Random Forest

clusterid	Metabolite	Networked
37	Unknown	Dioctyl phthalate
2251	Unknown	Monolinolenin (9c,12c,15c) cis-8,11,14-Eicosatrienoic acid 9Z,11E,13E-Octadecatrienoic acid methyl ester

58

Table 4. Identified PLS-DA and Random forest features of Figure 4 a) The identified features of Figure 4a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 4a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 4b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 4b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued

d)

| clusterid	P-value	Box Plot
37	HAND-HIV 7.715943e-01 MCI-HIV 3.202714e-12 Normal-HIV 1.223999e-15 MCI-HAND 9.473453e-10 Normal-HAND 4.021084e-12 Normal-MCI 7.715943e-01	
2251	HAND-HIV 1.754926e-01 MCI-HIV 9.857829e-14 Normal-HIV 3.459246e-14 MCI-HAND 1.715312e-08 Normal-HAND 1.715312e-08 Normal-MCI 6.188313e-01	

The legend of Table 4 is described below.

        Cluster ID signifies the ID # of the metabolite denoted by the GNPS job stated in Figure 4. All Cluster IDs displayed in the table were derived from the PLS-DA and Random Forest plots denoted in Figure 4. While other Cluster IDs were found in the previously stated plots, some were excluded if the Cluster ID failed to generate both a direct and indirect annotation. The term Metabolite in relation to the table describes the direct annotation that the Cluster ID is associated with in the GNPS job. Additionally, the term "Networked" refers to the annotated metabolites found in the molecular network of the stated metabolite. Furthermore, the term P-Value on the table is the value derived from the Dunn's test performed on the stated groups of interest to determine whether or not separation could be considered significant (to which significance was determined as a value below .05). In a similar vein, the boxplots on the rightmost side of the table serve as visual displays of the separation between groups in relation to their abundance. It should be noted that grouping nomenclature remains the same as it was in Figure 4.

Figure 5. CSF HIV vs HAND a) Partial Least Square Discriminant Analysis (PLS-DA) of the HIV, and HAND groups. The top metabolites (features) driving separation between the different statuses are listed in their order of importance (VIP score) b) Random Forest results expressing the top 15 contributing metabolites with the Mean Decrease Accuracy (MDA) representing their level of accuracy.

a) PLS-DA
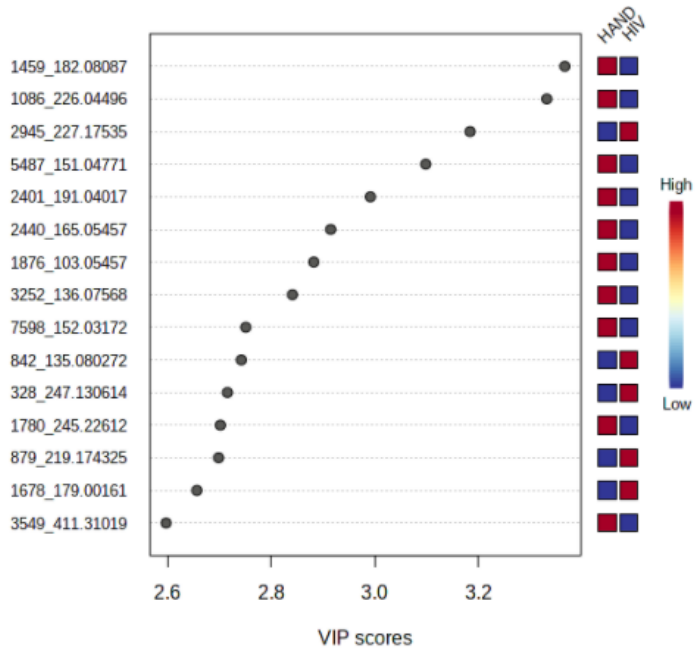
b) Random F
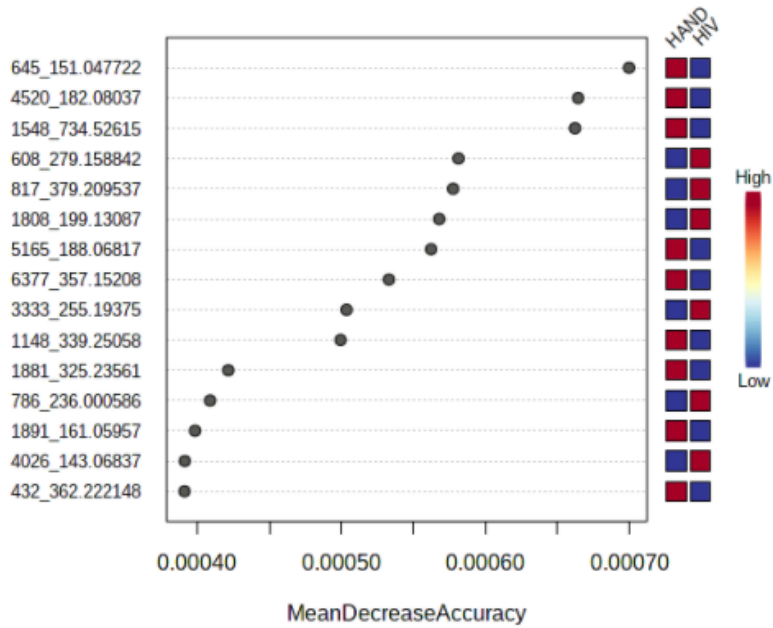
The legend for Figure 5 is described below:

 HIV represents the HIV positive patients with no cognitive impairment and HAND represents HIV positive patients with cognitive impairment.

 The top features driving the separation in our patient groups can be found with the VIP and MDA (mean decrease accuracy scores) values in the figure above. The higher VIP score signifies the increased importance of that metabolite for the separation in the patient groups for the PLS-DA plot and the lower the MDA score denotes the importance of that metabolite for the Random Forest model. The rightmost side of both PLS-DA and Random Forest figures denote the relative abundance of the metabolite in relation to the stated groups via the color gradient chart. The leftmost side of both the PLS-DA and Random Forest figures denotes the Cluster ID (which can be discerned as the collection of numbers prior to the "_") found in the GNPS job which is attached below.

https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=76a37187625a4a1c9841c6c269c915f6

Table 5. Identified PLS-DA and Random forest features of Figure 5 a) The identified features of Figure 5a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 5a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 5b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 5b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups.

a) PLS-DA

clusterid	Metabolite	Networked
1459	L-Tyrosine	3-METHOXYTYROSINE L-Tyrosine
2440	L-Tyrosine	
842	Unknown	Hydrocinnamic acid Phthalic anhydride
1780	unknown	cis-7,10,13,16-Docosatetraenoic acid (-)-Perillyl alcohol 3,5-Dimethyladamantan-1-amine 2-Butanone, 4-(2,6,6-trimethyl-2-cyclohexen-1-yl)- Amantadine 5.alpha.-Androsterone alpha.-Pinene oxide

Table 5. Identified PLS-DA and Random forest features of Figure 5 a) The identified features of Figure 5a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 5a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 5b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 5b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued

clusterid	Metabolite	Networked
879	Pinolenic acid ethyl ester	Self Loop
7735	Monolinolein (9c,12c,15c)	cis-8,11,14-Eicosatrienoic acid 9Z,11E,13E-Octadecatrienoic acid methyl ester
4520	L-Tyrosine	3-METHOXYTYROSINE
4345	9Z,11E,13E-Octadecatrienoic acid methyl ester	Monolinolenin (9c,12c,15c) cis-8,11,14-Eicosatrienoic acid

Table 5. Identified PLS-DA and Random forest features of Figure 5 a) The identified features of Figure 5a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 5a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 5b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 5b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued

b)

clusterid	P - Value	Box Plot
1459	HAND-HIV 6.664516e-05	
2440	HAND-HIV 0.009104421	
842	HAND-HIV 0.006226779	
1780	HAND-HIV 0.1088707	
879	HAND-HIV 0.003517763	
7735	HAND-HIV 0.00471221735635266	
4520	HAND-HIV 0.00562586023202411	
4345	HAND-HIV 0.0511019021105961	

Table 5. Identified PLS-DA and Random forest features of Figure 5 a) The identified features of Figure 5a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 5a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 5b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 5b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued

c) Random F

clusterid	Metabolite	Networked
4520	L-Tyrosine	3-METHOXYTYROSINE L-Tyrosine
6377	Unknown	Monolinolenin (9c,12c,15c) cis-8,11,14-Eicosatrienoic acid 9Z,11E,13E-Octadecatrienoic acid methyl ester

d)

clusterid	P - Value	Box Plot
4520	HAND-HIV 0.00562586	
6377	HAND-HIV 0.008212262	

The legend of Table 5 is described below.

   Cluster ID signifies the ID # of the metabolite denoted by the GNPS job stated in Figure 5. All Cluster IDs displayed in the table were derived from the PLS-DA and Random Forest plots denoted in Figure 5. While other Cluster IDs were found in the previously stated plots, some were excluded if the Cluster ID failed to generate both a direct and indirect annotation. The term Metabolite in relation to the table describes the direct annotation that the Cluster ID is associated with in the GNPS job. Additionally, the term "Networked" refers to the annotated metabolites found in the molecular network of the stated metabolite. Furthermore, the term P-Value on the table is the value derived from the Dunn's test performed on the stated groups of interest to determine whether or not separation could be considered significant (to which significance was determined as a value below .05). In a similar vein, the boxplots on the rightmost side of the table serve as visual displays of the separation between groups in relation to their abundance. It should be noted that grouping nomenclature remains the same as it was in Figure 5.

Figure 6. CSF Alzheiers vs Normal a) Partial Least Square Discriminant Analysis (PLS-DA) of the Normal and Alzheimer's groups. The top metabolites (features) driving separation between the different statuses are listed in their order of importance (VIP score) b) Random Forest results expressing the top 15 contributing metabolites with the Mean Decrease Accuracy (MDA) representing their level of accuracy.

a) PLS-DA
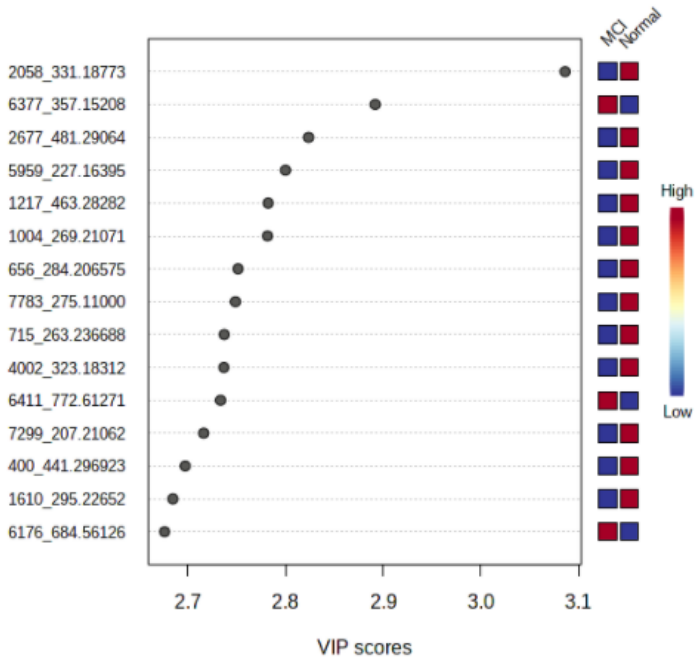
b) Random F
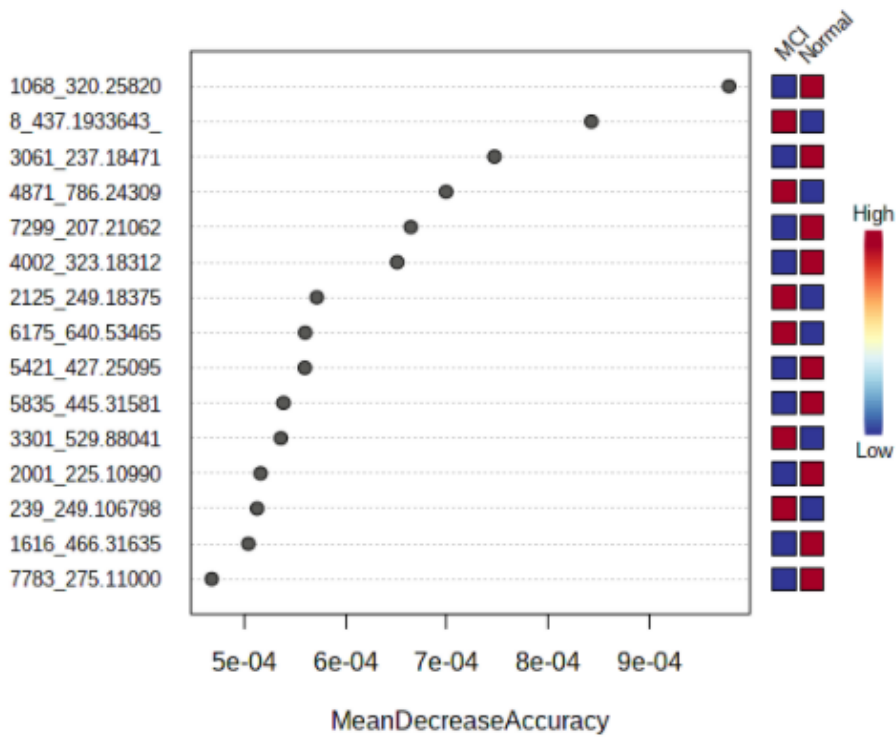
The legend for Figure 6 is described below:

Normal represents HIV negative patients with no cognitive impairment and MCI represents HIV negative patients with cognitive impairment.

The top features driving the separation in our patient groups can be found with the VIP and MDA (mean decrease accuracy scores) values in the figure above. The higher VIP score signifies the increased importance of that metabolite for the separation in the patient groups for the PLS-DA plot and the lower the MDA score denotes the importance of that metabolite for the Random Forest model. The rightmost side of both PLS-DA and Random Forest figures denote the relative abundance of the metabolite in relation to the stated groups via the color gradient chart. The leftmost side of both the PLS-DA and Random Forest figures denotes the Cluster ID (which can be discerned as the collection of numbers prior to the "_") found in the GNPS job which is attached below.

https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=76a37187625a4a1c9841c6c269c915f6

Table 6. Identified PLS-DA and Random forest features of Figure 6 a) The identified features of Figure 6a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 6a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 6b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 6b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups.

a) PLS-DA

clusterid	Metabolite	Networked
6377	Unknown	Monolinolenin (9c,12c,15c) cis-8,11,14-Eicosatrienoic acid 9Z,11E,13E-Octadecatrienoic acid methyl ester
715	Conjugated linoleic acid (9E,11E)	Conjugated linoleic acid (9E,11E) cis-11,14-Eicosadienoic acid cis-5,8,11-Eicosatrienoic acid 9Z,11E,13E-Octadecatrienoic acid methyl ester 12(S)-Hydroxy-16-heptadecynoic acid cis,cis-9,12-Octadecadien-1-ol Ricinoleic acid methyl ester 2,4-dihydroxyheptadecyl acetate Neodihydroprotolichesterinic acid 9(10)-EpOME

Table 6. Identified PLS-DA and Random forest features of Figure 6 a) The identified features of Figure 6a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 6a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 6b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 6b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued

clusterid	Metabolite	Networked
4002	Unknown	Monolinolenin (9c,12c,15c) cis-8,11,14-Eicosatrienoic acid 9Z,11E,13E-Octadecatrienoic acid methyl ester

Table 6. Identified PLS-DA and Random forest features of Figure 6 a) The identified features of Figure 6a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 6a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 6b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 6b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued

clusterid	Metabolite	Networked
5959	1,11-Undecanedicarboxylic acid	Dodecanedioic acid
		3-Hydroxydodecanoic acid
		Mononervonin (15c)
		Monoelaidin
		Myristoleic acid
		Palmitelaidic acid
		Cis-9-Hexadecenoic acid
		Elaidic acid
		Phytomonic acid
		3-Hydroxyoctadecanoic Acid
		Monopalmitolein (9c)
		8-hydroxy-8-(3-octyloxiran-2 -yl)octanoic acid [IIN-based on

Table 6. Identified PLS-DA and Random forest features of Figure 6 a) The identified features of Figure 6a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 6a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 6b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 6b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued

clusterid	Metabolite	Networked
6183	Monolinolenin (9c,12c,15c)	cis-8,11,14-Eicosatrienoic acid 9Z,11E,13E-Octadecatrienoic acid methyl ester
260	Unknown	Octadecanamide Palmitamide

Table 6. Identified PLS-DA and Random forest features of Figure 6 a) The identified features of Figure 6a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 6a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 6b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 6b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued

b)

clusterid	Trend	Box Plot
6377	Normal-MCI 0.03726561	
715	Normal-MCI 0.01252045	
4002	Normal-MCI 0.001766091	
5959	Normal-MCI 0.007619782	
6183	Normal-MCI 0.0060375773650213	

Table 6. Identified PLS-DA and Random forest features of Figure 6 a) The identified features of Figure 6a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 6a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 6b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 6b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued
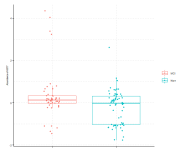
clusterid	Trend	Box Plot
260	Normal-MCI 0.0588320136714975	
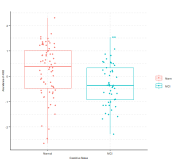
c) Random Forest

clusterid	Metabolite	Networked
2125	Unknown	Monolinolenin (9c,12c,15c) cis-8,11,14-Eicosatrienoic acid 9Z,11E,13E-Octadecatrienoic acid methyl ester
4002	Unknown	Monolinolenin (9c,12c,15c) cis-8,11,14-Eicosatrienoic acid 9Z,11E,13E-Octadecatrienoic acid methyl ester

Table 6. Identified PLS-DA and Random forest features of Figure 6 a) The identified features of Figure 6a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 6a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The ident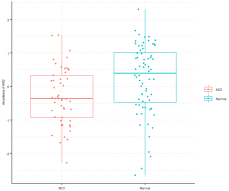ified features of Figure 6b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 6b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued

clusterid	Metabolite	Networked
3061	Unknown	Conjugated linoleic acid (9E,11E)
		cis-11,14-Eicosadienoic acid
		cis-5,8,11-Eicosatrienoic acid
		9Z,11E,13E-Octadecatrienoic acid methyl ester
		12(S)-Hydroxy-16-heptadecynoic acid
		cis,cis-9,12-Octadecadien-1-ol
		Ricinoleic acid methyl ester
		2,4-dihydroxyheptadecyl acetate
		Neodihydroprotolichesterinic acid
		9(10)-EpOME

Table 6. Identified PLS-DA and Random forest features of Figure 6 a) The identified features of Figure 6a are displayed in relation to their annotated name and molecular network, and P-Value. b) Represents the identified metabolites of Figure 6a in conjunction to a boxplot representing their relative abundances in addition to a p-value test between the stated groups. c) The identified features of Figure 6b are displayed in relation to their annotated name and molecular network, and P-Value. d) Represents the identified metabolites of Figure 6b in conjunction to a boxplot representing their relative abundances in addition to a P-value test between the stated groups. Continued

d)

clusterid	P - Value	Box Plot
2125	Normal-MCI 0.9515414	
4002	Normal-MCI 0.00176609112160751	
3061	Normal-MCI 0.1524652	

The legend of Table 6 is described below.

Cluster ID signifies the ID # of the metabolite denoted by the GNPS job stated in Figure 6. All Cluster IDs displayed in the table were derived from the PLS-DA and Random Forest plots denoted in Figure 6. While other Cluster IDs were found in the previously stated plots, some were excluded if the Cluster ID failed to generate both a direct and indirect annotation. The term Metabolite in relation to the table describes the direct annotation that the Cluster ID is associated with in the GNPS job. Additionally, the term "Networked" refers to the annotated metabolites found in the molecular network of the stated metabolite. Furthermore, the term P-Value on the table is the value derived from the Dunn's test performed on the stated groups of interest to determine whether or not separation could be considered significant (to which significance was determined as a value below .05). In a similar vein, the boxplots on the rightmost side of the table serve as visual displays of the separation between groups in relation to their abundance. It should be noted that grouping nomenclature remains the same as it was in Figure 6.

Discussion

To conduct the supervised analysis, the previous classifications denoted by 90 were utilized to define the distinctive traits between our groups. Using the Metaboanalyst platform[15], we were able to conduct Partial Least Squares Discriminant Analysis (PLS-DA), a linear regression model that determines the level of importance of various metabolites in driving separation between the experimental groupings. Additionally, the Metaboanalyst platform allows us to utilize Random Forest, a learning method that implements a decision tree to assess the strength of metabolites in their role for predicting groupings. Random Forest serves to corroborate the validity of the metabolites of interest identified by PLS-DA. The features found were ranked by their mean decrease accuracy scores

The PLS-DA and Random Forest Analysis were carried out initially by comparing all 4 categories of interest at the same time as seen in Figure 1 in respect to the sample sources of plasma and CSF. Afterwards, the groups were separated into smaller groups and analyzed again as seen in Figure 2 where we compare only the HIV and HAND groups. Separation of the groups allowed for identification of metabolites that were unique to the pairs of HIV to HAND progression and that of normal to Alzheimer's progression. Separation in the PLS-DA results were validated using the Q2 value in conjunction with a Dunns test. The Q2 value denotes the predictive value of the PLS-DA models while the Dunns test assesses whether or not the separation between the groups were found to be significant via its p-value score. The random forest analysis was assessed in a similar fashion via the OOB score and Dunns test. The OOB score means Out-of-Bag score and serves as the generalized error rate of the plot.

After generating a list of metabolites from both plots, we first look to find annotations by referencing the metabolite's cluster id with the GNPS library. Once an annotation was found, the metabolite's molecular network was observed using cytoscape and recorded as well.

In respect to the plasma samples, our data has found multiple metabolites of interest that have been previously discussed in previous publications relating to HAND. Kynurenine, a derivative of tryptophan, is one of these metabolites in questions. The kynurenine pathway is associated with a myriad of biological functions, most notably for the purpose of this study, the dilation of blood vessels in the brain during inflammation [19]. Increased activation of the kynurenine pathway has been found in previous studies as an indicator of HAND due to a derivative of kyunerine known as Quinolinic acid contributing to the neurotoxicity of the brain[19]. In our study, kynurenine was found to be an important metabolite for our PLS-DA plots when focusing on the separation of the HIV and HAND patients. We consider the metabolite to be important though kyunerine having one of the top 15 highest VIP scores derived from the PLS-DA plot in conjunction with significant separation of the groups being below .05 for their p-value. Furthermore, the PLS-DA data also indicated a higher abundance of kyunerine in patients with HAND in relation to patients with only HIV.

Another metabolite that was found in our data, while not strictly related to HAND pathogenesis was that of arachidonic acid. Arachidonic acid is a fatty acid that functions in cellular signaling and the up-regulation of Arachidonic acid is associated with the increase of amyloid beta plaques which are associated with pathogenesis of alzheimers [20]. Furthermore, the arachidonic acid pathway is another contributor to neural inflammation as well [21] which may contribute to the pathogenesis of HAND. In our data arachidonic acid has been found in higher quantities in our HAND group in comparison to HIV.

One major note of relevance to our study was that of the significance of bile acids in the plasma data. Several bile acids were found to be significant in the separation of our groups for both HIV and Alzheimer's. The primary molecular network identified includes: Taurocholic

acid, Glycochenodeoxycholate, glycocholate, glycoursodeoxycholic acid, glycocholic acid, and taurohyodeoxycholic acid. Additionally, while not in this network itself, other identified metabolites found in our study with associations to this network include Deoxycholate, Cholic acid, and cholesterol. Overall, while HAND and Alzheimer's groups shared some similarities in their primary and secondary bile acid acid profiles, there were still differences between the groups that were unique to each profile. Thus we can reasonably come to the conclusion that the microbiome plays a role of some kind to the cognitive impairments of both HAND and Alzheimer's in different ways. What makes this significant is that while bile acids have been previously found to play a role in studies relating to Alzheimer's disease [22, 23], there is a deficit in publications in the profiles of bile acids in HIV.

In regards to our CSF data, our data did not indicate any identical overlap within the profiles of our metabolites of interest in relation to the plasma profiles. Nevertheless, the data has shown that fatty acids and fatty amides played a significant role in discerning our group categorizations. Fatty acids such as cis-8,11,14-Eicosatrienoic acid, Monolinolein, Elaidic acid, and Linoleic acid were found to be decreased in abundance among our Alzheimer's group in relation to their control while cis-8,11,14-Eicosatrienoic acid and Monolinolein were upregulated in our HAND groups in relation to HIV without cognitive impairement. Additionally, the fatty amides of Palmitamide and Octadecanamide were upregulated in the Alzheimer's group in relation to the control. What I reasonably hypothesize with this information is that the alteration of the bile acid profiles in addition to that of the microbiome community may have a downstream effect on absorption of certain fatty acids and amides. This is due to the bile acids having a significant role in the emulsification process of fat globules which is essential for the digestion of

fatty acids and amides. Some studies in mice even indicate that specific bile acids play a role in the uptake of fatty acids [24].

Conclusion and Perspective

Overall the future prospects look promising with the discovery of bile acids having an apparent influence on the pathogenesis of HAND. The study provides future studies relating to HAND multiple avenues of exploration in regards to the microbiome and bile acids as they connect to neurocognitive disorders. As the sample size of our study was limited, it would be interesting to explore whether the metabolite profiles remain the same given a larger sample size in addition to assisting with the accuracy of the PLS-DA and Random forest scores. Another aspect of exploration would be that of therapeutics. With continued testing, bile acids may serve the role of being an indicator for the progression of cognitive impairment in patients as some studies indicate are possible with that of Alzheimer's Disease. Additionally, clinical trials focused on restoring the bile acid profiles to that of pre-cognitive impairment levels may help with reducing symptoms of the various neurocognitive disorders. In a similar fashion, fatty amides and fatty acids could be studied as well for their influence on neurocognitive disorders.

Likewise, the role of tyrosine may be another route of exploration as well. As we found altered levels of tyrosine and methoxy tyrosine in the csf profiles of alzheimer's patients, there may be alterations in the neurotransmitters that are produced in their metabolite pathways that could affect certain regions of the brain that have receptors for them. This would be interesting as alterations in other neurotransmitters such as acetylcholine and glutamate have been previously documented in Alzheimer's studies, the inclusion of tyrosine derived neurotransmitters would be another step forward towards understanding neurocognitive disorders.

References

87

1. Deeks SG, Lewin SR, Havlir DV. The end of AIDS: HIV infection as a chronic disease. Lancet. 2013;382(9903):1525-1533. doi:10.1016/S0140-6736(13)61809-7.

2. Rumbaugh, Jeffrey A, and William Tyor. "HIV-associated neurocognitive disorders: Five new things." Neurology. Clinical practice vol. 5,3 (2015): 224-231. doi:10.1212/CPJ.0000000000000117

3. Gendelman, Howard E. "Predictive biomarkers for cognitive decline during progressive HIV infection." EBioMedicine vol. 51 (2020): 102538. doi:10.1016/j.ebiom.2019.10.064

4. Chan DC, Kim PS. HIV entry and its inhibition. Cell. 1998 May 29;93(5):681-4. doi: 10.1016/s0092-8674(00)81430-0. PMID: 9630213.

5. Cunningham AL, Donaghy H, Harman AN, Kim M, Turville SG (2010). "Manipulation of dendritic cell function by viruses". Current Opinion in Microbiology. 13 (4): 524–529. doi:10.1016/j.mib.2010.06.002. PMID 20598938.

6. Lloyd, A. "HIV infection and AIDS." Papua and New Guinea medical journal vol. 39,3 (1996): 174-80.

7. Garg H, Mohl J, Joshi A (November 9, 2012). "HIV-1 induced bystander apoptosis". Viruses. 4 (11): 3020–43. doi:10.3390/v4113020. PMC 3509682. PMID 23202514.

8. Clifford DB, Ances BM. HIV-associated neurocognitive disorder. Lancet Infect Dis. 2013;13(11):976-986. doi:10.1016/S1473-3099(13)70269-X

9. Thomas S, Mayer L, Sperber K (2009). "Mitochondria influence Fas expression in gp120-induced apoptosis of neuronal cells". The International Journal of Neuroscience. 119 (2): 157–65. doi:10.1080/00207450802335537. PMID 19125371. S2CID 25456692.

10. Saylor, D., Dickens, A. M., Sacktor, N., Haughey, N., Slusher, B., Pletnikov, M., Mankowski, J. L., Brown, A., Volsky, D. J., & McArthur, J. C. (2016). HIV-associated neurocognitive disorder--pathogenesis and prospects for treatment. Nature reviews. Neurology, 12(4), 234–248. https://doi.org/10.1038/nrneurol.2016.27

11. Cohen, R. A., Harezlak, J., Schifitto, G., Hana, G., Clark, U., Gongvatana, A., Paul, R., Taylor, M., Thompson, P., Alger, J., Brown, M., Zhong, J., Campbell, T., Singer, E., Daar, E., McMahon, D., Tso, Y., Yiannoutsos, C. T., & Navia, B. (2010). Effects of nadir CD4 count and duration of human immunodeficiency virus infection on brain volumes in the

highly active antiretroviral therapy era. Journal of neurovirology, 16(1), 25–32. https://doi.org/10.3109/13550280903552420

12. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV, Meehan MJ, Liu WT, Crüsemann M, Boudreau PD, Esquenazi E, Sandoval-Calderón M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu CC, Floros DJ, Gavilan RG, Kleigrewe K, Northen T, Dutton RJ, Parrot D, Carlson EE, Aigle B, Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J, Murphy BT, Gerwick L, Liaw CC, Yang YL, Humpf HU, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, Klitgaard A, Larson CB, P CAB, Torres-Mendoza D, Gonzalez DJ, Silva DB, Marques LM, Demarque DP, Pociute E, O'Neill EC, Briand E, Helfrich EJN, Granatosky EA, Glukhov E, Ryffel F, Houson H, Mohimani H, Kharbush JJ, Zeng Y, Vorholt JA, Kurita KL, Charusanti P, McPhail KL, Nielsen KF, Vuong L, Elfeki M, Traxler MF, Engene N, Koyama N, Vining OB, Baric R, Silva RR, Mascuch SJ, Tomasi S, Jenkins S, Macherla V, Hoffman T, Agarwal V, Williams PG, Dai J, Neupane R, Gurr J, Rodríguez AMC, Lamsa A, Zhang C, Dorrestein K, Duggan BM, Almaliti J, Allard PM, Phapale P, Nothias LF, Alexandrov T, Litaudon M, Wolfender JL, Kyle JE, Metz TO, Peryea T, Nguyen DT, VanLeer D, Shinn P, Jadhav A, Müller R, Waters KM, Shi W, Liu X, Zhang L, Knight R, Jensen PR, Palsson BO, Pogliano K, Linington RG, Gutiérrez M, Lopes NP, Gerwick WH, Moore BS, Dorrestein PC, Bandeira N. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat Biotechnol. 2016 Aug 9;34(8):828-837. doi: 10.1038/nbt.3597. PMID: 27504778; PMCID: PMC5321674.

13. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003 Nov;13(11):2498-504. doi: 10.1101/gr.1239303. PMID: 14597658; PMCID: PMC403769.

14. Nothias LF, Petras D, Schmid R, Dührkop K, Rainer J, Sarvepalli A, Protsyuk I, Ernst M, Tsugawa H, Fleischauer M, Aicheler F, Aksenov AA, Alka O, Allard PM, Barsch A, Cachet X, Caraballo-Rodriguez AM, Da Silva RR, Dang T, Garg N, Gauglitz JM, Gurevich A, Isaac G, Jarmusch AK, Kameník Z, Kang KB, Kessler N, Koester I, Korf A, Le Gouellec A, Ludwig M, Martin H C, McCall LI, McSayles J, Meyer SW, Mohimani H, Morsy M, Moyne O, Neumann S, Neuweger H, Nguyen NH, Nothias-Esposito M, Paolini J, Phelan VV, Pluskal T, Quinn RA, Rogers S, Shrestha B, Tripathi A, van der Hooft JJJ, Vargas F, Weldon KC, Witting M, Yang H, Zhang Z, Zubeil F, Kohlbacher O, Böcker S, Alexandrov T, Bandeira N, Wang M, Dorrestein PC. Feature-based molecular networking in the GNPS analysis environment. Nat Methods. 2020 Sep;17(9):905-908.

doi: 10.1038/s41592-020-0933-6. Epub 2020 Aug 24. PMID: 32839597; PMCID: PMC7885687.

15. Chong J, Wishart DS, Xia J. Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. Curr Protoc Bioinformatics. 2019 Dec;68(1):e86. doi: 10.1002/cpbi.86. PMID: 31756036.

16. Ruiz-Perez D, Guan H, Madhivanan P, Mathee K, Narasimhan G. So you think you can PLS-DA? BMC Bioinformatics. 2020 Dec 9;21(Suppl 1):2. doi: 10.1186/s12859-019-3310-7. PMID: 33297937; PMCID: PMC7724830.

17. Banerjee, Priyanka; Ghosh, Sanghamitra; Dutta, Mainak; Subramani, Elavarasan; Khalpada, Jaydeep; RoyChoudhury, Sourav; Chaudhury, Koel . (2015): Important features identified by PLS-DA and VIP scores.. PLOS ONE. Figure. https://doi.org/10.1371/journal.pone.0080940.g004

18. Sarica A, Cerasa A, Quattrone A. Random Forest Algorithm for the Classification of Neuroimaging Data in Alzheimer's Disease: A Systematic Review. Front Aging Neurosci. 2017 Oct 6;9:329. doi: 10.3389/fnagi.2017.00329. PMID: 29056906; PMCID: PMC5635046.

19. Kandanearatchi A, Brew BJ. The kynurenine pathway and quinolinic acid: pivotal roles in HIV associated neurocognitive disorders. FEBS J. 2012 Apr;279(8):1366-74. doi: 10.1111/j.1742-4658.2012.08500.x. Epub 2012 Mar 27. PMID: 22260426.

20. Giil LM, Midttun Ø, Refsum H, Ulvik A, Advani R, Smith AD, Ueland PM. Kynurenine Pathway Metabolites in Alzheimer's Disease. J Alzheimers Dis. 2017;60(2):495-504. doi: 10.3233/JAD-170485. PMID: 28869479.

21. Amtul Z, Uhrig M, Wang L, Rozmahel RF, Beyreuther K. Detrimental effects of arachidonic acid and its metabolites in cellular and mouse models of Alzheimer's disease: structural insight. Neurobiol Aging. 2012 Apr;33(4):831.e21-31. doi: 10.1016/j.neurobiolaging.2011.07.014. Epub 2011 Sep 13. Erratum in: Neurobiol Aging. 2018 Feb;62:247. PMID: 21920632.

22. Baloni P, Funk CC, Yan J, Yurkovich JT, Kueider-Paisley A, Nho K, Heinken A, Jia W, Mahmoudiandehkordi S, Louie G, Saykin AJ, Arnold M, Kastenmüller G, Griffiths WJ, Thiele I; Alzheimer's Disease Metabolomics Consortium, Kaddurah-Daouk R, Price ND. Metabolic Network Analysis Reveals Altered Bile Acid Synthesis and Metabolism in

Alzheimer's Disease. Cell Rep Med. 2020 Nov 17;1(8):100138. doi:
10.1016/j.xcrm.2020.100138. PMID: 33294859; PMCID: PMC7691449.

23. Ackerman, Hayley D, and Glenn S Gerhard. "Bile Acids in Neurodegenerative
Disorders." Frontiers in aging neuroscience vol. 8 263. 22 Nov. 2016,
doi:10.3389/fnagi.2016.00263

24. Nie B, Park HM, Kazantzis M, Lin M, Henkin A, Ng S, Song S, Chen Y, Tran H, Lai R,
Her C, Maher JJ, Forman BM, Stahl A. Specific bile acids inhibit hepatic fatty acid
uptake in mice. Hepatology. 2012 Oct;56(4):1300-10. doi: 10.1002/hep.25797. PMID:
22531947; PMCID: PMC3445775.

25. Nho K, Kueider-Paisley A, MahmoudianDehkordi S, Arnold M, Risacher SL, Louie G,
Blach C, Baillie R, Han X, Kastenmüller G, Jia W, Xie G, Ahmad S, Hankemeier T, van
Duijn CM, Trojanowski JQ, Shaw LM, Weiner MW, Doraiswamy PM, Saykin AJ,
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 Feb;15(2):232-244. doi: 10.1016/j.jalz.2018.08.012. Epub
2018 Oct 15. PMID: 30337152; PMCID: PMC6454538.

26. Di Ciaula A, Garruti G, Lunardi Baccetto R, Molina-Molina E, Bonfrate L, Wang DQ,
Portincasa P. Bile Acid Physiology. Ann Hepatol. 2017 Nov;16(Suppl. 1: s3-105.):s4-s14.
doi: 10.5604/01.3001.0010.5493. PMID: 29080336.