Pre-print version of the paper associated *doi.org/10.1007/s00216-020-02817-z*

Phospholipidomics of peripheral blood mononuclear cells (PBCM): the tricky case of children with autism spectrum disorder (ASD) and their healthy siblings

G. Ventura[1], C. D. Calvano[1,2], V. Porcelli[3], L. Palmieri[3,4], A. De Giacomo[5], Y. Xu[6], R. Goodacre[6], F. Palmisano[1,2] and T.R.I. Cataldi*[1,2]

[1]Department of Chemistry, and [2]Interdepartmental Research Center SMART University of Bari Aldo Moro, via Orabona 4, 70126, Bari, Italy

[3]Department of Biosciences, Biotechnologies and Biopharmaceutics, University of Bari Aldo Moro, via Orabona 4, 70126, Bari, Italy [4]CNR Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies (IBIOM), Bari, Italy

[5]Child Neuropsychiatry Unit, Department of Basic Medical Sciences, Neuroscience and Sense Organs, University of Bari Aldo Moro, Bari, Italy

[6]Department of Biochemistry, Institute of Integrative Biology, University of Liverpool, Biosciences Building, Crown Street, Liverpool L69 7ZB, UK

Number of Tables: 2

Number of Figures: 7

Number of Schemes: 1

Supplemental Material: Yes

Keywords: Autism Spectrum Disorder, lipidomics, HILIC-ESI-FTMS, PBMC, biomarkers, chemometrics

Author for correspondence, email: *tommaso.cataldi@uniba.it*

Abstract

Autism spectrum disorder (ASD) is a broad and heterogeneous group of neurological developmental disorders characterized by impaired social interaction and communication, restricted and repetitive behavioural patterns and altered sensory processing. Currently, no reliable ASD molecular biomarkers are available. Since immune dysregulation has been supposed to be related with ASD onset and dyslipidaemia has been recognised as an early symptom of biological perturbation, lipid extracts from peripheral blood mononuclear cells (PBMC), consisting primarily of lymphocytes (T cells, B cells and NK cells) and monocytes, of 38 children with ASD and their non-autistic siblings were investigated by hydrophilic interaction liquid chromatography (HILIC) coupled with electrospray ionization and Fourier-transform mass spectrometry (ESI-FTMS). Performances of two freeware software for data extraction and processing were compared to acquire reliable data regardless the used informatics. A reduction of variables from 1460 by the untargeted XCMS to 324 by the semi-untargeted Alex^{123} software was attained. All ion fragmentation (AIF) MS/MS scans along with Alex^{123} software were successfully applied to reveal the fatty acyl chains of glycerophospholipids occurring in PBMC. Principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) were explored to verify the occurrence of significant differences in the lipid pool composition of ASD children compared with 36 healthy siblings. After rigorous statistical validation we conclude that phospholipids extracted from PBMC of children affected by ASD do not exhibit diagnostic biomarkers. Yet interindividual variability comes forth from this study as the dominant effect in keeping with the existing phenotypic and etiological heterogeneity among ASD individuals.

1. INTRODUCTION

The American Psychiatric Association in their 5th edition of *"Diagnostic and Statistical Manual of Mental Disorders"* (DSM-5) [1] defined that autism spectrum disorder (ASD) as a broad and heterogeneous group of neurological developmental disorders characterized by several primary symptoms across several areas, such as difficulty with social abilities, stereotypical repetitive behaviors and interests, flawed language and conversation. Incidence rates of ASD are argued and dependent on diagnostic criteria. The Centers for Disease Control and Prevention has evaluated that the overall prevalence of ASD in children aged 8 years in the USA equaled 1 out of 68 children [2]. Despite thorough neurological, genetic and biochemical studies, the ASD aetiology is still largely mysterious. To add further complexity, environmental factors are also likely involved.

Due to lack of information regarding molecular mechanisms of the disorder, specific treatment and reliable diagnostic biomarkers are not available. Hence, diagnosis is currently made based on information gathered through children's interviews performed by specialized physicians and psychologists in which behaviour impairments are assessed along with psychiatric and developmental disorders [1]. DSM-5 diagnostic criteria include three functional raising levels defined on the base of needed *support* that a subject requires to relationship in the general community. At least 24 – 36 months of age are needed to diagnose ASD in children [3]. Clinically, individuals with ASD can differ substantially from each other in terms of the quality and severity of core symptoms, level of intellectual ability, co-occurring psychiatric symptoms, and developmental trajectories. Multiple neurocognitive and neurobiological abnormalities have been reported, but none seem to be shared by all individuals with ASD [4]. Yet, since the effect of early behavioral intervention is significant, the identification of diagnostic markers has gained considerable attention also due to the

increasing prevalence of ASD [5]. Interestingly, several lines of evidence indicate that diverse genetic as well as environmental risk factors may converge on a smaller number of interacting molecular pathways, including (Ca^{2+}) homeostasis [6,7], mitochondrial function [8,9] and immune response [10,11], which in turn impact brain circuit development and function [12].

Growing attention is being paid to immune dysregulation that may lead to impairments in neurodevelopment as numerous findings of altered immune system function in ASD children have been described [13]. An extensive search has shown that a subgroup of individuals with ASD show immune dysregulation that may represent a comorbidity of ASD or it may play a straight role in the development of ASD via impairment of neurodevelopmental processes. Nonetheless, results of these studies appear confusing due to design issues or small sample sizes [14]. Starting in 1986, numerous investigations on immune cells from peripheral blood of ASD children have been carried out [15] and recently reviewed [13], demonstrating imbalanced ratios of helper/suppressor cell and abnormalities in the number of total lymphocytes. Human peripheral blood mononuclear cells (PBMC), consisting primarily of lymphocytes (T cells, B cells and NK cells) and monocytes, are extensively used for research of immune cell functions, identification of biomarkers and development of diagnostics and therapeutics for human diseases. Alterations of mRNA expressions in PBMC obtained from ASD subjects have been shown [16]. Furthermore, in a preliminary case-control study, proteomics has recently led to the identification of 41 differentially expressed proteins in ASD children as potential biomarkers for early diagnosis [17].

In the present work, the contents of phospholipids extracted from PBMC of ASD children and their healthy siblings were evaluated to identify putative biomarkers through a metabolomic approach based on liquid chromatography with electrospray ionization coupled with Fourier-transform mass spectrometry (LC-ESI-FTMS). Hydrophilic interaction liquid

chromatography (HILIC) was employed due to its good ability to separate complex phospholipid mixtures on the basis of their polar head [18,19]; by this approach lipid species from different classes with the same nominal mass, almost co-eluent in reverse phase chromatography (RPC) due to their side chains structural similarity, can be separated and without ambiguity identified by MS. Untargeted LC-ESI-FTMS-based metabolomics generates huge amounts of data and their processing is challenging [20]. Here, two well-known freely available software packages often used for untargeted and semi-untargeted analysis, XCMS [21] and Alex^{123} [22,23], respectively, were evaluated and used to obtain data matrix. To ensure reliable results, quality control (QC) samples along with randomization of data extraction and analysis was applied. These lipidomics data were examined by principal component analysis (PCA) and partial least-squares discriminant analysis (PLS-DA).

2. MATERIALS AND METHODS

2.1 Chemicals. LC–MS grade water, acetonitrile (ACN), methanol (MeOH) and HPLC grade chloroform, formic acid, and ammonium acetate were obtained from Sigma-Aldrich (Milan, Italy). Standard solutions for negative calibrations were purchased from Thermo Scientific (Waltham, MA, USA).

2.2 Lipid nomenclature. Lipids were named according to the comprehensive classification system for lipids [24,25], e.g. 1-tetradecanoyl-2-hexadecanoyl-*sn*-glycerophosphocholine is designated PC (14:0/16:0). When fatty acid chain composition could not be determined, the total number of carbons and double bonds of all fatty acyl chains are given, e.g. PC (30:0).

2.3 PBMC Samples. Lipidomic analyses have been performed on PBMC samples obtained

from 38 patients affected by ASD with a disease severity degree from 1 to 3 according to DSM-5 and from their unaffected brothers or sisters; autistic children did not carry out any pharmacological treatment. The use of PBMC isolated from fresh whole blood in this study was approved by the Local Committee at the Azienda Ospedaliera Universitaria (Bari, IT) (n.164, November 11, 2016). Parents and participants provided written informed consent and all experiments were performed in agreement with guidelines and regulations. About 10 mL of blood have been treated on a Ficol gradient to separate blood mononucleates (lymphocytes and monocytes predominantly) from plasma; obtained samples were stored at −80 °C until sample preparation for metabolomics analysis. Data concerning all donors under investigation are summarized in **Table 1.**

2.4 Sample Preparation. Lipids were extracted from PBMC following the Bligh & Dyer protocol [26]. Briefly, approximately 2×10^6 lymphocyte cells were dissolved in 400 µL of LC-MS grade water and 1.5 mL of methanol/chloroform (2:1, v/v) added to the solution and left for 1 h at room temperature. Then, 0.5 mL of chloroform was added, and the mixture was vortexed for 30 s. Finally, 0.5 mL of water was added, and the solution was shaken before being centrifuged for 10 min at 3000 x*g*. The lower phase containing lipids was dried under nitrogen; the residue was dissolved in 100 µL of methanol and then analysed by LC–MS in two different analytical batches.

2.5 Instrumentation and operating conditions. Quality control (QC) samples were prepared for each batch following protocols in [27] by pooling equal volumes of reconstituted samples and divided them in three aliquots. Samples were transferred into 2 mL glass vials

containing 100 µL glass inserts with polymer feet (Supelco). Vials were covered with preslit polytetrafluoroethylene (PTFE)/silicone screw caps.

Samples were analysed in a randomized fashion by using an Ultimate 3000 UHPLC system (Thermo Scientific, Waltham, MA, USA) coupled to a Q-Exactive mass spectrometer (Thermo Scientific, Waltham, MA, USA), including a quadrupole connected to an Orbitrap analyser. The column effluent was transferred into the Q-Exactive spectrometer through a heated electrospray ionization (HESI) interface. The main electrospray and ion optics parameters were the following: sheath gas flow rate, 35 arbitrary units (a.u.); auxiliary gas flow rate, 15 a.u.; spray voltage, ±3.5 kV (positive/negative polarity); capillary temperature, 320◦C; S-Lens RF Level, 100 a.u. MS spectra were acquired in the *m/z* range 200–2000, at a mass resolving power of 140000 (measured at *m/z* 200). The Orbitrap fill-time was set to 200 ms and the automatic gain control (AGC) level was set to 2.5×10^6.

To retrieve information on the separated phospholipids (PL), additional AIF-MS^2 acquisitions were performed during each chromatographic run using a resolving power of 70,000 (at *m/z* 200), an Orbitrap fill-time of 100 ms and an AGC value of 5×10^5. All ions fragmentation (AIF) with multiple dissociation techniques, i.e. in source collision induced dissociation (sid) and HCD, providing MS and MS/MS data were also employed to increase the amount of retrievable information. AIF spectra were acquired using an NCE value of 35% and the same resolving power, trap-fill time and AGC value adopted for MS acquisitions. The mass accuracy, after calibration using a solution containing caffeine, the MRFA peptide and Ultramark, provided by Thermo Scientific, ranged between 0.43 and 0.49 ppm in negative polarity.

Silica phase Ascentis Express HILIC column (150 × 2.1 mm id, 2.7 µm particle size) equipped with an Ascentis Express HILIC (5 × 2.1 mm id) security guard cartridge (Supelco,

Bellefonte, PA, USA) operating at a flow rate of 0.3 mL/min was used to perform chromatographic separations; sample injection (5 µL) was performed by a RS Autosampler (Thermo Scientific, Waltham, MA, USA). The following binary elution program, based on water and 2.5 mmol/L ammonium acetate (solvent A) and ACN (solvent B), both containing 0.1% (v/v) of formic acid, was adopted: 0–5 min, linear gradient from 97 to 88% solvent B; 5–10 min, isocratic at 88% solvent B; 10–11 min, linear gradient from 88 to 81% solvent B; 11–20 min, linear gradient from 81 to 70% solvent B; 20–22 min, linear gradient from 70 to 50% solvent B; 22–28 isocratic at 50% solvent B; 28–30 min, return to the initial composition, followed by a 5 min equilibration time.

Following 3 blank injections (solvent blanks), 5 QC sample injections were carried out at the beginning of each batch for column conditioning and every five samples throughout the analytical run to assess analytical reproducibility; two QC injections were performed at the end of the experiment to eliminate the impact on signal correction if one sample-injection or instrument failure accidently occurs; the absence of a QC sample at the end of the experiment significantly impacts on the QC-RLSC algorithm applied [27,28]. The analysis order is described in **Table S1**. LC–MS instrumentation control and first processing of data were performed by the Xcalibur software 2.2 SP1.48 (Thermo Scientific).

2.6 Data processing. Raw files were converted to mzXML format with MSconvert (a tool provided by ProteoWizard: http://proteowizard.sourceforge.net/tools.shtml). Two software packages, XCMS and Alex[123], were used to obtain the data matrix containing the list of all detected features, including information such as accurate measured masses and areas of chromatographic peaks. The list of the parameters used for XCMS software is available as Supplementary Material (see Table S2); full-MS and AIF-MS/MS spectra over each HILIC

chromatographic band were manually averaged and resulting txt files were used as input for Alex[123]; only attributions with accuracy lower than 5 ppm and intensities higher than 5000 counts were further analysed.

3. RESULTS AND DISCUSSION

3.1. Data processing of XCMS results.

Hydrophilic interaction liquid chromatography (HILIC) coupled to high-resolution Fourier-transform mass spectrometry (FTMS) is recognized as a powerful platform for lipid identification [29,30]. HILIC separations, wherein lipid elution order is decreed by the nature of the polar head, coupled with fast and sensitive MS detection systems enable quantitative measurements of hundreds of lipid species even within complex clinical specimens, such as human blood plasma [31] or cells [32,33]. In this study high-resolution, accurate-mass (HR/AM) ESI-Orbitrap MS analysis performed in negative-ion mode by source-induced dissociation (sid) was employed to investigate lipid extracts from lymphocytes samples. The resulting data were subsequently processed by two freely available software tools, XCMS and Alex[123] for peak detection and integration. Both these software packages generated 2D data matrices, including variable indices, sample names and peak areas. **Figure 1** shows the adopted strategy of lipidextraction, acquisition, data extraction, pre-processing and chemometrics analysis combined into a single data analysis workflow. XCMS Online [21,34] is a well-known LC-MS data analysis freely available platform, developed by the Scripps Center for untargeted metabolomics data (www.xcmsonline.xcripps.edu). This software incorporates nonlinear retention time alignment, matched filtration, peak detection and peak matching; no preliminary information on the investigated analytes are required and, since a completely untargeted approach is used, the risk of introduction of artefacts is very low. However, the

outcome of data processing strongly depends on setting parameters, such as tolerated *m/z* deviation in consecutive scans and maximum or minimum chromatographic peak width; if those factors are not carefully chosen, biased results may be expected. The software package called *Isotopologue Parameter Optimization* (IPO) [35] optimizes XCMS peak picking parameters by using natural, stable ^{13}C isotopic peaks to calculate a peak picking score [36]. Retention time correction was optimized by minimizing relative retention time differences within peak groups, while grouping ones were optimized by maximizing the number of peak groups that show one peak from each injection of a pooled sample. The different parameter settings were achieved by design of experiments and the resulting scores were evaluated using response surface models. So, IPO was applied to QC samples and results were implemented in XCMS to increase the reliability of metabolomics data (see **Table S1**).

Figure 2 shows the superimposition of all the total ion current (TIC) chromatographic profiles along with retention time deviation between runs obtained after the XCMS alignment procedure. HILIC-ESI-FTMS was applied for the lipid separation of 38 samples of autism spectrum disorder affected children, 36 control samples (unaffected siblings of studied subjects) and 21 pooled quality control samples (QC). CAMERA (Collection of Algorithms for MEtabolite pRofile Annotation) software package was employed to interpret and evaluate LC-MS data, including algorithms for annotation of isotope peaks and adducts [37] in order to obtain a data matrix of 95 samples x 2821 variables; only M+1 and M+3 isotopologue peaks were removed, because of the well-known problem of overlapping between the third isotopologue of a species with *X* carbon atoms and *Y* unsaturations and the first isotopologue peak of the species having *X* carbon atoms and *Y*-1 unsaturations [38]. Critical issues faced over time while performing metabolome profiling analysis [27,39–41] are related to instrumental sensitivity changes along with degradation of sample extracts, ion source

contamination and retention-time shifts. To this aim, quality control (QC) samples, obtained by pooling together small aliquots of each biological sample, were run every five biological samples to monitor signal intensity drift over the analysis time (in this study there was very little drift as discussed below) and to allow its correction considering that the relative standard deviation (RSD) of each variable was generally set to a maximum of 20% [40]. Upon removal of (i) signal intensity with RSD >20% in QC, (ii) peak signals with *m/z* lower than 400 and (iii) M+1 and M+3 isotopologues, the number of features decreased to 1460. Considering the significant amount of lipid species in a biological sample and the possibility of many potential adducts for each one (i.e. principally deprotonate, demethylate, formate, acetate or chlorinate adducts) along with the plausible presence of species not related to lipids (i.e. contaminants, but also different metabolites that could be extracted using Bligh & Dyer protocol) this number is not surprising. As described by Dunn *et al*. [27], to compensate difference in extraction yields, QC correction and normalization to the sum of all the signals found in the spectra were performed and multivariate analysis was applied to obtained data matrix. Since pre-treatment methods represent a crucial step to revealing hidden information in metabolomic analysis [42], four different data transformation were examined. As can be seen in Figure 3, QC samples (red circles) were always well clustered in the PCA score plots showing excellent analytical repeatability. This was irrespective of the data pre-processing used: *viz*., mean centering (plot A), autoscaling (plot B), level scaling (plot C) and $\log_{10}$-transformation (plot D), even reporting in the score plot principal components (PCs) of major order (data not shown). In the PCA plots samples belonging to different classes (i.e. healthy *vs* ASD children) were labelled by different colours and symbols: blue diamonds for patients and green squares for healthy siblings. Despite diverse pre-processing methods such as autoscaling of data, needed to ensure that less abundant but, perhaps, important lipid species

contribute to a separation of data, no clear trends or clustering were observed: in all graphs, ASD patient samples are spread in the whole principal component space even taking into account subsequent PCs or different pre-treatment data. Although some ASD children (i.e., PT 6, 7, 8, 9, 27, 28 and 32) appear to be more separated from control group for example in score plot of **Figure 3 A**, it does not seem to be a consistent reason linking their disease, and so we do not design this significant. Notably, they have not so much in common because ASD's severity, ranging from 1 to 3, age between 3 to 17, gender and cognitive delay are different (see Table 1). Thus, PCA score plots reporting sex and age do not bring to a clustering according to gender, and cluster or trends are not observed in age (**Figure S1**). Absence of obvious groups or clusters was also obtained when the plots were labelled according to ASD degree and cognitive delay (**Figure S2**).

Subject group 42 is a triad composed by three siblings: one healthy 9-year-old boy, his ASD affected sister, with the maximum ASD severity, and her monozygotic twin sister, both 8 years old. As the concordance rate in monozygotic twins is estimated to be approximately 90% [43], this combination is really interesting because it represents an example of two monozygotic twins where just one of them is affected by ASD. Regardless of differences in sex, age (only 1 year difference between siblings) and ASD severity, these three siblings (see **Figure 3, plots A-D**) are not very close each other in the score plot: monozygotic twins, even having a most common genome and sharing age and gender, are not close in the space of variables, thus confirming the system complexity.

Afterward, we attempted to generate ASD predictive models by applying partial least squares-discriminant analysis (PLS-DA) on XCMS data (**Figure 4**). The following plots were generated, centered (A), autoscaled (B), level scaled (C) and $\log_{10}$-transformed (D) also using the ASD status (yes or not) as the target variable: ASD was encoded in the *Y*-variable as '1' and

absence of ASD as '0'. To test the validity of any modelling, 1000 bootstraps with replacement were performed and all assessments were made on the 1000 test sets (i.e., not the data used to construct the models). We also used permutation testing where the *Y*-variable was scrambled to generate 1000 null distributions, again for the test sets. As can be seen in Figure 4, there was a large overlap of distributions and absence of statistically significant separation among all plots. According to the resulting confusion matrix (**Table 2**), the PLS-DA model showed predictions close to 0.5 and large *p*-values, again confirming the lack of discriminatory power in these PLS-DA models. In **Figure S3** PLS-DA score plots are also displayed, together with Q^2 and R^2 metrics to outline that their use for inference of class differences often provides an over-optimistic understanding of the separation between classes [44]; although in this case the Q^2 values are all insignificant. At first glance, it appears that a class separation was obtained, yet the metrics proved that the expected results were not achieved as already demonstrated by the distribution overlap of **Figure 4.** In addition, other variables were examined for discriminant analysis (i.e. histological types, cognitive delay, gender and age), using the above-mentioned transformation data. Modelling of the four histological types (healthy, n° 35; ASD severity 1, n° 13; ASD severity 2, n° 14; ASD severity 3, n° 6; **Figure S4** plot A in Supplementary Material) also produced no discrimination power between the several examined pairs, since less than 2% of the data were correctly assigned for the highest ASD degree of severity while the major part of them appeared as belonging to control group. Next, we used the cognitive delay status to try to discrimination between ASD affected children with or without pathology (i.e., 13 cognitive delayed ASD affected children *vs* 23 ASD affected children without cognitive delay). Yet, no significant discriminant power was statistically obtained. The whole data set was later modelled by PLS-DA taking the gender as classifier to understand the importance of this variable. Note that a significant limitation is present in the

data set since there was a skew in the patient group with less than the 25% of female subjects and, among them, only five ASD children. The present results demonstrate that the metabolic differences between two genders were not significant (see level scaled data of **Figure S4C**). Being present a large difference in the age of investigated subjects, they were grouped in three ranges: A) 3-6, B) 7-12, C) 13-17. Apparently, PLS-DA models showed better predictive power, however it was still clearly not enough to discriminate among lipidomics data (see PLS-DA on log_{10}-scaled data of **Figure S4 D**). Once again, age was not a dominating variability source in our data set; the lipidic profile of PBMC is seemingly less important than other uncontrolled sources of variation.

3.2. Data processing of Alex^{123} results.

It is possible that if only a few lipid classes are involved in disease the presence of uninformative variables results in increased noise and discrimination power is lost: for example, alterations only in amino-glycerophospholipids levels (e.g., phosphatidylethanolamine, PE and phosphatidylserine, PS) of children with autism have been reported in plasma samples [45]. Although XCMS shows great advantages for obtaining rapid elaborate data, the same is not true when there is the need to dismiss certain lipid classes. In addition, another issue with the XCMS workflow is that it is likely to embrace non-lipid related species, and these may also add noise into raw data.

Alex^{123} [22,23] is a powerful and reliable high-throughput tool (freely available at www.mslipidomics.info) for semi-untargeted analysis of lipids that searches compounds in a customizable database containing, for each lipid: name, molecular formula and mass, indication of several adducts or modifications. A deep spectral examination is required to build up a comprehensive database in which retention time windows of lipid classes and most

common ions for a certain lipid generated in the described experimental conditions must be known. Specific and class-related product ions can be promptly retrieved by using the all ion fragmentation (AIF) MS/MS scan, a feature provided by the FTMS system, without the need of isolating and fragmenting definite precursor ions. By exploiting AIF MS/MS, it is possible to recover product ions related to the polar head of each PL generated at relatively high collisional energies in the HCD cell. In **Figure S5** are compared XIC spectra and structures of (A) *m/z* 168.043 due to phosphatidylcholines (PC), lyso-PC (LPC) and sphingomyelin (SM), (B) *m/z* 224.069 associated with PC and LPC, and (C) *m/z* 196.038, due to PE and lyso-PE (LPE). From the information obtained by AIF spectra, elution windows of PC, LPC, SM, PE and LPE can be easily detected; ether phospholipids, lipids in which the *sn*-1 position of the glycerol backbone has a lipid attached by an ether bond, co-elute with the more common diacyl subclasses; plasmanyl-phospholipids (indicated using *o*-) have an ether bond in position *sn*-1 to an alkyl chain, while plasmenyl-phospholipids (*p*-) have an ether bond in position *sn*-1 to an alkenyl moiety. A lipid species *p*-Z *X*:*Y*, where Z is the class name, *X* the carbon atoms and *Y* the degree of unsaturations in the side chains, respectively, is isobaric to a *o*-Z *X*:(*Y*-1) species; MS/MS analysis carried out in positive ion mode can be used to discriminate between these two species [46]. Note that in the first part of the work this information was not crucial and plasmanyl- and plasmenyl-phospholipids were indifferently indicated as *–O* and fully characterized later in the phospholipidomics of ASD disease.

To create the Alex^{123} database, only unambiguous identifications were selected to obtain putatively annotated compounds on the basis of intra-laboratory class retention time and Orbitrap FTMS accurate mass [47]. For instance, source-induced dissociation (sid) enhances the generation of $[M\text{-}CH_3]^-$ ions in PL bearing a choline moiety in the polar head, i.e., PC, LPC, and SM (M represents the zwitterionic form of these PL) that otherwise ionize

mainly as formate $[M+HCOO]^-$ adducts often confusingly with alternative isobaric species: e.g., the formate adduct of PC 36:2 is isobaric with the acetate adduct of PC 35:2. Therefore, after RSD criterion application, 324 variables were obtained (*i.e.*, 17 Hex_2Cer, 12 LPC, 4 LPC-O, 20 LPE, 15 LPE-O, 39 PC, 23 PC-O, 49 PE, 40 PE-O, 27 PG, 18 PI, 33 PS and 27 SM). The score plot obtained by using all these variables and then a row normalization gave good clustering of QC in the middle of the PCA score plot, highlighting that the data processing did not introduce any artefacts into the data output and again confirmed the excellent analytical reproducibility of the used LC-ESI-FTMS approach. Despite some little differences (in **Figure 5**, plot A, biplot is reported), PCA showed that the centered data were rather like that obtained using XCMS for data matrix construction. Also using different pre-processing methods, the same groupings in PCA score plots were regularly obtained (data not shown). As already mentioned, one of the main advantages of Alex[123] is the possibility to obtain data linked to lipid classes under examination. As an example, biplot obtained for LPE signals, normalized and auto scaled, exhibited a mixed distribution (**Figure 5B**). The examination of other lipid classes, also using different processing methods, did not lead to a well definite separation. Nonetheless, PLS-DA was applied on data elaborated by Alex[123] using pathology as discriminant variable and different data pre-treatment. In **Figure 6A** and **6B**, are displayed the PLS-DA performed on the whole data set and on SM class, respectively, either centered or autoscaled data; as reported for the data after deconvolution using XCMS, no significant discrimination power was again obtained.

3.3. **All ion fragmentation MS/MS scan as a tool in metabolomic analysis**

Recently we have described the possibility of using data generated after HILIC separation and AIF MS/MS scan to obtain a snapshot of the fatty acyl composition of some

glycerophospholipid classes [31]. Here, the same approach was applied, and data quality was assessed by using a metabolomic approach. Under the seven main chromatographic bands, i.e. PI (#1), PE & PEO (#2), LPE & LPE-O (#3), PS (#4), PC & PC-O (#5), SM (#6) and LPC & LPC-O (#7), AIF MS/MS spectra were integrated and data were evaluated by Alex^{123}. Note that the AIF data integrated under the SM band cannot be included among suitable spectra because the instrumental variability was higher than biological one as demonstrated by QC samples not clustered in the middle of the score plot (see **Figure S6**). This result was somehow predictable since fragmentation of SM does not produce very intense fatty acyl signals [46]; so, AIF data related to FA of SM species were removed and not further analysed. Upon RSD criterion application and QC correction, up to 133 FA related signals were obtained, namely 17 for band #1, 30 for band #2, 23 for band #3, 22 for band #4, 29 for band #5, 12 for band #7. **Figure 5C** shows a centered PCA plot of all the extracted FA data matrix. As can be seen, signals of fatty acyl substituents mainly bounded to PC and PE discriminate between studied subjects (i.e. mainly FA 16:0, 18:1 and 20:4 in PC, FA 20:4 in PE class, but also FA 18:0 in LPE and in PS), in accordance with relative intensities of the considered lipid classes. Notwithstanding no clustering related to ASD, gender or age was established. Interestingly, these preliminary findings showed for the first time the possibility to exploit AIF data collected by HILIC also for lipidomics purposes. HILIC-ESI(-)AIF MS/MS scan can be used to collect data under each chromatographic band, thus demonstrating subtle differences in the composition of acyl chains; **Figure 5D** shows the PCA score plot of FA signals under band #5. PLS-DA was applied using the whole data set (**Figure 6C**) or only FA signals obtained in AIF MS/MS scan under band #2 (**Figure 6D**) using different pre-treatment methods (for example, level scaling for data in **Figure 6C** and $\log_{10}$-transformation in **Figure 6D**) but no significant discrimination power was obtained.

411

412 **3.4 Univariate test.**

413 To supplement the multivariate approaches discussed above and to try to overcome between-
414 subjects' variability, paired-sample t-test and Wilcoxon signed rank test for zero median were
415 explored as selected by Alex[123], thus comparing the level of each variable between ASD
416 children and their healthy sibling. It is worth mentioning that this is a pseudo-paired test as
417 the pairs are not the same children after some perturbation but pairs of siblings: one with ASD
418 and the other non-autistic. The level of significance was set initially at $P < 0.05$ and, to avoid
419 multiple testing problems, false discovery rates (FDR) of multiple-hypothesis were tested
420 applying the procedure described by Storey et al. [48]. Notably, different significant features
421 were obtained, and box plots were constructed reporting lipid levels *vs* ASD severity degree;
422 whether an ascending or descending trend was observed, the feature could be related to
423 autism. Some variables seemed to follow a specific trend but the between-subject variability
424 still remains the dominating effect; an example is given in **Figure 7** where box-plots of PI 38:4
425 (Alex[123] assignment, *m/z* 885.550) and FA 20:4 under PI band (*m/z* 303.323) are illustrated. As
426 can be seen, similar trends were obtained in both graphs as a confirmation of the utility of full
427 and AIF scan comparison: in principle, AIF data can be used to understand which fatty acyl
428 chains are involved in the aetiology of the considered pathology. MS/MS spectra confirmed
429 the attribution to PI as two isobaric species, namely PI 18:0/20:4 and PI 16:0/22:4 (not shown).
430 However, as already mentioned, interindividual variability seems to be the dominant
431 outcome. In an attempt to minimize the environmental factors, the level of each considered
432 lipid and the corresponding FA obtained through AIF MS/MS scan was plotted and compared
433 among siblings; no presence of systematic trend was still evidenced.

434

Discussion

Lipids represent a very broad group of molecules with a substantial structural diversity that is reflected in the variety and complexity of the physiological processes in which they are involved, from providing cell structure to energy storage for cell signalling [49–52]. Generally, any perturbation of a biological system is expected to alter the abundance and/or composition of the lipid pool of that system [53]. In the field of ASD biomarker discovery, very few studies have examined fatty acid metabolism with the underlying idea being that the abnormal membrane fatty acid composition is involved in neurodevelopmental and psychiatric disorders [54–57]. However, ambiguous or at least non-definitive results have been presented including higher levels of PUFA occurring in biological fluids of autistic subjects. Vancassel *et al*. [58] speculated that the total *n*-3 PUFA were significantly lower in the population of autistic patients compared to mentally retarded ones, yet arachidonic acid (AA, *i.e.* FA 20:4) and docosahexaenoic acid (DHA, *i.e.* FA 22:6) plasma levels were only moderately reduced. Likewise, Bell *et al*. [59,60] found significantly lower AA and *n*-6 PUFA levels in phospholipids of red blood cells (RBC) in the autistic test group compared to pair-matched developmentally delayed controls. Yet, these abnormalities were not replicated in the study of Bu *et al*. [61] as no strong evidences of PUFA differences between autistic individuals and age-matched controls were confirmed. Wiest and colleagues [54] found that within the phosphatidylcholine class, DHA was significantly lower in the autistic group than in the general population, while plasma AA levels in phospholipids were not significantly different between groups, although AA was found to be significantly lower in free fatty acids of ASD participants. An increase in most of the saturated fatty acids and a decrease in most of polyunsaturated fatty acids was reported in the plasma of a cohort of autistic patients from Saudi Arabia [62]. More recently, multivariate statistical analysis of the content of a dozen fatty acids, including AA and DHA,

suggested that unsaturated fatty acids in erythrocytes are not predictive of autism spectrum disorder [63]. A major limitation of these studies lies in the fact that fatty acid contents in plasma and to a lesser extent also in red blood cells are responsive to dietary habits [64]. Indeed, many children with ASD display restrictive food preference [64] and low intake of foods containing PUFA by individuals with ASD has been documented [65]. Most importantly, these studies have been carried out under the assumption that ASD children are largely biochemically homogenous and a single or even a small number of "marker" molecules could discriminate them from sex/age matched neurotypical peers. However, it is likely that ASD covers a bunch of biochemical phenotypes such as the large heterogeneity observed clinically. Therefore, more insight into ASD pathogenicity may result from analyzing large patterns.

Here, for the first time, we have used high-resolution mass spectrometry to perform a comprehensive phospholipidomic analysis of PBMCs from ASD children and their non-autistic siblings. Another strength of this study is the case-sibling approach undertaken to minimize genetic variability. In fact, it is believed that genetic variation explains over 50% of the risk of developing ASD [66] and the risk is increased 10 fold if a sibling reports the diagnosis; moreover, aggregates in families and early twin studies estimated the proportion of the phenotype variance due to genetic factors to be up to 90% [67]. Therefore, unaffected siblings are an ideal control group because they enable a more accurate assessment as to whether any observed differences are due to the autism phenotype because they control for shared genes and possibly also common epigenetic modifications induced early in life.

For its complex aetiology, the role of environmental factors in the onset of ASD is still largely unknown: physiological and chemical elements are the most commonly studied in association with ASD, while research on nutritional and social influences are limited [68]. Unfortunately, the case-sibling approach does not overcome the difficulty of disentangling the

effects of variable genetic risks, environmental exposures across development and the likely interactions between these factors in a population with considerable phenotypic and prognostic heterogeneity. Furthermore, the fatty acid composition of human immune cells can be modified by altering oral intakes of certain fatty acids [69]. All those factors, together with differences in sex, age and lifestyles, could explain the observed variance in the present study and the lack of ability to separate ASD from paired healthy siblings.

Limitations of this study predominantly include the partial phenotyping, which, as noted previously, prevent us from addressing pressing questions about heterogeneity in ASD. The search for shared patterns of lipid composition associated with ASD may be more fruitful within ASD subgroups that reflect shared etiological and developmental factors. Detecting shared patterns of lipid composition associated with ASD subgroups requires a larger sample size and more extensive phenotyping and/or genotyping than the current dataset. All those factors, together with differences in sex, age and lifestyles, could explain the observed variance in the present study and the lack of ability to separate ASD from paired healthy siblings. We are therefore left wondering: are lymphocytes not a good choice for lipid levels comparison? We believe that further work is most likely needed to shed light on this tricky question.

CONCLUSIONS

In this study, lipid extracts obtained from isolated PBMC of children affected by ASD, along with samples of their healthy siblings were analysed by metabolomics. Deconvolution of the LC-ESI-FTMS data using untargeted (XCMS) and semi-targeted (Alex^{123}) approaches, were examined and compared. Interindividual variability is seemingly the most dominant factor as no significant differences were revealed by multivariate analysis for ASD *vs* healthy siblings

within the sampled lipid pools. It is possible that dietary habits and comorbidities contribute to the system complexity, together with differences in sex, age and/or ASD severity. The fact that the disease aetiology of ASD is unknown and likely multifactorial, confounds the current studies such as the present one. Further work with more defined ASD subgroups may shed some light on this complex disease.

ACKNOWLEDGMENTS

This work was supported by the following projects: (i) PONa3_00395/1 “BIOSCIENZE & SALUTE (B&H)” and (ii) SIR 2014, Identification and Characterization of Biomarkers for Autism Spectrum Disorder of “Ministero per l'Istruzione, l'Università e la Ricerca” (MIUR). G.V. during his Ph.D. period abroad carried out part of this work in Manchester at the Institute of Biotechnology.

Conflict-of-interest

The authors declare no conflict-of-interest.

REFERENCES

[1] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders, Fifth Edit, American Psychiatric Association, 2013. https://doi.org/10.1176/appi.books.9780890425596.

[2] E. Fonbomme, E. Fombonne, É. Fombonne, Epidemiological surveys of autism and other pervasive developmental disorders: An update, J. Autism Dev. Disord. 33 (2003) 365–382. https://doi.org/10.1023/A:1025054610557.

[3] A.M. Daniels, D.S. Mandell, Explaining differences in age at autism spectrum disorder diagnosis: A critical review, Autism. 18 (2014) 583–597. https://doi.org/10.1177/1362361313480277.

[4] T. Charman, Variability in neurodevelopmental disorders: Evidence from autism spectrum disorders, in: Neurodev. Disord. Res. Challenges Solut., Psychology Press, 2015: pp. 117–140. https://doi.org/10.4324/9781315735313-17.

[5] L. Shen, X. Liu, H. Zhang, J. Lin, C. Feng, J. Iqbal, Biomarkers in autism spectrum disorders: Current progress, Clin. Chim. Acta. 502 (2020) 41–54. https://doi.org/10.1016/j.cca.2019.12.009.

[6] R.L. Nguyen, Y. V. Medvedeva, T.E. Ayyagari, G. Schmunk, J.J. Gargus, Intracellular calcium dysregulation in autism spectrum disorder: An analysis of converging organelle signaling pathways, Biochim. Biophys. Acta - Mol. Cell Res. 1865 (2018) 1718–1732. https://doi.org/10.1016/j.bbamcr.2018.08.003.

[7] L. Palmieri, V. Papaleo, V. Porcelli, P. Scarcia, L. Gaita, R. Sacco, J. Hager, F. Rousseau, P. Curatolo, B. Manzi, R. Militerni, C. Bravaccio, S. Trillo, C. Schneider, R. Melmed, M. Elia, C. Lenti, M. Saccani, T. Pascucci, S. Puglisi-Allegra, K.-L. Reichelt, A.M. Persico, Altered calcium homeostasis in autism-spectrum disorders: evidence from biochemical and genetic studies of the mitochondrial aspartate/glutamate carrier AGC1, Mol. Psychiatry. 15 (2010) 38–52. https://doi.org/10.1038/mp.2008.63.

[8] L. Palmieri, A.M. Persico, Mitochondrial dysfunction in autism spectrum disorders: Cause or effect?, Biochim. Biophys. Acta - Bioenerg. 1797 (2010) 1130–1137. https://doi.org/10.1016/j.bbabio.2010.04.018.

[9] S. Rose, D.M. Niyazov, D.A. Rossignol, M. Goldenthal, S.G. Kahler, R.E. Frye, Clinical and Molecular Characteristics of Mitochondrial Dysfunction in Autism Spectrum Disorder, Mol. Diagn. Ther. 22 (2018) 571–593. https://doi.org/10.1007/s40291-018-0352-x.

[10] P. Ashwood, S. Wills, J. Van de Water, The immune response in autism: a new frontier for autism research, J. Leukoc. Biol. 80 (2006) 1–15. https://doi.org/10.1189/jlb.1205707.

[11] V. Napolioni, A.M. Persico, V. Porcelli, L. Palmieri, The Mitochondrial Aspartate/Glutamate

Carrier AGC1 and Calcium Homeostasis: Physiological Links and Abnormalities in Autism, Mol. Neurobiol. 44 (2011) 83–92. https://doi.org/10.1007/s12035-011-8192-2.

[12] M. Sahin, M. Sur, Genes, circuits, and precision therapies for autism and related neurodevelopmental disorders, Science. 350 (2015) aab3897–aab3897. https://doi.org/10.1126/science.aab3897.

[13] D. Gładysz, A. Krzywdzińska, K.K. Hozyasz, Immune Abnormalities in Autism Spectrum Disorder—Could They Hold Promise for Causative Treatment?, Mol. Neurobiol. 55 (2018) 6387–6435. https://doi.org/10.1007/s12035-017-0822-x.

[14] J. Mead, P. Ashwood, Evidence supporting an altered immune response in ASD, Immunol. Lett. 163 (2015) 49–55. https://doi.org/10.1016/j.imlet.2014.11.006.

[15] R.P. Warren, N.C. Margaretten, N.C. Pace, A. Foster, Immune abnormalities in patients with autism, J. Autism Dev. Disord. 16 (1986) 189–197. https://doi.org/10.1007/BF01531729.

[16] S.J. Glatt, M.T. Tsuang, M. Winn, S.D. Chandler, M. Collins, L. Lopez, M. Weinfeld, C. Carter, N. Schork, K. Pierce, E. Courchesne, Blood-Based Gene Expression Signatures of Infants and Toddlers With Autism, J. Am. Acad. Child Adolesc. Psychiatry. 51 (2012) 934-944.e2. https://doi.org/10.1016/j.jaac.2012.07.007.

[17] L. Shen, C. Feng, K. Zhang, Y. Chen, Y. Gao, J. Ke, X. Chen, J. Lin, C. Li, J. Iqbal, Y. Zhao, W. Wang, Proteomics Study of Peripheral Blood Mononuclear Cells (PBMCs) in Autistic Children, Front. Cell. Neurosci. 13 (2019) 105. https://doi.org/10.3389/fncel.2019.00105.

[18] I. Losito, R. Patruno, E. Conte, T.R.I. Cataldi, F.M. Megli, F. Palmisano, Phospholipidomics of Human Blood Microparticles, Anal. Chem. 85 (2013) 6405–6413. https://doi.org/10.1021/ac400829r.

[19] E. Cífková, M. Holčapek, M. Lísa, Nontargeted Lipidomic Characterization of Porcine Organs Using Hydrophilic Interaction Liquid Chromatography and Off-Line Two-Dimensional Liquid Chromatography–Electrospray Ionization Mass Spectrometry, Lipids. 48 (2013) 915–928. https://doi.org/10.1007/s11745-013-3820-4.

[20] H. Tsugawa, K. Ikeda, M. Arita, The importance of bioinformatics for connecting data-driven lipidomics and biological insights, Biochim. Biophys. Acta - Mol. Cell Biol. Lipids. 1862 (2017) 762–765. https://doi.org/10.1016/j.bbalip.2017.05.006.

[21] E.M. Forsberg, T. Huan, D. Rinehart, H.P. Benton, B. Warth, B. Hilmers, G. Siuzdak, Data processing, multi-omic pathway mapping, and metabolite activity analysis using XCMS Online, Nat. Protoc. 13 (2018) 633–651. https://doi.org/10.1038/nprot.2017.151.

[22] P. Husen, K. Tarasov, M. Katafiasz, E. Sokol, J. Vogt, J. Baumgart, R. Nitsch, K. Ekroos, C.S. Ejsing, Analysis of Lipid Experiments (ALEX): A Software Framework for Analysis of High-Resolution Shotgun Lipidomics Data, PLoS One. 8 (2013) e79736.

https://doi.org/10.1371/journal.pone.0079736.

[23] S.R. Ellis, M.R.L. Paine, G.B. Eijkel, J.K. Pauling, P. Husen, M.W. Jervelund, M. Hermansson, C.S. Ejsing, R.M.A. Heeren, Automated, parallel mass spectrometry imaging and structural identification of lipids, Nat. Methods. 15 (2018) 515–518. https://doi.org/10.1038/s41592-018-0010-6.

[24] G. Liebisch, J.A. Vizcaíno, H. Köfeler, M. Trötzmüller, W.J. Griffiths, G. Schmitz, F. Spener, M.J.O. Wakelam, Shorthand notation for lipid structures derived from mass spectrometry, J. Lipid Res. 54 (2013) 1523–1530. https://doi.org/10.1194/jlr.M033506.

[25] G. Liebisch, K. Ekroos, M. Hermansson, C.S. Ejsing, Reporting of lipidomics data should be standardized, Biochim. Biophys. Acta - Mol. Cell Biol. Lipids. 1862 (2017) 747–751. https://doi.org/10.1016/j.bbalip.2017.02.013.

[26] E.G. Bligh, W.J. Dyer, A rapid method of total lipid extraction and purification, Can. J. Biochem. Physiol. 37 (1959) 911–917. https://doi.org/10.1139/o59-099.

[27] W.B. Dunn, D. Broadhurst, P. Begley, E. Zelena, S. Francis-McIntyre, N. Anderson, M. Brown, J.D. Knowles, A. Halsall, J.N. Haselden, A.W. Nicholls, I.D. Wilson, D.B. Kell, R. Goodacre, Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry, Nat. Protoc. 6 (2011) 1060–1083. https://doi.org/10.1038/nprot.2011.335.

[28] M.A. Kamleh, T.M.D. Ebbels, K. Spagou, P. Masson, E.J. Want, Optimizing the Use of Quality Control Samples for Signal Drift Correction in Large-Scale Urine Metabolic Profiling Studies, Anal. Chem. 84 (2012) 2670–2677. https://doi.org/10.1021/ac202733q.

[29] S. Granafei, P. Azzone, V.A. Spinelli, I. Losito, F. Palmisano, T.R.I. Cataldi, Hydrophilic interaction and reversed phase mixed-mode liquid chromatography coupled to high resolution tandem mass spectrometry for polar lipids analysis, J. Chromatogr. A. 1477 (2016) 47–55. https://doi.org/10.1016/j.chroma.2016.11.048.

[30] S. Granafei, I. Losito, F. Palmisano, T.R.I. Cataldi, Identification of isobaric lyso-phosphatidylcholines in lipid extracts of gilthead sea bream (Sparus aurata) fillets by hydrophilic interaction liquid chromatography coupled to high-resolution Fourier-transform mass spectrometry, Anal. Bioanal. Chem. 407 (2015) 6391–6404. https://doi.org/10.1007/s00216-015-8671-9.

[31] G. Ventura, et all., HILIC-ESI-FTMS and all ion fragmentation as a tool for fast lipidome characterization, Manuscr. Prep. (2020).

[32] C.D. Calvano, A.M. Sardanelli, G. Ventura, M. Glaciale, L. Savino, I. Losito, F. Palmisano, T.R.I. Cataldi, Development and use of advanced mass spectrometry techniques for the characterization of cellular and mitochondrial lipidomic profiling in control fibroblasts and

Parkinson ' s disease patients, Trends Pharm. Biomed. Anal. 1 (2018) 1–10. https://doi.org/10.15761/TPBA.1000102.

[33] C.D. Calvano, G. Ventura, A.M.M. Sardanelli, L. Savino, I. Losito, G. De Michele, F. Palmisano, T.R.I. Cataldi, Searching for potential lipid biomarkers of parkinson's disease in parkin-mutant human skin fibroblasts by HILIC-ESI-MS/MS: Preliminary findings, Int. J. Mol. Sci. 20 (2019). https://doi.org/10.3390/ijms20133341.

[34] G.J. Patti, R. Tautenhahn, G. Siuzdak, Meta-analysis of untargeted metabolomic data from multiple profiling experiments, Nat. Protoc. 7 (2012) 508–516. https://doi.org/10.1038/nprot.2011.454.

[35] G. Libiseller, M. Dvorzak, U. Kleb, E. Gander, T. Eisenberg, F. Madeo, S. Neumann, G. Trausinger, F. Sinner, T. Pieber, C. Magnes, IPO: A tool for automated optimization of XCMS parameters, BMC Bioinformatics. 16 (2015) 1–10. https://doi.org/10.1186/s12859-015-0562-8.

[36] O.E. Albóniga, O. González, R.M. Alonso, Y. Xu, R. Goodacre, Optimization of XCMS parameters for LC–MS metabolomics: an assessment of automated versus manual tuning and its effect on the final results, Metabolomics. 16 (2020) 14. https://doi.org/10.1007/s11306-020-1636-9.

[37] C. Kuhl, R. Tautenhahn, C. Böttcher, T.R. Larson, S. Neumann, CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets, Anal. Chem. 84 (2012) 283–289. https://doi.org/10.1021/ac202450g.

[38] X. Han, Factors Affecting Accurate Quantification of Lipids, in: Wiley (Ed.), Lipidomics Compr. Mass Spectrom. Lipids, John Wiley & Sons, Inc, Hoboken, NJ, USA, 2016: pp. 335–352. https://doi.org/10.1002/9781119085263.ch15.

[39] W.B. Dunn, I.D. Wilson, A.W. Nicholls, D. Broadhurst, The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans, Bioanalysis. 4 (2012) 2249–2264. https://doi.org/10.4155/bio.12.204.

[40] D. Broadhurst, R. Goodacre, S.N. Reinke, J. Kuligowski, I.D. Wilson, M.R. Lewis, W.B. Dunn, Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies, Metabolomics. 14 (2018) 72. https://doi.org/10.1007/s11306-018-1367-3.

[41] J. Bartel, J. Krumsiek, F.J. Theis, Statistical methods for the analysis of high-throughput metabolomics data, Comput. Struct. Biotechnol. J. 4 (2013) e201301009. https://doi.org/10.5936/csbj.201301009.

[42] R.G. Brereton, Chemometrics, John Wiley & Sons, Ltd, Chichester, UK, 2003. https://doi.org/10.1002/0470863242.

[43] H. V. Bohm, M.G. Stewart, Brief Report: On the Concordance Percentages for Autistic Spectrum Disorder of Twins, J. Autism Dev. Disord. 39 (2009) 806–808. https://doi.org/10.1007/s10803-008-0683-2.

[44] P.S. Gromski, H. Muhamadali, D.I. Ellis, Y. Xu, E. Correa, M.L. Turner, R. Goodacre, A tutorial review: Metabolomics and partial least squares-discriminant analysis – a marriage of convenience or a shotgun wedding, Anal. Chim. Acta. 879 (2015) 10–23. https://doi.org/10.1016/j.aca.2015.02.012.

[45] V. Chauhan, A. Chauhan, I.L. Cohen, W.T. Brown, A. Sheikh, Alteration in amino-glycerophospholipids levels in the plasma of children with autism: A potential biochemical diagnostic marker, Life Sci. 74 (2004) 1635–1643. https://doi.org/10.1016/j.lfs.2003.08.024.

[46] R.C. Murphy, Tandem Mass Spectrometry of Lipids, Royal Society of Chemistry, Cambridge, 2015. https://doi.org/10.1039/9781782626350.

[47] L.W. Sumner, A. Amberg, D. Barrett, M.H. Beale, R. Beger, C.A. Daykin, T.W.-M. Fan, O. Fiehn, R. Goodacre, J.L. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A.N. Lane, J.C. Lindon, P. Marriott, A.W. Nicholls, M.D. Reily, J.J. Thaden, M.R. Viant, Proposed minimum reporting standards for chemical analysis, Metabolomics. 3 (2007) 211–221. https://doi.org/10.1007/s11306-007-0082-2.

[48] J.D. Storey, A direct approach to false discovery rates, J. R. Stat. Soc. Ser. B (Statistical Methodol. 64 (2002) 479–498. https://doi.org/10.1111/1467-9868.00346.

[49] M.R. Wenk, The emerging field of lipidomics, Nat. Rev. Drug Discov. 4 (2005) 594–610. https://doi.org/10.1038/nrd1776.

[50] T. Harayama, H. Riezman, Understanding the diversity of membrane lipid composition, Nat. Rev. Mol. Cell Biol. 19 (2018) 281–296. https://doi.org/10.1038/nrm.2017.138.

[51] X. Han, Lipids and Lipidomics, in: Lipidomics, John Wiley & Sons, Inc, Hoboken, NJ, USA, 2016: pp. 1–20. https://doi.org/10.1002/9781119085263.ch1.

[52] C.D. Calvano, F. Palmisano, T.R. Cataldi, Understanding neurodegenerative disorders by MS-based lipidomics, Bioanalysis. 10 (2018) 787–790. https://doi.org/10.4155/bio-2018-0023.

[53] K. Yang, X. Han, Accurate Quantification of Lipid Species by Electrospray Ionization Mass Spectrometry — Meets a Key Challenge in Lipidomics, Metabolites. 1 (2011) 21–40. https://doi.org/10.3390/metabo1010021.

[54] M.M. Wiest, J.B. German, D.J. Harvey, S.M. Watkins, I. Hertz-Picciotto, Plasma fatty acid profiles in autism: A case-control study, Prostaglandins, Leukot. Essent. Fat. Acids. 80 (2009) 221–227. https://doi.org/10.1016/j.plefa.2009.01.007.

[55] S.A. Brigandi, H. Shao, S.Y. Qian, Y. Shen, B.L. Wu, J.X. Kang, Autistic children exhibit decreased levels of essential fatty acids in red blood cells, Int. J. Mol. Sci. 16 (2015) 10061–10076.

https://doi.org/10.3390/ijms160510061.

[56] A.J. Richardson, M.A. Ross, Fatty acid metabolism in neurodevelopmental disorder: a new perspective on associations between attention-deficit/hyperactivity disorder, dyslexia, dyspraxia and the autistic spectrum, Prostaglandins, Leukot. Essent. Fat. Acids. 63 (2000) 1–9. https://doi.org/10.1054/plef.2000.0184.

[57] A. El-Ansary, S. Chirumbolo, R.S. Bhat, M. Dadar, E.M. Ibrahim, G. Bjørklund, The Role of Lipidomics in Autism Spectrum Disorder, Mol. Diagn. Ther. 24 (2020) 31–48. https://doi.org/10.1007/s40291-019-00430-0.

[58] S. Vancassel, G. Durand, C. Barthélémy, B. Lejeune, J. Martineau, D. Guilloteau, C. Andrès, S. Chalon, Plasma fatty acid levels in autistic children, Prostaglandins, Leukot. Essent. Fat. Acids. 65 (2001) 1–7. https://doi.org/10.1054/plef.2001.0281.

[59] J.G. Bell, E.E. MacKinlay, J.R. Dick, D.J. MacDonald, R.M. Boyle, A.C.A. Glen, Essential fatty acids and phospholipase A2 in autistic spectrum disorders, Prostaglandins, Leukot. Essent. Fat. Acids. 71 (2004) 201–204. https://doi.org/10.1016/j.plefa.2004.03.008.

[60] J. Gordon Bell, D. Miller, D.J. MacDonald, E.E. MacKinlay, J.R. Dick, S. Cheseldine, R.M. Boyle, C. Graham, A.E. O'Hare, The fatty acid compositions of erythrocyte and plasma polar lipids in children with autism, developmental delay or typically developing controls and the effect of fish oil intake, Br. J. Nutr. 103 (2010) 1160–1167. https://doi.org/10.1017/S0007114509992881.

[61] B. Bu, P. Ashwood, D. Harvey, I.B. King, J. Van de Water, L.W. Jin, Fatty acid compositions of red blood cell phospholipids in children with autism, Prostaglandins Leukot. Essent. Fat. Acids. 74 (2006) 215–221. https://doi.org/10.1016/j.plefa.2006.02.001.

[62] A.K. El-Ansary, A.G. Ben Bacha, L.Y. Al- Ayahdi, Plasma fatty acids as diagnostic markers in autistic patients from Saudi Arabia, Lipids Health Dis. 10 (2011) 62. https://doi.org/10.1186/1476-511X-10-62.

[63] D.P. Howsmon, J.B. Adams, U. Kruger, E. Geis, E. Gehn, J. Hahn, Erythrocyte fatty acid profiles in children are not predictive of autism spectrum disorder status: a case control study, Biomark. Res. 6 (2018) 12. https://doi.org/10.1186/s40364-018-0125-z.

[64] J. Neubronner, J.P. Schuchardt, G. Kressel, M. Merkel, C. von Schacky, A. Hahn, Enhanced increase of omega-3 index in response to long-term n-3 fatty acid supplementation from triacylglycerides versus ethyl esters, Eur. J. Clin. Nutr. 65 (2011) 247–254. https://doi.org/10.1038/ejcn.2010.239.

[65] Y.M. Al-Farsi, M.I. Waly, R.C. Deth, M.M. Al-Sharbati, M. Al-Shafaee, O. Al-Farsi, M.M. Al-Khaduri, S. Al-Adawi, N.W. Hodgson, I. Gupta, A. Ouhtit, Impact of nutrition on serum levels of docosahexaenoic acid among Omani children with autism, Nutrition. 29 (2013) 1142–1146.

https://doi.org/10.1016/j.nut.2013.03.009.

[66] S. De Rubeis, J.D. Buxbaum, Genetics and genomics of autism spectrum disorder: embracing complexity, Hum. Mol. Genet. 24 (2015) R24–R31. https://doi.org/10.1093/hmg/ddv273.

[67] S. Sandin, P. Lichtenstein, R. Kuja-Halkola, H. Larsson, C.M. Hultman, A. Reichenberg, The Familial Risk of Autism, JAMA. 311 (2014) 1770. https://doi.org/10.1001/jama.2014.4144.

[68] M. Ng, J.G. de Montigny, M. Ofner, M.T. Do, Environmental factors associated with autism spectrum disorder: a scoping review for the years 2003-2013., Heal. Promot. Chronic Dis. Prev. Canada Res. Policy Pract. 37 (2017) 1–23. https://doi.org/10.24095/hpcdp.37.1.01.

[69] P.C. Calder, The relationship between the fatty acid composition of immune cells and their function, Prostaglandins, Leukot. Essent. Fat. Acids. 79 (2008) 101–108. https://doi.org/10.1016/j.plefa.2008.09.016.

745 746 747 748 749

Table 1. Summary information of patients and healthy siblings involved in the present study. In the table, each row represents siblings of the same family; age and sex (F=female, M=male) together with the degree of severity established according to Diagnostic and Statistical Manual of Mental Disorders (DSM-5) are reported.

Family	Age	Severity	Cog. Delay	Sex
03	9/5	2	Yes	M/M
04	6/4	2	No	F/M
06	9/13	1	No	M/M
07	15/17	1	No	F/F
08	15/10	1	No	M/M
09	3/6	1	No	M/M
10	15/7	1	No	M/M
11	11/8	2	Yes	F/M
12	3/8	2	Yes	M/M
13	9/15	1	Yes	M/M
14	3/7	1	No	M/M
15	10/5	2	No	F/F
16	7/3	2	No	F/M
17	9/7	2	No	M/F
18	10/8	2	Yes	M/M
19	5/3	3	Yes	F/M
20	10/4	2	No	M/M
21	5/3	3	Yes	M/M
25	12/5	2	No	F/M
25	12/8	3	Yes	F/M
26	3/5	2	No	F/M
27	4/3	3	Yes	M/F
28	15/11	2	Yes	M/M
29	9/6	2	Yes	F/M
30	6/3	1	No	M/M
31	12/6	1	No	F/M
32	10/5	1	No	M/M
33	4/3	2	No	F/M
34	10/9	1	No	F/M
35	12/5	1	No	M/M
36	3/5	3	Yes	F/M
37	8/4	1	No	F/M
38	10/12	2	No	M/M
39	5/7	3	Yes	M/M
40	5/7	1	No	F/M
41	3/8	1	No	M/M
42	9/8	3	Yes	M/F
42*	8	-	-	F
43	4/16	1	No	M/M

750 * Monozygotic twin sister

751

752

753 **Table 2.** Average confusion matrices obtained upon PLSD-DA using 1000 iterations of bootstrapping
754 resampling of full, AIF Alex^{123} and XCMS data using (*a*) the incidence of pathology or (*b*) the degree
755 of pathology as discriminant variable. No discriminatory power was obtained.

		Alex^{123} FULL MS DATA		**Alex^{123} AIF MS/MS DATA**		**XCMS**	
		a					
		Healthy	*Patient*	*Healthy*	*Patient*	*Healthy*	*Patient*
Healthy		0.48	0.52	0.47	0.53	0.50	0.50
Patient		0.48	0.52	0.48	0.52	0.55	0.45

		b											
			ASD DoS[a]				*ASD DoS*				*ASD DoS*		
		Healthy	1	2	3	Healthy	1	2	3	Healthy	1	2	3
Healthy		0.64	0.16	0.19	0.02	0.82	0.07	0.11	0.01	0.68	0.14	0.17	0.01
ASD DoS	1	0.59	0.18	0.21	0.67	0.87	0.04	0.08	0.01	0.67	0.14	0.18	0.01
	2	0.65	0.14	0.19	0.69	0.91	0.03	0.06	0.00	0.69	0.12	0.17	0.01
	3	0.66	0.14	0.18	0.72	0.89	0.04	0.07	0.00	0.72	0.12	0.15	0.01

756 [a] Degree of severity (*ASD DoS*).

757

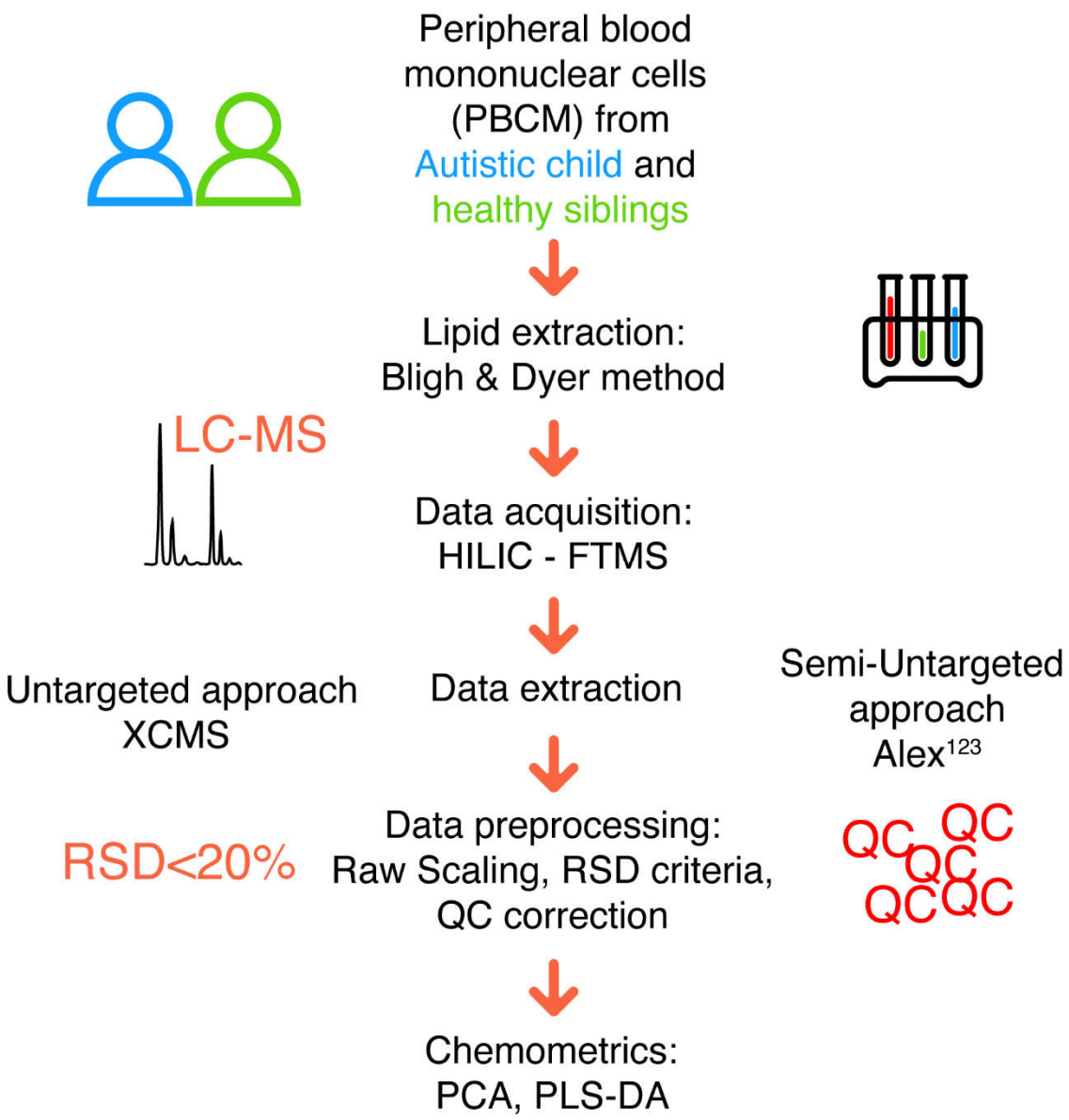
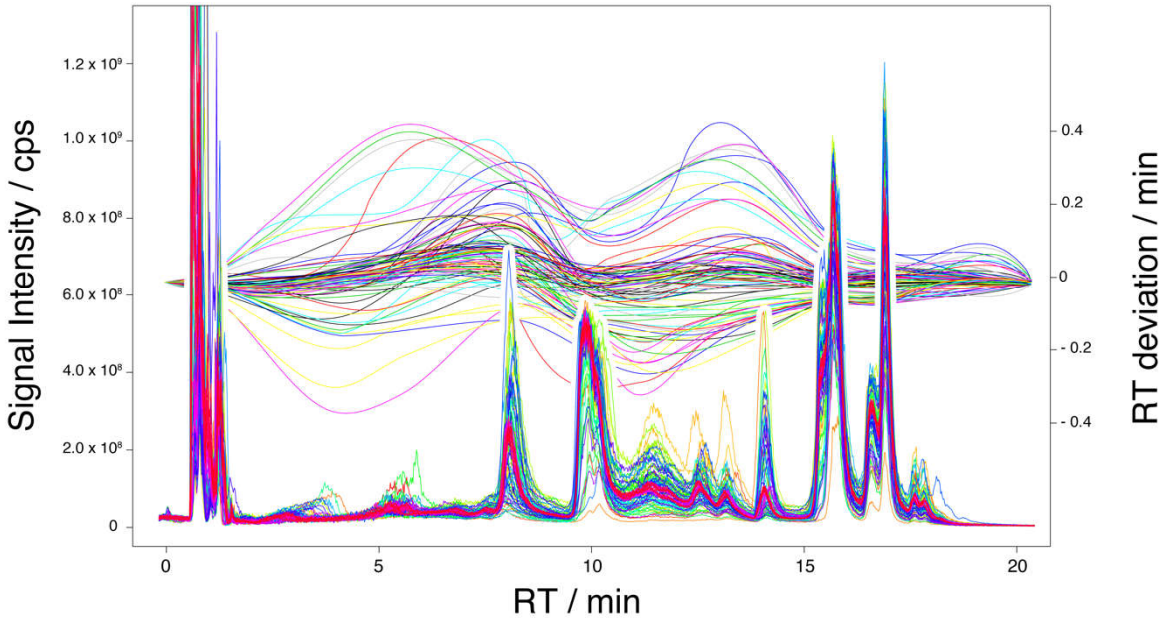


Figure 1. Adopted workflow: lipids from biological samples of PBMC have been extracted by using the procedure described by Bligh and Dyer: using a metabolomic approach, a HILIC column has been used to separate lipids according to their polar head and high resolution/accurate mass MS analysis has carried out using an ESI-Orbitrap spectrometer. XCMS and Alex[123] have been used to obtain data matrices; QC represent the quality controls and are useful for data reduction preceding the chemometrics analysis.

766



767

768 **Figure 2**. Superimposition of all the total ion current (TIC) chromatographic profiles obtained after XCMS

769 alignment procedure (y-axis 1) along with retention time deviation (y-axis 2) between runs; HILIC-ESI-MS

770 was applied for the lipid separation of 38 samples of autism spectrum disorder affected children, 36 control

771 samples (unaffected siblings of studied subjects) and 21 quality control samples (QC).

772

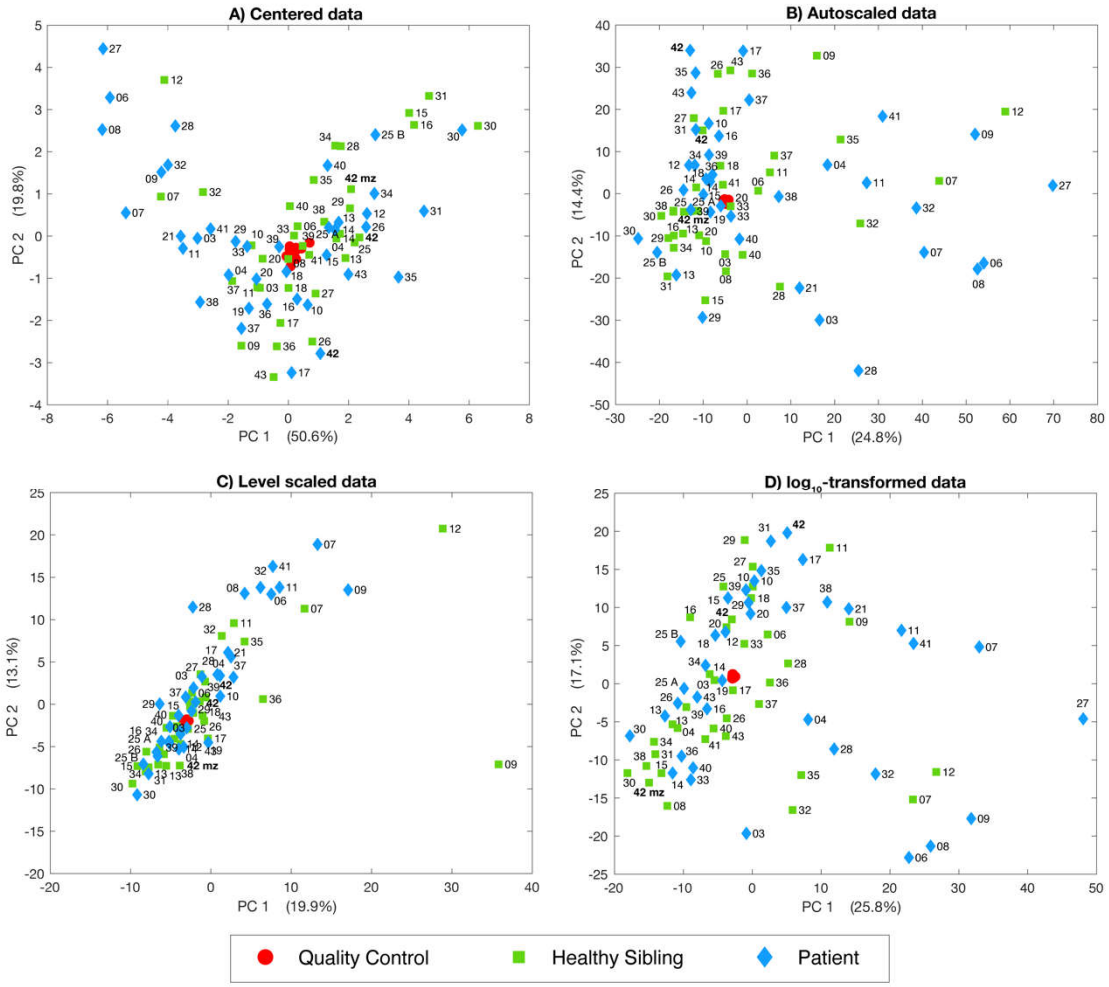
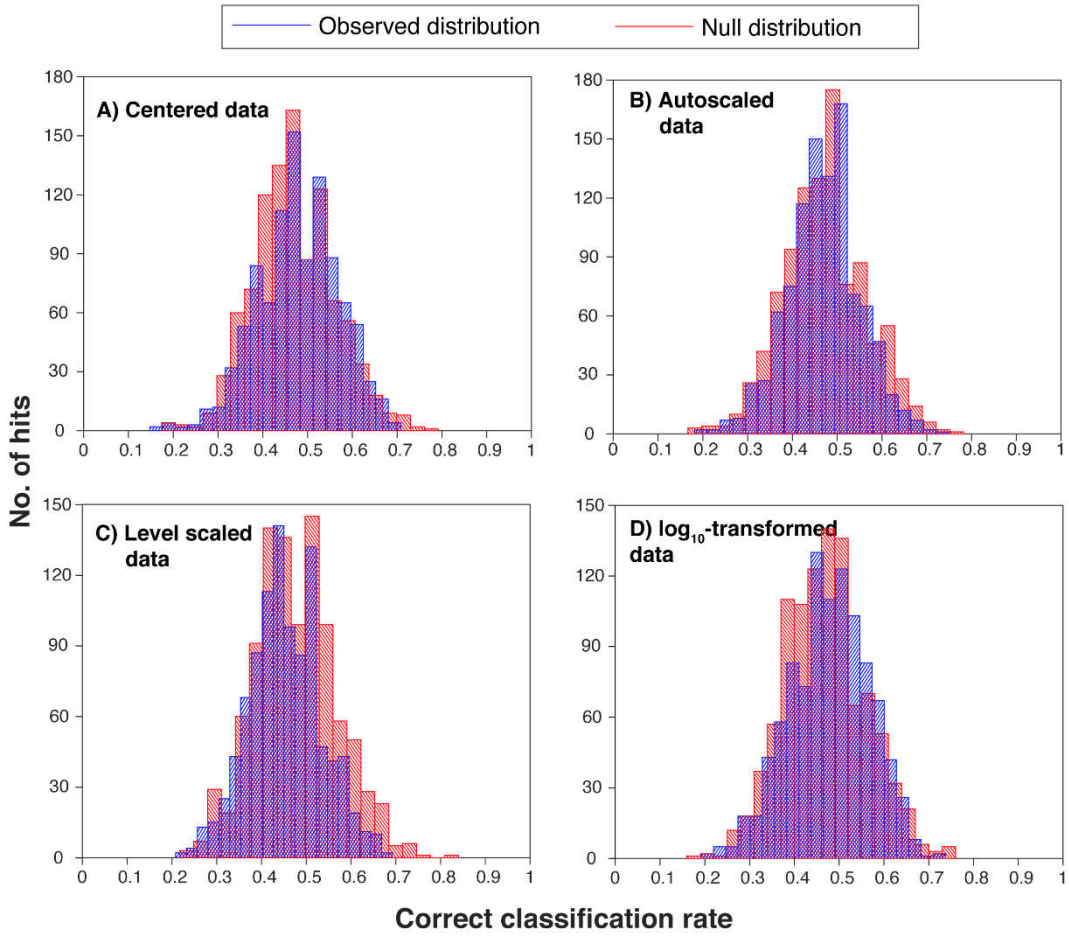


Figure 3. Comparison between Two-dimensional ordination plots of PC1 and PC2 scores from principal component analyses on data matrix obtained by using XCMS. Despite different pre-processing methods were used (i.e. (A) data centering, (B) autoscaling, (C) level scaling, (D) log_{10}-transformation) no obvious groupings/clusters or trends in the data were obtained. Note that the cluster of QC is tightly together showing that the analytical pipeline is robust.

779



780
781 **Figure 4.** PLS-DA on XCMS data matrix using pathology (yes (coded as '1') or not ('0')) as the Y-target
782 variable. Despite different pre-processing methods were used (i.e. (A) data centering, (B) autoscaling, (C)
783 level scaling, (D) $\log_{10}$-transformation), no statistically significant separation was obtained. These results
784 are from 1000 bootstraps (with replacement);blue histograms show the predictions from the 1000 test sets
785 and red histograms the null distributions from permutation testing.

786
787

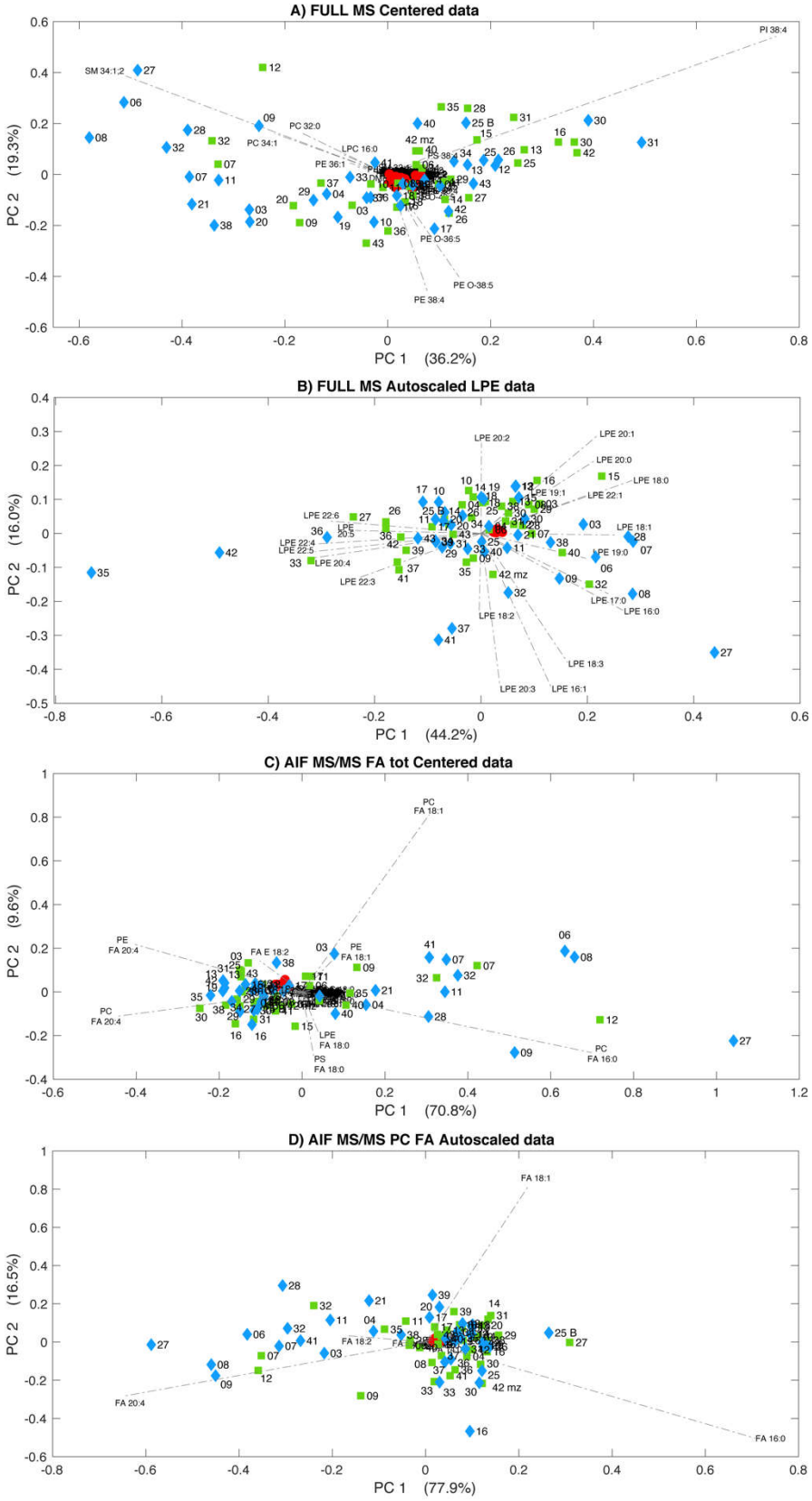


Figure 5. Principal component analysis biplot diagram obtained for (A) Alex Full MS data matrix after being mean-centered, (B) Alex LPE attributions, normalized and autoscaled; (C) Alex AIF MS/MS data matrix centered, (D) Alex AIF MS/MS data obtained under PC band. In all cases a mixed distribution has been obtained. Note that the QC cluster tightly together showing that the analytical pipeline is robust.

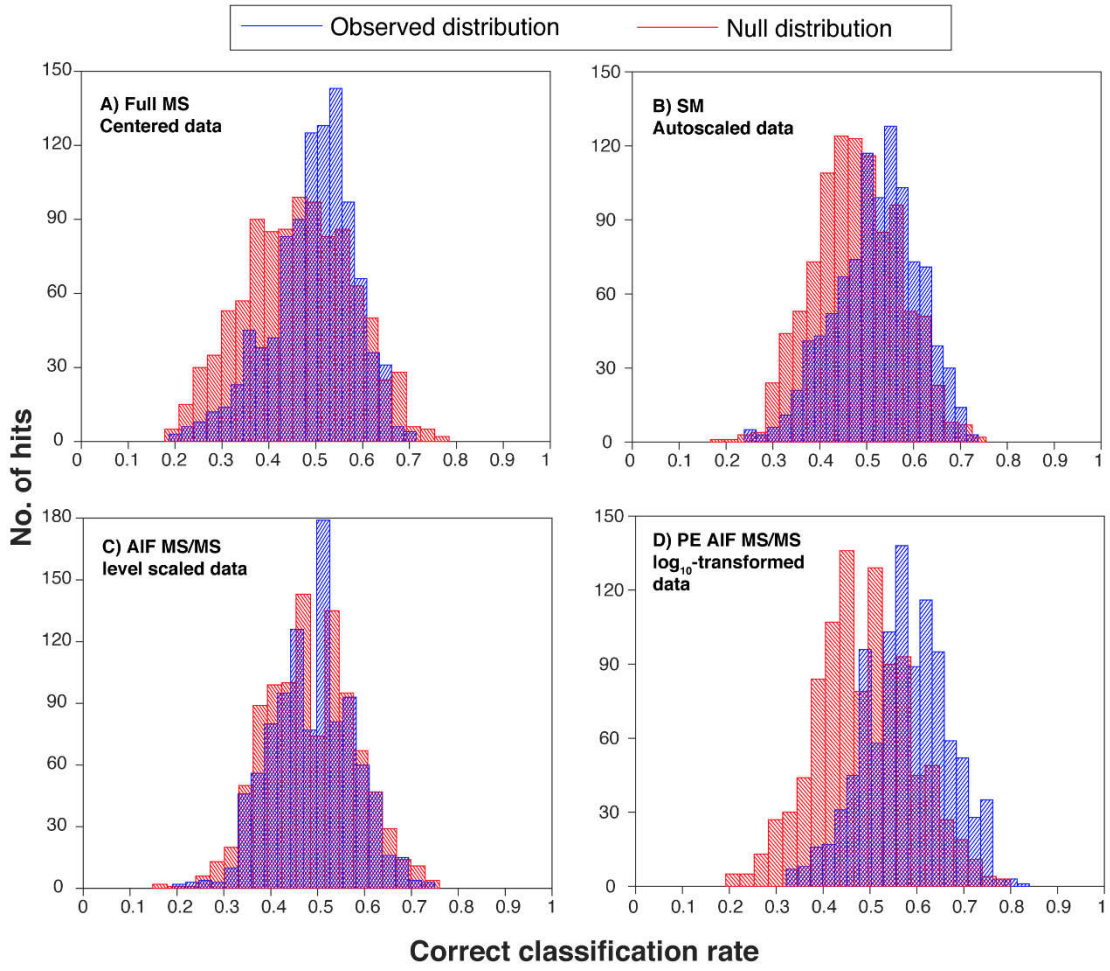


Figure 6. PLS-DA using pathology (yes or not) as target variable on Alex[123] data matrix obtained from (A) the Full MS spectrum (B) from the SM band MS spectrum (C) the whole AIF MS/MS spectrum and (D) FA signals obtained in AIF MS/MS modality under PE and PE-O band. Despite different pre-processing methods were used (i.e. (A) Data centering, (B) Autoscaling, (C) Level scaling, (D) $\log_{10}$-transformation), no statistically significant separation was obtained.

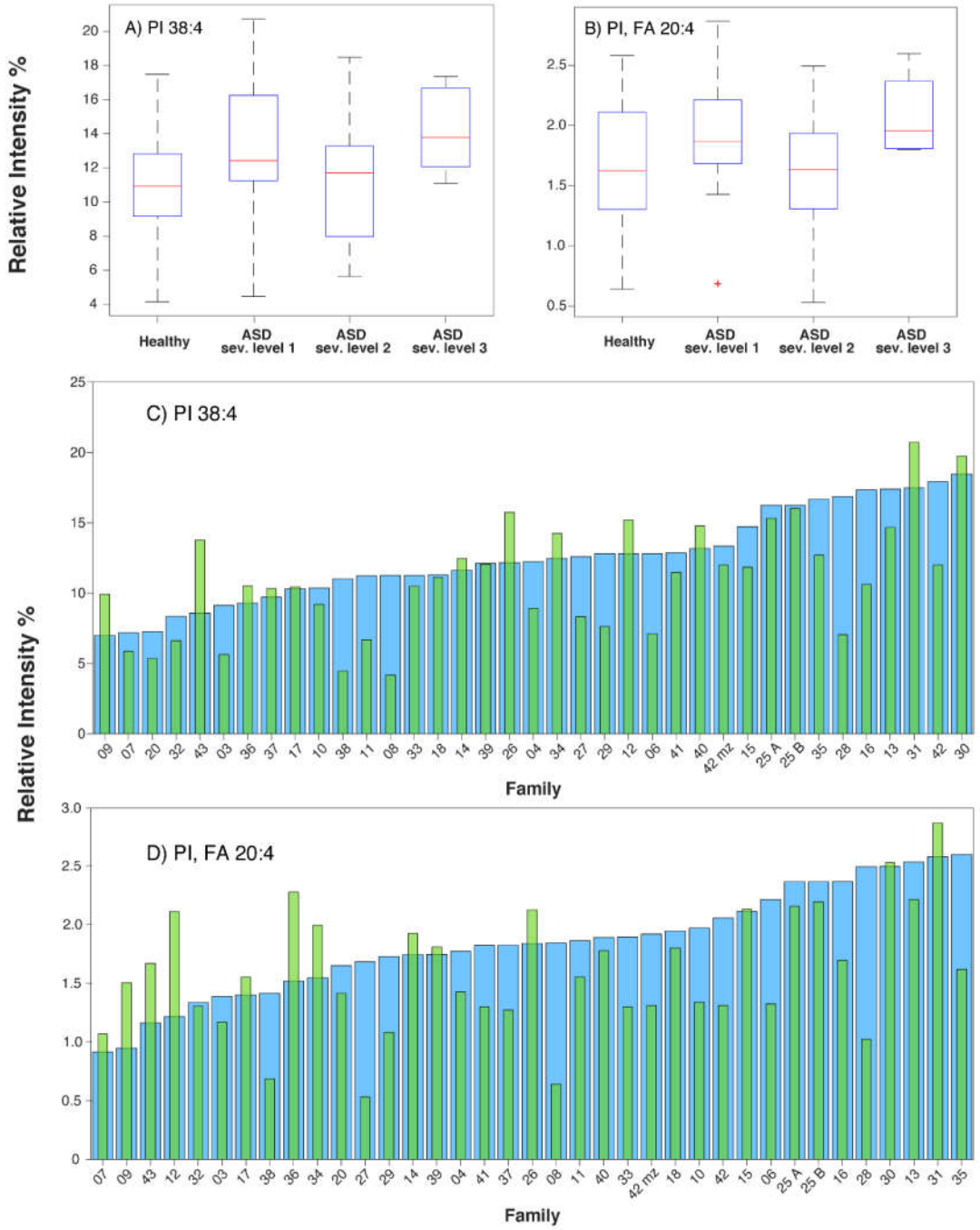


Figure 7. A) and B) Example of boxplots on two of the putative variables statistically different among healthy and ASD affected children (p value<0.05) showing a trend according to severity degree of autism obtained by using Alex^{123} software both on full MS and AIF MS/MS data (*i.e.* PI 38:4 and FA 20:4 generated from PI). C) and D) Comparison among siblings of the considered intact lipid and the corresponding FA obtained through AIF MS/MS analysis. Intrasubject variability is too large to find putative biomarkers that correlate/associate with autism.