BRAIN
ORIGINAL ARTICLE

Metabolomics identifies shared lipid pathways in independent amyotrophic lateral sclerosis cohorts

Stephen A. Goutman,[1,2,†] Kai Guo,[1,2,†] Masha G. Savelieff,[2,†] Adam Patterson,[1,2] Stacey A. Sakowski,[1,2] Hani Habra,[3] Alla Karnovsky,[3] Junguk Hur[4] and Eva L. Feldman[1,2]

[†]These authors contributed equally to this work.

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease lacking effective treatments. This is due, in part, to a complex and incompletely understood pathophysiology. To shed light, we conducted untargeted metabolomics on plasma from two independent cross-sectional ALS cohorts versus control participants to identify recurrent dysregulated metabolic pathways. Untargeted metabolomics was performed on plasma from two ALS cohorts (cohort 1, $n = 125$; cohort 2, $n = 225$) and healthy controls (cohort 1, $n = 71$; cohort 2, $n = 104$). Individual differential metabolites in ALS cases versus controls were assessed by Wilcoxon, adjusted logistic regression and partial least squares-discriminant analysis, while group lasso explored sub-pathway level differences. Adjustment parameters included age, sex and body mass index. Metabolomics pathway enrichment analysis was performed on metabolites selected using the above methods. Additionally, we conducted a sex sensitivity analysis due to sex imbalance in the cohort 2 control arm. Finally, a data-driven approach, differential network enrichment analysis (DNEA), was performed on a combined dataset to further identify important ALS metabolic pathways. Cohort 2 ALS participants were slightly older than the controls (64.0 versus 62.0 years, $P = 0.009$). Cohort 2 controls were overrepresented in females (68%, $P < 0.001$). The most concordant cohort 1 and 2 pathways centred heavily on lipid sub-pathways, including complex and signalling lipid species and metabolic intermediates. There were differences in sub-pathways that were enriched in ALS females versus males, including in lipid sub-pathways. Finally, DNEA of the merged metabolite dataset of both ALS and control cohorts identified nine significant subnetworks; three centred on lipids and two encompassed a range of sub-pathways. In our analysis, we saw consistent and important shared metabolic sub-pathways in both ALS cohorts, particularly in lipids, further supporting their importance as ALS pathomechanisms and therapeutics targets.

1 Department of Neurology, University of Michigan, Ann Arbor, MI, USA
2 NeuroNetwork for Emerging Therapies, University of Michigan, Ann Arbor, MI, USA
3 Department of Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, MI, USA
4 Department of Biomedical Sciences, University of North Dakota, Grand Forks, ND, USA

Correspondence to: Eva L. Feldman, MD, PhD
109 Zina Pitcher Place
Ann Arbor
MI 48109-2200, USA
E-mail: efeldman@umich.edu

Keywords: amyotrophic lateral sclerosis (ALS); metabolomics; lipidomics; sphingolipids; differential network enrichment analysis

Received June 29, 2021. Revised November 22, 2021. Accepted January 5, 2022. Advance access publication January 28, 2022

Introduction

Amyotrophic lateral sclerosis (ALS) is a fatal and progressive motor neuron disease,[1] which lacks effective treatments or cures. Therefore, understanding the disease mechanisms[2] is important to identifying novel therapeutic targets. ALS pathogenesis is complex and influenced by genetic,[3] epigenetic[4] and environmental[5,6] factors. An organism's metabolome derives from the cumulative effect of genetic, epigenetic, transcriptomic and proteomic forces, superimposed with environmental effects. Thus, metabolomics is a useful tool for investigating complex diseases that arise from multiple influences,[7] including ALS.[8] Moreover, the metabolome can reflect dysregulation from pathological processes, providing clues to potential treatment avenues. Specifically, in ALS, oxidative stress was identified as a disease characteristic through the observation of oxidized metabolites, e.g. nitric oxide and its toxic metabolite, peroxynitrite,[9] or oxidized lipids.[10] This line of research supported antioxidants as an ALS therapeutic, which led to clinical trials of edaravone,[11] which culminated in US Food and Drug Administration (FDA) approval, constituting one of only two drugs available for treating ALS.

To date, a handful of studies have employed untargeted metabolomics to identify differential metabolites and metabolic pathways in ALS versus control participants.[12–15] These studies uncovered multiple dysregulated pathways, including lipid,[14,16,17] amino acid,[14,18–21] and polyamine[18] metabolism. Although extremely insightful, these studies were limited to approximately 400 metabolites or less. Seeking further understanding, we recently utilized a commercial untargeted metabolomics platform of up to 3300 detectable compounds to yield insight into ALS mechanisms.[22] Our analysis included 899 metabolites, which spanned both novel and previously identified sub-pathways in ALS. Additionally, replication metabolomics studies are lacking in ALS, posing a significant roadblock, which prevents metabolomics applications in ALS from making truly meaningful advances.

In the current study, we sought to overcome this roadblock by performing a metabolomics analysis of an independent replication cohort using the same commercial platform as our initial cohort.[22] Replication cohorts are essential to understanding the reproducibility of metabolomics for identifying potential disease biomarkers for diagnostic applications[8] and pathomechanisms for drug development, which is especially pertinent for a complex, heterogeneous disease like ALS. We report the first ALS metabolomics replication study, to our knowledge, to identify potentially important, recurrent metabolic pathways and build prediction models from two independent cohorts to assess the feasibility of metabolomics for discovery in ALS. We identified shared and important pathways in the original and replication cross-sectional ALS cohorts using knowledge-based enrichment analysis, which centred on fatty acid and sphingomyelin metabolism as well as creatine and xanthine metabolism. We also combined the two datasets and implemented differential network enrichment analysis (DNEA),[23] a data-driven approach, which does not rely on known pathway annotation, allowing the method to uncover new potential metabolite correlations. Lastly, we generated prediction models leveraging metabolic data from the original cohort, which we employed to predict cases in the replication cohort. Collectively, these data confirm an association between distinct metabolite and lipid signatures in ALS and uncover new areas of research into ALS pathogenesis, biomarker identification and therapy development.

Materials and methods

Participants and biosamples

Our enrollment strategy is published.[6,22] Briefly, we recruited ALS patients older than 18 years and able to communicate in English seen at the University of Michigan Pranger ALS Clinic. Control participants were also recruited through the University of Michigan Institute for Clinical and Health Research. Participants provided their age, sex, height, and weight and underwent a clinical examination, including assessment of the ALS functional rating score–revised (ALSFRS-R) and other ALS characteristics. Participants also provided plasma samples, which were non-fasted, because it was deemed unethical to request ALS patients to fast. Studies show lack of dietary effects on the plasma lipidomics profile[24] and low intra-individual variation in non-fasted plasma[25]; thus, deep grained plasma metabolomics/lipidomics signatures of disease exist, independent of diet. We collected plasma samples from unfasted ALS participants via peripheral venipuncture, centrifuged at 2000*g* for 10 min at 4°C, aliquoted into cryovials and stored at −80°C, following good clinical practice. Samples were collected in exactly the same manner for cohort 1 and cohort 2, following a standard operating procedure conforming to the Centers for Disease Control and Prevention guidelines. Additionally, all samples were stored at −80°C, to minimize changes to sample metabolite composition.[26] Verbal and written informed consent were obtained from all participants and the study was approved by the University of Michigan institutional review board (HUM00028826).

Metabolomic profiling

Untargeted metabolomics profiling of plasma samples was performed by ultra-high performance liquid chromatography-tandem mass spectroscopy (UPLC-MS/MS) by Metabolon (Durham, NC).[27,28] Multiple recovery and internal standards were added to plasma samples for evaluating extraction efficiency and instrument performance, respectively, before the sample extraction process using methanol. Following sample extraction, metabolites were analysed by reverse-phase UPLC-MS/MS, in both positive and negative ion mode, and hydrophilic interaction chromatography UPLC-MS/MS. In addition to the spiked internal standards within each sample, a pooled 'technical replicate' generated from all study samples was periodically injected into the UPLC-MS/MS to assess instrument performance and calculate overall process and platform variability. Metabolites were identified by retention time/index, mass-to-charge ratio, and chromatographic data against authenticated standards and validated by Metabolon through data curation. Day-to-day variability was accounted for by rescaling the daily median for each metabolite to one and scaling that metabolite within each sample proportionately against the median. Missing values were replaced by the minimal value detected for that metabolite in the entire cohort, per Metabolon protocols.[27,29]

Metabolites detected in >80% of samples (missingness <20%) were included in downstream analyses. This differed from our recent publication of the original cohort,[22] which was more focused on the discovery of ALS mechanisms and hypothesis development, and included metabolites detected in >60% of samples (missingness <40%). In this present analysis, we were more interested in rigorous test reliability and thus employed more stringent missingness criteria, reanalysing the original cohort 1 and analysing the replication cohort 2 with the missingness cutoff <20%. Missingness was generally low (Supplementary Tables 1 and 2), and most metabolites were detected in 99.5–100% of samples

(Supplementary Fig. 1). Thus, despite the more stringent criteria of metabolites with missingness <20% in this analysis, the overlap with metabolites with missingness <40% from the previous analyses was high and similar metabolites were selected using the two missingness cutoffs (Supplementary Fig. 2).

Differential metabolite identification

Analytical methods

The statistical analysis plan in this report follows our recent publication,[22] but draws upon cases and controls from both the original cohort (cohort 1 ALS cases, cohort 1 controls)[22] and a new second replication cohort (cohort 2 ALS cases, cohort 2 controls). Thus, ALS case and control demographics were summarized and compared using chi-square and Wilcoxon rank-sum tests, as appropriate, across these four groups. Some missing demographic data from the initial publication have since been obtained, accounting for slight differences in the previous and current reported medians for cohort 1. Metabolite missingness was summarized across the original and replication datasets. As in our prior publication, we performed a series of case and control metabolite analyses and then compared the overlap in selected metabolites (Supplementary Fig. 3). An 'unadjusted' model used the non-parametric Wilcoxon rank-sum test to compare non-normally distributed metabolites with Benjamini–Hochberg *P*-value correction for multiple comparisons. An 'adjusted' model used logistic regression, adjusted for age, sex and body mass index (BMI) and regressed each natural log-transformed and standardized metabolite against case/control status. Participants were dropped from the analysis if they were missing a BMI value. *P*-values were adjusted for multiple comparisons using Benjamini–Hochberg correction.

Partial least squares-discriminant analysis (PLS-DA) was performed using the R package mixOmics[30] separately on the original and replication case and control cohorts (Supplementary Fig. 3). Differences in metabolites between cases versus controls were visualized with score plots of the variable importance in projection (VIP).[31,32] Metabolites with significant contributions to group separation had VIP > 1. The 10-fold cross-validation to select the tuning parameter for the PLS-DA analysis is shown in Supplementary Fig. 4A and B. To assess similarities in sub-pathways selected in the original and replication cohorts, we again used group lasso, adjusted for age, sex and BMI, using the gglasso R package with natural log-transformed and standardized metabolite data (Supplementary Fig. 3). Five-fold cross-validation was used to select the tuning parameter corresponding to a sparse model within 1 standard error (SE) of the minimum cross-validation error. Once the tuning parameter, corresponding to the group lasso penalty was finalized, group lasso was refit to the full dataset to obtain the final model. The five-fold cross-validation to select the tuning parameter for the group lasso analysis is shown in Supplementary Fig. 4C and D.

Overlapping metabolites and sub-pathways selected by each model (Wilcoxon, logistic regression, PLS-DA, group lasso) from each original cohort 1 and replication cohort 2 were represented in Venn diagrams.

Metabolite correlation with ALSFRS-R

Heatmaps were generated from the relative abundance of the 20 top differential metabolites in ALS versus controls, shared by both cohort 1 and cohort 2, as a function of the ALSFRS-R score at the time of plasma collection. Relative abundances were scaled by row. ALS participants were sorted by $\log_2$(ALSFRS-R), from high to low score.

Construction of case prediction models

To examine the feasibility of predicting metabolite-based ALS cases, machine learning classification models were constructed using PLS-DA, group lasso and random forest (RF). Prediction accuracy was calculated by the area under the curve (AUC) for each model, which were visualized through receiver operating characteristic (ROC) curves generated by the R package pROC.

Metabolism pathway analysis

We used the R package richR (https://github.com/hurlab/richR/) for pathway enrichment analysis (Supplementary Fig. 3). Sub-pathways were annotated by Metabolon. Over-represented sub-pathways were determined from the metabolites selected by unadjusted Wilcoxon, adjusted logistic regression, PLS-DA and group lasso models. A hypergeometric test was performed for each candidate sub-pathway. Sub-pathways with a *P*-value <0.05 were deemed significantly enriched.

Sex sensitivity analysis

To assess the impact of the sex imbalance, we performed a sex sensitivity analysis. Briefly, the original datasets were separated based on sex, each of which was analysed for differential metabolites and enriched pathways between case and control using the above analyses. Then, the results were compared between male and female at the metabolite- and pathway-levels.

Statistical software

All statistical and prediction analyses were performed using R statistical computing software.

Differential network enrichment analysis

To further understand the metabolic alterations underlying ALS, we merged the two datasets (cohort 1 and 2) and analysed them using the data-driven DNEA approach.

Metabolite selection and data treatment

The cohort 1 and 2 datasets, containing the same metabolite compound identifiers, were merged, numbering 954 total metabolites. Drug-related metabolites and metabolites with missingness <20% in either or both datasets were excluded, leaving 640 total metabolites. Missing values for these metabolites were imputed using each metabolite's minimum value. All measurements were subsequently log-transformed.

Covariate adjustment, autoscaling and dataset merging

We assessed batch effects within and between runs from both cohorts using principal components analysis (PCA). For each dataset separately, metabolites were linearly adjusted for age, sex and BMI. Out of 525 participants, only 15 had missing BMI values. Several imputation methods to approximate BMI values were tested, including simple (mean/median imputation) and machine-learning (linear regression, random forest, support vector regression) approaches. A linear model using the abundance of a select number of metabolites that correlated with BMI generated the best results. Thus, BMI values were imputed using linear models

based on the top 13 and 23 metabolites from cohort 1 and cohort 2, respectively, which correlated most strongly with BMI. After adjustment, each cohort 1 and cohort 2 dataset was separately autoscaled (mean centred and scaled by the standard deviation) to obtain an N(0,1) distribution for each metabolite. The adjusted and normalized datasets from the two original and replication cohorts were finally merged into a single dataset. PCA analysis demonstrates elimination of batch effects (Supplementary Fig. 5).

Differential network enrichment analysis method

We employed DNEA to identify metabolite subnetworks that differentiate ALS from control samples. DNEA methodology has been described previously.[23,33] Briefly, DNEA computes a partial correlation network using metabolomics data from two conditions (i.e. ALS and control) jointly. Due to an imbalance in the number of samples in ALS versus control groups, we employed a subsampling procedure coupled with partial correlation network estimation to obtain robust network edges, as described in Iyer *et al.*[23] The network was then clustered into densely-connected metabolite subnetworks. Next, we performed an enrichment analysis using the NetGSA method,[34] which takes into account differential metabolite abundances as well as the differences in network structure between cases and controls. DNEA results consist of computed subnetworks and their respective *P*-values and false discovery rate (FDR)-adjusted *q*-values, which correspond to significant subnetwork differences between cases and controls. We applied DNEA to the full dataset as opposed to each dataset separately to have a sufficient sample size, *n*, compared with the metabolite count, p, for the partial correlation network computation.

Data availability

Anonymized data will be shared by request from any qualified investigator.

Results

Original and replication ALS cohorts reflect typical ALS populations

The original cohort included 125 ALS and 71 control participants, and the replication cohort included 225 ALS and 104 control participants (Table 1). Age at plasma collection differed between groups overall ($P = 0.013$); replication ALS cases were slightly older than replication controls ($P = 0.009$). Replication ALS cases also had a lower proportion of females versus replication controls. There were no statistically significant differences in BMI or race between the four groups. There were no differences between the original and replication ALS groups in ALS family history, age at diagnosis, El Escorial criteria, onset segment, ALSFRS-R at plasma collection, time between symptom onset and diagnosis, time between diagnosis and blood draw and percentage of participants with feeding tubes. The original and replication ALS cohorts both reflect a typical ALS population, with median diagnostic ages of 62.0 and 63.0 years, intervals between symptom onset and diagnosis of 1.01 and 1.07 years and onset segment of bulbar 30.4 and 26.7%, cervical 30.4 and 38.8% and lumbar 39.2 and 34.9% for the original and replication cohorts, respectively. Plasma was collected within [median (25th–75th percentile)] 0.57 (0.36–0.75) years of diagnosis in the original cohort 1 and within 0.60 (0.36–1.12) years of diagnosis in the replication cohort 2.

Metabolite profiling in original and replication cohorts

In the original cohort 1, we identified and evaluated 1051 known metabolites by descriptive statistics by case/control status (Supplementary Table 1), of which 258 had missingness >20% and were removed from further analyses. In the replication cohort 2, we identified and evaluated 1019 known metabolites by descriptive statistics by case/control status (Supplementary Table 2), of which 315 had missingness >20% and were removed from further analyses. Metabolites identified in the original but not the replication cohort are listed in Supplementary Table 3. Conversely, metabolites unique to the replication cohort are listed in Supplementary Table 4. Metabolites were matched by ChemID, since Metabolon changed some metabolite names between the two cohorts. In our prior publication,[22] very few drug metabolites satisfied missingness criteria. Those that did, such as riluzole, only correlated weakly with other measured metabolites and therefore exerted a minimal impact on metabolite associations. Furthermore, another study demonstrated there were no differences in metabolites identified after washing out typical ALS drugs.[35] Thus, we further excluded any drug metabolites, which will not unduly bias case/control associations. A total of 793 and 703 metabolites for the cohorts 1 and 2, respectively, were used for differential metabolite and pathway analysis.

Differential metabolites in ALS cases versus controls

Wilcoxon identified 268 and 335 significant differential metabolites in the original and replication cohorts, respectively (adjusted *P*-value <0.05), visualized by volcano plot (Supplementary Fig. 6). Of these, 101 were unique to the original cohort and 168 to the replication cohort and 167 were shared (Supplementary Fig. 7A). Next, logistic regression, adjusted for age, sex and BMI, identified 289 and 317 metabolites in the original and replication cohorts, respectively, presented in Manhattan plots at the sub-pathway level (Supplementary Fig. 8). Of these, 99 were unique to the original cohort, 127 to the replication cohort, and 190 were shared (Supplementary Fig. 7B). There were differences among metabolites selected by Wilcoxon versus age-, sex- and BMI-adjusted logistic regression (Supplementary Fig. 9), suggesting that clinically important variables in ALS, such as age, sex and BMI, may significantly affect metabolite relationships between cases versus controls.

PLS-DA identified 275 and 230 metabolites in the original and replication cohorts (Supplementary Fig. 7C), respectively, with VIP >1 that separated cases from controls (Fig. 1A and B). Of these, 136 were unique to the original cohort and 91 to the replication cohort, whereas 139 were shared by both cohorts (Supplementary Fig. 7C). The top 50 metabolites, with the highest VIP and largest contribution to case/control separation, are presented in a VIP score plot (Fig. 1C and D). Among these top 50 metabolites, 19 were common between the original and replication cohorts. As previously reported,[22] we also performed group lasso because it considers sub-pathway structure to evaluate significant ALS and control metabolite differences. Age-, sex- and BMI-adjusted group lasso identified 251 and 299 differential metabolites [odds ratio (OR) $\neq 1$] in the original and replication cohorts, respectively (Supplementary Fig. 7D), which represented 35 and 42 sub-pathways, respectively. Of these metabolites, 127 were unique to the original cohort and 175 to the replication cohort, whereas 124 were shared (Supplementary Fig. 7D). Heatmaps visualize relative abundance

Table 1 Participant demographics

Covariate	Original Cohort 1		Replication Cohort 2		P-value
	ALS cases ($n = 125$)	Controls ($n = 71$)	ALS cases ($n = 225$)	Controls ($n = 104$)	
Age at plasma collection, years[a]	63.0 (53.3–69.6)	60.5 (53.4–64.8)	64.8 (56.7–71.3)	62.0 (55.7–67.5)	0.013
Sex[a]					<0.001
Female	51 (40.8)	29 (39.4)	101 (45)	71 (68.0)	
Male	74 (59.2)	42 (60.6)	124 (55)	33 (32.0)	
BMI at study entry, kg/m^2[b]	25.8 (23.0–29.7)	27.3 (24.6–31.1)	26.0 (22.6–30.2)	27.3 (24.3–30.4)	0.051
Race[a]					
Asian	0 (0)	0 (0)	3 (1.3)	3 (2.9)	
Black or African American	2 (1.6)	2 (2.8)	5 (2.2)	6 (5.8)	
Not reported				1 (0.9)	
White or Caucasian	123 (98.4)	69 (97.1)	217 (96.5)	94 (90.4)	
Family history of ALS					0.68
No	112 (89.6)		195 (86.7)		
Yes	10 (8.0)		25 (11.2)		
Unknown	2 (1.6)		4 (1.8)		
Missing	1 (0.8)		1 (0.4)		
ALSFRS-R at plasma collection	33 (27–37)		31 (25–36)		0.11
Age at diagnosis, years[a]	62.2 (52.7–68.7)		63.8 (54.6–70.7)		0.24
El Escorial criteria[a]					0.17
Suspected	3 (2.4)		13 (5.8)		
Possible	19 (15.2)		30 (13.0)		
Probable, LS	37 (29.6)		59 (26.3)		
Probable	42 (33.6)		60 (26.8)		
Definite	24 (19.2)		63 (28.1)		
Onset segment[a]					0.30
Bulbar	38 (30.4)		60 (26.7)		
Cervical	38 (30.4)		87 (38.8)		
Lumbar	49 (39.2)		78 (34.9)		
Time between diagnosis and blood draw, years[a]	0.57 (0.36–0.75)		0.61 (0.36–1.13)		0.33
Time between symptom onset and diagnosis, years[a]	1.01 (0.68–1.51)		1.07 (0.70–1.84)		0.53
PEG tube present, %	8 (6.4)		27 (12.0)		0.56

Table of descriptive statistics for the overall participant study population. Continuous variables represented as the median (25th–75th percentile) and for categorical variables as *n* (%). *P*-values correspond to Wilcoxon rank-sum test for continuous variables and chi-squared test for categorical variables. LS = lab supported; NA = not available; PEG = percutaneous endoscopic gastrostomy.

[a]Median, 25th percentile and 75th percentile are computed using all cases and controls (no missing subjects).

[b]Median, 25th percentile and 75th percentile are computed using 119 original cohort cases, 67 original cohort controls (six cases and four controls are missing), 218 replication cohort cases, 84 replication cohort controls (seven cases and three controls are missing).

differences between ALS versus controls for all group lasso metabolites in the original and replication cohorts (Supplementary Fig. 10). Among the top 50 metabolites, 14 were common between the original and replication cohorts.

Finally, Venn diagrams illustrate the unique number of metabolites for each analysis and for each cohort (Supplementary Fig. 9).

Differential metabolites correlate with ALSFRS-R

To assess whether metabolites correlate with clinical status, we plotted heatmaps of the relative abundance of the 20 top differential metabolites in ALS versus controls, shared by both cohort 1 and cohort 2, as a function of the ALSFRS-R score at the time of plasma collection (Fig. 2). When we sorted ALS participants by ALSFRS-R from high to low score, this generated an overall trend of decreasing metabolites with worsening ALS status. Interestingly, and aligned with the significant sub-pathways, several sphingolipids were among the metabolites that correlated with ALSFRS-R, such as lignoceroyl sphingomyelin (d18:1/24:0), sphingomyelin (d18:1/20:0, d16:1/22:0) and sphingomyelin (d18:1/14:0, d16:1/16:0), as well as acylcarnitines [e.g. lignoceroylcarnitine (C24)]. However, although sphingomyelin abundance is elevated in ALS cases versus participants, these specific sphingolipid species were lower in patients with more advanced disease.

Case prediction models

We leveraged our two independent cohorts to construct case prediction models. The original cohort 1 dataset was used to build PLS-DA, group lasso and RF prediction models to identify ALS cases in the replication cohort 2. Prediction accuracy was calculated by AUCs and visualized by ROC curves (Fig. 3). Metabolites that contributed to models are outlined in Supplementary Tables 5–7. Group lasso and PLS-DA had very similar AUCs (0.945 and 0.944, respectively), which were slightly higher versus RF (AUC of 0.903). As anticipated, creatine and creatinine were among the most strongly contributing metabolites to all three models, likely secondary to muscle loss and not to the causative disease process. Thus, they may not be specific to ALS.

Metabolites related to antioxidant defense and polyamine metabolism were high on the list in the group lasso model, amino acid metabolism in the PLS-DA model and xenobiotics and amino acids in the RF model. Additional strongly influential metabolites (top 10) include sphinganine and sphingadienine (members of 'Sphingolipid Synthesis') in the group lasso prediction model,

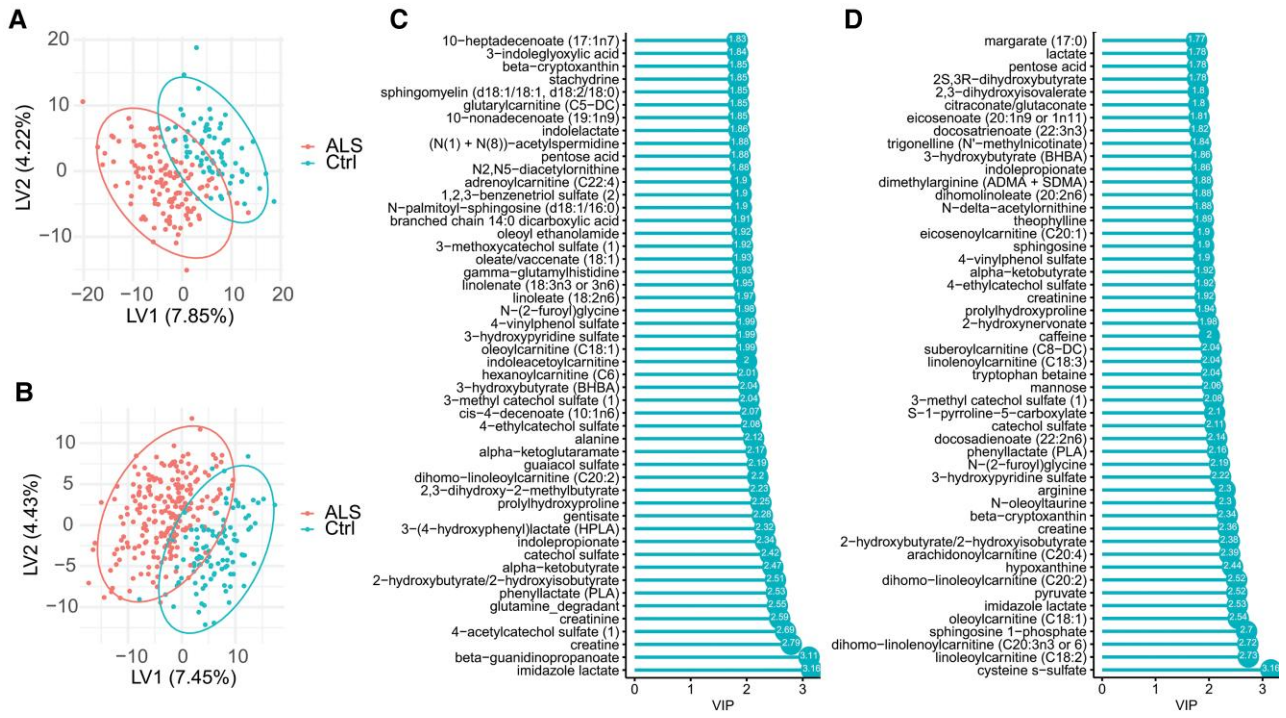

Figure 1 PLS-DA analysis of ALS cases versus controls for original and replication cohorts. (**A** and **B**) PLS-DA score plot of ALS cases (red) versus controls (blue) for (**A**) original cohort 1, and (**B**) replication cohort 2, individually; each dot represents an individual participant. (**C** and **D**) The VIP score plot of the top 50 PLS-DA metabolites, which most significantly separate ALS cases from controls for (**C**) original cohort 1 and (**D**) replication cohort 2, individually. A total of 275 (original) and 230 (replication) metabolites had VIP >1. Among the top 50 PLS-DA metabolites, 19 were shared between the original and replication cohorts. LV = latent variables.

which were less discriminating in the PLS-DA and RF models. Nevertheless, several sphingomyelins, e.g. sphingomyelin (d18:1/18:1, d18:2/18:0) and sphingomyelin (d18:2/16:0, d18:1/16:1), were in the top 100 metabolites of the PLS-DA and RF models. Thus, overall, all models performed well and could predict ALS cases with good accuracy. This supports metabolomics as a feasible approach for ALS biomarker discovery, although future studies will need to include disease mimics to assess specificity.

Differential sub-pathways in ALS cases versus controls

As before,[22] we next performed pathway enrichment analysis to identify over-represented sub-pathways based on differential metabolites identified by the various analysis methods. In the original cohort, Wilcoxon, adjusted logistic regression, PLS-DA and group lasso enriched 12, 13, 12 and 23 sub-pathways, respectively (Fig. 4). For the replication cohort, Wilcoxon, adjusted logistic regression, PLS-DA and group lasso enriched 16, 13, 14 and 25 sub-pathways, respectively (Fig. 4). There were few fully concordant sub-pathways that were selected across all four analytical methods in both the original and replication cohorts. However, the concordant sub-pathways selected by multiple methods in both cohorts centred heavily on lipid sub-pathways, including 'long chain saturated fatty acid', 'long chain polyunsaturated fatty acid (n3 and n6)', 'long chain monounsaturated fatty acid', 'fatty acid metabolism (acyl carnitine, polyunsaturated)' (Supplementary Fig. 11) and 'sphingomyelins'. With few exceptions, acylcarnitines, which are partially metabolized intermediates, of all chain lengths and saturation level were elevated in ALS cases versus controls. 'Xanthine' and 'creatine' were also repeatedly selected.

There were sub-pathways that were highly selected by multiple analytical methods but were mostly consigned to one cohort or the other. For instance, lipid 'ceramides', 'hexosylceramides', antioxidant 'gamma-glutamyl amino acid' and xenobiotics 'benzoate' were widely selected in the original cohort, whereas lipid 'sphingosines', 'fatty acid metabolism (acyl carnitine, hydroxyl)', 'fatty acid metabolism (acyl carnitine, dicarboxylate)' and energy 'TCA (tricarboxylic acid) cycle' were selected in the replication cohort. There was overlap in some of these sub-pathways between the two cohorts. The most significant sub-pathway was 'sphingomyelins' by group lasso of the original cohort, which was also selected in the replication cohort.

Sensitivity analysis for sex imbalances

The original cohort 1 was well-balanced for sex; however, females were overrepresented in the control group of the replication cohort 2, which led to differences in the sex composition of the original and replication cohort. Therefore, we conducted a sensitivity analysis to evaluate whether the imbalance contributed to uncertainty in the outcomes. Moreover, we[36,37] and others[38] have noted that sex is an important clinical factor in ALS; thus, analysis by sex may yield additional insight. We repeated the Wilcoxon, logistic regression, PLS-DA and group lasso analyses in male cohort 1 ($n = 74$ ALS, $n = 42$ control) and female cohort 1 ($n = 51$ ALS, $n = 29$) separately, which were compared to the combined original cohort 1, and additionally in male cohort 2 ($n = 123$ ALS, $n = 33$ control) and female cohort 2 ($n = 101$ ALS, $n = 70$ control) separately, which were compared to the combined replication cohort 2 (Fig. 5 and Supplementary Fig. 12).

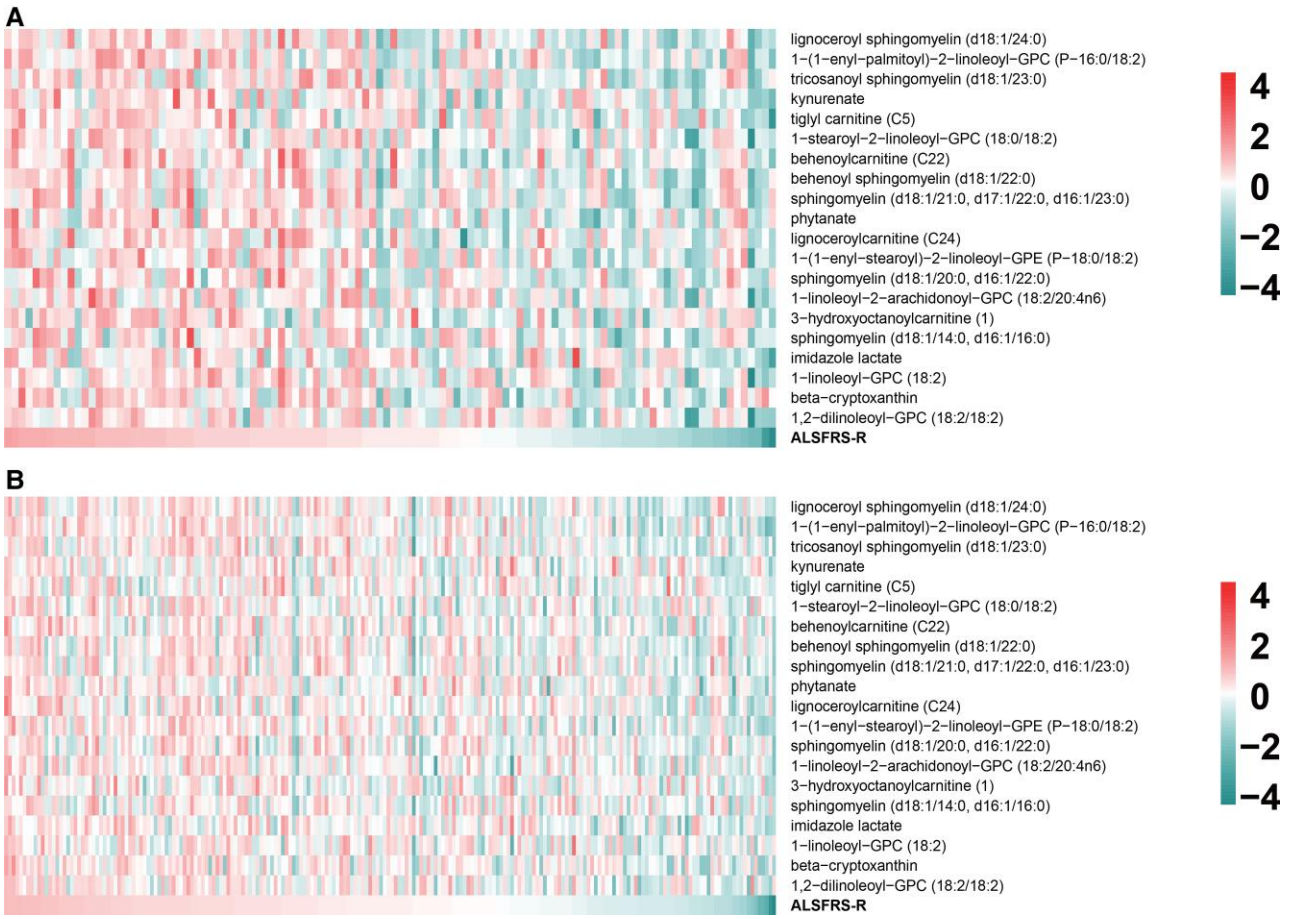

Figure 2 Correlation of metabolite abundance with ALSFRS-R. Cross-sectional heatmap visualization of the 20 top differential metabolites in ALS versus controls, shared by both cohort 1 and cohort 2, as a function of ALSFRS-R score at the time of plasma withdrawal. The relative abundance of the metabolites significantly correlated with ALSFRS-R in both (**A**) cohort 1 and (**B**) cohort 2. Relative abundances and $\log_2$(ALSFRS-R) were scaled by row. ALS participants were sorted by $\log_2$(ALSFRS-R), from high score (*left*, pink) to low score (*right*, green), i.e. progressive disease, which results in an overall trend of decreased metabolites.

Some interesting observations arose from the sex sensitivity analysis. Although 'fatty acid metabolism (acyl carnitine, polyunsaturated)' was almost invariably selected by all analysis methods, in both sexes, certain lipid focused sub-pathways were selected more frequently and/or significantly in ALS females, either in the original or replication cohorts, such as 'long chain saturated fatty acid', 'long chain polyunsaturated fatty acid (n3 and n6)', 'long chain monounsaturated fatty acid', 'sphingosines' and 'diacylglycerol'. One exception stood out in 'sphingomyelins', which was more often and/or significantly selected in ALS males from either the original or replication cohorts. These sex-dependent lipid differences, especially in sphingomyelins, were present irrespective of health or disease, when we analysed the control and ALS groups separately by male versus female, within both cohort 1 and cohort 2 (Supplementary Fig. 13). We found overall sphingomyelin abundance was lower in healthy males versus females,[39,40] which persisted in ALS, albeit to a lesser extent, presumably due to dysregulated sphingomyelin metabolism in ALS (Supplementary Fig. 14 and Supplementary Tables 8 and 9).

Another interesting finding was 'benzoate metabolism', which was selected in our original but not our replication cohort (Fig. 4). Examination of the sex analysis suggests this might arise due to sex imbalance in the replication cohort. When the original cohort was stratified by sex, benzoate metabolism was most frequently and/or significantly selected in ALS males (Fig. 5). Therefore, benzoate metabolism may not have been selected in the replication cohort because it was underrepresented in control males.

Differential network enrichment analysis

We employed DNEA to identify metabolite subnetworks that differentiate ALS from control samples. DNEA is a data-driven, functional enrichment approach, which facilitates discovery of novel, functionally active metabolite modules, without depending on prior knowledge of biochemical interactions. Additionally, it identifies biochemical interactions that are present in either healthy or ALS samples, or present in both. DNEA is especially useful for metabolites with incomplete pathway knowledge, e.g. lipids and exogenous metabolites. DNEA also eliminates batch effects between the two ALS cohorts ($n = 349$) and control cohorts ($n = 174$) by adjusting and autoscaling data. Once the datasets were merged, DNEA analysis identified a total of 15 different metabolite subnetworks, nine of which were significantly enriched at the 0.01 FDR cutoff (Supplementary Tables 10 and 11, Fig. 6 and Supplementary Figs 15–23).

The most significant subnetwork S1 comprised candidates from a variety of sub-pathways, encompassing 'benzoate metabolism' and 'food component/plant', both related to xenobiotics as well as

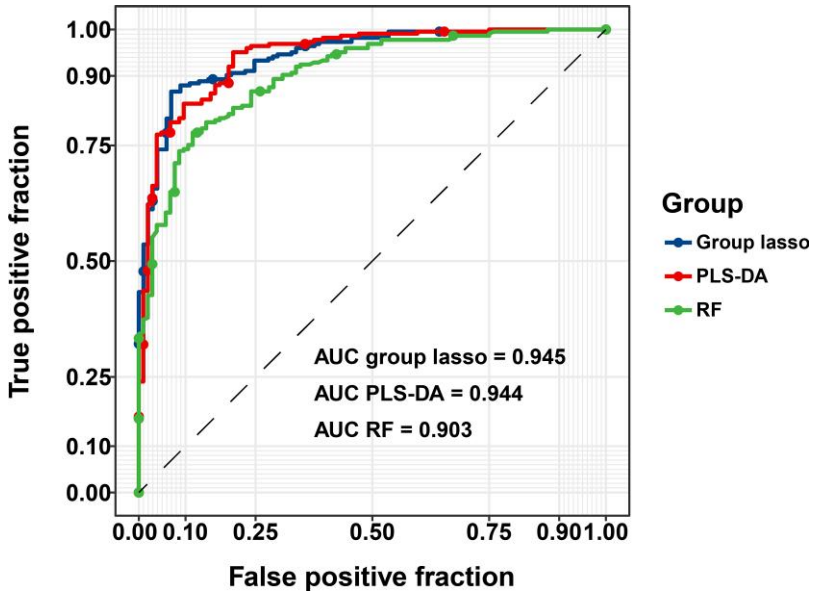

Figure 3 Case prediction models. ROC curves for ALS case prediction models generated by group lasso, PLS-DA and RF from cohort 1 applied to cohort 2. Prediction accuracy was calculated for each model by the AUC.

fatty acid and amino acid metabolites (Supplementary Fig. 15). The second most significant subnetwork S2 was primarily contributed by xenobiotics sub-pathways, 'benzoate metabolism' and 'xanthine metabolism' (Supplementary Fig. 16). Regarding lipid-centric transformations, long-chain and various fatty acid and complex lipid metabolic sub-pathways contributed the most metabolites to subnetworks S3, S8 and S9 (Fig. 6 and Supplementary Figs 17, 22 and 23). These subnetworks contained more ALS edges, indicating biochemical interactions that were present in the pathological condition. Subnetwork S8 contained a large number of complex lipids, mostly sphingomyelins, but also ceramides and dihydrosphingomyelins. Subnetwork S9 encompassed multiple species in signalling, e.g. diacylglycerols and complex lipid sub-pathways, e.g. phosphatidylcholines, phosphatidylethanolamines, phosphatidylinositols and lysophospholipids.

In contrast to these more focused aforementioned subnetworks, subnetworks S5 (Supplementary Fig. 19) and S7 (Supplementary Fig. 21) converged on metabolites from diverse super-pathways. Subnetwork S7 was particularly interesting, embodying several metabolic ALS characteristics, e.g. energy (TCA cycle), amino acid, antioxidants and creatine. Purine/pyrimidine[41] and xenobiotics metabolism were also featured. Subnetwork S5 similarly had a diverse profile.

Discussion

In the current study, we compared the metabolome of a second new ALS and control cohort to our previously published original cohort 1 ($n = 125$ ALS, $n = 71$ control), where we reported multiple ALS-associated metabolites and pathways.[22] Our goal was to identity recurrent dysregulated metabolites and pathways in ALS. The original investigation was a hypothesis-generating study; thus, it employed a less stringent missingness criteria of 60% cutoff.[22] Our replication effort reanalysing cohort 1 and analysing new cohort 2 ($n = 225$ ALS, $n = 104$ control) employed a more rigorous missingness criteria of 80% cutoff to identify the metabolites and pathways most strongly associated and replicated in ALS. We leveraged the two independent cohorts to examine recurrent metabolites and pathways, correlate metabolites to clinical status and build prediction models. As a final step, we merged the original and replications cohorts in a data-driven DNEA analysis to derive a more integrated view of plasma metabolome structure in ALS. This approach does not rely on prior biochemical knowledge, but rather clusters metabolites into subnetworks dictated solely by experimental measurements, in this case, by partial correlations between metabolites across samples.

When we examined the overlap in original and replication cohorts by metabolites, many were shared, but several also were not shared. For the PLS-DA and group lasso analyses, which rank metabolites by VIP and OR, the overlap between cohorts was 19 and 14 metabolites, respectively, when restricting our analysis to the top 50 metabolites. Additionally, when we examined the top 20 differential metabolites in ALS versus controls, which were shared by cohort 1 and cohort 2, we found that metabolite abundance decreased with decreasing ALSFRS-R score and advancing disease. Interestingly, this was the case for several sphingomyelin species, although sphingomyelin abundance generally correlated positively with ALS status. It is possible that a temporal element in sphingomyelin levels may exist,[42] with an initial rise followed by a dip, which might account for the pattern seen with ALSFRS-R score. Therefore, though only cross-sectional, this finding indicates that metabolomic profiles may correlate with clinical status in ALS, and future longitudinal studies could inform the importance of critical metabolites in the disease course.

Furthermore, even though metabolite overlap was not fully concordant between cohort 1 and cohort 2, the cohort 1 metabolite dataset could be used to construct several prediction models, which identify ALS cases in the replication cohort. Our AUCs are higher than previously reported in ALS studies,[13,14] possibly due to our larger metabolite dataset and sample numbers[13,14] or because we analysed metabolomics profiles from symptomatic ALS participants.[13] Our prediction model results support the viability of utilizing metabolomics for ALS biomarker discovery, shown, for the first time,

Figure 4 Pathway enrichment of Wilcoxon-, adjusted logistic regression-, PLS-DA- and group lasso-selected metabolites for original and replication cohorts. Significantly enriched sub-pathways from metabolites selected by Wilcoxon, adjusted logistic regression, PLS-DA and group lasso models illustrated in dot plots for original cohort 1 and replication cohort 2. Rich factor refers to the proportion of selected metabolites relative to total sub-pathway metabolites. Metabolite number (node size) refers to the number of sub-pathway metabolites. Node colour indicates the significance level according to $-\log_{10}$(P-value).

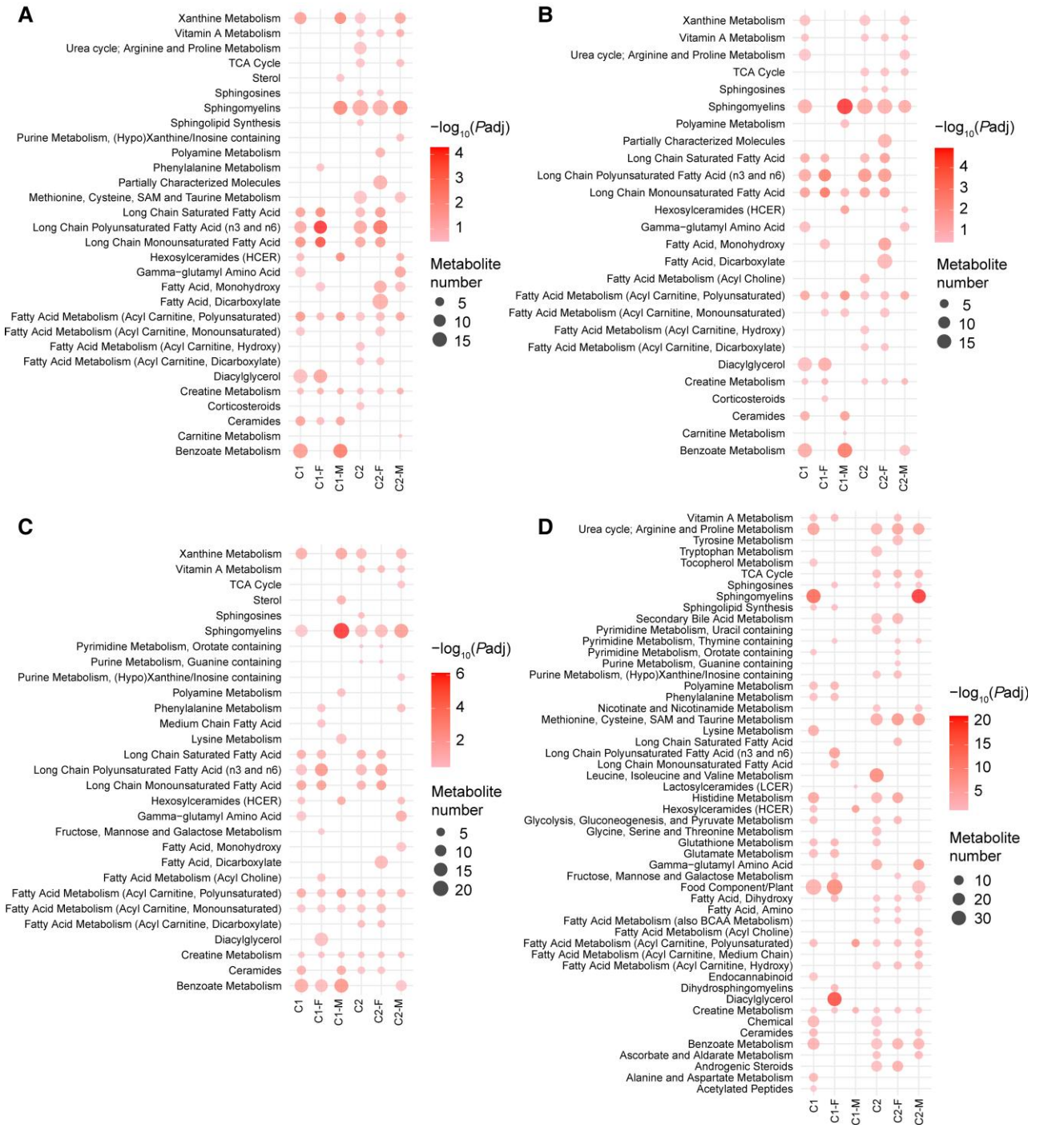

Figure 5 Pathway enrichment by sex of Wilcoxon-, adjusted logistic regression-, PLS-DA- and group lasso-selected metabolites for original and replication cohorts. Significantly enriched sub-pathways by sex from metabolites selected by (**A**) Wilcoxon, (**B**) adjusted logistic regression, (**C**) PLS-DA and (**D**) group lasso models illustrated in dot plots for original cohort 1 and replication cohort 2. Rich factor refers to the proportion of selected metabolites relative to total sub-pathway metabolites. Metabolite number (node size) refers to the number of sub-pathway metabolites. Node colour indicates the significance level according to $-\log_{10}$(*P*-value). C1 = cohort 1; C1-F = cohort 1 females; C1-M = cohort 1 males; C2 = cohort 2; C2-F = cohort 2 females; C2-M = cohort 2 males.

using two independent cohorts. Future research will need to include disease mimics to evaluate specificity for ALS versus possible differential diagnoses to determine the diagnostic utility in a real-world setting.

Although identifying individual differential metabolites sheds insight on pathogenesis and may identify new biomarkers, metabolites exist along a series of biochemical transformations, which contribute to a biological process. Therefore, pathway analysis of all significant differential metabolites emphasizes metabolic networks rather than discrete metabolites and provides more meaningful insight on entire pathway dysregulation in disease pathogenesis. As with individual metabolites, there were shared

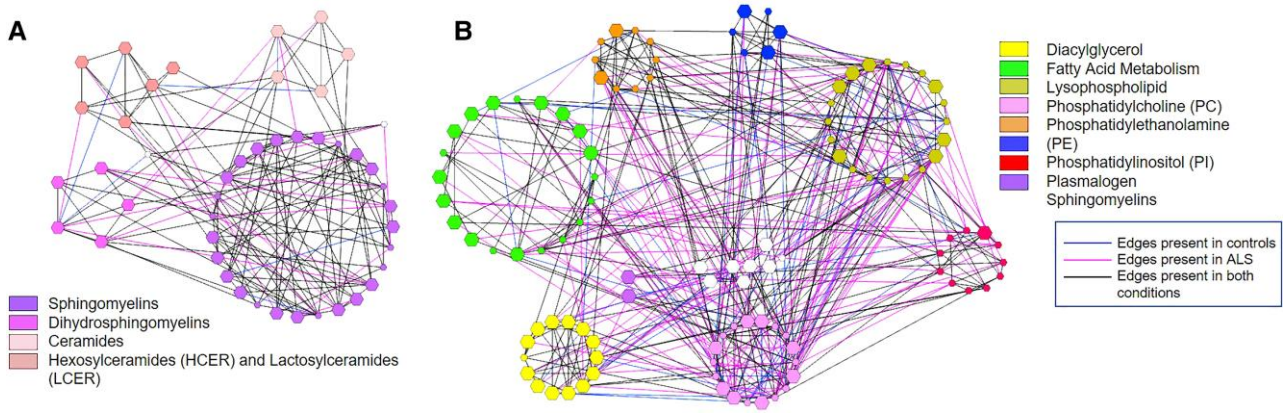

Figure 6 DNEA analysis subnetworks S8 and S9 overview. Subnetworks (**A**) S8 (Padj = 1.11×10^{-3}) and (**B**) S9 (Padj = 2.74×10^{-3}) from the data driven DNEA analysis. Nodes represent metabolites, which are colour-coded by sub-pathway; larger nodes indicate metabolites higher in ALS; smaller nodes indicate metabolites lower in ALS. Only metabolites from sub-pathways contributing three or more metabolites to the subnetworks are colour-coded; metabolites from sub-pathways contributing less than three metabolites are white. Edges present in controls, blue; edges present in ALS, pink; edges present in both, black. Fully annotated subnetworks S8 and S9 are available in Supplementary Figs 22 and 23 and member metabolites and their direction change and significance are in Supplementary Table 10.

sub-pathways identified within the original and replication cohorts, but also sub-pathways that were unique to either cohort.

Sub-pathways consistently selected in both the original and replication cohort centred heavily on lipid metabolism. The most significant sub-pathway 'sphingomyelin', selected by group lasso in the original cohort as well as the three other methods, continued to be selected in the replication cohort, but not by group lasso. Other recurrent lipid pathways in the original and replication cohorts included various long-chain fatty acids, acyl intermediates and ceramides. Similarly, in the DNEA analysis, subnetworks S3, S8 and S9 were highly populated by lipid species.

With regards to subnetwork S8, this subnetwork contained mostly elevated sphingomyelins, dihydrosphingomyelins, ceramides and hexosylceramides (significant) in ALS, which is aligned with other studies of ALS participant plasma,[13,14,35] spinal cord[43,44] and CSF,[16] and ALS mouse studies.[17,43–45] Impaired sphingolipid metabolism is a central, and at this point relatively well-validated, aspect of ALS pathogenesis, although the details and underlying aetiology, especially in sporadic disease, remain incompletely understood. The increase in ceramides and glucosylceramides in ALS has been linked to enhanced glucocerebrosidase activity in mutant superoxide dismutase 1 ($SOD1^{G93A}$) mice,[44] although conversely glucosylceramide synthase expression is upregulated in $SOD1^{G86R}$ mouse muscle.[42] These differences may arise from distinct model systems, but also from natural evolution during the disease course.[42] Nevertheless, this dysfunction in sphingolipid synthesis could be integral and potentially causative to disease progression, as seen in a model of monogenic childhood-onset ALS with mutant serine palmitoyltransferase subunit 1 (*SPTLC1*).[46] In an induced pluripotent stem cell ALS model, allele specific mutant *SPTLC1* knock-in increases sphinganine and ceramide levels versus wild-type allele.[46] Indeed, we observed elevated sphinganine and ceramides in our sporadic ALS cohort, indicating impaired sphingolipid metabolism could be central to ALS pathogenesis. Ceramides are pro-apoptotic and potentially excito-[43] or neurotoxic,[47] contributing to neurodegeneration (Fig. 7). Additionally, high fatty acid levels (see below) increase ceramides and possibly dihydroceramide intermediates[48] and sphingomyelins,[49,50] as observed in both the original and replication cohorts and the combined ALS dataset.

When examining subnetwork S9, this subnetwork contained various mostly downregulated complex lipids in ALS, though only some phosphatidylcholines and lysophospholipids were significantly decreased. Our findings agree with some reports,[45] but conflict with others,[13,16] which may arise from the specific phosphatidylcholine species[16] or the stage in ALS development.[13] Phosphatidylcholines and phosphatidylethanolamines are primarily synthesized from diacylglycerols through the Kennedy pathway[51] and phosphatidylinositols from diacylglycerol intermediates.[52] Phosphatidylcholines, phosphatidylethanolamines and phosphatidylinositols are important membrane constituents; loss of phosphatidylethanolamines is especially central to mitochondrial dysfunction through impaired mitochondrial membrane curvature, fission/fusion and bioenergetics.[53] In addition to their role in membrane structure, phosphatidylcholines and phosphatidylethanolamines are precursors to diverse signalling molecules, e.g. diacylglycerols,[51] which control various biological processes, such as proliferation, survival and migration.[54] Our study emphasizes the importance of continued research in the area of complex and bioactive lipids and lipid signalling in ALS.

Subnetwork S3 comprised free long-chain fatty acids of all saturation levels (saturated, monounsaturated, polyunsaturated; highly significant), which were universally elevated in ALS participants. Subnetwork S9 encompassed increased fatty acid intermediate acylcarnitines (significant), linked to β-oxidation, along with raised diacylglycerols (significant) in ALS plasma. Though noted in some studies,[13,17] these observations regarding free fatty acids, acylcarnitines, and diacylglycerols in symptomatic sporadic ALS, are relatively novel and align with our growing understanding of ALS 'hypermetabolism'.[55]

ALS 'hypermetabolism' is characterized by elevated resting energy expenditure,[55] possibly related to glucose uptake,[56] and low BMI and fat-free mass; thus, β-oxidation may be very relevant to ALS pathogenesis (Fig. 7). Increased levels of non-metabolized free fatty acids and partially metabolized intermediate acylcarnitines could indicate dysfunctional or at capacity β-oxidation,[57] which ties in with impaired fatty acid uptake and utilization as well as mitochondrial dysfunction. A study found $SOD1^{G93A}$ mice cleared triacylglycerol from plasma post feeding to a greater extent versus control mice,[58] which could elevate plasma free fatty acids,

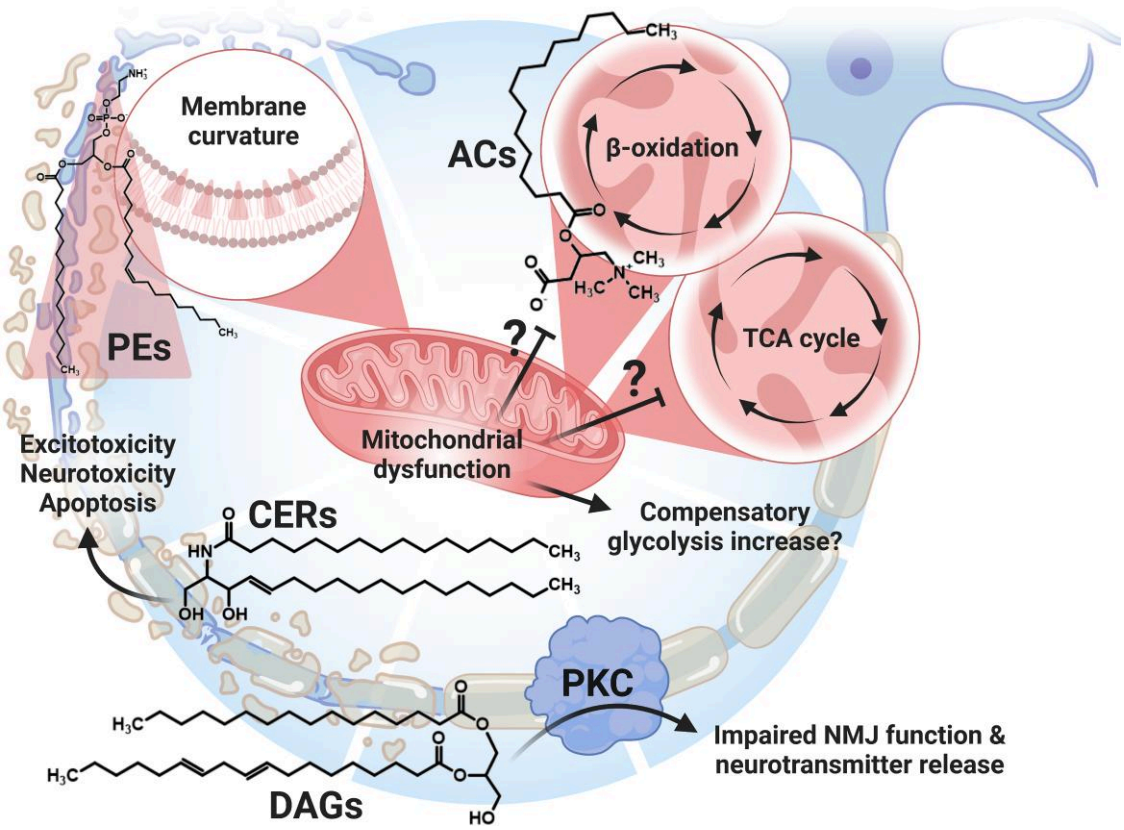

Figure 7 Potential mechanisms of lipid and energy dysregulation in ALS. Plasma metabolomics in two independent ALS cohorts revealed recurrent dysregulation in pathways related to lipid metabolism and energy. Lipids encompassed classes associated with multiple biological processes. Acylcarnitines (ACs) were mostly and significantly upregulated in ALS and could be indicative of impaired β-oxidation, which also feeds into the TCA cycle. TCA metabolites also differentiated ALS from control plasma; however, there was no pattern of up versus downregulation across specific TCA species. Impaired β-oxidation and TCA metabolism could lead to a compensatory increase in glycolysis. Dysfunctional sphingolipid metabolism leads to elevated ceramides (CERs), which are excito- and neurotoxic and trigger apoptosis. Various phospholipids, important membrane constituents, were mostly downregulated in ALS plasma, though only phosphatidylcholines attained significance. Phosphatidylethanolamines (PEs) regulate membrane curvature, and hence mitochondrial function, and were non-significant subnetwork S9 members. Diacylglycerols (DAGs), which can be synthesized from phosphatidylcholines and phosphatidylethanolamines, were universally upregulated and mostly significant in ALS plasma. Diacylglycerols activate protein kinase C (PKC), which has diverse biological properties, including important roles in the peripheral nervous system related to synaptic transmission and neuromuscular junction function. Future investigation will be required to evaluate these pathways in ALS relevant tissues, i.e. motor nerves, spinal cord, frontal and temporal brain lobes, and longitudinally, to assess whether these transformations are causative or downstream of pathogenesis. NMJ = neuromuscular junction. Created, in part, with ACD/ChemSketch and BioRender.com.

as we observed, if tissues do not compensate by enhanced uptake. In both our original and replication cohorts, acylcarnitines of all lengths and saturation levels were universally elevated in ALS, with few exceptions, indicating an overall lack of catabolic β-oxidation. Enhanced muscle lipolysis is reported in SOD1^{G86R} mice,[59] along with increases in mRNA[59] and protein[60] levels of β-oxidation enzymes, such as carnitine palmitoyltransferases (CPTs) and fatty-acid-binding proteins. In a mutant TAR DNA-binding protein 43 (TDP-43) fly model, acylcarnitine accumulation was observed, along with differential carnitine palmitoyltransferase expression.[61] It is possible that hypermetabolism may revolve around glycolysis as a compensatory mechanism[62] to impaired β-oxidation.

Subnetworks S5 and S7 contained metabolites from several subpathways related to ALS pathogenesis, including energy metabolism, which, as noted above for β-oxidation, appears to be a central ALS theme. Among them were metabolites related to TCA and glycolysis energy metabolism,[13,14,16,17,35,63] as well as amino acid metabolism, e.g. histidine, lysine, 'leucine, isoleucine and valine' (otherwise known as branched-chain amino acid)[18,20,21] and creatine.[14,19,64] Purine/pyrimidine[41] and xenobiotics metabolism were also featured. Members of the glycolysis, gluconeogenesis and pyruvate metabolism sub-pathways were mostly elevated in ALS, including glucose, lactate and pyruvate metabolites, whereas the level of TCA substrates and intermediates did not follow very significant or evident trends within this S7 subnetwork. The literature also reports elevated glycolysis,[35,65] although it is less clear regarding TCA metabolism.[66] From our own results, it is not possible to deduce a pathway direction for these interrelated metabolites in glycolytic and TCA metabolism. It may be difficult to infer shifts in metabolic processes from steady state metabolite evaluations by metabolomics, and fluxomics are needed to elucidate TCA metabolite flux in ALS.[67] Additionally, some amino acid sub-pathways in subnetwork S7 may also eventually feed into energy metabolism, e.g. glycine, serine and threonine, tyrosine and phenylalanine metabolism.[68]

Creatine was significantly increased, and creatinine significantly decreased in ALS, as in the literature,[14,19] but is likely linked to muscle loss as part of the disease process. Oxidative stress is an extremely prominent ALS characteristic,[69] which led to one of only

two FDA-approved ALS therapies, edaravone.[11] Gamma-glutamyl amino acid, glutathione metabolism and vitamin A were selected in both the original and replication cohorts and were mostly downregulated in ALS. Deficits in antioxidants have previously been noted in ALS, such as for glutathione,[70,71] tocopherol[14] and ascorbate.[19,63] However, since clinical trials of antioxidant therapies have been unsuccessful[72] and edaravone only minimally slows ALS progression, antioxidant dysregulation is likely not causative and may occur downstream of disease initiation, like creatine/creatinine.

Sex is an important clinical variable in ALS.[73] We recently reported that elevated neutrophil counts correlated with ALS progression, but that the trend was especially pronounced in female participants.[36] We also found that sex influenced the association of the natural killer cell cytotoxicity marker, NKp30, with ALS progression by ALSFRS-R score.[37] These findings implicate sex differences in immune system aspects of ALS. We took advantage of the sex imbalance in the control group of cohort 2 in the replication to conduct a sensitivity analysis and examine potential sex differential effects on plasma metabolomics in ALS. We found that although lipid metabolism pathways continued to be selected in both males and females in both cohorts, certain sub-pathways were more significant in females, such as long-chain fatty acid metabolism, diacylglycerols and sphingosines, whereas others were more important in males, e.g. sphingomyelin. Several xenobiotics pathways, such as benzoate and xanthine metabolism, were also more often and/or significantly selected in male ALS participants. Metabolic differences among ALS males and females could explain variation in published studies and may be important considerations for any therapeutics targeting metabolism in ALS. These differences could also yield new avenues of research to better understand why the incidence of ALS is highest in males.[74]

Our study has strengths and limitations. Among the strengths are the extensive number of detectable and identifiable metabolites in the metabolomics platform and large size of both cohorts. Additional strengths include the validation design of the original study in this replication effort, especially with the more stringent missingness criteria, and analyses using several statistical methods, including the data driven DNEA approach. Despite numerous strengths, our study has weaknesses. Most salient to a neurodegenerative disease is the tissue issue, since plasma may not necessarily reflect the peripheral and central nervous system milieu. Furthermore, since plasma collection was cross-sectional, the study does not establish causality among any of the observed metabolic sub-pathways. The study also does not correlate the metabolome with clinical progression, although this is a future area of investigation, which is required to delineate correlative from causative shifts in metabolic profiles in ALS. However, the long prodromal phase in ALS poses significant challenges. From a technical perspective, both our cohorts consisted mainly of White participants and our replication study was not well balanced for sex, although this limitation prompted us to conduct analyses for sex differences. Additionally, plasma samples were not collected from fasted participants because it was deemed an unethical request. Finally, batch effects were present since samples were run a year apart, although data were autoscaled for the DNEA analysis, removing any batch influence.

Overall, our metabolomics replication study highlighted recurrently dysregulated lipid metabolism in multiple sub-pathways, indicative of altered β-oxidation, mitochondrial bioenergetics and complex lipid signalling. Targeted lipidomics looking at specific lipid classes and lipid species along the ALS continuum could provide new insight into disease pathogenesis. Future directions could also address metabolic flux to move past steady state evaluations. Preclinical work along a time continuum could shed additional insight on early metabolic changes and in motor nerves and spinal cord tissue. The concurrent dysregulation in transcriptome and epigenome in ALS may also advocate a multi-omics approach.

Acknowledgements

We are indebted to the study participants that provided samples. We thank Crystal Pacut, Blake Swihart, Jayna Duell, RN, Daniel Burger and Amanda Williams.

Funding

Centers for Disease Control and Prevention / Agency for Toxic Substances and Disease Registry / National ALS Registry (1R01TS000289); Centers for Disease Control and Prevention / Agency for Toxic Substances and Disease Registry / National ALS Registry CDCP-DHHS-US (CDC/ATSDR 200-2013-56856); National Institute of Environmental Health Sciences K23ES027221; NIEHS R01ES030049; National Cancer Institute 1U01CA235487; the NeuroNetwork for Emerging Therapies, the NeuroNetwork Therapeutic Discovery Fund, the Peter R. Clark Fund for ALS Research, the Sinai Medical Staff Foundation and Scott L. Pranger, University of Michigan; National Center for Advancing Translational Sciences at the National Institutes of Health (UL1TR002240).

Competing interests

S.A.G. has acted as a medical advisor for Biogen and ITF Pharma and served on a DSMB. The other authors report no competing interests.

Supplementary material

Supplementary material is available at *Brain* online.

References

1. Goutman SA. Diagnosis and clinical management of amyotrophic lateral sclerosis and other motor neuron disorders. *Continuum (Minneap Minn)*. 2017;23(5):1332–1359.
2. Chia R, Chiò A, Traynor BJ. Novel genes associated with amyotrophic lateral sclerosis: diagnostic and clinical implications. *Lancet Neurol*. 2018;17(1):94–102.
3. Goutman SA, Chen KS, Paez-Colasante X, Feldman EL. Emerging understanding of the genotype-phenotype relationship in amyotrophic lateral sclerosis. *Handb Clin Neurol*. 2018;148:603–623.
4. Paez-Colasante X, Figueroa-Romero C, Sakowski SA, Goutman SA, Feldman EL. Amyotrophic lateral sclerosis: mechanisms and therapeutics in the epigenomic era. *Nat Rev Neurol*. 2015;11(5):266–279.
5. Goutman SA, Boss J, Patterson A, Mukherjee B, Batterman S, Feldman EL. High plasma concentrations of organic pollutants negatively impact survival in amyotrophic lateral sclerosis. *J Neurol Neurosurg Psychiatry*. 2019;90(8):907–912.
6. Su FC, Goutman SA, Chernyak S, *et al*. Association of environmental toxins with amyotrophic lateral sclerosis. *JAMA Neurol*. 2016;73(7):803–811.

7. Baharum SN, Azizan KA. Metabolomics in systems biology. *Adv Exp Med Biol*. 2018;1102:51–68.
8. Blasco H, Patin F, Madji Hounoum B, *et al*. Metabolomics in amyotrophic lateral sclerosis: how far can it take us? *Eur J Neurol*. 2016;23(3):447–454.
9. Cassina P, Peluffo H, Pehar M, *et al*. Peroxynitrite triggers a phenotypic transformation in spinal cord astrocytes that induces motor neuron apoptosis. *J Neurosci Res*. 2002;67(1):21–29.
10. Dodge JC, Yu J, Sardi SP, Shihabuddin LS. Sterol auto-oxidation adversely affects human motor neuron viability and is a neuropathological feature of amyotrophic lateral sclerosis. *Sci Rep*. 2021;11(1):803.
11. Yoshino H, Kimura A. Investigation of the therapeutic effects of edaravone, a free radical scavenger, on amyotrophic lateral sclerosis (Phase II study). *Amyotroph Lateral Scler*. 2006;7(4):247–251.
12. Rozen S, Cudkowicz ME, Bogdanov M, *et al*. Metabolomic analysis and signatures in motor neuron disease. *Metabolomics*. 2005;1(2):101–108.
13. Bjornevik K, Zhang Z, O'Reilly ÉJ, *et al*. Prediagnostic plasma metabolomics and the risk of amyotrophic lateral sclerosis. *Neurology*. 2019;92(18):e2089–e2100.
14. Lawton KA, Brown MV, Alexander D, *et al*. Plasma metabolomic biomarker panel to distinguish patients with amyotrophic lateral sclerosis from disease mimics. *Amyotroph Lateral Scler Frontotemporal Degener*. 2014;15(5–6):362–370.
15. Krokidis MG. Transcriptomics and metabolomics in amyotrophic lateral sclerosis. *Adv Exp Med Biol*. 2020;1195:205–212.
16. Blasco H, Veyrat-Durebex C, Bocca C, *et al*. Lipidomics reveals cerebrospinal-fluid signatures of ALS. *Sci Rep*. 2017;7(1):17652.
17. Chaves-Filho AB, Pinto IFD, Dantas LS, *et al*. Alterations in lipid metabolism of spinal cord linked to amyotrophic lateral sclerosis. *Sci Rep*. 2019;9(1):11642.
18. Patin F, Corcia P, Vourc'h P, *et al*. Omics to explore amyotrophic lateral sclerosis evolution: the central role of arginine and proline metabolism. *Mol Neurobiol*. 2017;54(7):5361–5374.
19. Wuolikainen A, Jonsson P, Ahnlund M, *et al*. Multi-platform mass spectrometry analysis of the CSF and plasma metabolomes of rigorously matched amyotrophic lateral sclerosis, Parkinson's disease and control subjects. *Mol Biosyst*. 2016;12(4):1287–1298.
20. Blasco H, Nadal-Desbarats L, Pradat PF, *et al*. Biomarkers in amyotrophic lateral sclerosis: combining metabolomic and clinical parameters to define disease progression. *Eur J Neurol*. 2016;23(2):346–353.
21. Kumar A, Bala L, Kalita J, *et al*. Metabolomic analysis of serum by (1) H NMR spectroscopy in amyotrophic lateral sclerosis. *Clin Chim Acta*. 2010;411(7–8):563–567.
22. Goutman SA, Boss J, Guo K, *et al*. Untargeted metabolomics yields insight into ALS disease mechanisms. *J Neurol Neurosurg Psychiatry*. 2020;91:1329–1338.
23. Iyer GR, Wigginton J, Duren W, *et al*. Application of differential network enrichment analysis for deciphering metabolic alterations. *Metabolites*. 2020;10(12):479.
24. Rhee EP, Cheng S, Larson MG, *et al*. Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. *J Clin Invest*. 2011;121(4):1402–1411.
25. Niewczas MA, Sirich TL, Mathew AV, *et al*. Uremic solutes and risk of end-stage renal disease in type 2 diabetes: metabolomic study. *Kidney Int*. 2014;85(5):1214–1224.
26. Wuolikainen A, Hedenström M, Moritz T, Marklund SL, Antti H, Andersen PM. Optimization of procedures for collecting and storing of CSF for studying the metabolome in ALS. *Amyotroph Lateral Scler*. 2009;10(4):229–236.
27. Dehaven CD, Evans AM, Dai H, Lawton KA. Organization of GC/MS and LC/MS metabolomics data into chemical libraries. *J Cheminform*. 2010;2(1):9.
28. Evans AM, Bridgewater B, Liu Q, *et al*. High resolution mass spectrometry improves data quantity and quality as compared to unit mass resolution mass spectrometry in high-throughput profiling metabolomics. *Metabolomics*. 2014;4(2):1000132.
29. Do KT, Wahl S, Raffler J, *et al*. Characterization of missing values in untargeted MS-based metabolomics data and evaluation of missing data handling strategies. *Metabolomics*. 2018;14(10):128.
30. Rohart F, Gautier B, Singh A, Le Cao KA. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Comput Biol*. 2017;13(11):e1005752.
31. Galindo-Prieto B, Eriksson L, Trygg J. Variable influence on projection (VIP) for orthogonal projections to latent structures (OPLS). *J Chemom*. 2014;28(8):623–632.
32. Cho HW, Kim SB, Jeong MK, *et al*. Discovery of metabolite features for the modelling and analysis of high-resolution NMR spectra. *Int J Data Min Bioinform*. 2008;2(2):176–192.
33. Ma J, Karnovsky A, Afshinnia F, *et al*. Differential network enrichment analysis reveals novel lipid pathways in chronic kidney disease. *Bioinformatics*. 2019;35(18):3441–3452.
34. Ma J, Shojaie A, Michailidis G. Network-based pathway enrichment analysis with incomplete network information. *Bioinformatics*. 2016;32(20):3165–3174.
35. Lawton KA, Cudkowicz ME, Brown MV, *et al*. Biochemical alterations associated with ALS. *Amyotroph Lateral Scler*. 2012;13(1):110–118.
36. Murdock BJ, Goutman SA, Boss J, Kim S, Feldman EL. Amyotrophic lateral sclerosis survival associates with neutrophils in a sex-specific manner. *Neurol Neuroimmunol Neuroinflamm*. 2021;8(2):e953.
37. Murdock BJ, Famie JP, Piecuch CE, *et al*. Natural killer cells associate with amyotrophic lateral sclersois in a sex- and age-dependent manner. *JCI Insight*. 2021;6:e147129.
38. Pape JA, Grose JH. The effects of diet and sex in amyotrophic lateral sclerosis. *Rev Neurol (Paris)*. 2020;176(5):301–315.
39. Beyene HB, Olshansky G AATS, *et al*. High-coverage plasma lipidomics reveals novel sex-specific lipidomic fingerprints of age and BMI: Evidence from two large population cohort studies. *PLoS Biol*. 2020;18(9):e3000870.
40. Ishikawa M, Maekawa K, Saito K, *et al*. Plasma and serum lipidomics of healthy white adults shows characteristic profiles by subjects' gender and age. *PLoS One*. 2014;9(3):e91806.
41. Veyrat-Durebex C, Bris C, Codron P, *et al*. Metabo-lipidomics of fibroblasts and mitochondrial-endoplasmic reticulum extracts from ALS patients shows alterations in purine, pyrimidine, energetic, and phospholipid metabolisms. *Mol Neurobiol*. 2019;56(8):5780–5791.
42. Henriques A, Croixmarie V, Priestman DA, *et al*. Amyotrophic lateral sclerosis and denervation alter sphingolipids and up-regulate glucosylceramide synthase. *Hum Mol Genet*. 2015;24-(25):7390–7405.
43. Cutler RG, Pedersen WA, Camandola S, Rothstein JD, Mattson MP. Evidence that accumulation of ceramides and cholesterol esters mediates oxidative stress-induced death of motor neurons in amyotrophic lateral sclerosis. *Ann Neurol*. 2002;52(4):448–457.
44. Dodge JC, Treleaven CM, Pacheco J, *et al*. Glycosphingolipids are modulators of disease pathogenesis in amyotrophic lateral sclerosis. *Proc Natl Acad Sci U S A*. 2015;112(26):8100–8105.
45. Henriques A, Croixmarie V, Bouscary A, *et al*. Sphingolipid metabolism is dysregulated at transcriptomic and metabolic levels

in the spinal cord of an animal model of amyotrophic lateral sclerosis. *Front Mol Neurosci*. 2018;10:433.
46. Mohassel P, Donkervoort S, Lone MA, *et al*. Childhood amyotrophic lateral sclerosis caused by excess sphingolipid synthesis. *Nat Med*. 2021;27(7):1197–1204.
47. Wang G, Bieberich E. Sphingolipids in neurodegeneration (with focus on ceramide and S1P). *Adv Biol Regul*. 2018;70:51–64.
48. Chaurasia B, Summers SA. Ceramides - Lipotoxic inducers of metabolic disorders. *Trends Endocrinol Metab*. 2015;26(10):538–550.
49. Choi S, Snider AJ. Sphingolipids in high fat diet and obesity-related diseases. *Mediators Inflamm*. 2015;2015:520618.
50. Deevska GM, Nikolova-Karakashian MN. The twists and turns of sphingolipid pathway in glucose regulation. *Biochimie*. 2011;93(1):32–38.
51. Gibellini F, Smith TK. The Kennedy pathway–De novo synthesis of phosphatidylethanolamine and phosphatidylcholine. *IUBMB Life*. 2010;62(6):414–428.
52. Tracey TJ, Steyn FJ, Wolvetang EJ, Ngo ST. Neuronal lipid metabolism: Multiple pathways driving functional outcomes in health and disease. *Front Mol Neurosci*. 2018;11:10.
53. Ball W B, Neff JK, Gohil VM. The role of nonbilayer phospholipids in mitochondrial structure and function. *FEBS Lett*. 2018;592(8):1273–1290.
54. Geraldes P, King GL. Activation of protein kinase C isoforms and its impact on diabetic complications. *Circ Res*. 2010;106(8):1319–1331.
55. Desport JC, Preux PM, Magy L, *et al*. Factors correlated with hypermetabolism in patients with amyotrophic lateral sclerosis. *Am J Clin Nutr*. 2001;74(3):328–334.
56. Bauckneht M, Lai R, Miceli A, *et al*. Spinal cord hypermetabolism extends to skeletal muscle in amyotrophic lateral sclerosis: a computational approach to [18F]-fluorodeoxyglucose PET/CT images. *EJNMMI Res*. 2020;10(1):23.
57. van Eunen K, Simons SM, Gerding A, *et al*. Biochemical competition makes fatty-acid beta-oxidation vulnerable to substrate overload. *PLoS Comput Biol*. 2013;9(8):e1003186.
58. Fergani A, Oudart H, De Aguilar JLG, *et al*. Increased peripheral lipid clearance in an animal model of amyotrophic lateral sclerosis. *J Lipid Res*. 2007;48(7):1571–1580.
59. Dupuis L, Oudart H, René F, de Aguilar JLG, Loeffler JP. Evidence for defective energy homeostasis in amyotrophic lateral sclerosis: benefit of a high-energy diet in a transgenic mouse model. *Proc Natl Acad Sci U S A*. 2004;101(30):11159–11164.
60. Pharaoh G, Sataranatarajan K, Street K, *et al*. Metabolic and stress response changes precede disease onset in the spinal cord of mutant SOD1 ALS mice. *Front Neurosci*. 2019;13:487.
61. Manzo E, O'Conner AG, Barrows JM, Shreiner DD, Birchak GJ, Zarnescu DC. Medium-chain fatty acids, beta-hydroxybutyric acid and genetic modulation of the carnitine shuttle are protective in a drosophila model of ALS based on TDP-43. *Front Mol Neurosci*. 2018;11:182.
62. Manzo E, Lorenzini I, Barrameda D, *et al*. Glycolysis upregulation is neuroprotective as a compensatory mechanism in ALS. *eLife*. 2019;8:e45114.
63. Blasco H, Corcia P, Moreau C, *et al*. 1H-NMR-based metabolomic profiling of CSF in early amyotrophic lateral sclerosis. *PLoS One*. 2010;5(10):e13223.
64. Ito D, Hashizume A, Hijikata Y, *et al*. Elevated serum creatine kinase in the early stage of sporadic amyotrophic lateral sclerosis. *J Neurol*. 2019;266(12):2952–2961.
65. Valbuena GN, Rizzardini M, Cimini S, *et al*. Metabolomic analysis reveals increased aerobic glycolysis and amino acid deficit in a cellular model of amyotrophic lateral sclerosis. *Mol Neurobiol*. 2016;53(4):2222–2240.
66. Blasco H, Lanznaster D, Veyrat-Durebex C, *et al*. Understanding and managing metabolic dysfunction in Amyotrophic Lateral Sclerosis. *Expert Rev Neurother* 2020;20(9):907–919.
67. Steyn FJ, Li R, Kirk SE, *et al*. Altered skeletal muscle glucose-fatty acid flux in amyotrophic lateral sclerosis. *Brain Commun*. 2020;2(2):fcaa154.
68. Tefera TW, Borges K. Metabolic dysfunctions in amyotrophic lateral sclerosis pathogenesis and potential metabolic treatments. *Front Neurosci*. 2017;10:611.
69. D'Amico E, Factor-Litvak P, Santella RM, Mitsumoto H. Clinical perspective on oxidative stress in sporadic amyotrophic lateral sclerosis. *Free Radic Biol Med*. 2013;65:509–527.
70. Weiduschat N, Mao X, Hupf J, *et al*. Motor cortex glutathione deficit in ALS measured in vivo with the J-editing technique. *Neurosci Lett*. 2014;570:102–107.
71. D'Alessandro G, Calcagno E, Tartari S, Rizzardini M, Invernizzi RW, Cantoni L. Glutamate and glutathione interplay in a motor neuronal model of amyotrophic lateral sclerosis reveals altered energy metabolism. *Neurobiol Dis*. 2011;43(2):346–355.
72. Desnuelle C, Dib M, Garrel C, Favier A. A double-blind, placebo-controlled randomized clinical trial of alpha-tocopherol (vitamin E) in the treatment of amyotrophic lateral sclerosis. ALS riluzole-tocopherol Study Group. *Amyotroph Lateral Scler Other Motor Neuron Disord*. 2001;2(1):9–18.
73. Chiò A, Moglia C, Canosa A, *et al*. ALS phenotype is influenced by age, sex, and genetics: A population-based study. *Neurology*. 2020;94(8):e802–e810.
74. Yoshida S, Mulder DW, Kurland LT, Chu CP, Okazaki H. Follow-up study on amyotrophic lateral sclerosis in Rochester, Minn., 1925 through 1984. *Neuroepidemiology*. 1986;5(2):61–70.