Published in final edited form as:

Mol Biosyst. 2017 January 31; 13(2): 371–379. doi:10.1039/c6mb00600k.

Metabolic profiling of a myalgic encephalomyelitis/chronic fatigue syndrome discovery cohort reveals disturbances in fatty acid and lipid metabolism

Arnaud Germain^a, David Ruppert^b, Susan M. Levine^a, and Maureen R. Hanson^{a,*}

^aDepartment of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA

^bDepartment of Statistical Science and School of Operations Research and Information Engineering, Cornell University, Ithaca, NY 14853, USA

Abstract

Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) remains a continuum spectrum disease without biomarkers or simple objective tests, and therefore relies on a diagnosis from a set of symptoms to link the assortment of brain and body disorders to ME/CFS. Although recent studies show various affected pathways, the underlying basis of ME/CFS has yet to be established. In this pilot study, we compare plasma metabolic signatures in a discovery cohort, 17 patients and 15 matched controls, and explore potential metabolic perturbations as the aftermath of the complex interactions between genes, transcripts and proteins. This approach to examine the complex array of symptoms and underlying foundation of ME/CFS revealed 74 differentially accumulating metabolites, out of 361 ($P < 0.05$), and 35 significantly altered after statistical correction ($Q < 0.15$). The latter list includes several essential energy-related compounds which could theoretically be linked to the general lack of energy observed in ME/CFS patients. Pathway analysis points to a few pathways with high impact and therefore potential disturbances in patients, mainly taurine metabolism and glycerophospholipid metabolism, combined with primary bile acid metabolism, as well as glyoxylate and dicarboxylate metabolism and a few other pathways, all involved broadly in fatty acid metabolism. Purines, including ADP and ATP, pyrimidines and several amino acid metabolic pathways were found to be significantly disturbed. Finally, glucose and oxaloacetate were two main metabolites affected that have a major effect on sugar and energy levels. Our work provides a prospective path for diagnosis and understanding of the underlying mechanisms of ME/CFS.

Introduction

The disease known variously as Myalgic Encephalomyelitis or Chronic Fatigue Syndrome (ME/CFS) has historically been diagnosed from symptoms, often those described by Fukuda et al.¹ or the Canadian Consensus Criteria². Recently, the Institute of Medicine released a report offering guidance to the medical field and Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) patients, redefining the illness with a new name, Systemic Exertion Intolerance Disease (SEID), and new diagnostic criteria with the mission of

*Corresponding author: mrh5@cornell.edu.

promoting early detection³. ME/CFS/SEID is currently predicted to affect as many as 2.5 million Americans. One report indicates that 84% to 91% of patients remain undiagnosed, and many individuals are ill for years before receiving the proper diagnosis⁴. Objective tests for the disease are critically needed to identify patients, and may also provide fundamental information about the disruption that occurs in ME/CFS/SEID.

The study of metabolic networks and their influence on health and diseases has established metabolic signatures for various disorders, allowing for early diagnosis and development of therapeutics. While some metabolite studies of individuals with ME/CFS have examined urine^{5–10}, others have also focused on blood^{6, 8–16} and found that energy, amino acid, nucleotide, nitrogen, hormone and oxidative stress metabolisms are disturbed. However, the methodology of these studies was largely hypothesis-driven and as a result targeted a limited number of metabolites, such as amino acids and carbohydrates, for example.

We took advantage of advances in mass spectrometry to identify and quantify the status of metabolic networks in a pilot cohort. This broad approach allows for non-user-biased discovery and offers a new gateway into knowledge of the effects of ME/CFS on the metabolism of the body, a chance for future targeted screenings as a mean of diagnosis, and potentially some clues about the underlying causes of the illness. We analyzed blood metabolites in cases vs. controls to identify differences in levels that could distinguish between ill and healthy status.

Results

Cohort characteristics and metabolite measurements

The pilot cohort, 15 controls and 17 patients, originated from the same study population selected for our gut microbiome analysis¹⁷, and solely includes female subjects between the age of 42 and 68 (Table 1). All patients selected meet the 1994 Fukuda definition¹ for ME/CFS. Although the samples were collected before the publication of the IOM report about SEID, we believe that most of them, if not all, fit the SEID criteria. The values of Bell's disability scale¹⁸ for patients are also reported in Table 1.

We successfully identified and quantified 361 metabolites from plasma samples using the Q-Exactive ME (QE-MS) method¹⁹, for a total of 11,913 data points (Supplementary Data 1). QE-MS allows for the detection of a wide variety of compounds over diverse chemical classes, ranging from amino acids to polar lipids, with a straightforward workflow in order to minimize metabolite deterioration (Fig. S1). While some of the metabolites are known, others cannot be distinguished from one another and their identifiers are listed as chemical formulas.

Statistical analysis finds a disrupted metabolic network for 20% of the metabolites identified

An overview of the information for each metabolite is presented as box plots in Figure 1A and B as well as in Fig. S2. Box plots display the distribution pattern of the data and allow comparison between controls and patients and have been ordered according to the control medians. At first glance, although a majority of the metabolites are within the same range in

the two groups tested, it is evident that there is generally more variation among levels of individual metabolites within the patient cohort.

In order to test for significant difference between the two groups, two different statistical analysis were applied to the collected data, with either no distribution assumption using the non-parametric Kruskal-Wallis test or the assumption of a normal distribution using the parametric t-test after log10 transformation. The aim is to limit statistical assumptions about the dataset, especially since we are analyzing a small subset of the blood metabolome. The Kruskal-Wallis tests were initially performed and we found 65 metabolites out of 361 with significantly different accumulations at $P < 0.05$ (Table 2). Of the most significantly perturbed metabolites, we notice ATP and ADP, which are not only linked to the energy metabolism, but are also indispensable for many other metabolic pathways. A few other noteworthy compounds, such as acetylcarnosine and taurine, are known to be important in muscle tissue function or development. Carbohydrates, including D-glucose or L-erythrulose, also exhibit different levels.

In contrast, the t-tests found 65 metabolites with significantly different accumulations at $P < 0.05$ (Table 2). Importantly, 56 metabolites overlap with the t-test list, including the metabolites mentioned above. If we combine the metabolites found significant by at least one of the tests, we obtain a list of 74 metabolites out of 361, or 20% of the metabolic network tested.

The use of multiple comparisons performed during the statistical tests on numerous data points calls for statistical correction by adjusting the false discovery rate (FDR). Here we control the FDR by applying the Benjamini-Hochberg (BH) procedure to account for the proportion of discoveries that are actually false positives at a rate up to 15%, $Q < 0.15$. After FDR correction 35 metabolites detected in both tests remain significant (Table 2), including the particular metabolites mentioned above. The proportion of metabolites affected after FDR correction is almost 10% of those in the metabolic network tested. If we focus on the data distribution of these 35 metabolites, and as mentioned earlier, we notice that generally, the variation within the patient cohort is greater compared to controls but also that the values are lower for 31 of them in the patient cohort (Fig. 1A and Table 2).

The original data can be divided in three trends based on median value, with 5% remaining equal between controls and patients (Fig. S2A), 59% decreasing in patients compared to controls (Fig. 1A and B, and Fig. S2B, C, D and E) and 36% increasing in patients vs. controls (Fig. 1A and B, and Fig. S2F, G and H). After the Kruskal-Wallis and t-tests, but before FDR correction (Fig. 1A and B), the proportion remained similar with 72% and 28% respectively out of 74 metabolites. A striking observation is that out of the 39 metabolites removed by the FDR correction step, 44% of them were higher in the patients compared to the controls (Fig. 1B) while only 4 (11%) out of 35 are higher after FDR correction (Fig. 1A and Table 2).

Heat map and cluster analysis on subjects reveals statistical power to classify patients vs. controls using metabolite as biomarkers

A specific metabolic signature, potentially correlated with the health status of a subject, can be sought by performing a cluster analysis on this dataset. To do this, we used the interactive heat map function implemented with the function “d3heatmap”²⁰ in the R package by the same name. Because of the wide range of concentrations measured for each metabolite and to improve the clustering efficiency, the data was ranked within each metabolite, with lower and higher values receiving lower and higher ranks, respectively. Fig. 2 demonstrates the increased clustering efficiency of controls compared to patients in distinct branches, as we limit the number of metabolites considered for analysis, from 361, to 74 ($P<0.05$) and finally to 35 ($P<0.05$ and $Q<0.15$). In the end, we obtain two major branches with 92% controls and 90% patients respectively while the other two smaller branches are at 80% patients and 60% controls respectively. If larger numbers of metabolites are taken into account, we lose definition when trying to distinguish controls from patients, with major branches evenly split between controls and patients even though a minor branch reaches 100% of patients in one case (with 3 subjects only). The full heat maps along with the percentages can be found in Fig. S3A, B and C.

This analysis shows that there is a metabolic signature characteristic of healthy controls vs. patients and that we can improve its recognition by restricting the analysis to 10% of the blood metabolic network detected in this study.

Classification and regression tree (CART) analysis shows that four metabolites are sufficient to detect whether the subject is a control or a patient with 78.1% confidence

As a final statistical approach, we used a decision tree machine-learning approach to mine our dataset and determine if metabolite measurements could produce a predictive model by mapping the subjects into two distinct categories, namely healthy controls and patients.

Because there are significant differences between patients and controls in the concentrations of many metabolites, we investigated whether metabolite concentrations can be used to classify subjects. To do this, we used the random forest classifier Breiman²¹ implemented with the function “randomForest”²² in the R package by the same name. Random forest creates a large number of bootstrap resamples²³, e.g., 500 in our work. A decision tree is trained on each resample. Each tree classifies a subject as patient or control and then the random forest algorithm classifies that subject by a majority vote. Single decision trees are highly variable and greatly affected by small perturbations of the data or slight changes in the tuning parameters of the algorithm, but averaging over many decision trees reduces these problems and produces a stable algorithm²⁴.

It is well known that classification error rates are underestimated if one classifies the data used to train the algorithm. For that reason, error rates are estimated using validation data, i.e., data that are held out and so not used for training. For example, in leave-one-out cross-validation, each subject is classified using a tree trained with only the other data.

The random forest algorithm computes out-of-bag (OOB) error rates, which are similar to leave-one-out cross-validation error rates²⁴. A bootstrap resample is created by random

sampling, with replacement, from the original sample with each resample the same size as the original sample. Any subject will be in some resamples but not in others. OOB error rates are determined by classifying a subject using only the trees created from resamples not containing that subject.

All 361 metabolites were used in the random forest algorithm. The out-of-bag (OOB) confusion matrix is given in Table 3 and shows an overall error rate of $10/32 = 31.2\%$.

To see which metabolites are most useful for classifying subjects, a variable importance plot is shown on the left hand side of Fig. 3A. This plot shows the decrease in accuracy when the values of a variable are randomly permuted. The random permutation destroys the correlation between that variable and subjects' disease status, so that the variable no longer has predictive value. The effect is much like deleting the variable. Greater decreases in accuracy indicate more important variables. Similarly, the right hand side of Fig. 3A shows the decreases in Gini's index of node purity. Purer nodes mean greater classification accuracy, so the ordering of importance in the left and right hand plots are similar, although not identical. Interestingly, the sets of the four most important metabolites are identical: ADP (19), taurine (28), ATP (17), and 13'-carboxy-alpha-tocopherol (5).

Fig. 3B contains partial dependence plots for the four most important metabolites mentioned above. Each plot shows the log-odds of being a patient given the concentration of one metabolite with the effects of the other metabolites removed. For each of the four most important metabolites, as the concentration increases there is a sharp drop in the probability of being a patient. The sizes of these drops are somewhat small, which suggests that the metabolite concentrations might not have great predictive power. However, the small effect sizes are due to high correlations between the metabolite concentrations, particularly ADP, ATP, and taurine (Table 4). Because of these high correlation, any three of the metabolite concentrations can serve as a proxy for the fourth concentration.

Pathway analysis reveals three main affected pathway categories: lipids, purine and amino acids, and energy metabolism

As we are trying to understand the underlying networks within the significantly affected metabolites, we performed a pathway analysis on all of the 74 metabolites depicted in Table 2 ($P < 0.05$). The analysis was executed on 62 metabolites as it requires HMDB (Human Metabolome Database) identification that is not available for the six compounds whose name is only a chemical formula, nor for three additional ones (Table 2). The tool suite used (www.metaboanalyst.ca) was also not able to take into account another 10 metabolites, for a total of 19. Additionally, six metabolites with a double or triple HMDB identification, due to the inability to distinguish between them, were included in the analysis.

The output yielded 39 pathways, reported in Table 5, for which we decided to focus on those exhibiting a number of hits, ranging from five to one metabolite per pathway along with the calculated pathway impact from the pathway topology analysis. The latter value takes into account the impact of a change in a specific set of metabolites within a pathway.

The pathway with the highest impact factor is the taurine and hypotaurine metabolism with taurine (28) as the central compound of this pathway (Fig. S4A). A decrease in concentration of this metabolite might reflect a general effect on this pathway and the consequences associated with the general metabolism of the body. Of important note, taurine is also part of the primary bile acid biosynthesis pathway where three of the near-final products (16, 19 and 28) are found to be reduced in patients according to our measurements (Fig. S4B). The main function of bile acids is to physically support the digestion of dietary fats.

The glyoxylate and dicarboxylate metabolism is another pathway with high pathway impact, with two major metabolites (20 and 36) both reduced in patients compared to controls (Fig. S4C). Of note, the glyoxylate cycle also allows humans to use fats for the synthesis of carbohydrates, a function interrelated to bile acids. One of the two pathways with the highest number of hits is the glycerophospholipid metabolism with five metabolites affected (52, 18, 6 and 56 decreases while 61 increases, Fig. S4D). This is another example of a lipid metabolism pathway, in this case, mainly lipids involved in biological membrane composition. The fatty acid biosynthesis and metabolism pathways also showed disturbances but all three of the acids affected (72, 73 and 74) had higher concentrations in patients vs. controls (Fig. S4E).

Another major category of pathways affected in our pilot cohort was related to purine and amino acid metabolism. Of primary interest is the purine metabolism pathway with four metabolites reduced in ME/CFS patients (17, 19, 20, and 42, Fig. S4F). ATP (17) and ADP (19) are extremely important for numerous metabolic pathways in the human body, including DNA composition and energy metabolism. A majority of the amino acid metabolisms, biosynthesis or degradation pathways, appear to be impacted in the patients vs. controls, for a total of at least 34 hits, or almost 50% of all the hits for this pathway analysis. This includes tryptophan, histidine, arginine, proline, valine, leucine, isoleucine, tyrosine, alanine, phenylalanine, glycine, serine, threonine, cysteine, methionine, alanine, aspartate and glutamate.

Finally, the remaining pathways are related to the energy and sugar metabolism, with two major metabolites present in the following pathways, D-glucose (29) and oxaloacetate (36). The subsequent pathways have two hits: pentose phosphate pathway (29 and 21), ascorbate and aldarate metabolism (24 and 37), glycolysis and gluconeogenesis (29 and 36), while the remaining pathways have one hit: citrate cycle (36), starch and sucrose metabolism (29), galactose metabolism (29) and pyruvate metabolism (36).

Discussion

Despite the fact that this study was conducted on a small-scale pilot cohort restricted to female subjects, the diverse statistical analysis of the abundant data obtained by explorative QE-MS provided valuable information, with the identification of 74 metabolites affected at $P < 0.05$ in patients compared to matching controls and 35 metabolites that remained relevant after applying further statistical correction ($Q < 0.15$, Table 2 and Fig. 1A and 1B). These compounds are all components of human primary metabolism. Pathway analysis pointed to

major pathways with several hits for each pathway, some of them consistent with discoveries from previous studies using targeted methodology, namely amino acids, nucleotides and energy-related molecules. While we found a number of metabolites whose concentration is disturbed in patients compared to healthy controls, we do not discuss each individual metabolite in favor of assembling our findings in the hope of identifying trends that can be further explored in future research utilizing a larger sample size. The major metabolites affected are related to energy metabolism, not only ATP and ADP, but also numerous sugars from various different pathways.

The novel pathways of interest mainly revolve around several aspects of lipid metabolism, including possibly lower digestion capabilities of dietary fats, lower potential carbohydrate synthesis from fats, and reduced membrane lipids, while other categories of lipids were elevated in patients compared to controls. Notably, the major carbohydrate glucose was negatively altered in patients, as were two other major energy molecules, namely ADP and ATP, which were identified by all statistical methods used in this study (Table 2 and Figs. 1A, 3 and 4). Similarly, taurine was found to be significantly lower in patients.

Taurine, which has already been found to be present in lower concentrations in the plasma of patients⁸, is already used as a nutritional supplement by many ME/CFS patients because of possibility of symptom relief, although no rigorous research has yet proven its effectiveness. Taurine has a number of functions in skeletal muscle, the retina and the central nervous system and its concentrations may be relevant to the altered amounts of lipid metabolism compounds we observed in ME/CFS patients. Taurine conjugates with bile acids to form bile salts, necessary for proper lipid absorption during digestion, and the bile acid biosynthesis pathway was found to be negatively affected in our patient cohort.

Primary bile acids such as sulfoglycolithocholate (13) are synthesized in the liver and the major bile salts result from its conjugation with taurine (28) and glycine, forming taurochenodeoxycholate and glycochenodeoxycholate (21), respectively, which are all numbered metabolites found to be significantly reduced in ME/CFS patients. Reduction in these compounds is suggestive of damage to the liver. A study from the FDA National Center for Toxicological Research (NCTR) was able to identify liver injury biomarkers as the result of drug-induced hepatotoxicity in rats²⁵. Strikingly, several other metabolites identical to our findings were also identified in their report, namely 5-guanidino-2-oxopentanoic acid (12, also named 2-oxoarginine), sebacic acid (60), along with energy metabolites from the glyoxylate and dicarboxylate metabolism. These biomarkers could be used to define a serum metabolic signature by creating a panel for hepatotoxicity prediction²⁵. Whether ME/CFS patients actually have liver damage as a result of the disease is not known, as it also could be a consequence of medication toxicity in this patient group, which often uses prescription or over-the-counter pain relievers. Abnormally elevated levels of fatty acids 72, 73 and 74 (Table 2) could be a sign of hyperlipidemia and could also be related to a deficiency in liver activity.

Another hypothesis that could be drawn from pathway analysis concerns the glycerophospholipid metabolism, as we observed five altered metabolites involved in biological membrane composition. Phosphoinositides are phospholipids that play critical

roles in the brain and the spinal cord and peripheral nerves, by being involved in cell regulation and membrane dynamics. For example, they are implicated in a recently hypothesized link with leukodystrophy²⁶, a rare condition that expresses itself with an abnormal or diminished white matter in the brain, a condition also observed in ME/CFS patients²⁷. Conspicuously, one of the four main metabolites isolated by CART analysis is 13'-carboxy-alpha-tocopherol (5), a dehydrogenation carboxylate product of 13'-hydroxy-alpha-tocopherol with vitamin E antioxidant activity. Compounds from this family are synthesized by plants and require intestinal fat absorption for final delivery of vitamin E to tissues and eventual transfer to lipoproteins via phospholipids. The supply of alpha-tocopherols to plasma and tissues is liver-dependent, thus this compound could also be relevant to the discussion above, regarding liver activity.

Elevated oxidative stress levels in ME/CFS cases have been documented for many years^{28–30} and this finding has contributed to various hypotheses about potential symptom relief and consequences for the patients. One theory about ME/CFS patients' chronically activated immunological response is that the disease causes damage to lipid membrane components³¹. In this scenario, disrupted lipids would become immunogenic; our results show that palmitate (73) was increased. We did not see evidence for a reduction in anti-oxidative stress defenses; instead we observed an increase in spermidine (63) which is part of the glutathione metabolism, an important antioxidant in the living kingdom. However, the overall decrease observed in lipid metabolism discussed above could explain a slower repair rate of membranes and an overall status of damage with possible repercussions throughout the cell signaling system.

The results of our pilot study on a small cohort is promising with regard to the possible use of plasma metabolite profiling for diagnosis of ME/CFS. After our manuscript was submitted for review, two other independent studies appeared with results that concur with our findings, while sampling distinctive populations. Indeed, Naviaux et al.³² found a similar trend of hypometabolic state in their study, with over 80% of significantly different metabolites being decreased in patients vs. controls. Although the compounds they measured are not all identical to the ones we detected, the major pathways determined to be affected by ME/CFS agree with our conclusions, and primarily include lipid metabolism and amino acids. A summarized comparison with their female cohort is presented in Table 6. An additional study³³ with stronger statistical power, thanks to a larger cohort along with a validation dataset, also points to deficiencies we noticed, namely the urea cycle which takes place in the liver, and the TCA cycle which in part generates energy from lipids and sugars to make ATP, all compounds/pathways we found to be affected in our CFS cohort when compared to healthy controls. An even larger study with the latest methods for discovery metabolomics might reveal a larger set of metabolites that can distinguish cases from controls with greater accuracy, as well as allowing us to generate additional hypotheses regarding metabolic dysregulation in ME/CFS.

Conclusions

Metabolomics is playing an increasingly important role in medicine to explore, predict and diagnose diseases, using an approach meant to be unbiased and not dependent on

assumptions about the condition studied. Our metabolomics analysis implicates an aspect of metabolism—lipid metabolism—not previously linked to ME/CFS, even though lipid replacement³⁴ has been one of the many ways explored to reduce fatigue symptoms in fatiguing illnesses. This pilot study has provided considerable new information which begs for verification and extension with a large cohort and a larger set of metabolites.

Experimental

Sample collection

The cohort was entirely comprised of females with non-significantly different BMI (t-test; $P=0.469$) and age, within 2 years, between controls and patients (Table 1). Blood samples were drawn into EDTA BD Vacutainer® collection tubes from an antecubital vein, while subjects were seated, at the medical office of Dr. Susan Levine in Manhattan, NY. Tubes were shipped overnight to Cornell University, Ithaca campus, where they were centrifuged at 4,000 r.p.m. for 30 min to pellet blood cells and collect the supernatant plasma, which was then stored at -80°C for later analysis. This study was approved by the Cornell University Institutional Review Board, thus performed in compliance with the relevant laws and institutional guidelines, and informed consent was obtained from all subjects.

Sample preparation and mass spectrometry

All blood samples were prepared for Mass Spectrometry analysis using a polar metabolite extraction protocol. A thawed serum sample of 20 μL was diluted in 80 μL of ddH₂O and placed on dry ice in a 1.5 mL Eppendorf tube. Subsequently, 400 μL of extraction solvent (100% MeOH) was added and the tube were placed in the -80°C freezer for 15 min. After centrifugation at 20,000 r.c.f. for 10 min at 4°C , the supernatant was split into 2 Eppendorf tubes and dried in a SpeedVac. (approximately an hour for 0.5 mL). Dried pellets were stored at -20°C until mass spectrometry analysis which was performed as described in Liu et al.¹⁹, as was the metabolite identification.

Data analysis

The raw data extracted from the peaks is available in Supplementary Data 1. Although the detection limit of the method is 10^3 , we decided not to filter out those signals and keep their value at the arbitrary defined 10^3 , so as to reflect the low amount of metabolite and not count it as missing data.

Most of the statistical analysis and data display was performed in R computing language (www.r-project.org) while the pathway analysis approach was achieved through the metaboanalyst tool suite (www.metaboanalyst.ca).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by Cornell University and NIH NIAID grant 1R21AI101614 to MRH. We thank Jason Locasale and Xiaojing Liu (previously Cornell University, currently Duke University) for performing the mass spectrometry and providing advice about metabolomic analysis, and Lin Lin for excellent technical support. The authors declare no conflict of interest.

References

1. Fukuda K, Straus SE, Hickie I, Sharpe MC, Dobbins JG, Komaroff A. Ann Intern Med. 1994; 121:953–959. [PubMed: 7978722]
2. Carruthers BM, Jain AK, De Meirleir KL, Peterson DL, Klimas NG, Lerner AM, Bested AC, Flor-Henry P, Joshi P, Powles ACP, Sherkey JA, Van de Sande ML. J Chronic Fatigue Syndr. 2003; 11:7–115.
3. I. o. M. (IOM). Beyond myalgic encephalomyelitis/chronic fatigue syndrome: Redefining an illness. The National Academies Press; Washington, D.C: 2015.
4. Jason LA, Bell DS, Rowe K, Van Hoof ELS, Jordan K, Lapp C, Gurwitt A, Miike T, Torres-Harding S, De Meirleir K. J Chronic Fatigue Syndr. 2006; 13:1–44.
5. Eaton KK, Hunnisett A. J Nutr Environ Med. 2004; 14:95–101.
6. Jones MG, Cooper E, Amjad S, Goodwin CS, Barron JL, Chalmers RA. Clin Chim Acta. 2005; 361:150–158. [PubMed: 15992788]
7. McGregor NR, Dunstan RH, Zerbis M, Butt HL, Roberts TK, Klineberg IJ. Biochem Mol Med. 1996; 58:85–92. [PubMed: 8809350]
8. Niblett SH, King KE, Dunstan RH, Clifton-Bligh P, Hoskin LA, Roberts TK, Fulcher GR, McGregor NR, Dunsmore JC, Butt HL, Klineberg I, Rothkirch TB. Exp Biol Med (Maywood). 2007; 232:1041–1049. [PubMed: 17720950]
9. Armstrong CW, McGregor NR, Lewis DP, Butt HL, Gooley PR. Metabolomics. 2015; 11:1626–1639.
10. Demitrack MA, Dale JK, Straus SE, Laue L, Listwak SJ, Kruesi MJ, Chrousos GP, Gold PW. J Clin Endocrinol Metab. 1991; 73:1224–1234. [PubMed: 1659582]
11. Armstrong CW, McGregor NR, Sheedy JR, Butfield I, Butt HL, Gooley PR. Clin Chim Acta. 2012; 413:1525–1531. [PubMed: 22728138]
12. Georgiades E, Behan WM, Kilduff LP, Hadjicharalambous M, Mackie EE, Wilson J, Ward SA, Pitsiladis YP. Clin Sci (Lond). 2003; 105:213–218. [PubMed: 12708966]
13. Suarez A, Guillo E, Roig T, Blazquez A, Alegre J, Bermudez J, Ventura JL, Garcia-Quintana AM, Comella A, Segura R, Javierre C. J Womens Health (Larchmt). 2010; 19:1073–1077. [PubMed: 20469961]
14. Pall ML. J Chronic Fatigue. 2002; 10:37–41.
15. Korszun A, Sackett-Lundeen L, Papadopoulos E, Brucksch C, Masterson L, Engelberg NC, Haus E, Demitrack MA, Crofford L. J Rheumatol. 1999; 26:2675–2680. [PubMed: 10606381]
16. Kuratsune H, Yamaguti K, Lindh G, Evengard B, Takahashi M, Machii T, Matsumura K, Takaishi J, Kawata S, Langstrom B, Kanakura Y, Kitani T, Watanabe Y. Int J Mol Med. 1998; 2:51–56. [PubMed: 9854142]
17. Giloteaux L, Goodrich JK, Walters WA, LSM, Ley RE, Hanson MR. Microbiome. 2016
18. Bell, DS. The Doctor's Guide to Chronic Fatigue Syndrome. Reading, Mass: Addison-Wesley; 1995.
19. Liu X, Ser Z, Locasale JW. Anal Chem. 2014; 86:2175–2184. [PubMed: 24410464]
20. Cheng, J; Galili, T; Rstudio, I; Bostock, M; Palmer, J. d3heatmap: Interactive Heat Maps using “htmlwidgets” and “D3.js”. R package, version 0.6.1.1. 2016. <https://github.com/rstudio/d3heatmap>
21. Breiman L. Mach Learn. 2001; 45:5–32.
22. Liaw, A. randomForest. R package 4.6–12. 2015. <http://www.r-project.org/>
23. Efron, B, Tibshirani, RJ. An Introduction to the Bootstrap. Chapman & Hall/CRC; London, UK: 1993.

24. Hastie, T, Tibshirani, R, Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer; 2009.
25. Sun J, Ando Y, Ahlbory-Dieker D, Schnackenberg LK, Yang X, Greenhaw J, Pence L, Qian F, Salminen WF, Mendrick DL, Beger RD. J Mol Biomark Diagn. 2013:S1.
26. Baskin JM, Wu X, Christiano R, Oh MS, Schauder CM, Gazzero E, Messa M, Baldassari S, Assereto S, Biancheri R, Zara F, Minetti C, Raimondi A, Simons M, Walther TC, Reinisch KM, De Camilli P. Nat Cell Biol. 2016; 18:132–138. [PubMed: 26571211]
27. Zeineh MM, Kang J, Atlas SW, Raman MM, Reiss AL, Norris JL, Valencia I, Montoya JG. Radiology. 2015; 274:517–526. [PubMed: 25353054]
28. Kennedy G, Spence VA, McLaren M, Hill A, Underwood C, Belch JJ. Free Radic Biol Med. 2005; 39:584–589. [PubMed: 16085177]
29. Fulle S, Mecocci P, Fano G, Vecchiet I, Vecchini A, Racciotti D, Cherubini A, Pizzigallo E, Vecchiet L, Senin U, Beal MF. Free Radic Biol Med. 2000; 29:1252–1259. [PubMed: 11118815]
30. Vecchiet J, Cipollone F, Falasca K, Mezzetti A, Pizzigallo E, Bucciarelli T, De Laurentis S, Affaitati G, De Cesare D, Giamberardino MA. Neurosci Lett. 2003; 335:151–154. [PubMed: 12531455]
31. Maes M, Mihaylova I, Leunis JC. Neuro Endocrinol Lett. 2006; 27:615–621. [PubMed: 17159817]
32. Naviaux RK, Naviaux JC, Li K, Bright AT, Alaynick WA, Wang L, Baxter A, Nathan N, Anderson W, Gordon E. Proc Natl Acad Sci U S A. 2016; 113:E5472–5480. [PubMed: 27573827]
33. Yamano E, Sugimoto M, Hirayama A, Kume S, Yamato M, Jin G, Tajima S, Goda N, Iwai K, Fukuda S, Yamaguti K, Kuratsune H, Soga T, Watanabe Y, Kataoka Y. Sci Rep. 2016; 6:34990. [PubMed: 27725700]
34. Nicolson GL, Ellithorpe R. J Chronic Fatigue Syndr. 2006; 13:57–68.

Fig. 1. Distribution of logged metabolic scores for metabolites significantly different between controls and patients (A) Contains the 35 metabolites with $Q < 0.15$ by the Kruskal-Wallis test (B) Contains the 74 metabolites with $P < 0.05$ by the t-test or the Kruskal-Wallis test. The identity of the numbered metabolites can be found in Table 2.

Fig. 2. Dendrograms derived from the heat map analysis, with decreasing number of metabolites used for analysis as statistical criteria are applied to the dataset.

Fig. 3. Relative importance of metabolites. (A) Variable importance plots. All 361 metabolites were used but only the 20 most important are shown. The identity of the numbered metabolites can be found in Table 2. (B) Partial dependence plots. Each plot shows the log-odds for being a patient given the concentration of a single metabolite adjusted for the other 360 metabolites.

Table 1

Characteristics of the study population.

		CONTROLS	PATIENTS
Gender	Female	15	17
Age	Mean $\pm$ SD	51.9 $\pm$ 6.2	53.9 $\pm$ 8.6
	Median (range)	50 (42–61)	50 (43–68)
BMI	Mean $\pm$ SD	26.6 $\pm$ 5.6	25.2 $\pm$ 5.5
	Median (range)	28.3 (17–36)	24 (16–38)
Bell's disability scale	10–20		7
	30–40	ND	8
	50–60		2

Table 2

List of metabolites found to be significantly different between controls and patients according to the t-tests and Kruskal-Wallis tests. Paper ID is the number used throughout the paper for metabolite identification. The same list is found in Supplementary Data 1, along with the remaining metabolites. HMDB ID is the Human Metabolome Database ID used for the pathway analysis. Grey boxes, “???” and ID not in bold were not used for the pathway analysis. Bold *P* and *Q* values (corrected *P* values) meet the *P*<0.05 and *Q*<0.15 criteria used in this study. The trend column with arrows pointing down, displays the metabolites with lower quantity in the patient cohort compared to the controls; while the metabolites without arrows are increased in patients vs. controls. These arrows are a summary of the detailed data presented in Figure 1.

Paper ID	HMDB ID	Metabolite name	Kruskal-Wallis test		Trend	T-test	
			<i>P</i> value	<i>Q</i> value		<i>P</i> value	<i>Q</i> value
9	HMDB12881	Acetylcarnosine	0.0014	0.1075	↓	0.0014	0.1293
17	HMDB00538	ATP	0.0027	0.1075	↓	0.0024	0.1293
19	HMDB01341	ADP	0.0024	0.1075	↓	0.0034	0.1293
21	HMDB00637	Glycochenodeoxycholate	0.0024	0.1075	↓	0.0044	0.1293
26		C3H4O2.3	0.0027	0.1075	↓	0.0050	0.1293
22	HMDB02266	2-Methylglutaconic acid	0.0027	0.1075	↓	0.0097	0.1304
14		C20H34O4.4	0.0027	0.1075	↓	0.0032	0.1293
28	HMDB00251	Taurine	0.0014	0.1075	↓	0.0073	0.1293
5	HMDB12555	13'-carboxy-alpha-tocopherol	0.0018	0.1075	↓	0.0027	0.1293
33	HMDB02231	cis-11-Eicosenoate	0.0062	0.1116		0.0069	0.1293
15	HMDB01014	4-Imidazolone-5-propanoate	0.0049	0.1116	↓	0.0023	0.1293
13	HMDB02639	Sulfoglycolithocholate(2-)	0.0055	0.1116	↓	0.0058	0.1293
27		C4H6O3.3	0.0062	0.1116	↓	0.0074	0.1293
31	HMDB12880	Acetamidopropanal	0.0055	0.1116	↓	0.0164	0.1739
30	HMDB00162	L-proline	0.0055	0.1116	↓	0.0164	0.1739
32		C20H30O2.4	0.0044	0.1116		0.0041	0.1293
29	HMDB00122	D-glucose	0.0055	0.1116	↓	0.0091	0.1293
25	HMDB06293	L-erythrulose	0.0049	0.1116	↓	0.0093	0.1293
35	HMDB02183	Cervonic acid C22:6(n-3), Docosahexaenoate	0.0044	0.1116		0.0042	0.1293
8	???	2,3-epoxy-alpha-tocopherylquinone,5,6-epoxy-alpha-tocopherylquinone	0.0062	0.1116	↓	0.0082	0.1293
24	HMDB03290	L-gulonate	0.0078	0.1167	↓	0.0091	0.1293
23	HMDB00625	D-gluconate	0.0078	0.1167	↓	0.0091	0.1293
11		C20H32O4.21	0.0078	0.1167	↓	0.0034	0.1293

Paper ID	HMDB ID	Metabolite name	Kruskal-Wallis test		Trend	T-test	
			P value	Q value		P value	Q value
3	HMDB00698	Glycolithocholate	0.0078	0.1167	↓	0.0075	0.1293
6	HMDB01413	CDP-choline	0.0087	0.1205	↓	0.0141	0.1640
16	HMDB00138	Glycocholate	0.0087	0.1205	↓	0.0110	0.1374
4	HMDB60126	25-hydroxyvitamin D3-26,23-lactone	0.0097	0.1295	↓	0.0075	0.1293
1	HMDB00761	Lithocholate	0.0102	0.1316	↓	0.0182	0.1873
20	HMDB00119	Glyoxylate	0.0108	0.1345	↓	0.0073	0.1293
34	HMDB01527	3-(methylthio)propionate	0.0120	0.1448		0.0299	0.2297
18	HMDB01565	Choline phosphate(1-)	0.0134	0.1464	↓	0.0085	0.1293
10	HMDB61717	Succinylcarnitine	0.0134	0.1464	↓	0.0271	0.2220
12	HMDB06815	5-guanidino-2-oxopentanoic acid	0.0134	0.1464	↓	0.0050	0.1293
2	HMDB60079	[2R,3S,4R,5R)-5-(2,4-dioxo-1,2,3,4-tetrahydropyrimidin-1-yl)-3,4-dihydroxyoxolan-2-yl]methyl [(3R,4S,5S,6R)-3-trihydroxy-6-(hydroxymethyl)oxan-2-yl]phosphonato-oxy-phosphonate-UDP-D-galactose	0.0361	0.2413	↓	0.0110	0.1374
7	HMDB06868	S-(2-Methylpropanoyl)-dihydrolipoamide	0.0183	0.1691	↓	0.0064	0.1293
52	HMDB00224	Ethanolamine phosphate	0.0149	0.1578	↓	0.0201	0.1911
45	???	S-[2-Carboxy-1-(1H-imidazol-4-yl)ethyl]-L-cysteine	0.0183	0.1691	↓	0.0153	0.1728
74	HMDB00207	Trans-vaccenate-elaidate-oleate	0.0165	0.1691		0.0375	0.2460
46	HMDB60137	3alpha,7alpha,12alpha,25-Tetrahydroxy-5beta-cholestane-24-one/3alpha,7alpha,12alpha-Trihydroxy-5beta-cholestanolate	0.0183	0.1691	↓	0.0136	0.1631
43	HMDB00859	O-methylhippurate	0.0183	0.1691	↓	0.0903	0.3691
55	HMDB00684/HMDB12948	L-kynurenine/Formyl-5-hydroxykynurenamine	0.0223	0.1875	↓	0.0201	0.1911
44	HMDB60278	Fructoseglycine	0.0223	0.1875	↓	0.0332	0.2405
36	HMDB00223	Oxaloacetate	0.0223	0.1875	↓	0.0396	0.2466
67	HMDB60268	N-acetylputrescinium	0.0223	0.1875		0.0598	0.2955
49		C27H46O4.3	0.0246	0.2022	↓	0.0259	0.2219
53	HMDB00763	5-Hydroxyindoleacetate	0.0272	0.2179	↓	0.0328	0.2405
69	HMDB00079	5,6-Dihydrothymine	0.0299	0.2297		0.0259	0.2219
59	HMDB00301	Urocanate	0.0296	0.2297		0.0392	0.2466
62	HMDB06476	N2-Formyl-N1-(5-phospho-D-ribose)glycinamide	0.0361	0.2413		0.0361	0.2454
66	HMDB60287	4-Hydroperoxy-2-nonenal	0.0361	0.2413		0.0264	0.2219
48	HMDB00425	2-Keto-3-deoxy-D-glycero-D-galacto-nononic acid	0.0361	0.2413	↓	0.0430	0.2587
61	HMDB00126	Glycero-3-phosphate	0.0329	0.2413		0.0333	0.2405

Paper ID	HMDB ID	Metabolite name	Kruskal-Wallis test		Trend	T-test	
			P value	Q value		P value	Q value
57	HMDB00172/HMDB00687	L-Isoleucine, L-Leucine	0.0361	0.2413	↓	0.0482	0.2674
56	HMDB00895/HMDB06831	Acetylcholine, 4-(trimethylammonio)butanoate	0.0361	0.2413	↓	0.1438	0.4351
71	HMDB60102	Arachidonate, Eicosatetraenoic acid	0.0396	0.2552		0.0240	0.2219
51	HMDB12557	13-Carboxy-gamma-tocopherol	0.0396	0.2552	↓	0.0475	0.2674
64	HMDB00795	Pristanic acid, pristanate	0.0474	0.2634		0.0290	0.2297
37	HMDB03976	D-Glucuronate 1-phosphate	0.0434	0.2634	↓	0.0195	0.1911
70	HMDB02259	Margarate	0.0434	0.2634		0.0742	0.3283
73	HMDB00220	Palmitate	0.0474	0.2634		0.1172	0.4069
72	HMDB00827	Stearate	0.0474	0.2634		0.0806	0.3424
54	HMDB01406	Nicotinamide	0.0474	0.2634	↓	0.0396	0.2466
60	HMDB00792	Sebacic acid	0.0474	0.2634		0.0562	0.2955
39	HMDB60176	(S)-3-sulfonatolactate(2-)	0.0474	0.2634	↓	0.4017	0.6664
40	HMDB03501/HMDB00707	Trans-caffeate/(3,4-hydroxyphenyl)pyruvate	0.0434	0.2634	↓	0.3189	0.6135
47	???	3-oxo-8(R)-hydroxy-hexadeca-6E10Z-dienoate_3-oxo-8(S)-hydroxy-hexadeca-6E10Z-dienoate	0.0518	0.2750	↓	0.0264	0.2219
58	HMDB00012	Deoxyuridine	0.0639	0.2915		0.0296	0.2297
68	HMDB06528/ HMDB60113	Clupanodonic acid docosa-4,7,10,13,16-pentaenoic acid	0.0616	0.2850		0.0341	0.2416
38	HMDB00030	Biotin	0.0518	0.2750	↓	0.0363	0.2454
65	HMDB00020/HMDB00669 /HMDB03791	4-Hydroxyphenylacetate/2-Hydroxyphenylacetate/3,4-Dihydroxyphenylacetaldehyde	0.0518	0.2750		0.0430	0.2587
50	HMDB61189	3-Hydroxyisovalerylcarnitine	0.0791	0.3104	↓	0.0460	0.2661
41	HMDB00285	UTP	0.0616	0.2850	↓	0.0461	0.2661
63	HMDB01257	Spermidine	0.0616	0.2850		0.0464	0.2661
42	HMDB01439	5-amino-1-(5-phospho-D-ribose)imidazole-4-carboxamide	0.1171	0.3493	↓	0.0367	0.2454

Table 3

Confusion matrix from the random forest algorithm using all 361 metabolites.

True	Predicted categories		Classification error
	CONTROLS	PATIENTS	
CONTROLS	10	5	0.333
PATIENTS	5	12	0.294

Table 4

Correlation matrix of the logarithms of the four most important metabolite concentrations.

Metabolites	19	5	28	17
19	1.000	0.296	0.800	0.798
5	0.296	1.000	0.344	0.287
28	0.800	0.344	1.000	0.632
17	0.798	0.287	0.632	1.000

Table 5

Pathway analysis output from www.metaboanalyst.ca with match status, which includes the number of hits compared to the number of metabolites in the whole pathway and pathway impact. The pathways highlighted in orange can be grouped under the larger group of lipid metabolism, the ones in green as amino-acid metabolism and the ones in blue as energy metabolism.

Pathway name	Match Status	Impact
Glycerophospholipid metabolism	5/39	0.1921
Purine metabolism	4/92	0.09849
Tryptophan metabolism	3/79	0.04639
Arginine and proline metabolism	3/77	0.12723
Tyrosine metabolism	3/76	0.10175
Aminoacyl-tRNA biosynthesis	3/75	0
Pyrimidine metabolism	3/60	0.06534
Fatty acid biosynthesis	3/49	0
Primary bile acid biosynthesis	3/47	0.0266
Amino sugar and nucleotide sugar metabolism	2/88	0.00686
Pentose and glucuronate interconversions	2/53	0.31847
Glyoxylate and dicarboxylate metabolism	2/50	0.24855
Phenylalanine metabolism	2/45	0.01045
Ascorbate and aldarate metabolism	2/45	0.00784
Histidine metabolism	2/44	0.13561
Valine, leucine and isoleucine degradation	2/40	0.02232
Pentose phosphate pathway	2/32	0.08639
Glycolysis or Gluconeogenesis	2/31	0
Valine, leucine and isoleucine biosynthesis	2/27	0.0265
Cysteine and methionine metabolism	1/56	0
Fatty acid metabolism	1/50	0.02959
Starch and sucrose metabolism	1/50	0.01703
Glycine, serine and threonine metabolism	1/48	0
Nicotinate and nicotinamide metabolism	1/44	0.03827
Galactose metabolism	1/41	0.00276
Nitrogen metabolism	1/39	0
Glutathione metabolism	1/38	0.03608
Ubiquinone and other terpenoid-quinone biosynthesis	1/36	0.04886
Pyruvate metabolism	1/32	0
Glycerolipid metabolism	1/32	0.00714
Beta-Alanine metabolism	1/28	0.06625
Phenylalanine, tyrosine and tryptophan biosynthesis	1/27	0
Fatty acid elongation in mitochondria	1/27	0
Sphingolipid metabolism	1/25	0.01288

Pathway name	Match Status	Impact
Alanine, aspartate and glutamate metabolism	1/24	0.0057
Caffeine metabolism	1/21	0
Taurine and hypotaurine metabolism	1/20	0.33094
Citrate cycle (TCA cycle)	1/20	0.13145
Biotin metabolism	1/11	0.20325

Table 6

Comparison of pathways found to be affected by Naviaux et al.³² and our results.

612	Metabolites	361
44	Female cohort size	32
37 metabolites	Significance (Q<0.15)	35 metabolites
84% decreased	Hypometabolic state	89% decreased
Affected Pathways		
Common		
Sphingolipids	Phospholipids	Other AA
Glycosphingolipids	Purines	Energy metabolism
Microbiome	P5C, Argininw, Proline	Sugar metabolism
Cholesterol	Branch Chain AA	Fatty acid metabolism
Vitamin B2 (Riboflavin)	Fatty Acid Oxidation	Vitamin B7 (Biotin)
Vitamin C/Collagen	Biles Acids	Other metabolites (e.g. Taurine, Vitamin E pathway)
Endocannabinoids	Amino Sugars	
Vitamin B12		