



OPEN

## Deep phenotyping of myalgic encephalomyelitis/chronic fatigue syndrome in Japanese population

Toshimori Kitami<sup>1,12</sup>✉, Sanae Fukuda<sup>2,3,4,12</sup>, Tamotsu Kato<sup>1,12</sup>, Kouzi Yamaguti<sup>2</sup>, Yasuhito Nakatomi<sup>2,5</sup>, Emi Yamano<sup>2,4,6</sup>, Yosky Kataoka<sup>2,4,6,7</sup>, Kei Mizuno<sup>2,4,6</sup>, Yuuri Tsuboi<sup>8</sup>, Yasushi Kogo<sup>9</sup>, Harukazu Suzuki<sup>1</sup>, Masayoshi Itoh<sup>9</sup>, Masaki Suimye Morioka<sup>1</sup>, Hideya Kawaji<sup>1,9</sup>, Haruhiko Koseki<sup>1</sup>, Jun Kikuchi<sup>8,10,11,13</sup>, Yoshihide Hayashizaki<sup>9,13</sup>, Hiroshi Ohno<sup>1,11,13</sup>, Hirohiko Kuratsune<sup>2,3,4,5,6,13</sup> & Yasuyoshi Watanabe<sup>2,4,6,13</sup>✉

Myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) is a complex and debilitating disease with no molecular diagnostics and no treatment options. To identify potential markers of this illness, we profiled 48 patients and 52 controls for standard laboratory tests, plasma metabolomics, blood immuno-phenotyping and transcriptomics, and fecal microbiome analysis. Here, we identified a set of 26 potential molecular markers that distinguished ME/CFS patients from healthy controls. Monocyte number, microbiome abundance, and lipoprotein profiles appeared to be the most informative markers. When we correlated these molecular changes to sleep and cognitive measurements of fatigue, we found that lipoprotein and microbiome profiles most closely correlated with sleep disruption while a different set of markers correlated with a cognitive parameter. Sleep, lipoprotein, and microbiome changes occur early during the course of illness suggesting that these markers can be examined in a larger cohort for potential biomarker application. Our study points to a cluster of sleep-related molecular changes as a prominent feature of ME/CFS in our Japanese cohort.

Myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) is a complex and debilitating disease with a spectrum of symptoms including unexplained fatigue, post-exertional malaise, impaired memory, pain, gastrointestinal and immune dysfunction, and sleep disturbance<sup>1,2</sup>. About 0.2 to 2.6% of the population, 75% being female, are estimated to be affected by ME/CFS<sup>3,4</sup> with no treatment option, resulting in depression, absence from work, and social isolation. ME/CFS is currently diagnosed based on symptoms and there are currently no accepted molecular diagnostic tools<sup>5</sup>. This poses challenges to medical practitioners who often rely on molecular diagnostic tools and physical signs of illness for making diagnostic decisions.

In order to identify potential diagnostic markers for ME/CFS, researchers have relied on omics technologies to systematically profile molecular changes associated with ME/CFS. One of the first omics technologies to be applied was microarray, which measures the gene expression changes across thousands of genes in the genome. This technology led to the identification of gene expression markers for ME/CFS, particularly in peripheral blood mononuclear cells (PBMCs)<sup>6,7</sup>. However, a validation study in a different cohort failed to robustly separate patients from controls<sup>8</sup>. Additionally, a study in twin pairs has shown that no gene expression difference could be detected between ME/CFS patients and controls when controls were genetically matched<sup>9</sup>, suggesting that gene expression changes do not strongly reflect the disease state of ME/CFS.

The use of omics technologies in ME/CFS studies have recently shifted to metabolomics and microbiome analysis. Metabolomics studies using gas chromatography–, liquid chromatography–, or capillary electrode–mass

<sup>1</sup>RIKEN Center for Integrative Medical Sciences, Kanagawa, Japan. <sup>2</sup>Osaka City University Graduate School of Medicine, Osaka, Japan. <sup>3</sup>Kansai University of Welfare Sciences, Osaka, Japan. <sup>4</sup>RIKEN Center for Biosystems Dynamics Research, Hyogo, Japan. <sup>5</sup>Nakatomi Fatigue Care Clinic, Osaka, Japan. <sup>6</sup>RIKEN Compass to Healthy Life Research Complex Program, Hyogo, Japan. <sup>7</sup>RIKEN Baton Zone Project, RIKEN-JEOL Collaboration Center, Hyogo, Japan. <sup>8</sup>RIKEN Center for Sustainable Resource Sciences, Kanagawa, Japan. <sup>9</sup>RIKEN Preventive Medicine and Diagnosis Innovation Program, Saitama, Japan. <sup>10</sup>Graduate School of Bioagricultural Sciences, Nagoya University, Aichi, Japan. <sup>11</sup>Graduate School of Medical and Life Sciences, Yokohama City University, Kanagawa, Japan. <sup>12</sup>These authors contributed equally: Toshimori Kitami, Sanae Fukuda and Tamotsu Kato. <sup>13</sup>These authors jointly supervised this work: Jun Kikuchi, Yoshihide Hayashizaki, Hiroshi Ohno, Hirohiko Kuratsune and Yasuyoshi Watanabe. ✉email: toshimori.kitami@riken.jp; yywata@riken.jp

spectrometry (GC–MS, LC–MS, or CE–MS) in plasma samples have identified sets of metabolites that can robustly distinguish ME/CFS patients from healthy controls<sup>10–14</sup>. However, more studies are needed to identify common sets of metabolites that are altered in ME/CFS patients as the number of studies are still few. In comparison to metabolomics study, a larger number of fecal microbiome analysis have been conducted for ME/CFS, which is summarized in a recent review<sup>15</sup>. Among the key differences detected, a decrease in *Faecalibacterium* appears to be shared across several studies of ME/CFS. In addition to these technologies, global cytokine profiling<sup>16–19</sup> and immunophenotyping<sup>20,21</sup> have been performed in plasma and cerebrospinal fluid samples to identify immune biomarkers of ME/CFS. More recently, a proteomics survey of extracellular vesicles (EVs) was performed to identify EV-specific protein biomarker of ME/CFS<sup>22</sup>. All of these technologies survey biological samples from the periphery, which are readily accessible to medical practitioners.

In parallel to the advancement in omics technologies targeting the periphery, advances in imaging technologies have revealed key insights into molecular changes that occur in the brain of ME/CFS patients<sup>23–29</sup>. These studies have uncovered neuroinflammation<sup>30</sup> related to the microglial activation by cytokines, neurotransmitter abnormalities<sup>31,32</sup>, costly and less efficient performance of frontal cortex<sup>33</sup>, as well as ‘brain fog’<sup>34</sup>. These imaging technologies require trained specialists to perform the study and omics technologies still represent a more accessible platform for molecular diagnostics in most laboratories.

Despite the accumulation of omics data in ME/CFS, key challenges remain in integrating and interpreting these data. First, the performance of omics platforms cannot be compared to each other as the studies are conducted in different sets of cohorts. Second, one cannot relate changes in molecular profile at one omics level, such as metabolite, to another omics level such as microbiome, when patients differ between studies. Therefore, in order to address these challenges, we set out to perform a deep phenotyping of ME/CFS using five molecular profiling platforms, accompanied by questionnaire and quantitative measures of fatigue. From our study, we identified 26 potential markers that distinguished ME/CFS patients from healthy controls. We found that markers from immunophenotyping, microbiome analysis, and lipoprotein profiling performed best in distinguishing patients from controls. When combined, these markers did not completely separate patients from controls suggesting limitations of our profiling technologies. We also uncovered strong correlation between sleep disruption, lipoprotein profiles, and microbiome abundance suggesting that these markers form one of the core networks of ME/CFS. These changes were evident during the early course of illness suggesting that these markers may be studied in a larger cohort for biomarker application. Our study points to a cluster of sleep-related molecular changes as a prominent feature of ME/CFS in our Japanese cohort.

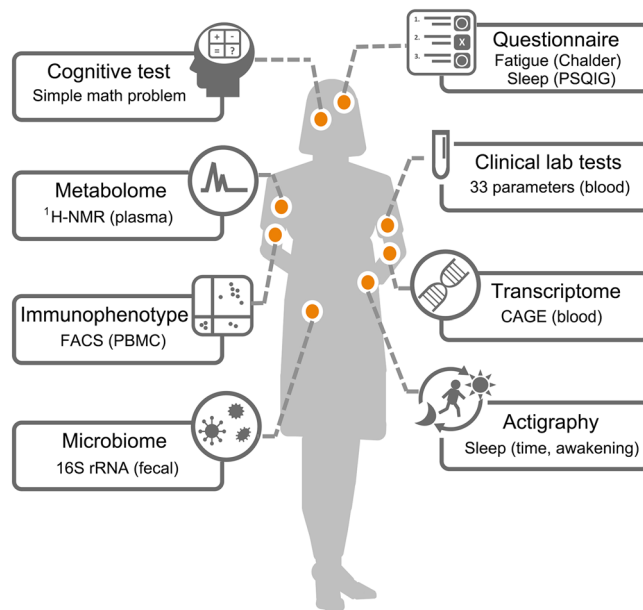
## Results

**Overview of the study.** The goal of our study was twofold: (1) identify potential markers of myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) across multiple omics platforms; and (2) uncover relationships between these markers for insights into the syndrome. We recruited 48 ME/CFS patients and 52 healthy controls (Table S1) matched for age, gender, and BMI. ME/CFS patients were diagnosed based on the 1994 Center for Disease Control clinical criteria (Fukuda criteria)<sup>1</sup> and the International Consensus Criteria<sup>2</sup>. These cohorts underwent questionnaires, activity measurements, simple cognitive tests, as well as multi-omics profiling of blood and fecal samples (Fig. 1). To assess the severity of fatigue, we used Chalder fatigue scale<sup>35</sup> and quality of sleep using Pittsburgh Sleep Quality Index Global (PSQIG) score<sup>36</sup>. To quantitatively measure sleep, we also performed 1-week actigraphy measurements<sup>37</sup>. For cognitive assessment, we administered simple mathematical problems. For molecular profiling, blood samples were taken in the morning after overnight fast and assessed for standard clinical laboratory tests, <sup>1</sup>H-NMR metabolomics of plasma, FACS-based immunophenotyping of peripheral blood mononuclear cell (PBMC), and transcriptome analysis of whole blood. Fecal samples were collected at home and were sent for analysis of microbiota composition using 16S rRNA sequencing. Given the technical difficulties in several of the omics measurements as well as difficulties in obtaining enough patient materials, some of the measurements could not be completed (Fig. S1). For each platform, we used all available data to identify markers that distinguished ME/CFS patients from healthy controls. To compare markers across platforms, we restricted the analysis to individuals with complete data. In total, we evaluated 33 standard clinical laboratory tests (Table S1), 20 types of immune cell (Fig. S2), 31 metabolite profiles (Table S2), eight lipoprotein fractions (Table S3), expression levels of 820 gene sets, and relative abundance of 20 different bacterial genera (Fig. S3), resulting in over 70,000 data points.

**Molecular phenotyping.** We first searched for molecular differences between ME/CFS patients and healthy controls for each of the platforms (Fig. 2). We identified differences in nine blood-based clinical laboratory tests (Fig. 2, Table S1), eight lipoprotein fractions (Fig. 2, Table S3), two immune cell types (Fig. 2, Fig. S2), and six different bacterial genera (Fig. 2, Fig. S3A) after adjusting for false discovery rate to 0.20 for each of the molecular platform. For microbiome analysis, alpha diversity did not significantly differ between patients and controls (Fig. S3B).

For whole blood transcriptomics, given that the expression level of individual genes is subject to technical and biological noise, we searched for sets of genes that were differentially expressed between patients and controls. This method provides greater statistical power and ease of biological interpretation as gene sets are defined using prior knowledge of pathways<sup>38,39</sup>. We used gene set annotations from pathway databases as well as list of genes that were differentially expressed in five prior studies of ME/CFS<sup>8,9,40–42</sup>. One gene set showed significant increase in ME/CFS patients compared to controls after adjusting for false discovery rate to 0.20 (Fig. S4). This gene set was previously found to be upregulated in chronic fatigue syndrome patients<sup>41</sup>.

Among the 26 molecular differences we identified, several of the markers were also discovered in other studies of ME/CFS. These included a decrease in uric acid<sup>12</sup> and HDL<sup>43</sup>, and an increase in triglyceride<sup>13,43</sup> for ME/



**Figure 1.** Deep phenotyping of myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). Schematics of the datasets collected from myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) patients and healthy controls.

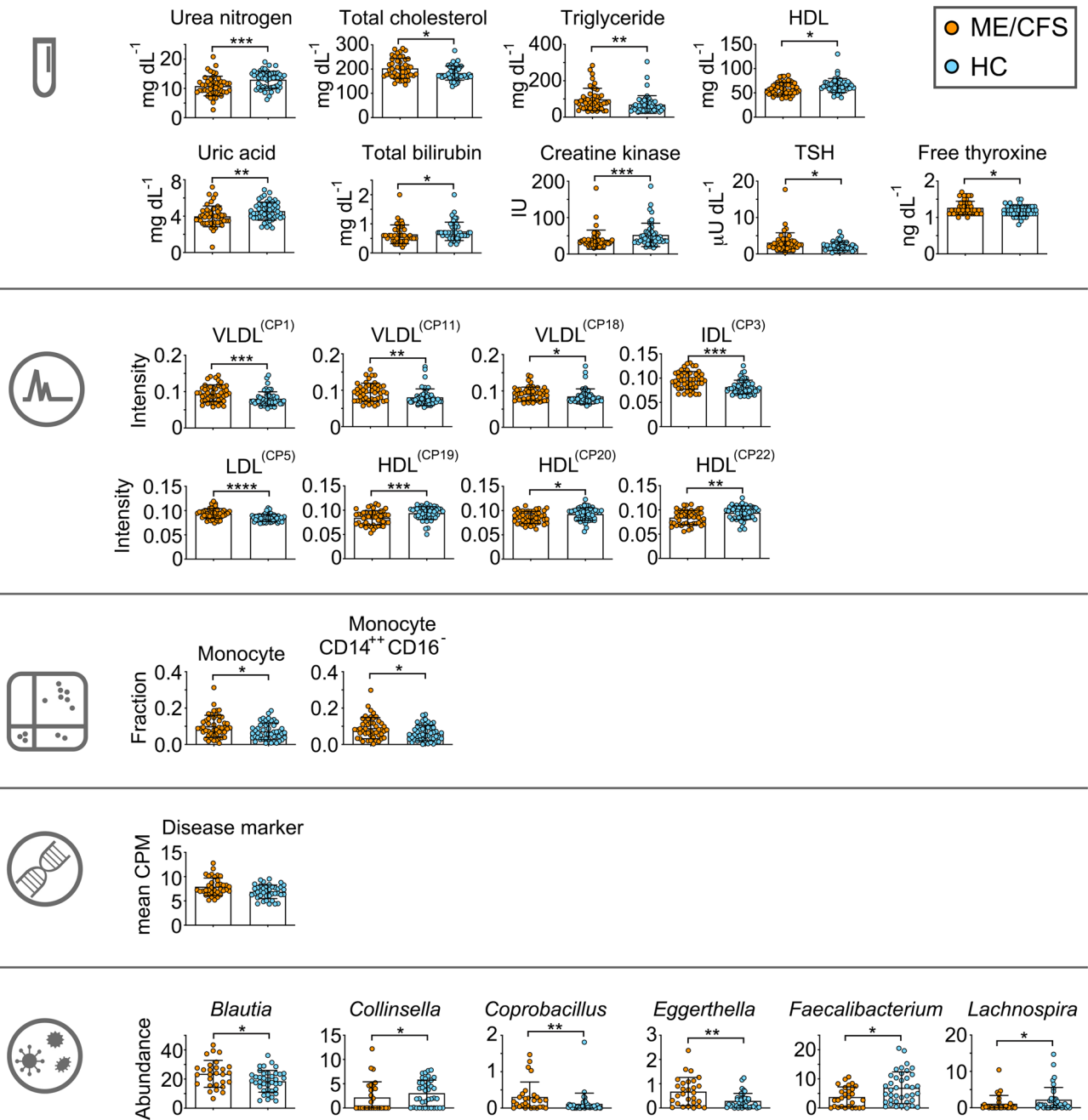
CFS patients. An increase in monocyte number in ME/CFS patients were also identified in one study<sup>21</sup> although several other studies did not find such difference<sup>44,45</sup>. Additionally, a decrease in *Faecalibacterium*<sup>46–48</sup> and an increase in *Coprobacillus*<sup>46,48</sup> were found in ME/CFS patients in previous studies as well as in our current study. Overall, we identified 26 potential markers of ME/CFS from five different molecular platforms including those identified in prior studies.

**Multi-marker analysis.** We next examined whether a combination of 26 markers can help distinguish ME/CFS patients from healthy controls. We used partial least square discriminatory analysis (PLS-DA) to project maximum separation between patients and controls using our 26 input markers. We found that while a subset of patients can be separated from controls (Fig. 3A), about half of the patients could not be separated from controls suggesting limitations of our molecular profiling platforms or limitations of biological signals present in peripheral blood or fecal samples.

We next searched for molecular markers that best separated patients from controls using variable importance in projection (VIP) analysis (Fig. 3B). VIP scores estimate the importance of each marker in the PLS-DA projection. Among the 26 markers we identified, monocyte number, abundance of *Coprobacillus*, *Eggerthella*, and *Blautia*, and levels of triglyceride and VLDL performed best in distinguishing patients from controls (Fig. 3B). In particular, three out of six markers from microbiome analysis appeared at the top of the list. This may result from some of the bacterial genera being almost undetectable in one group but present in another group, thus giving rise to a robust separation between patients and controls. Overall, we found that while patients and controls could not be completely separated using our 26 molecular markers, the top markers came from immunophenotype, microbiome, and metabolomic platforms.

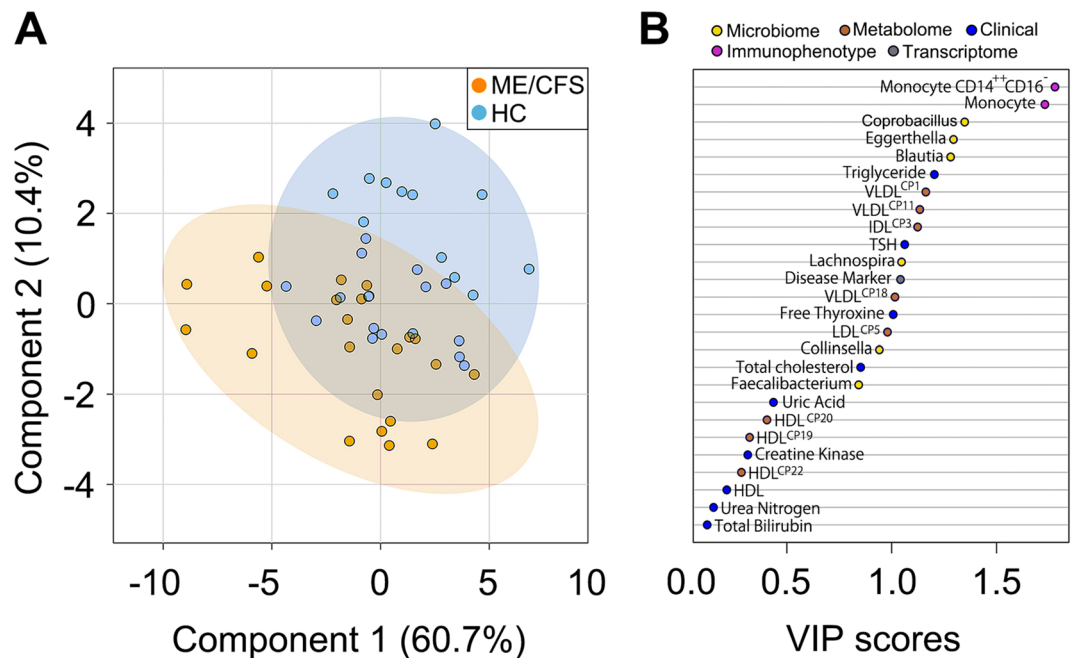
**Non-molecular measurements of fatigue.** In order to understand the relationship between fatigue and our 26 molecular markers, we quantified fatigue by Chalder fatigue scale, a questionnaire-based method that measures both physical and psychological fatigue<sup>35</sup>. We also assessed quality of sleep using questionnaire-based instrument, PSQIG score<sup>36</sup>. As expected, questionnaire-based assessment of fatigue and sleep quality showed the largest separation between patients and controls (Fig. S5), much more so than any of the molecular and non-molecular measurements we made. This is likely because ME/CFS is diagnosed based on self-reported symptoms of fatigue including unrefreshing sleep. To more objectively and quantitatively assess symptoms related to ME/CFS, we measured patterns of sleep and physical activity using actigraphy<sup>37</sup>. Among the four main measurements from actigraphy, we found the largest difference in number of awakening during total sleep and total sleep time within an average 24 h period (Fig. S5). To quantify cognitive performance, we administered a mathematical test consisting of simple addition of two numbers. We observed a difference in average time to solve math problem but not for percent of correct answer (Fig. S5). We therefore selected these measurements of fatigue-related parameters in our subsequent correlation analysis.

**Correlation network.** To understand the relationships between fatigue-related measurements and our 26 molecular markers, we performed Spearman rank correlation analysis (Fig. 4). We uncovered some of the



**Figure 2.** Molecular markers of ME/CFS. Top 26 molecular markers of myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) across five platforms (from the top: clinical lab tests, metabolome, immunophenotype, transcriptome, microbiome). For transcriptome data, gene sets with significant difference between ME/CFS patients and healthy controls (HC) (Fig. S4) are represented with geometric mean of the gene set expression level for illustrative purpose. *P* values were determined by two-tailed Mann–Whitney U-test. \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001, \*\*\*\**P* < 0.0001. *P*-values were corrected for multiple testing using Benjamini–Hochberg false discovery rate (FDR) method after FDR adjustment at 0.20. *P* values for transcriptome data using Gene Set Enrichment Analysis (GSEA) are indicated in Fig. S4. The number of ME/CFS patients and controls for each platform are summarized in Fig. S1.

expected correlations including positive correlation between triglyceride (clinical laboratory test) and LDL (metabolomics), negative correlation between HDL and LDL, and positive correlation between HDL from clinical laboratory test and HDL from <sup>1</sup>H-NMR metabolomics. We also identified strong positive correlation between *Coprobacillus* and *Eggerthella* from the microbiome platform. Therefore, a handful of our 26 molecular markers captured overlapping information rather than 26 distinct sets of information. We also observed strong positive correlations between questionnaire-based scores (Chalder, PSQIG) and measures of sleep (total sleep awakening, total sleep time) and cognitive performance (time to solve math problem) (Fig. 4). This suggested



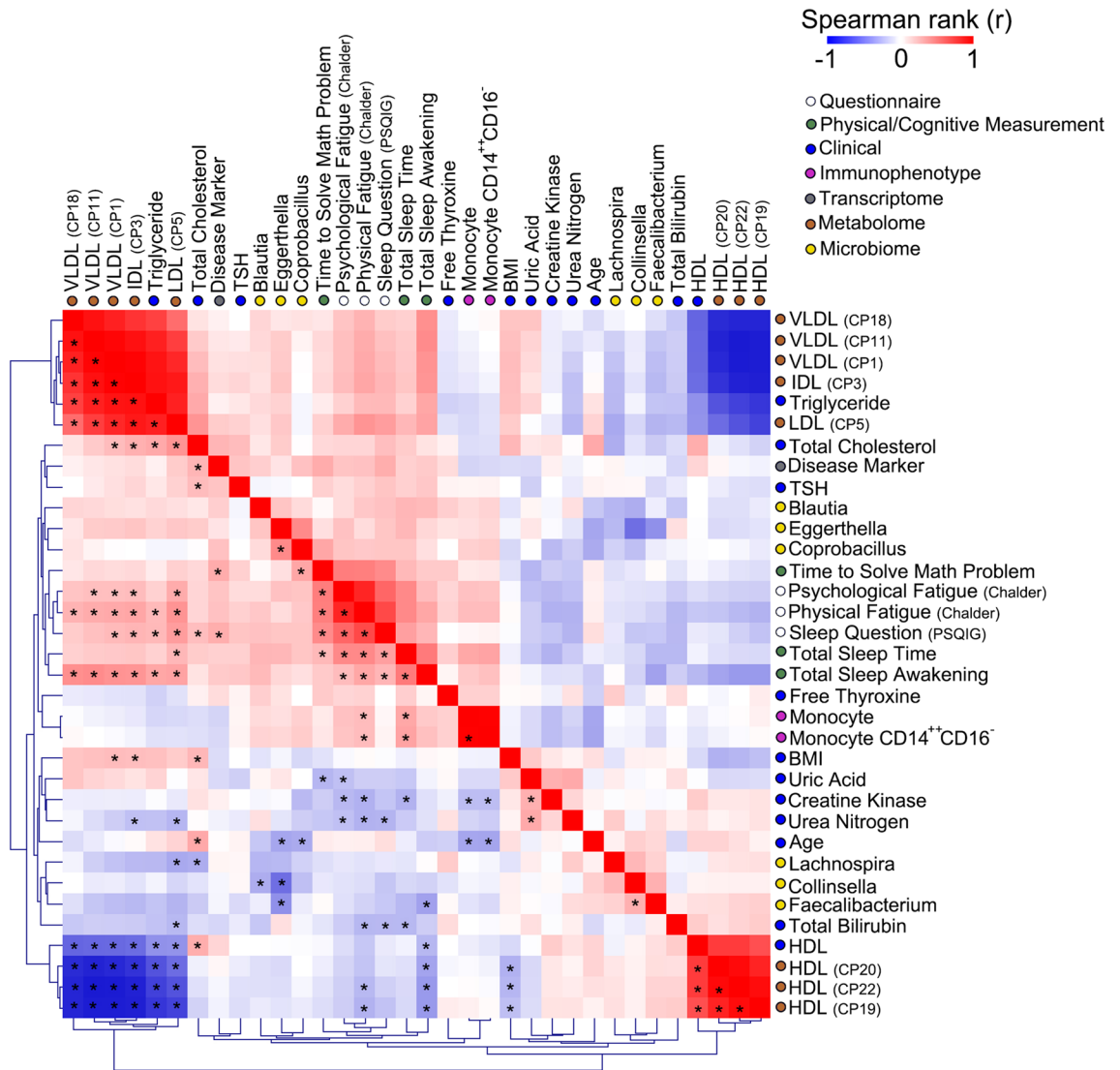
**Figure 3.** Combinatorial analysis of molecular markers. Combination of top 26 molecular markers for distinguishing myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) patients from healthy controls (HC). **(A)** Partial least squares discriminant analysis (PLS-DA) of top 26 molecular markers. **(B)** Variable importance of projection (VIP) scores for distinguishing ME/CFS patients from HC based on component 1. ME/CFS patients (n = 22) and HC (n = 29) with complete molecular profiling across five platforms (clinical lab tests, metabolome, immunophenotype, transcriptome, microbiome) were used for the analysis.

that self-assessment of fatigue and sleep quality were reflected in quantitative measures of sleep and cognitive performance using actigraphy and simple math tests.

Next, we asked which of the 26 molecular markers more closely correlated with measurements of sleep and cognitive performance (Fig. 5A). These two measures represent more objective assessment of symptoms associated with ME/CFS compared to questionnaire-based measures of sleep quality and psychological fatigue. We found that lipoprotein levels most closely correlated with number of awakening during sleep while monocyte number most closely correlated with the length of sleep (Fig. 5B). These results suggested that different molecular markers were associated with different aspects of sleep. Prior studies have shown that sleep apnea or insufficient sleep were associated with decreased HDL level<sup>49,50</sup> suggesting that lipoprotein levels may reflect patterns of sleep disruption. We also observed strong negative correlation between *Faecalibacterium* abundance and number of awakening during sleep (Fig. 5B). For cognitive performance (time to solve math problems), a different set of markers showed strong correlation compared to those correlated with sleep. These included *Coprobacillus* and gene set from prior chronic fatigue syndrome study. These results suggested that lipoprotein levels, monocyte number, and *Faecalibacterium* abundance may reflect sleep-related changes in ME/CFS patients.

**Medication and duration of illness.** We examined whether any of the 26 molecular markers in our study were affected by the use of medications. Among the ME/CFS patients with information on medication, those taking antidepressants (n = 9) showed no significant difference in 25 of the molecular markers when compared to patients not taking antidepressants (n = 36) (Mann–Whitney U-test;  $P > 0.05$ ). The exception was a gene set that was previously elevated in chronic fatigue syndrome (CFS) patients (Fig. S6A). ME/CFS patients on antidepressants, when compared to patients not on antidepressants, showed higher expression level of the CFS marker gene set. Given that the original study in which the gene set was derived allowed patients to continue taking their prescribed medication<sup>45</sup>, the gene set may reflect the use of antidepressants among ME/CFS patients rather than the changes resulting from the syndrome itself. Antidepressant use did not affect sleep parameters or cognitive performance (Fig. S6B). We also found that ME/CFS patients on sleeping pill (n = 9) showed lower total cholesterol level compared to patients not taking sleeping pill (n = 36) (Fig. S6C). Interestingly, the actigraphy measurement of sleep did not differ with the use of sleeping pill suggesting that the medication did not improve abnormal sleep but only restored total cholesterol level to healthy control level. We did not observe differences in the remaining 25 molecular markers based on the use of sleeping pill. We note that our analysis of medication use is limited as the number of patients on each medication was few.

To assess the diagnostic potential of our molecular markers, we examined whether our 26 molecular markers still showed significant difference during short-duration of ME/CFS. We separated patients into two groups ( $\leq 3$  years,  $> 3$  years) similar to prior biomarker study of ME/CFS<sup>18</sup>. We found that lipoprotein markers as well

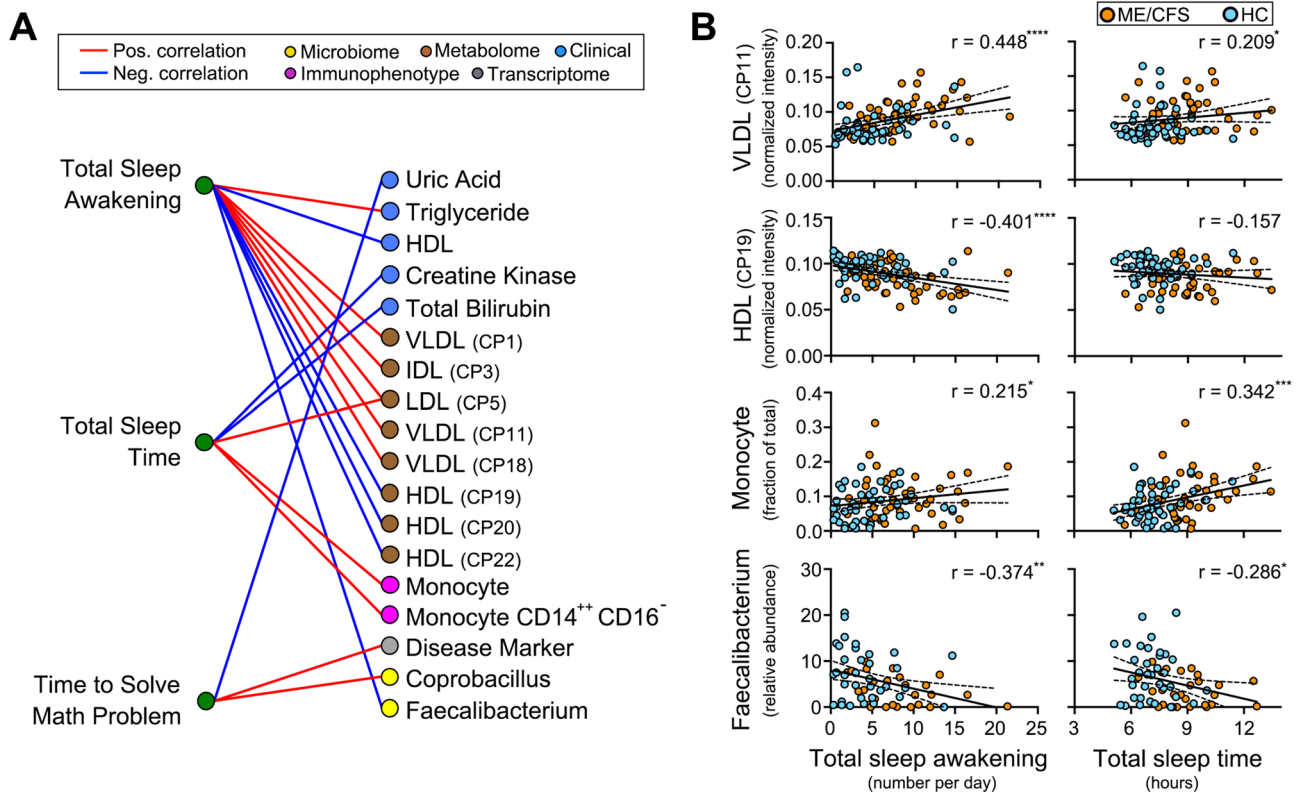


**Figure 4.** Correlation between top markers of ME/CFS. Blue and red colors indicate Spearman rank correlation value between a pair of markers. Stars (\*) denote Spearman rank *P*-value of *P* < 0.05 after adjusting for multiple hypothesis testing using Benjamini–Hochberg false discovery rate (FDR) method set at FDR of 0.05. Circles denote type of data and molecular platform used. Correlation values were clustered using average linkage hierarchical clustering. The number of samples available for a given pair of measurement platforms, as described in Fig. S1, were used for the analysis.

as sleep and cognitive measurements were already different during short-duration ( $\leq 3$  years) of ME/CFS and were still different during long-duration ( $> 3$  years) of illness (Fig. S7). We also found that microbiome changes were much more prominent during the short-duration of illness ( $\leq 3$  years) compared to long-duration of illness ( $> 3$  years) suggesting that microbiome markers may reflect temporary changes during the early phase of ME/CFS (Fig. S7). For monocyte number, we did not observe significant changes during the short-duration of illness ( $\leq 3$  years). Overall, we found that sleep, lipoprotein, and microbiome changes occurred during the early course of illness, an important criterion for future biomarker application.

### Discussion

To gain insights into molecular markers that best distinguish myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) patients from healthy controls, we performed deep phenotyping of cases and controls using five molecular platforms, sleep and cognitive measurements, and questionnaire-based assessment of fatigue. Our analysis identified 26 potential molecular markers of ME/CFS. Among these, monocyte number, microbiome profiles, and lipoprotein profiles provided the highest information value in distinguishing patients from controls. Although these 26 markers could not completely separate patients from controls, a subset of these molecular platforms such as lipoprotein profiles and microbiome analysis, showed difference even during the early phases of illness ( $\leq 3$  years). These platforms can be extended to a larger cohort with longitudinal design to assess the diagnostic performance of markers. Given that lipoprotein profiling is part of a routine clinical test and that



**Figure 5.** Correlation between measured phenotypes and molecular markers of ME/CFS. **(A)** Spearman rank correlation between three measures related to fatigue (total sleep awakening, total sleep time, time to solve math problem) and molecular markers. Spearman rank correlation with  $P < 0.05$  are indicated with red (positive correlation) or blue line (negative correlation). **(B)** Pairwise plot of sleep parameters versus molecular markers. Solid lines are regression lines and dotted lines are 95% confidence interval for the slope. Spearman rank correlation value ( $r$ ) and corresponding  $P$  values are indicated.  $^*P < 0.05$ ,  $^{**}P < 0.01$ ,  $^{***}P < 0.001$ ,  $^{****}P < 0.0001$ .  $P$ -values were corrected for multiple testing using Benjamini–Hochberg false discovery rate (FDR) method after FDR adjustment at 0.20. The number of samples available for a given pair of measurement platforms, as described in Fig. S1, were used for the analysis.

microbiome analysis is starting to be offered as a diagnostics service, these two platforms may be a promising diagnostic tool for ME/CFS.

One of the key advantages of deep phenotyping is that we can now start to uncover relationships between markers across platforms to gain insights into the syndrome. We found that self-assessment of fatigue (Chalder fatigue scale) was most closely correlated with more objective measurements of sleep (actigraphy) and cognitive performance (simple mathematical test). Interestingly, some of our top markers including elevated LDL were more closely correlated with sleep abnormality (sleep awakening) compared to cognitive performance. Previous study has shown that statin therapy, which is used to lower LDL level, is associated with decreased number of awakening during sleep<sup>51</sup> suggesting that there may be a mechanistic link between sleep and lipoprotein profile. With unrefreshing sleep being one of the symptoms and major complaints of ME/CFS patients, molecular markers that reflect sleep or altered circadian rhythm<sup>52</sup> may serve as additional markers of ME/CFS. One note of caution is that the mathematical test in our study is an unvalidated tool and future study should include validated neuropsychological tools in correlating molecular changes to cognitive measurements.

In relation to past studies, we uncovered several molecular markers that were previously reported in ME/CFS patients. These include decreased uric acid<sup>12</sup>, HDL<sup>43</sup>, and *Faecalibacterium*<sup>46–48</sup>, and increased triglyceride<sup>13,43</sup>, monocyte number<sup>21</sup>, and *Coprobacillus*<sup>46,48</sup>. However, there were also markers that were not identified in our study. Aside from the differences in cohort, these differences could result from the differences in the technological platforms and source of biological samples used in our study. One main difference is we used whole blood instead of PBMC for transcriptome analysis which could elevate the contribution of non-immune cell, mainly red blood cells, to the gene expression profile. Another difference is we used <sup>1</sup>H-NMR for metabolomics while many other studies use GC–MS, LC–MS, or CE–MS method. More studies are needed to assess the impact of technological platforms in detecting some of these reported markers.

One of the main limitations of our study is that some of the markers we identified reflect comorbidity associated with ME/CFS rather than ME/CFS itself. In fact, several of the markers identified in our study are also markers of major depressive disorder including decreased uric acid, urea nitrogen, and total bilirubin<sup>53</sup>, and increased monocyte number<sup>54</sup>. Given that patients suffering from ME/CFS can sometimes develop depression as a result of unresolved debilitating fatigue<sup>55</sup>, it is possible that some of the markers captured comorbid depression.

However, molecular markers that were different during the early phase of illness ( $\leq 3$  years), such as decreased HDL, were not part of markers previously identified in major depressive disorder. HDL is elevated in patients with major depressive disorder<sup>53</sup> while it is decreased in our ME/CFS cohort. It is important to note that depression which results from ME/CFS does not resolve with antidepressant fluoxetine and may represent a different form of depression than those of major depressive disorder<sup>56</sup>. Since the ME/CFS patients who joined the present study did not show a typical depression, future study of these markers in a larger cohort should include a considerable number of patients with comorbid depression to assess the specificity of our markers for ME/CFS.

Another important comorbidity to note regarding our study is that ME/CFS is often accompanied by irritable bowel syndrome, which alters the composition of gut microbiota. Although we observed changes in the microbiome profile during the early phase of illness ( $\leq 3$  years), a previous study found that a decrease in *Faecalibacterium* was a top marker for ME/CFS with irritable bowel syndrome while a decrease in *Bacteroides vulgatus* was a top biomarker for ME/CFS without irritable bowel syndrome<sup>48</sup>. Therefore, several of the markers from our microbiome analysis may also reflect irritable bowel syndrome comorbidity. Future study of our microbiome markers should include a larger cohort with assessment of irritable bowel syndrome.

ME/CFS represents a collection of heterogeneous illness with diagnostic criteria meant to capture patients at the extreme end of the fatiguing illness<sup>5,23</sup>. Our deep phenotyping of ME/CFS in Japanese cohort revealed several molecular markers of illness that strongly correlated with sleep abnormality. Given that ME/CFS spans multiple symptoms<sup>1,2,23</sup>, some of the molecular markers identified in our study may reflect a subset of symptoms that were more prevalent in our cohort. This could influence whether potential biomarkers identified in our cohort can be applied to another cohort for ME/CFS. Although it remains unclear whether sleep abnormality is an over-represented symptom in our Japanese cohort compared to other cohorts, these molecular markers can still serve as measures of sleep-related changes in ME/CFS patients. Patients positive for these sleep-related molecular markers can then be recommended for more extensive analysis of sleep using actigraphy or polysomnography to objectively assess symptom related to unrefreshing sleep. In addition, with our extensive knowledge through PET molecular imaging and neurofunctional imaging<sup>23–34</sup>, the functional and molecular abnormality in the brain could be connected with these peripheral molecular markers. In our on-going studies, we have started the correlation studies among brain dysfunction, neuroinflammation detected by PET, and these molecular markers. Future studies that quantitatively measure symptoms related to fatigue and functional deterioration, including cognitive dysfunction and unrefreshing sleep, can help clarify the relationship between newly identified molecular markers and ME/CFS. These in turn could help untangle heterogeneity inherent in ME/CFS, which could aid in tailoring potential therapeutics for patient subgroups.

## Materials and methods

**Study design.** Study subjects included 48 myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) patients and 52 healthy controls, recruited at the Osaka City University Hospital Fatigue Clinical Center (Osaka, Japan). The study was approved by the ethics committees of Osaka City University Graduate School of Medicine (Approval No. 1498, 2151, 1499), and of RIKEN (Approval No. KOBE-IRB-11-13, YOKOHAMA-IRB-H24-17), and was conducted in accordance with the Declaration of Helsinki. All subjects, ME/CFS patients (n = 48) and healthy individuals (n = 52), provided written informed consent for participation in the study before enrolment. Healthy individuals were confirmed not to have abnormal results on any major clinical laboratory tests (hemoglobin, CRP, albumin, triglycerides, glucose, AST, ALT, or cholesterol, etc.), not to be in the BMI range of  $\geq 30$  or  $< 17$ , not to have subjective sleep problems, problems in daily life by fatigue, or be a shift worker. ME/CFS patients who visited the outpatient clinic of Osaka City University Hospital were randomly enrolled into the study. ME/CFS patients were diagnosed based on meeting both the 1994 Center for Disease Control clinical criteria (Fukuda criteria)<sup>1</sup> and the International Consensus Criteria<sup>2</sup> by the specialists at the Osaka City University Hospital. All patients also fulfilled the diagnostic criteria of Systemic Exertion Intolerance Disease (SEID, 29). All subjects were non-smokers. Degree of physical and psychological fatigue were assessed using Chalder fatigue scale<sup>35</sup>. The quality of sleep was assessed using Pittsburgh Sleep Quality Index Global (PSQIG) score<sup>36</sup>. Chalder fatigue scale and PSQIG were not used as part of the diagnosis process.

**Sample collection and processing.** Case and control subjects were fasted after 9:00 p.m. of the last night of the clinical test day (for at least 12 h) and peripheral blood was drawn between 9:00 am and 12:00 p.m. and collected into EDTA blood collection tubes. Blood samples were analyzed on automatic biochemical analyzer for biochemical parameters listed in Table S1. Whole blood was used for transcriptome analysis using Cap Analysis of Gene Expression (CAGE). For FACS and metabolome analysis, 5 mL of blood was diluted with 30 mL of Dulbecco's Phosphate Buffered Saline (D-PBS) in 50 mL Falcon tube and 15 mL of Ficoll Paque PLUS (GE Healthcare) were gently added, then centrifuged for 35 min at  $400 \times g$  at 20 °C. Top layer (plasma) was transferred to a new tube for metabolomics. For immunophenotype analysis, the middle layer containing lymphocytes and monocytes were transferred to a new 50 mL Falcon tube, filled up to 50 mL with D-PBS, centrifuged at  $300 \times g$  for 10 min at 20 °C. Supernatant was removed and the tube was refilled with D-PBS and centrifugation step was repeated. The resulting pellet was resuspended with 1 mL of 2% fetal calf serum (FCS)/D-PBS for antibody staining.

**Actigraphy.** Activity and sleep patterns were monitored using actigraphy method<sup>57</sup>. Activity was monitored in ME/CFS patients for seven days and healthy controls for three days by having subjects wear ActiGraph (Ambulatory Monitoring, Inc., USA) on non-dominant hand. The actigraph software using Cole-Kripke algorithm<sup>58</sup> was used to calculate number of awakening during total sleep, total sleep time, total sleep efficiency, and total activity time within an average 24 h period.



**Simple mathematical tests.** Simple mathematical tasks consisting of addition of two single digit numbers were administered to case and control subjects for 5 min. Percentage of total correct answers and average time to solve mathematical problems were measured.

**Immunophenotype.** To 50  $\mu$ L of lymphocytes and monocytes in 2% FCS/D-PBS, antibodies (Table S4) were added at the indicated volume for 30 min at 4 °C. 1 mL of 2% FCS/D-PBS were added to stained cells, centrifuged at 1200 rpm for 5 min at 4 °C, washed with 2% FCS/D-PBS twice, and resuspended in 300  $\mu$ L of 2% FCS/D-PBS, and analyzed using FACS Aria III (BD Bioscience). The gating strategy is shown in Fig. S8. Fraction of cells within each gating scheme was used for subsequent analysis.

**Transcriptome analysis.** RNA from whole blood was obtained using Ribopure blood kit (Ambion). CAGE libraries were prepared as described previously<sup>39</sup>. CAGE libraries were sequenced with the 50 bases single-end mode on the Illumina HiSeq 2500 platform according to the manufacturer's instructions (Illumina). The raw reads were processed in MOIRAI pipeline (Version 20121120)<sup>60</sup> system as follows: ligation adaptor sequences were trimmed; rRNA-derived reads and a base called 'N' were discarded by rRNA dust program, the processed reads were aligned to human reference genome (hg19) using BWA (Version: 0.7.10-r789)<sup>61</sup>, poorly mapped reads (mapping quality < 20) were discarded using SAMtools (Version: 0.1.18). The robust TSS sets identified in the FANTOM5 project were used as TSS reference, and the count of 5'-end of remaining CAGE reads mapped on TSS regions were used as raw signal of promoter expression. The expression signals were normalized by relative log expression method in the edgeR package<sup>62</sup>. Promoter expression with at least 1 cpm for all subjects were selected for subsequent analysis. Gene set enrichment analysis was performed as described before<sup>39</sup> using gene set annotation from curated pathways (v7.0: KEGG, BioCarta, Reactome). To obtain list of genes that were differentially expressed in previous studies of ME/CFS, we obtained gene list from publication<sup>8,41</sup>, or obtained dataset from NCBI GEO (GSE98139, GSE16059, GSE14577)<sup>9,40,42</sup>. For the three datasets from NCBI GEO, we filtered out genes that did not meet our filtering criteria. The criteria were normalized count of > 0 for GSE98139, expression value of > 6.6 (corresponding to Affymetrix expression value of 100) for GSE16059, and expression value of > 100 for GSE14577 for all samples within the dataset. Signal-to-noise ratios were calculated between cases and controls and top 100 up- and down-regulated genes were obtained as gene set. For GSEA, samples were permuted 1000 times using signal-to-noise ratio ranking to estimate *P*-value. For genes with multiple promoters, promoter with highest expression value was selected so that one gene corresponded to one gene expression (promoter) data. For correlation analysis, geometric mean for genes assigned as "core enrichment" from GSEA were calculated to obtain a single numerical representation of gene expression levels within a gene set.

**Metabolome analysis.** For metabolome analysis, 50  $\mu$ L of deuterium oxide containing 5 mM DSS-*d*<sub>6</sub> reference material (Wako) was added to 450  $\mu$ L of plasma. Prepared samples were measured on an NMR spectrometer (Bruker Avance II 700; Bruker Biospin) at 298 K and <sup>1</sup>H-NMR was measured using a Bruker standard program (noesypr1d) with 32,768 data points, 32 scans, 4 dummy scans, a 16-ppm spectral width, and 2 s relaxation delay as described<sup>63</sup>. Annotation of signals were made with two-dimensional *J*-resolved NMR measurement using Bruker standard program (jresgpprgf) with 32 data points for F1 and 16,384 data points for F2, 16 scans, 16 dummy scans, a 50-Hz spectral width for F1, a 18-ppm spectral width for F2, and a 2 s relaxation delay as described<sup>64</sup>. For two-dimensional *J*-resolved NMR measurements, intensity values were normalized to total intensity, both across subjects and metabolites. Signals were annotated with SpinCouple program<sup>65</sup> with reference to the Human Metabolome Database<sup>66</sup>. For annotation of lipoproteins, the diffusion-edited pulse program (ledbpgppr2s1d) was conducted with 66,560 data points, 64 scans, 4 dummy scans, a 16-ppm spectral width, and 1 s relaxation delay. Then the diffusion-edited spectra were divided into 30 fractions in the -CH<sub>3</sub> regions<sup>67</sup> and fraction corresponding to HDL was assigned based on correlation to the clinical laboratory-based test result of HDL value using peak assignment in Table S3.

**Fecal microbiome analysis.** Stool specimens were collected and sent to Osaka City University Hospital from the patients and healthy volunteers and stored at -20 °C until the transfer of all fecal samples to RIKEN Yokohama Institute. DNA extraction, 16S rRNA sequencing, and analysis were performed as described before<sup>68</sup>. Freeze-dried fecal samples were suspended in 10% sodium dodecyl sulfate, 10 mM Tris-HCl, and 1 mM EDTA (pH 8.0), then disrupted with 0.1-mm zirconia/silica beads (BioSpec Products) by shaking at 1500 rpm for 10 min. After centrifugation, bacterial DNA was purified using 25:24:1 phenol-chloroform-isoamyl alcohol and precipitated by ethanol and sodium acetate. The resulting DNA was treated with RNase A and precipitated by polyethylene glycol. The V1-V2 variable region of the 16S rRNA was PCR amplified using 27Fmod-338R primer pairs for 22 cycles, indexed using Nextera XT index primers, and sequenced on MiSeq (Illumina). 16S rRNA sequencing data were processed using QIIME software package<sup>69</sup>. An operational taxonomic unit (OTU) was defined at 97% similarity. OTU relative abundance below 0.05% were filtered to remove noise. OTU taxonomy was assigned based on comparison to Greengenes Database using RDP classifier with confidence level set at 0.5<sup>70</sup>.

**Statistical analysis.** Differences between ME/CFS patients and healthy controls were analyzed using Mann-Whitney U test (GraphPad Prism) and corrected for multiple testing using Benjamini-Hochberg correction method<sup>71</sup>. For comparison of three or more sample groups, Kruskal-Wallis test was performed followed by Dunn's multiple comparison post-test (GraphPad Prism). Power calculation was not performed prior to the initiation of the study as accepted methods to assess statistical power across multiple omics measurements did

not exist at the time. For correlation between markers, Spearman rank correlation were calculated (GraphPad Prism) and markers were clustered based on similarity in correlation value using average linkage hierarchical clustering. For multi-marker analysis, data were log transformed and normalized by sample median, and auto scaled. Separation between ME/CFS patients and healthy controls were visualized with partial least squares discriminant analysis (PLSDA) and importance of markers for separating the two groups were evaluated using variable importance in projection (VIP) score. Multi-marker analysis were performed using MetaboAnalyst 4.0<sup>72</sup>.

## Data availability

All data associated with this study are presented in the paper or the Supplementary Information.

Received: 27 July 2020; Accepted: 6 November 2020

Published online: 16 November 2020

## References

1. Fukuda, K. *et al.* The chronic fatigue syndrome: A comprehensive approach to its definition and study. International Chronic Fatigue Syndrome Study Group. *Ann. Intern. Med.* **121**, 953–959 (1994).
2. Carruthers, B. M. *et al.* Myalgic encephalomyelitis: International consensus criteria. *J. Intern. Med.* **270**, 327–338 (2011).
3. Jason, L. A. *et al.* A community-based study of chronic fatigue syndrome. *Arch. Intern. Med.* **159**, 2129–2137 (1999).
4. Nacul, L. C. *et al.* Prevalence of myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) in three regions of England: A repeated cross-sectional study in primary care. *BMC Med.* **9**, 91 (2011).
5. Prins, J. B., van der Meer, J. W. M. & Bleijenberg, G. Chronic fatigue syndrome. *Lancet* **367**, 346–355 (2006).
6. Kaushik, N. *et al.* Gene expression in peripheral blood mononuclear cells from patients with chronic fatigue syndrome. *J. Clin. Pathol.* **58**, 826–832 (2005).
7. Kerr, J. R. *et al.* Gene expression subtypes in patients with chronic fatigue syndrome/myalgic encephalomyelitis. *J. Infect. Dis.* **197**, 1171–1184 (2008).
8. Frampton, D., Kerr, J., Harrison, T. J. & Kellam, P. Assessment of a 44 gene classifier for the evaluation of chronic fatigue syndrome from peripheral blood mononuclear cell gene expression. *PLoS ONE* **6**, e16872 (2011).
9. Byrnes, A. *et al.* Gene expression in peripheral blood leukocytes in monozygotic twins discordant for chronic fatigue: no evidence of a biomarker. *PLoS ONE* **4**, e5805 (2009).
10. Fluge, O. *et al.* Metabolic profiling indicates impaired pyruvate dehydrogenase function in myalgic encephalopathy/chronic fatigue syndrome. *JCI Insight* **1**, e89376 (2016).
11. Nkinsin, A., Ruppert, D., Levine, S. M. & Hanson, M. R. Metabolic profiling of a myalgic encephalomyelitis/chronic fatigue syndrome discovery cohort reveals disturbances in fatty acid and lipid metabolism. *Mol. Biosyst.* **13**, 371–379 (2017).
12. Naviaux, R. K. *et al.* Metabolic features of chronic fatigue syndrome. *Proc. Natl. Acad. Sci. USA* **113**, E5472–E5480 (2016).
13. Nagy-Szakal, D. *et al.* Insights into myalgic encephalomyelitis/chronic fatigue syndrome phenotypes through comprehensive metabolomics. *Sci. Rep.* **8**, 10056 (2018).
14. Yamano, E. *et al.* Index markers of chronic fatigue syndrome with dysfunction of TCA and urea cycles. *Sci. Rep.* **6**, 34990 (2016).
15. Newberry, F., Hsieh, S. Y., Wileman, T. & Carding, S. R. Does the microbiome and virome contribute to myalgic encephalomyelitis/chronic fatigue syndrome?. *Clin. Sci. (Lond.)* **132**, 523–542 (2018).
16. Hornig, M. *et al.* Cytokine network analysis of cerebrospinal fluid in myalgic encephalomyelitis/chronic fatigue syndrome. *Mol. Psychiatry* **21**, 261–269 (2016).
17. Montoya, J. G. *et al.* Cytokine signature associated with disease severity in chronic fatigue syndrome patients. *Proc. Natl. Acad. Sci. USA* **114**, E7150–E7158 (2017).
18. Hornig, M. *et al.* Distinct plasma immune signatures in ME/CFS are present early in the course of illness. *Sci. Adv.* **1**, e1400121 (2015).
19. Klimas, N. G., Broderick, G. & Fletcher, M. A. Biomarkers for chronic fatigue. *Brain. Behav. Immun.* **26**, 1202–1210 (2012).
20. Brenu, E. W. *et al.* Natural killer cells in patients with severe chronic fatigue syndrome. *Auto Immun. Highlights* **4**, 69–80 (2013).
21. Hardcastle, S. L. *et al.* Characterisation of cell functions and receptors in Chronic Fatigue Syndrome/Myalgic Encephalomyelitis (CFS/ME). *BMC Immunol.* **16**, 35 (2015).
22. Eguchi, A. *et al.* Identification of actin network proteins, talin-1 and filamin-A, in circulating extracellular vesicles as blood biomarkers for human myalgic encephalomyelitis/chronic fatigue syndrome. *Brain. Behav. Immun.* **84**, 106–114 (2020).
23. Institute of Medicine of the National Academies. *Beyond Myalgic Encephalomyelitis/Chronic Fatigue Syndrome* (National Academies Press, Washington, 2015).
24. Tanaka, M., Ishii, A. & Watanabe, Y. Neural mechanisms underlying chronic fatigue. *Rev. Neurosci.* **24**, 617–628 (2013).
25. Tanaka, M. *et al.* Frontier studies on fatigue, autonomic nerve dysfunction, and sleep-rhythm disorder. *J. Physiol. Sci.* **65**, 483–498 (2015).
26. Tanaka, M. & Watanabe, Y. A new hypothesis of chronic fatigue syndrome: Co-conditioning theory. *Med. Hypotheses* **75**, 244–249 (2010).
27. Watanabe, Y. PET/SPECT/MRI/fMRI studies in the myalgic encephalomyelitis/chronic fatigue syndrome. In *PET and SPECT in Psychiatry* (eds Dierckx, R. A. J. O. *et al.*) (Springer, New York, in press).
28. Evengard, B. *et al.* (eds) *Fatigue Science for Human Health* (Springer, New York, 2008).
29. Watanabe, A., Kuratsune, H. & Kajimoto, O. Biochemical indices of fatigue for anti-fatigue strategies and products. In *The Handbook of Operator Fatigue* (ed. Matthews, G.) 209–224 (Ashgate Publishing, Farnham, 2012).
30. Nakatomi, Y. *et al.* Neuroinflammation in patients with chronic fatigue syndrome/myalgic encephalomyelitis: An (1)(1)C-(R)-PK11195 PET Study. *J. Nucl. Med.* **55**, 945–950 (2014).
31. Yamamoto, S. *et al.* Reduction of [<sup>11</sup>C](+)-3-MPB binding in brain of chronic fatigue syndrome with serum autoantibody against muscarinic cholinergic receptor. *PLoS ONE* **7**, e51515 (2012).
32. Yamamoto, S. *et al.* Reduction of serotonin transporters of patients with chronic fatigue syndrome. *NeuroReport* **15**, 2571–2574 (2004).
33. Mizuno, K. *et al.* Less efficient and costly processes of frontal cortex in childhood chronic fatigue syndrome. *Neuroimage Clin.* **9**, 355–368 (2015).
34. Hornig, M. Can the light of immunometabolism cut through “brain fog”? *J. Clin. Invest.* **130**, 1102–1105 (2020).
35. Chalder, T. *et al.* Development of a fatigue scale. *J. Psychosom. Res.* **37**, 147–153 (1993).
36. Buysse, D. J., Reynolds, C. F. 3rd, Monk, T. H., Berman, S. R. & Kupfer, D. J. The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. *Psychiatry Res.* **28**, 193–213 (1989).
37. Blackwell, T. *et al.* Comparison of sleep parameters from actigraphy and polysomnography in older women: The SOF study. *Sleep* **31**, 283–291 (2008).

38. Mootha, V. K. *et al.* PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273 (2003).
39. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
40. Gow, J. W. *et al.* A gene signature for post-infectious chronic fatigue syndrome. *BMC Med. Genomics* **2**, 38 (2009).
41. Vernon, S. D., Unger, E. R., Dimulescu, I. M., Rajeevan, M. & Reeves, W. C. Utility of the blood for gene expression profiling and biomarker discovery in chronic fatigue syndrome. *Dis. Markers* **18**, 193–199 (2002).
42. Nguyen, C. B. *et al.* Whole blood gene expression in adolescent chronic fatigue syndrome: An exploratory cross-sectional study suggesting altered B cell differentiation and survival. *J. Transl. Med.* **15**, 102 (2017).
43. Tomic, S., Brkic, S., Maric, D. & Mikic, A. N. Lipid and protein oxidation in female patients with chronic fatigue syndrome. *Arch. Med. Sci.* **8**, 886–891 (2012).
44. LaManca, J. J. *et al.* Immunological response in chronic fatigue syndrome following a graded exercise test to exhaustion. *J. Clin. Immunol.* **19**, 135–142 (1999).
45. Mawle, A. C. *et al.* Immune responses associated with chronic fatigue syndrome: A case–control study. *J. Infect. Dis.* **175**, 136–141 (1997).
46. Giloteaux, L. *et al.* Reduced diversity and altered composition of the gut microbiome in individuals with myalgic encephalomyelitis/chronic fatigue syndrome. *Microbiome* **4**, 30 (2016).
47. Giloteaux, L., Hanson, M. R. & Keller, B. A. A pair of identical twins discordant for myalgic encephalomyelitis/chronic fatigue syndrome differ in physiological parameters and gut microbiome composition. *Am. J. Case Rep.* **17**, 720–729 (2016).
48. Nagy-Szakal, D. *et al.* Fecal metagenomic profiles in subgroups of patients with myalgic encephalomyelitis/chronic fatigue syndrome. *Microbiome* **5**, 44 (2017).
49. Tan, K. C. *et al.* HDL dysfunction in obstructive sleep apnea. *Atherosclerosis* **184**, 377–382 (2006).
50. Aho, V. *et al.* Prolonged sleep restriction induces changes in pathways involved in cholesterol metabolism and inflammatory responses. *Sci. Rep.* **6**, 24828 (2016).
51. Broncel, M. *et al.* Sleep changes following statin therapy: A systematic review and meta-analysis of randomized placebo-controlled polysomnographic trials. *Arch. Med. Sci.* **11**, 915–926 (2015).
52. Davies, S. K. *et al.* Effect of sleep deprivation on the human metabolome. *Proc. Natl. Acad. Sci. USA* **111**, 10761–10766 (2014).
53. Peng, Y. F., Xiang, Y. & Wei, Y. S. The significance of routine biochemical markers in patients with major depressive disorder. *Sci. Rep.* **6**, 34402 (2016).
54. Seidel, A. *et al.* Major depressive disorder is associated with elevated monocyte counts. *Acta Psychiatr. Scand.* **94**, 198–204 (1996).
55. Wessely, M., Hotopf, M. & Sharpe, M. *Chronic Fatigue and Its Syndromes* (Oxford University Press, Oxford, 1998).
56. Vercoleyen, J. H. *et al.* Randomised, double-blind, placebo-controlled study of fluoxetine in chronic fatigue syndrome. *Lancet* **347**, 858–861 (1996).
57. Marino, M. *et al.* Measuring sleep: Accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. *Sleep* **36**, 1747–1755 (2013).
58. Cole, R. J., Kripke, D. E., Gruen, W., Mullaney, D. J. & Gillin, J. C. Automatic sleep/wake identification from wrist activity. *Sleep* **15**, 461–469 (1992).
59. Murata, M. *et al.* Detecting expressed genes using CAGE. *Methods Mol. Biol.* **1164**, 67–85 (2014).
60. Hasegawa, A., Daub, C., Carninci, P., Hayashizaki, Y. & Lassmann, T. MOIRAI: A compact workflow system for CAGE analysis. *BMC Bioinform.* **15**, 144 (2014).
61. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
62. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
63. Motegi, H. *et al.* Identification of reliable components in multivariate curve resolution-alternating least squares (MCR-ALS): A data-driven approach across metabolic processes. *Sci. Rep.* **5**, 15710 (2015).
64. Misawa, T., Wei, F. & Kikuchi, J. Application of two-dimensional nuclear magnetic resonance for signal enhancement by spectral integration using a large data set of metabolic mixtures. *Anal. Chem.* **88**, 6130–6134 (2016).
65. Kikuchi, J. *et al.* SpinCouple: Development of a web tool for analyzing metabolite mixtures via two-dimensional J-Resolved NMR database. *Anal. Chem.* **88**, 659–665 (2016).
66. Wishart, D. S. *et al.* HMDB 4.0: The human metabolome database for 2018. *Nucleic Acids Res.* **46**, D608–D617 (2018).
67. Suna, T. *et al.* 1H NMR metabolomics of plasma lipoprotein subclasses: Elucidation of metabolic clustering by self-organising maps. *NMR Biomed.* **20**, 658–672 (2007).
68. Kato, T. *et al.* Multiple omics uncovers host–gut microbial mutualism during prebiotic fructooligosaccharide supplementation. *DNA Res.* **21**, 469–480 (2014).
69. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
70. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).
71. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Methodol.* **57**, 289–300 (1995).
72. Chong, J. *et al.* MetaboAnalyst 4.0: Towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res.* **46**, W486–W494 (2018).

## Acknowledgements

Authors thank Kunihiro Tanaka for data organization, Naoko Takahashi, Ayumi Ogawa, Makis Motakis, Piero Carninci, GeNAS (sequence facility) for CAGE analysis.

## Author contributions

Y.H., H.O., and Y.W. conceived and supervised the project. S.F., K.Y., Y.N., E.Y., Y.Kataoka, K.M., H.Kuratsune, and Y.W. performed patient diagnosis, collected patient samples, performed clinical laboratory tests, administered questionnaires, actigraphy measurements, and simple math tests. S.F. and H.Kuratsune organized the clinical data. H.S., M.I., M.S.M., H.Kawaji, and Y.H. performed CAGE gene expression analysis. T.Kato and H.O. performed FACS immunophenotyping and microbiome analysis. Y.T. and J.K. performed metabolite profiling. H.Koseki and Y.Kogo organized the project protocols. J.K., T.Kato, and T.Kitami strategized data analysis and T.Kitami performed the overall data analysis. T.Kitami and Y.W. wrote the manuscript, and T.Kitami, S.F., Y.Kataoka, H.S., H.Kawaji, J.K., and Y.W. revised the manuscript.

## Funding

This work was supported by RIKEN President's Fund.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-77105-y>.

**Correspondence** and requests for materials should be addressed to T.K. or Y.W.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



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

© The Author(s) 2020