Use of single-nucleotide polymorphisms (SNPs) to distinguish gene expression subtypes of chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME)

Nana Shimosako,[1] Jonathan R Kerr[1,2]

▶ Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/jclinpath-2014-202597).

[1]CFS Group, Department of Cellular & Molecular Medicine, St George's University of London, London, UK
[2]Escuela de Medicina y Ciencias de Salud, Universidad del Rosario, Bogota, Colombia

Correspondence to
Dr Jonathan R Kerr, Escuela de Medicina y Ciencias de Salud, Universidad del Rosario, Quinta de Mutis, Bogota 111221, Colombia; jonathan@ssl-mail.com

Received 30 July 2014
Revised 2 September 2014
Accepted 4 September 2014
Published Online First 19 September 2014

ABSTRACT

Aims We have reported gene expression changes in patients with chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME) and the fact that such gene expression data can be used to identify subtypes of CFS/ME with distinct clinical phenotypes. Due to the difficulties in using a comparative gene expression method as an aid to CFS/ME disease and subtype-specific diagnosis, we have attempted to develop such a method based on single-nucleotide polymorphism (SNP) analysis.

Methods To identify SNP allele associations with CFS/ME and CFS/ME subtypes, we tested genomic DNA of patients with CFS/ME (n=108), patients with endogenous depression (n=17) and normal blood donors (n=68) for 504 human SNP alleles located within 88 CFS-associated human genes using the SNP Genotyping GoldenGate Assay (Illumina, San Diego, California, USA). 360 ancestry informative markers (AIM) were also examined.

Results 21 SNPs were significantly associated with CFS/ME compared with depression and normal groups. 148 SNP alleles had a significant association with one or more CFS/ME subtypes. For each subtype, associated SNPs tended to be grouped together within particular genes. AIM SNPs indicated that 4 subjects were of Asian origin while the remainder were Caucasian. Hierarchical clustering of AIM data revealed the relatedness between 2 couples of patients with CFS only and confirmed the overall heterogeneity of all subjects.

Conclusions This study provides evidence that human SNPs located within CFS/ME associated genes are associated with particular genomic subtypes of CFS/ME. Further work is required to develop this into a clinically useful subtype-specific diagnostic test.

INTRODUCTION

Chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME) is a disease characterised by severe and debilitating fatigue, sleep abnormalities, impaired memory and concentration, and musculoskeletal pain.[1] In the Western world, the population prevalence is estimated to be of the order of 0.5%.[2 3] Research studies have identified various features relevant to the pathogenesis of CFS/ME such as viral infection, immune abnormalities and immune activation, exposure to toxins, chemicals and pesticides, stress, hypotension, lymphocyte abnormalities and neuroendocrine dysfunction. However, the precise underlying disease mechanisms and means by which these abnormalities inter-relate in patients with CFS/ME remain to be clarified.[4 5]

We have previously reported the differential expression of 88 human genes in CFS/ME that were identified by a comprehensive microarray study with confirmation of gene targets using TaqMan real-time PCR.[6] These data were later confirmed in a follow-up study.[7] Clustering of gene expression data revealed 8 genomic subtypes of CFS/ME with distinct differences in Short Form-36 (SF-36) scores, clinical phenotypes, severity and geographical distribution.[7]

It has been recognised for some time that subtypes of CFS/ME exist, and it has been thought that such subtypes may, at least in part, reflect particular aetiological factors.[8] A symptom-based approach has had some success in identifying musculoskeletal, inflammatory and neurological subtypes;[9] however, these groups had only minor differences in overall functional severity in contrast to those of our gene expression studies.[6 7]

Although it is intriguing that such subtypes may be identified using gene expression methods, it is unlikely that such an approach could be useful in clinical practice to identify such subtypes as it is difficult to define reference ranges for each gene in each subtype. Therefore, we set out to achieve such a test using single-nucleotide polymorphisms (SNPs) located within the 88 CFS/ME-associated genes. This study demonstrates that there were 148 human SNPs that had significant associations with one or more CFS/ME genomic subtype. Although this study does not present a robust subtype-differentiating test, the results do provide evidence that such an approach may be feasible.

METHODS

Subject enrolment and clinical characterisation

Patients with CFS/ME (n=108) were diagnosed according to Fukuda diagnostic criteria for CFS/ME.[1] All of these suffered from idiopathic CFS/ME except 6 patients from Birmingham, who suffered from CFS/ME that had been triggered by laboratory-documented Q fever. Patients with major psychiatric disease and those with evidence of drug or alcohol abuse were excluded using the Minnesota International Neuropsychiatric Interview (MINI). These patients lived in Birmingham (n=6), Bristol (n=14), Leicester (n=1), London (n=12), New York (n=47) and Dorset (n=28). These patients had all been included in a previously published study on gene expression and had genomic subtypes assigned.[7]

Patients suffering from endogenous depression (n=17) were enrolled from Bristol, UK, and surrounding areas. These patients fulfilled DSM-IV criteria, had not smoked within the previous year and had not taken antidepressants in the previous year.

To cite: Shimosako N, Kerr JR. *J Clin Pathol* 2014;**67**:1078–1083.

Healthy normal blood donors enrolled from the Dorset National Blood Service (NBS) (n=68) were used as a comparison group. Restrictions imposed by the NBS on those allowed to donate blood are outlined elsewhere.[7]

For all patient groups, individuals who smoked in the previous year, who abused alcohol or other drugs, were currently taking (or were within 3 months of taking) antibiotics, steroids, cytotoxic drugs or antidepressants were excluded from the study.

For all enrolled subjects (patients and controls), according to the recommendations of the International CFS/ME Study Group,[10] severity of physical and mental fatigue was assessed using the Chalder fatigue scale;[11] level of disability was assessed using the Medical Outcomes Survey SF-36; accompanying symptoms were characterised using the Somatic and Psychological Health Report; sleep abnormalities were assessed using the Pittsburgh Sleep Questionnaire; and assessment of type and severity of pain was performed using the McGill Pain Questionnaire.

Patients and controls gave informed written consent according to guidance of the Wandsworth Research Ethics Committee (approval number 05/Q0803/137). For the New York patients, approval of the local Institutional Review Board was obtained. The human experimentation guidelines of the US Department of Health and Human Services were followed in this study.

Blood sampling and DNA extraction

In total, 5 ml blood, anticoagulated using EDTA, was taken from all subjects, and genomic DNA was extracted using the blood mini-kit (Qiagen) according to the instructions of the manufacturer. DNA quality and amount were confirmed by microspectrophotometry (Nanodrop, Rockland, Delaware, USA).

Whole-genome amplification

In total, 1 µl of extracted genomic DNA per sample (concentration range 5–500 ng/µl) was used in REPLI-g UltraFast Mini (Qiagen, Cat. no. 150035), following the manufacturer's instructions. Final dsDNA products were quantified using Quant-iT PicoGreen dsDNA (Invitrogen, P7581) and fluorescence emission was measured by Synergy HT Multi-Mode Microplate Reader (BioTek), following protocols from manufacturers.

Ancestry informative markers

In order to prescreen DNA sample for assay performance prior to conducting genotyping studies and to obtain some information on the ancestry of each subject, we used the DNA Test Panel (Illumina, San Diego, California, USA), which consists of 360 SNP loci selected from the Linkage IV panel distributed across the human genome. These loci have been chosen as genomic controls not likely to be associated with disease.

SNPs within CFS/ME-associated human genes

We searched for SNPs located within the promoter and coding regions (including introns) for 88 human genes that we have previously shown are differentially expressed in patients with CFS/ME as compared with normal blood donors.[6] Thus, we selected 504 SNPs for analysis, which comprised 1–8 SNPs per CFS/ME-associated human gene.[6]

Illumina SNP genotyping GoldenGate Assay

The list of intended SNPs to be tested (360 ancestry informative markers (AIM) SNPs and 504 CFS/ME-associated gene SNPs) was assessed by Illumina using the Assay Design Tool to determine the likely success of developing an assay to each SNP. Only SNPs with a good probability for success in developing such an assay were further targeted.

A total of 500 ng of genomic DNA per subject was used for Illumina SNP genotyping at either the Penn Microarray Facility, University of Pennsylvania, or the Biomics Centre, St George's University of London, using the Illumina BeadArray platform and GoldenGate Assay following the manufacturer's protocol. The fluorescence images of an array matrix carrying Cy3- and Cy5-labeled beads were generated with the two-channel scanner. Raw hybridisation intensity data processing, clustering and genotype calling were performed using the genotyping module in the GenomeStudio package (Illumina). Illumina developed a self-normalisation algorithm that relies on information contained in each array. This algorithm adjusts for channel-dependent intensity variations, differences in the background between the channels and possible crosstalk between the dyes. The normalisation procedure implemented in the GenomeStudio genotyping module includes outlier removal and background correction and scaling (details of this proprietary normalisation algorithm could be obtained from Illumina). Before genotype calling, the trimmed mean intensities were calculated from the normalised intensity values obtained for each bead type on the array by rejecting outliers to ensure high quality of genotype data. Genotype calls were generated using the GenCall software incorporated into the GenomeStudio package.[12] This algorithm uses a Bayesian model to assign normalised intensity values to one of the three possible homozygous and heterozygous genotype clusters.

Possible influence of CFS/ME-associated SNPs on previous gene expression array results

We examined each of the SNPs that were found to be associated with CFS/ME for their position relative to the so-called interrogation position on the Affymetrix U133+2 gene expression array probe sets using AffyMAPSdetector software.[13] This interrogation position is the nucleotide that is purposely mutated in the mismatch probe sequence compared with the perfect match probe. The possibility exists that for an SNP in this position, that the mismatch probe binds with greater affinity than the perfect match probe, and thus gives a higher value.

Comparative sequence analysis for detection of altered transcription factor binding sites

Each of the 21 SNP loci that differentiated CFS/ME/depression/normal or CFS/ME/normal was analysed for possible localisation within binding sites of CFS/ME-associated transcription factors and for possible effects on CFS/ME-associated transcription factor (TF) binding to CFS/ME-associated genes. Briefly, for each SNP of interest, the flanking sequences (50 bp either side) including the SNP (101 db in total) were obtained from the dbSNP database. This wild-type sequence was inserted into zpicture (http://www.zpicture.dcode.org),[14] along with its mutated sequence (by mutation of the SNP loci of interest) and analysed for predicted alteration of binding of CFS/ME-associated TFs (i.e., NFKB1, NHLH1, GABPA, REPIN1, EGR1, EGR3, ETS1)[6] using the rVISTA 2.0 tool portal (http://rvista.dcode.org).[15]

High-dimensional data analysis and visualisation

For the patients with CFS, CFS-associated SNP haplotypes were converted to numbers (AA=1, AB=2, BB=3, where A is allele A and B is allele B) and analysed by both self-organising graph (SOG) and t-distributed stochastic neighbour embedding (t-SNE) in that order within VisuMap 3.0.843 software (VisuMap Technologies, Calgary, Alberta, Canada) using default settings. SOG is an algorithm based on self-organising map

using an artificial neural network to model the data in high-dimensional space; t-SNE is a variation of SNE and is intended to fragment and unfold complex structure within a high-dimensional dataset so that details within cluster structures are visualised in a non-overlapping manner.[16]

Statistical analysis

To determine the significance of association of SNP alleles between CFS/ME, depression and normal groups, we used either a 2-column or 3-column χ^2 test. To determine the significance of association for SNP alleles with particular CFS/ME subtypes, we first used an 8-column χ^2 test. Then for each SNP, we compared the allele distribution in one subtype with the distribution in all other subtypes combined, and tested the significance using a χ^2 test. This was repeated for all SNPs in all subtypes. A p value of ≤ 0.05 was taken to be significant.

RESULTS

Subjects

This study included 108 patients with CFS/ME, 17 patients with endogenous depression and 68 normal blood donors. Subject demographics are shown in table 1. Patients with CFS/ME were typical of the CFS/ME population, and patients with depression were typical of the endogenous depression population. Normal blood donors showed very low levels of CFS/ME-associated symptoms, as would be expected.

Ancestry informative markers

Data were obtained on AIM SNP alleles for 108 patients with CFS/ME, 17 patients with endogenous depression and 29 of 68 normal blood donors. These data indicated that all normal blood donors and patients with depression were of European origin; all patients with CFS/ME were of European origin, except for four who were of Asian origin.

These alleles were converted to numerical values and were clustered to determine the degree of relatedness between the study subjects. This analysis revealed that except for 6 subjects these subjects were not closely related to each other, having a similarity index of >0.63. These 6 patients consisted of 2 pairs of subjects with CFS/ME who were highly similar (similarity of >0.95) and were known to be related and 1 pair consisting of one normal and one patient with depression who had a similarity of approximately 0.87 and were not known to be related.

Table 1 Subject demographics for patients with CFS/ME and endogenous depression, and normal blood donors

Clinical parameter	All patients with CFS/ME (n=108)	Patients with endogenous depression (n=17)	Normal blood donors (n=68)
Gender (M:F)	29:82	6:11	34:34
Mean age (years)	41.1	39.94	42.8
Mean duration of disease (years:months)	3.1	0:6	N/A
Symptoms/signs			
Headache	56	7	2
Sore throat	56	2	2
Poor memory/concentration	76	13	5
Muscle pain	79	6	3
Muscle weakness	67	4	2
Joint pain	93	8	2
Postexertional malaise	101	9	4
Sleep problem	68	8	6
Gastrointestinal problems	71	7	2
Fainting/dizziness	70	8	1
Numbness/tingling	49	2	0
Tender lymphadenopathy	49	0	0
Mean scores			
Physical fatigue (Chalder)	15.23	14.67	6.59
Mental fatigue (Chalder)	7.76	7.16	4.97
McGill pain questionnaire	16.45	9.03	3.35
Sphere questionnaire	11.22	11.98	2.46
SF-36 questionnaire	42.38	45.81	84.44
Pittsburgh sleep quality index	10.08	12.36	5.62

CFS/ME, chronic fatigue syndrome/myalgic encephalomyelitis; SF-36, Short Form-36.

CFS/ME-associated gene SNP alleles in CFS/ME, depression and normal groups

We found that 21 CFS/ME-associated gene SNPs had a frequency distribution across CFS/ME, depression and normal groups, which would not be expected by chance and which was significant (table 2). Interestingly, 7 of these SNPs were within the BMP2K gene and 2 were within the IL6ST gene. Regarding CFS/ME-associated gene SNPs, which showed an unexpected distribution between CFS/ME and normal, there were 10 of these that were significant, all of which were included in the list of the above 21 SNPs (table 2).

SNP alleles in CFS/ME subtypes

There were 148 SNPs that had a distribution across the CFS/ME subtypes and that was not expected by chance. All 148 were identified by comparing the allele distribution in one subtype with the distribution in all other subtypes combined (2×2 χ^2 test). In total, 27 of these 148 were detected using an 8-column χ^2 test (table 3). The distribution of particular SNP alleles across particular CFS/ME subtypes is interesting and shows that each subtype is associated with particular SNP alleles and that these SNP alleles are grouped within particular genes for each CFS/ME subtype, such that, for example, SNP alleles within ACTR3 are associated with subtype F, while SNP alleles within AKAP10 are associated with subtype G (see online supplementary table).

Possible influence of CFS/ME-associated SNPs on previous gene expression array results

We found that only one SNP was sited in the interrogation position of a probe sequence of one of the 88 CFS/ME-associated human genes. This SNP was rs2228431 within the ARSD gene (NM_001669); Affy probe set ID 206831_s_at; interrogation position 1767; sequence, AGCAGTTTTCCATGAGCAACAT CCT. A review of the previously published dataset on gene expression in CFS/ME versus normal controls[6] revealed that this probe did not influence these results as it was excluded from the analysis due to its lack of specificity to the ARSD gene.

Comparative sequence analysis for detection of altered transcription factor binding sites

Comparative sequence analysis, as outlined, predicted that 4 of 21 CFS/ME-associated SNPs had altered CFS/ME-associated transcription factor binding for CFS/ME-associated human genes (table 4).

High-dimensional data analysis and visualisation

SNP data converted into numerical data were analysed by both SOG and t-SNE, in that order, and the data displayed within VisuMap V.3.0.843 software. Figure 1 shows the result of this processing and shows that using the SNP haplotype data there

Table 2 Single-nucleotide polymorphisms (SNPs) showing a frequency distribution that was significantly different between CFS/ME, depression and normal, and between CFS/ME and normal

RefSNP ID	Alleles	Gene	SNP allele associated with CFS/ME	CFS/ME Allele (%)	Normal Allele%	Depression Allele%	χ^2 test, p value for distribution between CFS/ME, depression and normals	χ^2 test, p value for distribution between CFS/ME and normals
rs11895568	AG	FAM126B	A	99.1	100.0	94.1	0.011	0.011
rs1860661	AG	TCF3	A	77.5	95.5	79.4	<0.0001	<0.0001
rs10787901	AG	EIF3A	A	56.1	51.1	54.2	<0.0001	<0.0001
rs2071167	AG	UBTF	A	31.9	19.9	21.9	0.036	0.024
rs3752411	AG	METTL3	A	13.1	5.9	2.9	0.032	0.031
rs3737529	CT	SORL1	C	94.8	99.3	100.0	0.038	0.028
rs7719246	AT	IL6ST	A	81.94	91.9	82.4	0.030	NS
rs540516	CT	PNPLA6	C	87.5	77.2	67.6	0.0034	NS
rs12796043	CT	SORL1	C	67.6	51.5	52.9	0.0067	NS
rs3775525	CT	BMP2K	C	4.2	5.2	17.6	0.0072	NS
rs3775513	AG	BMP2K	A	95.3	95.5	82.4	0.0081	NS
rs3822106	AC	BMP2K	A	27.8	26.1	47.1	0.048	NS
rs6850116	GT	BMP2K	G	95.8	95.5	79.4	0.00059	NS
rs1426137	AT	BMP2K	A	96.2	95.5	84.4	0.016	NS
rs2228431	CT	ARSD	C	9.3	5.2	20.6	0.018	NS
rs306772	CT	GSN	C	80.7	84.3	61.8	0.013	NS
rs11549467	AG	HIF1A	A	0	0.7	5.9	0.0017	NS
rs3775516	AG	BMP2K	A	4.9	5.4	21.4	0.010	0.0025
rs1426139	AT	BMP2K	A	5.6	5.1	17.6	0.021	0.0091
rs1373998	CT	IL6ST	C	88.0	91.9	74.3	0.016	0.013
rs3802758	CT	PEX16	C	31.9	11.8	38.2	<0.0001	<0.0001

NS, not significant.
CFS/ME, chronic fatigue syndrome/myalgic encephalomyelitis.

was some degree of separation of the different CFS/ME subtypes using this method within this 2-dimensional figure. These subgroups had 73% correlation with the gene expression subtypes (data not shown).

DISCUSSION

This study was undertaken to identify CFS/ME-associated SNPs occurring within the promoter and coding sequences of 88 CFS/ME-associated human genes. These 88 human genes were originally identified by analysis of peripheral blood of 55 CFS/ME and 75 normal blood donors using a comprehensive gene expression microarray (Affymetrix U133+2), the results of which were confirmed by TaqMan real-time PCR testing.[6]

It is important to confirm that there was no significant degree of relatedness among the subjects used in the study that would confound the results, and this was confirmed using the AIM SNP allele data from the DNA test panel (Illumina). It is also important to ensure that any SNP allele associations found are not the result of a significant over-representation of one race compared with the others in one group, and this has also been ruled out by the results of the AIM SNP data.

We have shown that the allele frequency for each of the 21 SNPs has a statistically significant and unexpected distribution across CFS/ME/depression / normal groups and that for 10 of these 21 SNPs, particular alleles are associated with CFS/ME compared with normal (table 2). Although SNP allele associations have been reported in CFS/ME previously, this low yield (21 of 504) may seem surprising given that these SNP loci occurred within genes found to be differentially expressed in CFS/ME. However, this observation is probably explained by the known heterogeneity of CFS/ME generally and the existence of subtypes in this CFS/ME cohort, which were revealed using gene expression data clustering.[7] These subtypes had distinct clinical phenotypes and levels of severity. This assumption is further supported by the finding of 148 SNP alleles that were associated with one or more CFS/ME subtypes. The distribution of particular SNP alleles between particular CFS/ME subtypes is interesting (see online supplementary table). However, this bore no discernible relationship to the gene expression data for each of these subtypes (data not shown).[7]

It is reassuring that our analysis of the location of these SNPs relative to the probe sets of the Affymetrix U133+2 gene array, which was originally used to identify the differential expression of these genes,[6] did not support a significant influence on the gene expression data in general.

The fact that four of the SNP alleles found to be associated with CFS/ME were predicted to exhibit reduced binding of CFS/ME-associated transcription factors (table 4) suggests that such mutations may well influence the gene expression signatures found in patients with CFS/ME, particularly, as in 3 of these 4, the allele was associated with the entire group of CFS/ME and not simply one subtype (table 4) and as these transcription factors have numerous binding sites within the promoter regions of these genes.[6]

One possible application of the knowledge that these SNPs have significantly different distributions across CFS/ME subtypes is the development of a subset of them for a subtype-specific diagnostic test. This may be useful once the existence of the previously published gene expression CFS/ME subtypes[7] has been independently confirmed and when we know more about them in terms of natural history, triggering factors, response to different treatments, etc. Such an SNP-based test is highly desirable

Table 3 p Values of χ^2 tests of the allele distribution between 8 CFS/ME subtypes for 27 CFS/ME subtype-associated single-nucleotide polymorphisms (SNPs) with an eight-column χ^2 test, p value ≤0.05

				CFS/ME gene expression subtype (no. of patients tested in each subtype)							
RefSNP ID	Alleles	Gene	8-column χ^2 test	A 22	B 6	C 19	D 5	E 20	F 13	G 19	H 4
rs11658169	CT	AKAP10	0.036	C:0.041						C:0.026	
rs2515194	CT	ATP6V1C1	0.00034	T:0.038		T:0.042		C:0.013		C:0.0003	
rs12687359	AT	BCOR	0.00032					T:0.031			T:< 0.0001
rs5917933	AG	BCOR	0.0054				A:< 0.0001				
rs1373998*	CT	IL6ST	0.0063			C:0.023		T:0.00087			
rs3752411*	AG	METTL3	0.040			A:0.0062					
rs1139130	AG	METTL3	0.0083			G:0.0083		G:0.030			A:0.041
rs7115	AG	MRPS6	0.013		A:0.018			G:0.00084			
rs2834384	GT	MRPS6	0.013		G:0.022			T:0.0014			
rs11621566	AG	PAPOLA	0.016	G:0.014					A:0.037		
rs2274795	CT	PAPOLA	0.016	T:0.014					C:0.037		
rs9654453	CT	PDCD6	0.0083				C:< 0.0001				
rs3802758*	CT	PEX16	0.0016			T:0.032		C:0.0020			
rs540516*	CT	PNPLA6	0.029					T:0.0081			
rs1904298	CT	PPP2R5C	0.0067			T:0.034					T:0.0004
rs11686919	AG	PUM2	0.040							G:0.0091	
rs157476	AG	SFXN1	0.0061					A:0.028		G:0.014	
rs925197	AG	SFXN1	0.013			A:0.022		G:0.015			
rs2662170	CT	SFXN1	0.024			T:0.0046		C:0.044			
rs937353	CG	SFXN1	0.011		C:0.05	G:0.007		C:0.038			
rs2834378	CT	SLC5A3	0.013		C:0.018			G:0.00084			
rs1860661*	AG	TCF3	0.0079			A:0.0014	G:0.033			G:0.050	
rs1061026	GT	TOX4	0.037			G:0.0057					
rs13128884	AG	USP38	0.014		A:0.0083						G:0.021
rs28470858	AT	USP38	0.016		A:0.0020						
rs34461753	AG	USP38	0.0063		G:0.00093					G:0.013	
rs4690779	AG	USP38	0.014		G:0.0083						A:0.021

The table shows results of both eight-column χ^2 test and χ^2 test of the distribution of alleles in one subtype versus that in all the remaining subtypes combined.
*SNP also included in table 1.
CFS/ME, chronic fatigue syndrome/myalgic encephalomyelitis.

compared with a gene expression-based test as it is robust and easily reproducible unlike the comparative gene expression method that was used to identify these 88 human genes.[7] However, this would require extensive further testing and validation before it could be recommended for such a purpose.

In conclusion, in this paper, we report 21 SNP alleles that are significantly associated with patients with CFS/ME compared with depression and normal controls, and 148 SNP alleles that are associated with one or more CFS/ME subtypes, 27 of which showed a p value ≤0.05 in the 8-column χ^2 test. The sample size of this study, and especially the sample sizes of each gene expression subtype, are too small to draw any firm conclusions. However, assuming these results can be replicated, it may be possible to develop a subtype-specific diagnostic test using an SNP subset, which may aid in the investigation and clinical management of patients with CFS/ME.

Table 4 CFS/ME-associated gene single-nucleotide polymorphisms (SNPs) that are localised within binding sites of CFS/ME-associated transcription factors

SNP	Table no. in this study	CFS/ME-associated gene	Alleles	CFS/ME-associated allele	CFS/ME-associated transcription factor	Sequence* (from dbSNP database)	Mutation resulting in no binding (predicted)
rs3802758	2, 3	PEX16	TC	C	NHLH1 (HEN1)	ggcTCGCTCATCTGC**T**GCCctt	T>C
rs1904298	3	PPP2R5C	CT	T	GABPA	tg**C**GCTTCTTGt	C>T
rs11218304	2	SORL1	AG	G	REPIN1 (AP4)	acaga**C**TGCTGTTCTCag	A>G
rs1426137	2	BMP2K	AT	A	ETS1	at**TT**CCtc	T>A

*Highly conserved positions are indicated by capitals; the site of mutation is indicated in boldface.
CFS/ME, chronic fatigue syndrome/myalgic encephalomyelitis.

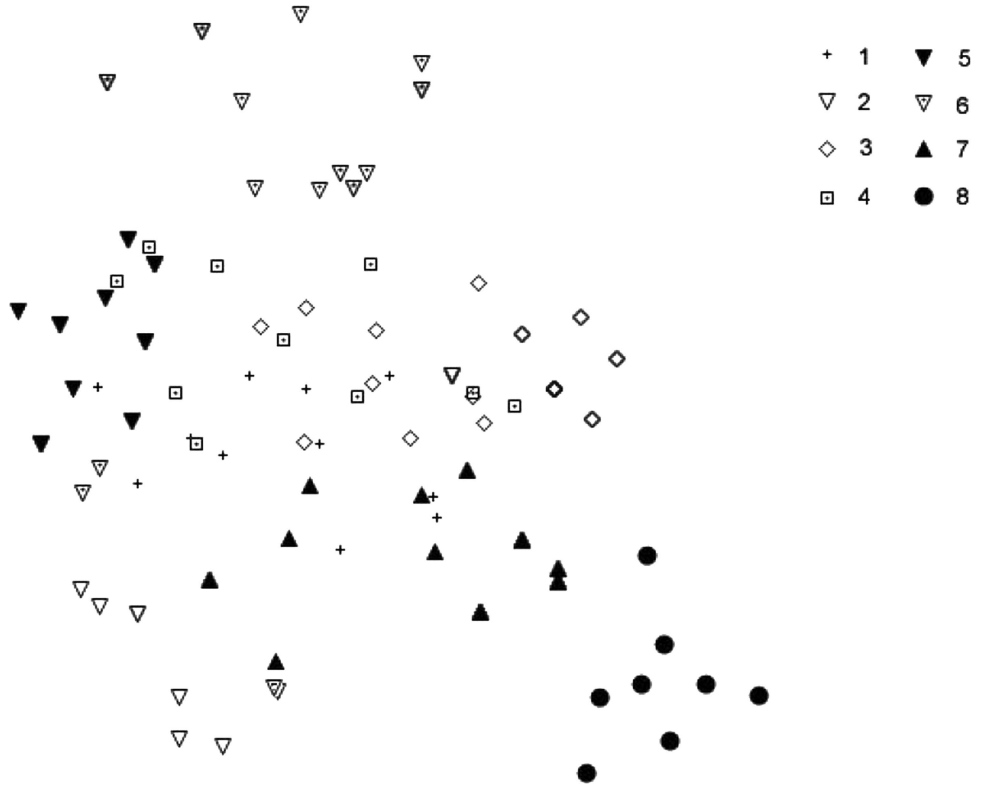

Figure 1 High-dimensional data analysis and visualisation using self-organising graph followed by t-distributed stochastic neighbour embedding within VisuMap 3.0.843 (VisuMap Technologies, Calgary, Alberta, Canada).

Take home messages

- Chronic fatigue syndrome (CFS)/myalgic encephalomyelitis is a heterogeneous disease with multiple aetiologies and different presentations.
- Some single-nucleotide polymorphisms (SNPs) within genes that are differentially expressed in CFS are significantly associated with the disease and with particular CFS subtypes.
- This study has attempted to differentiate gene expression subgroups of CFS based on SNP analysis, and this has had a degree of success.
- Reproducible methods to differentiate subgroups of CFS do not exist at present, but they are needed in order to develop CFS subgroup-specific treatments.

Acknowledgements We thank ME Research UK for funding of the project and the salary of Nana Shimosako and Sir Joseph Hotung for funding of the salary of Jonathan Kerr. The Illumina GoldenGate SNP arrays were performed at the Penn Microarray Facility, University of Pennsylvania, USA, and at the Medical Biomics Centre, St George's University of London, UK. We thank the following clinicians for participating in the study; Derek Enlander, Selwyn Richards, David Nutt, Janice Main and David Honeybourne.

Contributors JRK designed the study, obtained funding from ME Research UK, oversaw the research work, analysis of results and wrote the paper. NS performed the practical work, analysed the data and reviewed and corrected the manuscript.

Funding Sir Joseph Hotung (N/A), ME Research UK (N/A).

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Fukuda K, Straus SE, Hickie I, *et al*. The chronic fatigue syndrome: a comprehensive approach to its definition and study. International Chronic Fatigue Syndrome Study Group. *Ann Intern Med* 1994;121:953–9.
2 A Report of the CFS/ME Working Group. Department of Health, January 2002. http://www.doh.gov.uk/cmo/cfsmereport/
3 Papanicolaou DA, Amsterdam JD, Levine S, *et al*. Neuroendocrine aspects of chronic fatigue syndrome. *Neuroimmunomodulation* 2004;11:65–74.
4 Komaroff AL, Buchwald D. Chronic fatigue syndrome: an update. *Annu Rev Med* 1998;49:1–13.
5 Devanur LD, Kerr JR. Chronic fatigue syndrome. *J Clin Virol* 2006;37:139–50.
6 Kerr JR, Petty R, Burke B, *et al*. Gene expression subtypes in patients with chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME). *J Infect Dis* 2008;197: 1171–84.
7 Zhang L, Goudh J, Christmas D, *et al*. Microbial infections in eight genomic subtypes of chronic fatigue syndrome/myalgic encephalomyelitis. *J Clin Pathol* 2009;63:156–64.
8 Jason LA, Corradi K, Torres-Harding S, *et al*. Chronic fatigue syndrome: the need for subtypes. *Neuropsychol Rev* 2005;15:29–58.
9 Janal MN, Ciccone DS, Natelson BH. Sub-typing CFS patients on the basis of 'minor' symptoms. *Biol Psychol* 2006;73:124–31.
10 Reeves WC, Lloyd A, Vernon SD, *et al*. Identification of ambiguities in the 1994 chronic fatigue syndrome research case definition and recommendations for resolution. *BMC Health Serv Res* 2003;3:25.
11 Chalder T, Berelowitz G, Pawlikowska T, *et al*. Development of a fatigue scale. *J Psychosom Res* 1993;37:147–53.
12 Illumina Inc. Illumina GenCall Data Analysis Software. Technology Spotlight. 2005.
13 Kumari S, Verma LK, Weller JW. AffyMAPSDetector: a software tool to characterize Affymetrix GeneChip expression arrays with respect to SNPs. *BMC Bioinformatics* 2007;8:276.
14 Ovcharenko I, Loots GG, Hardison RC, *et al*. zPicture: dynamic alignment and visualization tool for analyzing conservation profiles. *Genome Res* 2004;14: 472–7.
15 Loots GG, Ovcharenko I, Pachter L, *et al*. rVista for comparative sequence-based discovery of functional transcription factor binding sites. *Genome Res* 2002;12:832–9.
16 Platzer A. Visualization of SNPs with t-SNE. *PLoS ONE* 2013;8:e56883.