(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 February 2007 (01.02.2007)

PCT

(10) International Publication Number
WO 2007/012879 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/GB2006/002849

(22) International Filing Date: 31 July 2006 (31.07.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0515685.6 29 July 2005 (29.07.2005) GB

(71) Applicant *(for all designated States except US)*: **IMPERIAL COLLEGE INNOVATIONS LIMITED** [GB/GB]; Electrical and Electronic Engineering Building, Level 12, Imperial College, Ehibition Road, London SW7 2AZ (GB).

(71) Applicant *(for US only)*: **SCOTT, Frances** (legal representative of the deceased inventor) [GB/GB]; Spinney Hill, Quarry Lane, Exeter EX2 5PS (GB).

(72) Inventor: **TYRRELL, David** (deceased).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **KERR, Jonathan, Richard** [GB/GB]; Department of Cellular and Molecular Medicine, St George's, University of London, Cranmer Terrace, London SW17 0RE (GB). **HOLGATE, Stephen, Townley** [GB/GB]; Infection, Inflammation and Repair Division, Mail Point 810, Level F, South Block, Southampton General Hospital, Southampton, Hampshire SO16 6YD (GB). **RICHARDS, Selwyn, Charles, Morgan** [GB/GB]; Rheumatology, Poole Hospital, Poole BH15 2JB (GB). **HARRISON, Tim** [GB/GB]; Centre for Hepatology, Department of Medicine, Royal Free Campus, Rowland Hill Street, Hampstead, London NW3 2PF (GB).

(74) Agents: **CHAPMAN, Paul, William** et al.; Kilburn & Strode, 20 Red Lion Street, London WC1R 4PJ (GB).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) **Title:** METHOD, KIT AND MICROARRAY FOR DIAGNOSING CHRONIC FATIGUE SYNDROME (CFS) OR MYALGIC ENCEPHALOMYELITIS (ME)

(57) **Abstract:** The present invention provides methods for diagnosing chronic fatigue syndrome (CFS) or myalgic encephalomyelitis (ME), by measuring the level of expression of one or more genes associated with these conditions.

WO 2007/012879 A1

METHOD, KIT AND MICROARRAY FOR DIAGNOSING CHRONIC FATIGUE SYNDROME (CFS) OR MYALGIC ENCEPHALOMYELITIS (ME)

The present invention provides a diagnostic test for myalgic encephalomyelitis or Chronic fatigue syndrome.

Chronic fatigue syndrome (CFS) is a multisystem disease, the pathogenesis of which remains undetermined. CFS/myalgic encephalomyelitis (ME) is a disease that is characterised by severe and debilitating fatigue, sleep abnormalities, impaired memory and concentration, and musculoskeletal pain. In the Western world, the population prevalence is estimated to be in the order of 0.5%. Although CFS is now recognised as a genuine clinical entity, a considerable research effort has failed to identify quantifiable parameters that consistently exhibit abnormal results in well documented cases. Therefore, the pathological basis for CFS remains poorly understood. Although there is no known aetiology and no known diagnostic marker, a large number of diverse factors such as viral infection; immune activation; exposure to toxins, chemicals, and pesticides; stress; hypotension; lymphocyte abnormalities; and neuroendocrine dysfunction have been proposed as factors in the pathogenesis of CFS.

Previous reports have studied gene expression in the peripheral blood of patients with CFS. Although there was little agreement between these studies as to the genes identified, only one used quantitative polymerase chain reaction (PCR) to confirm the initial findings and the initial method in this study was differential display. Therefore it is still unclear which genes are potentially involved in CFS.

To address this discrepancy and to investigate the hypothesis that abnormalities of gene regulation occur in CFS, gene expression in peripheral blood mononuclear cells (PBMC) of patients with CFS and normal blood donors was

investigated by the inventors using a microarray. Taqman real time PCR was used to confirm those genes identified as being differentially expressed between the groups. Sixteen genes were confirmed as having an expression profile associated with the CFS.

Thus in the first aspect the present invention provides a method of diagnosing chronic fatigue syndrome (CFS) or myalgic encephalomyelitis (ME) comprising:

(a) measuring the level of expression of one or more genes which are differentially expressed in CFS or ME;

(b) comparing the level of expression with that of a normal healthy control.

Genes which are differentially expressed in CFS or ME include ABCD4, PRKCL1, MRPL23, IL10RA, CD2BP2, GSN, NTE, POLR2G, PEX16, EIF2B4, EIF4G1, ANAPC11, PDCD2, KHSRP, BRMS1, and GABARAPL1. Preferably the gene is IL10RA.

As used herein "differentially expressed" refers to genes which are either upregulated or downregulated in CFS or ME, compared to normal healthy controls.

Step (a) can be carried out using RT-PCR. RNA can be extracted from a sample obtained from a patient. Suitable primers can be designed that hybridise to those genes known to be expressed differentially in CFS or ME. Pairs of primers are designed to hybridise and thus amplify the genes of interest. Methods for designing such primers, and carrying out RT-PCR are well known to the person skilled in the art. The level of expression for the gene

of interest can be determined from the amount of product created during the amplification process.

Alternatively the method can be carried out using a microarray. As used herein the term "microarray" refers to a set of oligonucleotide probes arranged on a solid matrix, such as a microscope slide or silicon wafer. The oligonucleotide probes are 10-50 nucleotides in length, preferably 15-40 nucleotides, more preferably 20-30 nucleotides, most preferably 24 nucleotides. Each probe has a defined locus. The microarray preferably comprises oligonucleotide probes corresponding to those genes known to be differentially expressed in CFS or ME. cRNA derived from a sample obtained from the patient can be labelled and hybridised to the oligonucleotide probes on the microarray to detect the level of gene expresion.

In another aspect the present invention provides the use of a microarray comprising one or more oligonucleotide probes corresponding to those genes which are differentially expressed in CFS or ME in the diagnosis of CFS or ME. Preferably, the microarray comprises or consists of oligonucleotide probes corresponding to the genes ABCD4, PRKCL1, MRPL23, IL10RA, CD2BP2, GSN, NTE, POLR2G, PEX16, EIF2B4, EIF4G1, ANAPC11, PDCD2, KHSRP, BRMS1, and GABARAPL1.

In a further aspect the present invention provides a microarray comprising or consisting of oligonucleotide probes corresponding to the genes ABCD4, PRKCL1, MRPL23, IL10RA, CD2BP2, GSN, NTE, POLR2G, PEX16, EIF2B4, EIF4G1, ANAPC11, PDCD2, KHSRP, BRMS1, and GABARAPL1. Preferably the microarray consists of oloigonucleotides corresponding to the genes ABCD4, PRKCL1, MRPL23, IL10RA, CD2BP2, GSN, NTE, POLR2G, PEX16, EIF2B4, EIF4G1, ANAPC11, PDCD2, KHSRP, BRMS1, and

GABARAPL1, and one or more control oligonucleotide probes. Examples of such control oligonucleotide probes include mismatch probes which differ from oligonucleotides which are a perfect match for the desired gene by a single nucleotide. Oligonucleotide probes which allow for internal calibration can also be used as control oligonucleotide probes.

In another aspect the present invention provides a kit for diagnosing CFS or ME comprising one or more pairs of suitable primers to carry out RT-PCR to measure the level of expression of one or more genes which are differentially expressed in CFS or ME patients.

In a further aspect, the invention provides a kit for diagnosing CFS or ME comprising a microarray comprising one or more oligonucleotide probes corresponding to those genes which are differentially expressed in CFS or ME in the diagnosis of CFS or ME.

Preferably, the microarray comprises or consists of oligonucleotide probes corresponding to the genes ABCD4, PRKCL1, MRPL23, IL10RA, CD2BP2, GSN, NTE, POLR2G, PEX16, EIF2B4, EIF4G1, ANAPC11, PDCD2, KHSRP, BRMS1, and GABARAPL1.

The invention will now be described with reference to the following figures:

Figure 1 shows hierarchical clustering experiment of differentially expressed gene profiles among patients with chronic fatigue syndrome (CFS; n = 25) and normal persons (n = 25) identified by analysis in BRB Array Tools. The figure was generated using Genepilot software. Each column represents the expression profile for each of the 35 genes. Each row represents a single gene, with its GenBank accession number to the right hand side of the figure.

Coloured pixels represent the magnitude of the response for any gene. Shades of dark grey and light grey represent induction and repression, respectively, relative to the mean value for each respective gene among the normal persons. This figure shows a cluster of 18 subjects (from P29–P9) consisting of predominantly patients with CFS (n = 17) and one normal person, who have a similar profile of expression of these 35 genes. P, patients; C, controls.

Figure 2 (A) Bar chart showing the fold difference in gene expression between test and control groups by microarray (shaded) and real time polymerase chain reaction (solid black) for 16 genes that are differentially expressed in chronic fatigue syndrome (CFS). (B) Bar chart showing the mean relative quantity of mRNA transcripts in test (shaded) and control (solid black) groups for 16 genes that are differentially expressed in CFS. Error bars indicate the standard deviation from the mean in each case. All values for the mean relative quantity mRNA transcript are shown on the left y axis, except those for NTE and EIF4G1, which are shown on the right y axis.

Example 1

Subject enrolment

Patients with CFS (n = 25) were enrolled from the Dorset CFS service in South East England. These cases were diagnosed according to the criteria of Fukuda and colleagues. Additional clinical information was recorded and is presented in table 1. This includes measurements of physical and mental fatigue using the Chalder Fatigue Scale and measurements of anxiety and depression using the Hospital Anxiety and Depression scales. None of these patients had undergone previous treatment for psychiatric disorders. Patients were sampled at two time points, six months apart, between which their symptoms did not vary significantly; the first sample from each patient was used for microarray analysis whereas the second was used for real time PCR. This approach

provides an additional safeguard against attaching particular importance to genes that may be differentially expressed at a single time point, but not reproducibly.

Normal blood donors were enrolled from the East Dorset division of the National Blood Service (NBS); 25 age and sex matched normal blood donors were used as a comparison group for the microarray part of the study, and 21 normal blood donors (age and sex matched as a group) were used as a comparison group for the real time PCR part of the study. The NBS restricts donors to those who fulfil the following criteria: aged between 17 and 59 years; have not given blood in the previous 16 weeks; are not currently suffering from an infection; are not pregnant; are not currently taking (or within three months of taking) antibiotics, steroids, or antidepressants; have not had hepatitis, jaundice, body piercing, acupuncture, or blood transfusion in the past year; do not have two family members who suffered from Creutzfeldt-Jakob Disease; are not known to be positive for human immunodeficiency virus, hepatitis B or C; are not currently or previously abusers of injecting or body building drugs; are not post-vaccination; and do not suffer a chronic illness including malaria. Patients and controls gave written consent according to the guidance of the ethics committees of both the East Dorset , and the Royal Brompton, Harefield & NHLI.

Sample collection and processing

A 20 ml sample of blood was immediately placed in cell preparation tubes containing density gradient solution and EDTA (BD Biosciences, Manchester, UK). PBMCs were isolated by density gradient centrifugation. Total RNA was extracted using Trizol (Invitrogen, Carlsbad, California, USA), washed in phosphate buffered saline, and the quality and amount confirmed by microspectrophotometry (Nanodrop, Rockland, Delaware, USA). Total RNA

samples had an absorbance ratio (A260/280) of 1.85–1.95. A 20 ml sample of peripheral blood is estimated to contain a total of approximately $15x10^6$ PBMCs and buffy coat from this sample yielded 10–25 mg total RNA, which was shipped to Nimblegen, USA, at room temperature in ethanol, for array testing. Total RNA was converted to double stranded cDNA by the SuperScript Choice System (Invitrogen) with an oligo dT primer containing the T7 RNA polymerase promoter sequence :
5'-GGCCAGTGAATTGTAATACGACTCACTATAGGGAGGCGG-T24-3'.

A 15 μg aliquot of total RNA was converted into cDNA. *In vitro* transcription was used to produce biotin labelled cRNA from cDNA using the Ambion MEGAscript T7 kit (Ambion, Austin, Texas, USA). Before hybridisation, cRNA was fragmented to an average size of 50– 200 bp by incubation in 100mM potassium acetate, 30mM magnesium acetate, and 40mM Tris/acetate at 94°C for 35 minutes. Fragmentation was checked by gel electrophoresis in 1% agarose.

Microarray design and analysis

A custom microarray was manufactured by Nimblegen (Madison, Wisconsin, USA) using maskless array synthesis (http://www.nimblegen.com). The human genes on this design (n = 9522) were selected from the Homo sapiens entries in the RefSeq collection of sequences as of August 2002. Each gene was compared with all others using the BLAST program to remove redundancies. Ten probe pairs for each target were selected from the 3' 1 kb of each target. Probes were spaced evenly over the length of the target region (≤1 kb), so that the exact spacing depended on the length of the target sequence. Each probe was 24 nucleotides in length. For each perfect match probe there was also a mismatch probe, which differed by a single nucleotide.

Labelled cRNA was hybridised to the oligonucleotide probes on the microarray. After washing, arrays were stained with streptavidin–cy3 conjugate (Amersham Biosciences, Piscataway, New Jersey, USA) for 25 minutes at room temperature, followed by washing and a blow dry step using high pressure grade 5 Argon (Badger Welding, Madison, Wisconsin, USA). Slides were scanned using a GenePix 4000B microarray scanner (Axon Instruments, Union City, California, USA), and the feature intensities extracted from the TIF files were calculated by the scanner software using a proprietary application developed at NimbleGen (Madison, Wisconsin, USA). This application calculates mean signal intensities for the pixels that define each feature (3x3 grid of pixels). The intensities for each gene are calculated by taking the mean of the intensities for the perfect match probes specific to each target minus the mean of the intensity of the mismatch probes. Probes that differed from the mean for the set by more than 3 SD were removed from the set and the mean recalculated. Average differences (recalculated mean) were used for subsequent analysis.

Data analysis was performed using BRB ArrayTools version 3.02 (Molecular Statistics and Bioinformatics Section, National Cancer Institute, Bethesda, Maryland, USA) developed by Dr R Simon and A Peng (http://linus.nci.nih.gov/BRB-ArrayTools.html). Average difference values were normalised to median over the array. The data were filtered so that only those genes that were adequately measured on 75% of the arrays were included. A class comparison protocol was used to identify genes whose degree of expression differed significantly by ≥ 1.5 fold between the two groups. This consisted of a multivariate permutation test, which was computed based on 1000 random permutations using the following parameters: nominal significance level = 0.001; confidence level of false discovery rate assessment = 50%; maximum allowed number of false positive genes = 10; maximum

allowed proportion of false positive genes = 0.1. Values for differentially expressed genes were used to cluster all 50 subjects using Genepilot software (http://www.genepilot. com) (TG Services, El Sobrante, California, USA).

Taqman real time PCR

Taqman real time PCR (Applied Biosystems, Foster City, California, USA) was used to confirm the importance of genes identified by array experiments in the same group of CFS cases (n = 17) and a different group of normal controls (n = 21); the controls were age and sex matched. cDNA was prepared from total RNA using the random hexamer method of reverse transcription, according to the instructions of the kit manufacturer (Applied Biosystems, Warrington, UK). Experiments were performed in triplicate in a custom 384 well low density array format using the ABI PRISM 7900HT instrument (Applied Biosystems) incorporating 38 target gene assays (including three instances where two assays were used for a single gene to include transcript variants) (table 2) along with endogenous controls, namely: HMBS (hydroxymethylbilane synthase), HPRT-1 (hypoxanthine phosphoribosyltransferase 1), GAPDH (glyceraldehyde-3-phosphate dehydrogenase), and eukaryotic 18S rRNA. Fluorogenic probes were 5' labelled with 6-carboxyfluorescein (FAM) and 3' labelled with MGB non-fluorescent quencher. Reactions were performed in a 1 μl reaction volume in Taqman Universal PCR Mastermix (Applied Biosystems UK). For each card channel (48 PCR reactions), 50 ng total RNA from PBMCs converted to cDNA was used as inoculum. Fluorescent signal detection used ''ROX'' as the internal passive reference dye.

The inoculum (100 μl) was a mixture of Taqman universal PCR master mix (Applied Biosystems) (50 ml), RNase/DNase free sterile water (45 μl), and cDNA (5 μl). Cycling times and temperatures were as follows. Initial

denaturation was carried out for 10 minutes at 95°C, followed by 40 cycles of denaturation at 95°C for 15 seconds and combined primer annealing/extension at 60°C for one minute. Data was displayed using SDS 2.1 software (Applied Biosystems). GAPDH gave the least variable results in all samples and was used as the endogenous reference control. The threshold cycle (Ct) for each gene/sample pair was compared with a calibrator sample and a DCt value used to calculate a relative quantity of gene expression compared with the calibrator. RQ values were finally normalised to GAPDH expression. The F test for equality of standard deviations was used to compare mean RQ values for each gene in test versus control groups.

RESULTS

CFS disease phenotype

All 25 patients with CFS were diagnosed according to the Centers for Disease Control criteria of Fukuda *et al.* Table 1 summarises the patient and clinical details. Our study included patients with CFS whose disease was severe and necessitated bed rest for much of the day, and also several patients whose disease was of a milder nature (Table 1).

Microarray analysis

Analysis of microarray data identified 35 genes that showed significantly different expression in patients with CFS compared with normal controls (Table 2). Hierarchical clustering of subjects on the basis of their degree of expression of these 35 genes revealed a cluster of 18 subjects, comprising 17 CFS patients and one normal person, whose expression profiles were very similar but distinctly different from the other patients and controls (Fig 1). However, this patient cluster did not differ significantly with regard to other clinical variables shown in Table 1.

Taqman real time PCR

TaqMan real time PCR analysis was used to confirm the importance of genes that were identified using microarray analysis. Significantly different expression, with the same profile as in gene arrays, was confirmed for 16 of 33 genes in the 17 patients tested. This method revealed upregulation of 15 genes and downregulation of one gene (Table 2). Figure 2 illustrates the concurrent fold difference in gene expression between test and control groups for the 16 differentially expressed genes for both microarray and real time PCR analysis. The standard deviation from the mean is also shown in Figure 2B and provides an indirect measure of the probable reproducibility of differential expression for each gene. In general, the standard deviation of these 16 genes in normal persons is very much lower than in patients with CFS, except for IL-10RA, in which the opposite is the case.

Table 3 provides information on the chromosomal location, expression, subcellular localisation, and function of these 16 genes. Although these genes do not fit neatly into known metabolic pathways, several broad themes are apparent. For example, T cell activation and neuronal and mitochondrial function.

The expression of 16 genes was significantly different in patients compared with controls in both microarray analysis and real time PCR. These genes may be important in the pathogenesis of CFS and can be grouped according to immune, neuronal, mitochondrial, and other functions that have particular relevance to the present knowledge of the epidemiology of CFS (table 4). Our present study has certain parallels with two published studies in this area, summarised in table 4. T cell activation is suggested by upregulation of CD2BP2 and downregulation of IL-10RA; in addition, PRKCL1 plays a role in the immune response. Genes that are active in the immune response have been

found to be differentially expressed in all studies of gene expression in CFS (table 4). Furthermore, genes that are crucial for T cell activation have also been found to be upregulated in all three studies, namely: CD2BP2 and IL-10RA (present study); moesin and cathepsin C4; ITGA and NFATC3.5 These findings are consistent with previous work showing that patients with CFS have evidence of immune activation, such as increased numbers of activated T cells and cytotoxic T cells, and raised circulating cytokine concentrations.

A neuronal component is suggested by the upregulation of PRKCL1, NTE, GSN, GABARAPL1, KHSRP, and EIF2B4. Protein kinase C family members are implicated in various psychiatric and affective disorders, and have been implicated in previous gene studies of CFS. NTE is a target for organophosphates and chemical warfare agents, both of which may precipitate CFS, on the basis of a neuropathy resulting from inactivation of serine esterase activity. GSN regulates cell growth and plays a role in amyloidosis (Finnish type), which may result in dysfunction of neurones, skeletal muscle, and thyroid gland. GABARAPL1 is a microtubule associated anchor protein with increased expression in neuronal cells. KHSRP facilitates splicing of the N1 exon of the SRC protooncogene in neuronal but not other cells. EIF2B4 is a mitochondrial translation initiation factor and one of the EIF2B family, within which mutations have been shown to be associated with central nervous system hypomyelination and encephalopathy. Upregulation of an EIF2B3 gene homologue (BQ580379) has been reported. These findings are interesting in that abnormalities in the white matter of the frontal lobes have been found in patients with CFS using magnetic resonance imaging and have been suggested to account for the cognitive defect in CFS. Neuronal gene involvement in CFS has also been reported. Mitochondrial involvement is suggested by the upregulation of EIF2B4, EIF4G1 (see above), and MRPL23. Mitochondrial gene upregulation has also been reported.

The cell cycle is implicated by upregulation of ANAPC11, which regulates the onset of anaphase by mediation of degradation of mitotic cyclins. Upregulation of MAD1L1, which prevents the onset of anaphase until all chromosomes are aligned at the metaphase plate, has been reported.

Transcriptional perturbation is suggested by the upregulation of POLR2G and BRMS1. The upregulation of genes homologous with POLR1B (BQ580386) and RCOR3 (BQ580388), which are each involved in transcriptional regulation has been reported.

Upregulated peroxisomal function is suggested by the upregulation of ABCD4 and PEX16, which may suggest enhanced defence to oxidative stress in CFS. Oxidative stress has already been suggested as a disease mechanism in CFS. Persistent virus infection is a recognised feature of CFS, which is interesting in the light of our finding of upregulation of EIF4G1 transcript variant 5, a mitochondrial translation initiation factor. This finding has also been reported in patients with CFS who have rapid (?triggered by virus infection) as compared with insidious onset. EIF4G1 is a component of the protein complex, EIF4F, which is crucial in translation through its involvement in the recognition of the mRNA cap, ATP dependent unwinding of 59 terminal secondary structure, and recruitment of mRNA to the ribosome. Various viruses have developed strategies to divert EIF4G1 from its utilisation by the cellular machinery to facilitate production of viral proteins. The best characterised example is that of poliovirus, but this has also been demonstrated to occur with coxsackie virus, rhinoviruses, rotavirus, influenza virus, adenovirus, vesicular stomatitis virus, and human immunodeficiency virus 1. Therefore, the upregulation of EIF4G1 identified in our present study may represent a common host response to persistent infection with several different

viruses. The vulnerability of EIF4G1 to virus modification may have particular importance for the development of CFS after an acute virus infection. In conclusion, we report the differential expression of 16 human genes in patients with CFS compared with normal controls. The involvement of genes from several disparate pathways suggests a complex pathogenesis involving T cell activation and abnormalities of neuronal and mitochondrial function, and suggests possible molecular bases for the recognised contributions of organophosphate exposure and virus infection, respectively.

Table 1 Patient information including summary of aspects used for the CFS case definition and scores for physical fatigue, mental fatigue, anxiety and depression.

Clinical parameter	**Proportion**
Gender	36% male / 64% female
Age	Mean = 40.6 years
Duration of disease (years)	Mean = 2.3 years
CFS began with flu-like illness	60%
Impaired memory	88%
Impaired concentration	84%
Sore throat	68%
Myalgia	80%
Arthralgia	52%
Headache	76%
Unrefreshing sleep	84%
Post-exertional malaise	92%
Tender lymphadenopathy	72%
Autonomic features	80%
Atopic features	52%
Physical fatigue (Chalder) (most severe = 21)	Mean = 17.2; Range 12-21
Mental fatigue (Chalder) (most severe = 12)	Mean = 9.28; Range 5-12
Anxiety score (HAD) (most severe = 21)	Mean = 9.88; Range 3-18
Depression score (HAD) (most severe = 21)	Mean = 7.96; Range 1-20

Table 2 Genes showing a significant >1.5 fold change between cases and controls in microarray experiments

GenBank Accession No.	Homo sapiens gene name (mRNA)	Gene symbol[D]	Fold difference (array)	P value	Taqman assay ID	Fold difference (PCR)	P value
NM_020325	ATP-binding cassette, sub-family D (ALD), member 4, tv-4	**ABCD4**	3.398	0.00257	Hs00245340_m1	1.825	0.0019
NM_002741	Protein kinase C-like 1	**PRKCL1**	2.288	0.00438	Hs00177028_m1	2.669	1.09×10^{-5}
NM_021134	Mitochondrial ribosomal protein L23	**MRPL23**	2.175	0.00135	Hs00221699_m1	4.032	1.25×10^{-6}
NM_001558	Interleukin 10 receptor, alpha	**IL10RA**	0.252	0.00225	Hs00387004_m1	0.395	2.34×10^{-12}
NM_006110	CD2 antigen (cytoplasmic tail) binding protein 2	**CD2BP2**	2.311	0.00158	Hs00272036_m1	2.411	2.6×10^{-4}
NM_000177	Gelsolin (amyloidosis, Finnish type)	**GSN**	2.141	0.00105	Hs00609276_m1	1.83	6.23×10^{-7}
NM_006702	Neuropathy target esterase	**NTE**	2.877	0.00043	Hs00198648_m1	6.409	1.31×10^{-12}
NM_002696	polymerase (RNA) II (DNA directed) polypeptide G	**POLR2G**	2.264	0.00372	Hs00275738_m1	3.407	3.69×10^{-6}
NM_004813	Peroxisomal biogenesis factor 16, tv-1	**PEX16**	3.004	0.00213	Hs00191337_m1	1.758	0.0126
NM_015636	Eukaryotic translation initiation factor 2B, subunit 4 delta, tv-1	**EIF2B4**	2.414	0.00125	Hs00248984_m1	1.882	1.8×10^{-5}
NM_004953	Eukaryotic translation initiation factor 4 gamma, 1, tv-5	**EIF4G1**	3.081	0.00177	Hs00191933_m1	2.964	7.63×10^{-13}
NM_016476	APC11 anaphase promoting complex subunit 11 homolog	**ANAPC11**	3.278	0.00219	Hs00212858_m1	2.366	3.5×10^{-6}

NM_002598	Programmed cell death 2, tv-1	**PDCD2**	2.052	0.00266	Hs00751277_sH	1.887	1.3×10^{-4}
NM_003685	KH-type splicing regulatory protein (FUSE binding protein 2)	**KHSRP**	2.366	0.00139	Hs00269352_m1	1.64	0.0022
NM_015399	Breast cancer metastasis-suppressor 1	**BRMS1**	2.246	0.00219	Hs00363036_m1	1.598	0.003
NM_031412[B]	GABA(A) receptor-associated protein like 1	**GABARAPL1**	2.358	0.00046	Hs00744468_s1	1.838	0.0097
NM_004655	Axin 2 (conductin, axil)	AXIN2	0.351	0.00056	Hs00610344_m1	0.780	0.023
NM_019051[A]	Mitochondrial ribosomal protein L50	MRPL50	0.427	0.00409	Hs00745120_s1	0.815	0.0019
NM_005103[C]	Fasciculation and elongation protein zeta 1 (zygin I), tv-1	FEZ1	0.336	0.00007	Hs00192714_m1	3.485	2.3×10^{-6}
NM_002093	Glycogen synthase kinase 3 beta	GSK3B	2.294	0.00238	Hs00275656_m1	0.763	0.0015
NM_005098	Musculin (activated B-cell factor-1)	MSC	0.374	0.00121	Hs00231955_m1	1.118	0.036
NM_138325	Paired basic amino acid cleaving system 4, tv-6	PACE4	0.473	0.00414	Hs00159844_m1	2.093	1.2×10^{-4}
NM_003584	Dual specificity phosphatase 11	DUSP11	0.366	0.00005	Hs00186058_m1	1.725	0.134
NM_002483	Carcinoembryonic antigen-related cell adhesion molecule 6	CEACAM6	0.301	0.00048	Hs00366002_m1	1.585	0.182
NM_019051[A]	Mitochondrial ribosomal protein L50	MRPL50	0.427	0.00409	Hs00747929_m1	1.136	0.178
NM_030575	Hypothetical protein MGC10334	MGC10334	2.584	0.00118	Hs00257998_s1	1.724	0.067
NM_031412[B]	GABA(A) receptor-associated protein like 1	GABARAPL1	2.358	0.00046	Hs00740588_mH	0.93	0.127
NM_007175	Chromosome 8 open reading frame 2	C8orf2	2.554	0.00206	Hs00200360_m1	1.198	0.052
NM_032118	Hypothetical protein FLJ12953	FLJ12953	1.931	0.00309	Hs00259557_m1	1.139	0.203
NM_004879	Etoposide induced 2.4 mRNA	EI24	0.45	0.00181	Hs00747550_m1	1.03	0.085
NM_079834	Secretory carrier membrane protein 4	SCAMP-4	2.8	0.00069	Hs00365263_m1	1.586	0.173
NM_022145	Leucine zipper protein FKSG14	FKSG14	0.345	0.00205	Hs00259557_m1	1.212	0.345

NM_005103[C]	Fasciculation and elongation protein zeta 1 (zygin I), tv-1	FEZ1	0.336	0.00007	Hs00363763_m1	1.062	0.41
NM_003409	Zinc finger protein 161 homolog	ZFP161	0.415	0.00146	NA	NT	NT
NM_007075	JM5 protein	JM5	2.383	0.00172	NA	NT	NT
NM_005345	Heat shock 70kDa protein 1A	HSPA1A	3.198	0.00003	NA	NT	NT
NM_017616	Hypothetical protein FLJ20004	FLJ20004	2.679	0.00237	NA	NT	NT
NM_020663	Ras homolog gene family, member J	ARHJ	2.277	0.00346	NA	NT	NT

Table 3 Chronic fatigue syndrome associated genes: chromosomal location, expression, subcellular localisation, and function

GenBank Accession No.	*Homo sapiens* gene name (mRNA)	Gene symbol*	Chromosomal location	Expression	Subcellular localisation	Gene function
NM_020325	ATP-binding cassette, sub-family D (ALD), member 4, tv-4	ABCD4	14q24.3	Ubiquitous	Peroxisomal membrane	ALDP expression
NM_002741	Protein kinase C-like 1	PRKCL1	19p13.1-p12	Ubiquitous	Cytoplasm	Regulation of cell motility
NM_021134	Mitochondrial ribosomal protein L23	MRPL23	11p15.5	Widely	Mitochondrial matrix	Protein biosynthesis
NM_001558	Interleukin 10 receptor, alpha	IL10RA	11q23.3	-	Unknown	Cytokine signalling
NM_006110	CD2 antigen (cytoplasmic tail) binding protein 2	CD2BP2	16p12.1	-	Unknown	-
NM_000177	Gelsolin (amyloidosis, Finnish type)	GSN	9q33.3	-	Unknown	Severing and capping of actin
NM_006702	Neuropathy target esterase	NTE	19p13.3-p13.2	Nervous system	Plasma membrane	Neurodegenerative disease
NM_002696	polymerase (RNA) II (DNA directed) polypeptide G	POLR2G	11q13.1	Nervous system	Nucleus	Transcription from Pol II
NM_004813	Peroxisomal biogenesis factor 16, tv-1	PEX16	11p11.11	-	Peroxisomal membrane	Peroxisomal biogenesis
NM_015636	Eukaryotic translation initiation factor 2B, subunit 4 delta, tv-1	EIF2B4	2p23.3	Ubiquitous	Mitochondrion	Initiation of translation

NM_004953	Eukaryotic translation initiation factor 4 gamma, 1, tv-5	EIF4G1	3q27-qter	Ubiquitous	Mitochondrion	Initiation of translation
NM_016476	APC11 anaphase promoting complex subunit 11 homolog	ANAPC11	17q25.3	Brain, heart, pancreas, lung	Nucleus	Ubiquitin ligase activity
NM_002598	Programmed cell death 2, tv-1	PDCD2	6q27	Ubiquitous	Cytoplasm	Apoptosis and regulation of cell proliferation
NM_003685	KH-type splicing regulatory protein (FUSE binding protein 2)	KHSRP	19p13.3	Nervous system	Nucleus	Neurone-specific splicing of the N1 exon of SRC; assembly of other proteins
NM_015399	Breast cancer metastasis-suppressor 1	BRMS1	14q13.1	Brain	Unknown	-
NM_031412[2]	GABA(A) receptor-associated protein like 1	GABARAPL1	12p12.3	Ubiquitous	Unknown	-

Table 4 Possible mechanisms of disease in CFS based on three gene expression studies, each on patients with CFS and normal controls.

Mechanism	Present study Microarray (9,522 genes) / qPCR – 25CFS / 25 Normal	Powell and colleagues Differential display / qPCR – 7 CFS / 4 Normal	Vernon and colleagues Filter array (1764 genes) – 5 CFS – 17 Normal
T-cell activation	CD2BP2, IL-10RA,	Moesin, CTSC	ITGA, NFATC3,
Immune response	PRKCL1	TNF, MAIL	IL-8, CMRF35, ICAM2, ITGB, IER2, PLA2G2A, LCP-1, PRKCL
Neurone	PRKCL1, GSN, KHSRP, NTE, GABARAPL1		HD, PRKCL, ataxin, ADRA2A, DCTN1
Mitochondrion	EIF4G1, EIF2B4, MRPL23	SLC25A16, MCFP	
Skeletal muscle	GSN		DCTN1
Thyroid	GSN	THRAP2, SLC25A16	GDC
Cell cycle	ANAPC11	MAD1L1	
Apoptosis	PDCD2		
Transcription	POLR2G, BRMS1	RCOR3, POLR1B	
Peroxisome	ABCD4, PEX16		

CLAIMS

1. A method of diagnosing chronic fatigue syndrome (CFS) or myalgic encephalomyelitis (ME) comprising:

(a) measuring the level of expression of one or more genes which are differentially expressed in CFS or ME;

(b) comparing the level of expression with that of a normal healthy control.

2. A method as claimed in claim 1 wherein said gene is selected from ABCD4, PRKCL1, MRPL23, IL10RA, CD2BP2, GSN, NTE, POLR2G, PEX16, EIF2B4, EIF4G1, ANAPC11, PDCD2, KHSRP, BRMS1, and GABARAPL1.

3. A method as claimed in claim 1 or claim 2 wherein step (a) is carried out using RT-PCR.

4. A method as claimed in claim 1 or claim 2 carried out on a microarray.

5. A method as claimed in claim 4 wherein said microarray comprises or consists of oligonucleotide probes corresponding to the genes ABCD4, PRKCL1, MRPL23, IL10RA, CD2BP2, GSN, NTE, POLR2G, PEX16, EIF2B4, EIF4G1, ANAPC11, PDCD2, KHSRP, BRMS1, and GABARAPL1.

6. A microarray comprising or consisting of oligonucleotide probes corresponding to the genes ABCD4, PRKCL1, MRPL23, IL10RA, CD2BP2, GSN, NTE, POLR2G, PEX16, EIF2B4, EIF4G1, ANAPC11, PDCD2, KHSRP, BRMS1, and GABARAPL1.

7. The microarray of claim 6, wherein said microarray consists of oloigonucleotides corresponding to the genes ABCD4, PRKCL1, MRPL23, IL10RA, CD2BP2, GSN, NTE, POLR2G, PEX16, EIF2B4, EIF4G1, ANAPC11, PDCD2, KHSRP, BRMS1, and GABARAPL1, and one or more control oligonucleotide probes.

8. The use of a microarray comprising one or more oligonucleotide probes corresponding to those genes which are differentially expressed in CFS or ME in the diagnosis of CFS or ME.

9. The uses as claimed in claim 8 wherein said microarray is a microarray as defined in claim 6 or claim 7.

10. A kit for diagnosing CFS or ME comprising one or more pairs of suitable primers to carry out RT-PCR to measure the level of expression of one or more genes which are differentially expressed in CFS or ME patients.

11. A kit for diagnosing CFS or ME comprising a microarray comprising one or more oligonucleotide probes corresponding to those genes which are differentially expressed in CFS or ME in the diagnosis of CFS or ME.

12. A kit as claimed in claim 10 or claim 11 wherein said gene is selected from ABCD4, PRKCL1, MRPL23, IL10RA, CD2BP2, GSN, NTE, POLR2G, PEX16, EIF2B4, EIF4G1, ANAPC11, PDCD2, KHSRP, BRMS1, and GABARAPL1.

13. A kit as claimed in claim 11 comprising a microarray as defined in claim 6 or claim 7.

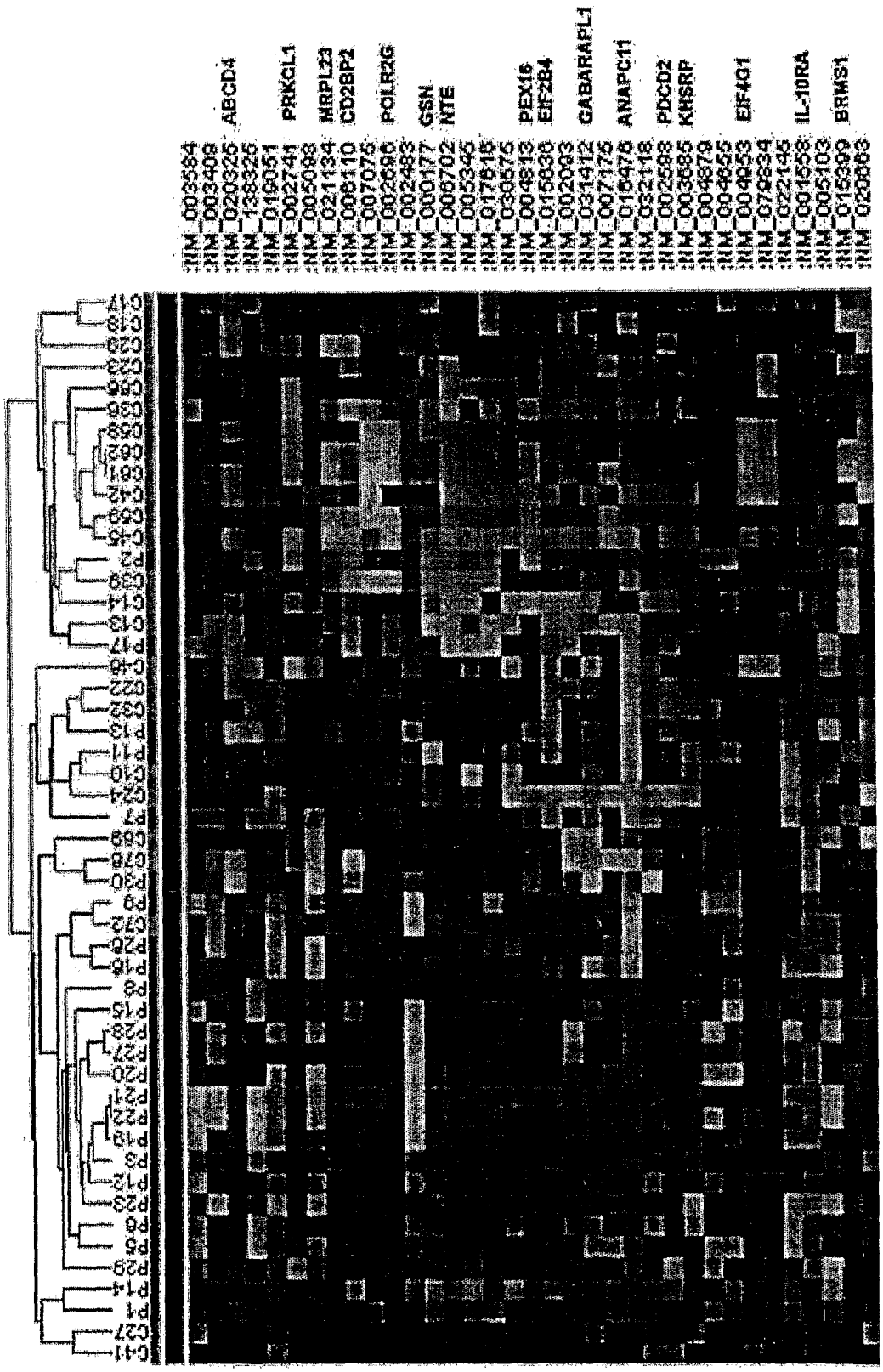

NM_003384
NM_003409
NM_020325
NM_138325
NM_019051
NM_002741
NM_021134
NM_005088
NM_006110
NM_007075
NM_002696
NM_002483
NM_000177
NM_006702
NM_005345
NM_017816
NM_030575
NM_004813
NM_015830
NM_002043
NM_031412
NM_007175
NM_018478
NM_032118
NM_002588
NM_003585
NM_004879
NM_004855
NM_004953
NM_079834
NM_022145
NM_001558
NM_005103
NM_015399
NM_020063
ABCD4
PRKCL1
MRPL23
CD28P2
POLR2G
GSN
NTE
PEX16
EIF2B4
GABARAPL1
ANAPC11
PDCD2
KHSRP
EIF4G1
IL-10RA
BRMS1

Figure 1

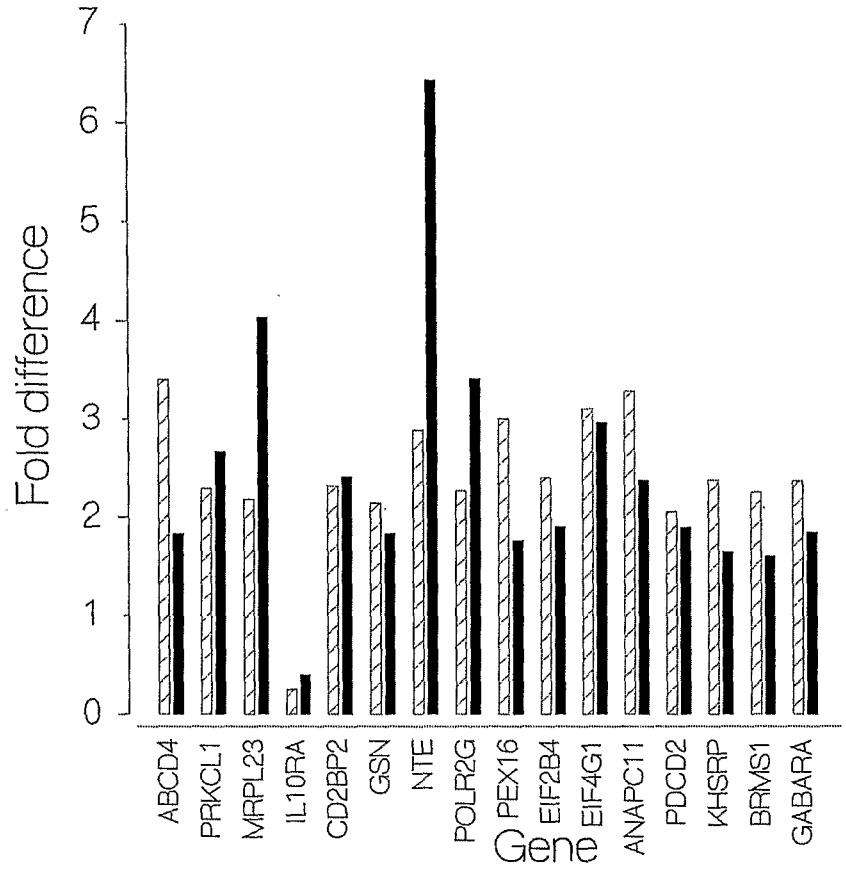

Fold difference
0
1
2
3
4
5
6
7
ABCD4
PRKCL1
MRPL23
IL10RA
CD2BP2
GSN
NTE
POLR2G
PEX16
EIF2B4
EIF4G1
ANAPC11
PDCD2
KHSRP
BRMS1
GABARA
Gene

Figure 2A

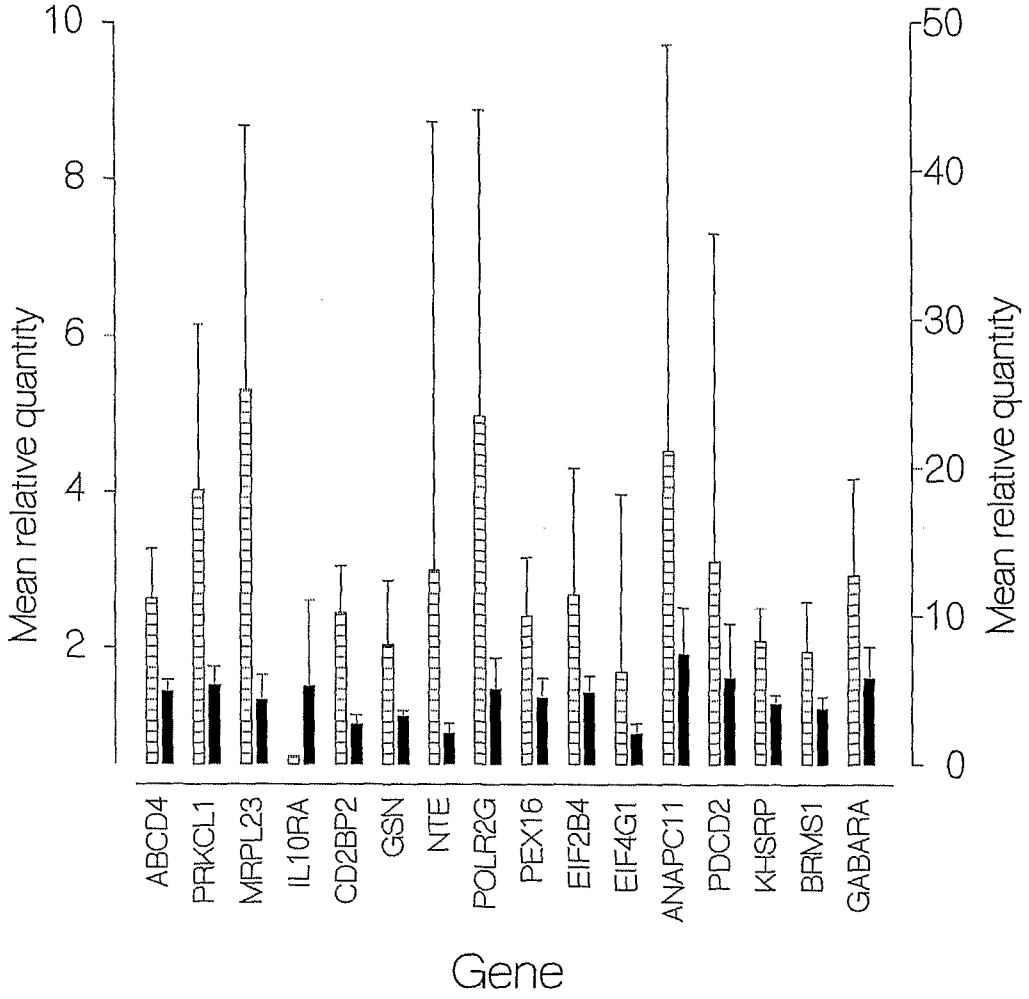

10
8
6
4
2
Mean relative quantity
50
40
30
20
10
0
Mean relative quantity
ABCD4
PRKCL1
MRPL23
IL10RA
CD2BP2
GSN
NTE
POLR2G
PEX16
EIF2B4
EIF4G1
ANAPC11
PDCD2
KHSRP
BRMS1
GABARA
Gene

Figure 2B

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2006/002849

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, EMBASE, BIOSIS, WPI Data, PAJ

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2004/108899 A (US GOVERNMENT [US]; NICHOLSON AINSLEY [US]; VERNON SUZANNE D [US]) 16 December 2004 (2004-12-16) page 33, line 20 - page 34, line 2 page 35, line 15 - line 33 example 5 claims 13,14	1,3,4,8, 10,11
Y		5,9
	-/--	

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
10 October 2006	26/10/2006

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Ulbrecht, Matthias

4

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	VERNON S D ET AL: "Utility of the blood for gene expression profiling and biomarker discovery in chronic fatigue syndrome" DISEASE MARKERS, WILEY, CHICHESTER, GB, vol. 18, no. 4, 2003, pages 193-199, XP002377238 ISSN: 0278-0240 abstract page 194, column 1, paragraph 2 - page 195, column 1, paragraph 1 table 1	1,3,4,8, 10,11
Y		5,9
X	STEINAU M ET AL: "Differential-display PCR of peripheral blood for biomarker discovery in chronic fatigue syndrome" JOURNAL OF MOLECULAR MEDICINE, SPRINGER VERLAG, DE, vol. 82, no. 11, 14 October 2004 (2004-10-14), pages 750-755, XP002377237 ISSN: 0946-2716 the whole document	1,3,4,8, 10,11
Y		5,9
X	WHISTLER T ET AL: "Intergration of gene expression, clinical, and epidemiologic data to characterize Chronic Fatigue Syndrome" JOURNAL OF TRANSLATIONAL MEDICINE, BIOMED CENTRAL, LONDON, GB, vol. 1, no. 10, 1 December 2003 (2003-12-01), pages 1-8, XP002377239 ISSN: 1479-5876 abstract page 2, column 1, paragraph 4 - page 4, column 2, paragraph 1	1,3,4,8, 10,11
Y		5,9
X	KOSARI FARHAD ET AL: "Clear cell renal cell carcinoma: gene expression analyses identify a potential signature for tumor aggressiveness." CLINICAL CANCER RESEARCH : AN OFFICIAL JOURNAL OF THE AMERICAN ASSOCIATION FOR CANCER RESEARCH. 15 JUL 2005, vol. 11, no. 14, 15 July 2005 (2005-07-15), pages 5128-5139, XP002401772 ISSN: 1078-0432 abstract	6,7, 11-13
Y		5,9

-/--

4

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2006/002849

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with Indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WANG ZHINING ET AL: "Identification and utilization of inter-species conserved (ISC) probesets on Affymetrix human GeneChip platforms for the optimization of the assessment of expression patterns in non human primate (NHP) samples." BMC BIOINFORMATICS [ELECTRONIC RESOURCE]. 26 OCT 2004, vol. 5, 26 October 2004 (2004-10-26), page 165, XP002401773 ISSN: 1471-2105 abstract	6,7, 11-13
Y		5,9
P,X	KAUSHIK N ET AL: "GENE EXPRESSION IN PERIPHERAL BLOOD MONONUCLEAR CELLS FROM PATIENTS WITH CHRONIC FATIGUE SYNDROME" JOURNAL OF CLINICAL PATHOLOGY, LONDON, GB, vol. 58, no. 8, August 2005 (2005-08), pages 826-832, XP009065410 ISSN: 0021-9746 the whole document	1-13

4

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/GB2006/002849

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2004108899 A	16-12-2004	AU 2004245998 A1	16-12-2004
		CA 2528162 A1	16-12-2004