Assessment of a 44 Gene Classifier for the Evaluation of Chronic Fatigue Syndrome from Peripheral Blood Mononuclear Cell Gene Expression

Daniel Frampton[1], Jonathan Kerr[2], Tim J. Harrison[3], Paul Kellam[1,4]*

[1] Department of Infection, Division of Infection and Immunity, University College London, London, United Kingdom, [2] Division of Clinical Sciences, St George's University of London, London, United Kingdom, [3] Department of Internal Medicine, University College London Medical School, London, United Kingdom, [4] The Wellcome Trust Genome Campus, The Wellcome Trust Sanger Institute, Cambridge, United Kingdom

Abstract

Chronic fatigue syndrome (CFS) is a clinically defined illness estimated to affect millions of people worldwide causing significant morbidity and an annual cost of billions of dollars. Currently there are no laboratory-based diagnostic methods for CFS. However, differences in gene expression profiles between CFS patients and healthy persons have been reported in the literature. Using mRNA relative quantities for 44 previously identified reporter genes taken from a large dataset comprising both CFS patients and healthy volunteers, we derived a gene profile scoring metric to accurately classify CFS and healthy samples. This metric out-performed any of the reporter genes used individually as a classifier of CFS. To determine whether the reporter genes were robust across populations, we applied this metric to classify a separate blind dataset of mRNA relative quantities from a new population of CFS patients and healthy persons with limited success. Although the metric was able to successfully classify roughly two-thirds of both CFS and healthy samples correctly, the level of misclassification was high. We conclude many of the previously identified reporter genes are study-specific and thus cannot be used as a broad CFS diagnostic.

Citation: Frampton D, Kerr J, Harrison TJ, Kellam P (2011) Assessment of a 44 Gene Classifier for the Evaluation of Chronic Fatigue Syndrome from Peripheral Blood Mononuclear Cell Gene Expression. PLoS ONE 6(3): e16872. doi:10.1371/journal.pone.0016872

Editor: Raya Khanin, Memorial Sloan Kettering Cancer Center, United States of America

Received July 28, 2010; **Accepted** January 7, 2011; **Published** March 30, 2011

Funding: This study was funded by a research grant from the Chronic Fatigue Syndrome Research Foundation (http://www.cfsrf.com). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: pk5@sanger.ac.uk

Introduction

Chronic fatigue syndrome (CFS) is a clinically defined illness with a broad range of symptoms including severe and debilitating fatigue, muscle pain, sleep disruption, difficulties with concentration, memory impairment and headaches. It is estimated to affect 0.4% of the population in Europe and North America [1] and cost $9 billion annually in lost productivity in the USA alone [2]. The cause and pathogenesis of CFS remain poorly understood, although various infectious triggers have been proposed.

There are currently no specific laboratory-based tests that provide a robust diagnosis of CFS. However, previous studies indicate significant differences in the patterns of gene expression in peripheral blood leukocytes between patients with CFS and healthy individuals [3–10]. Although many of these studies have not detailed predictive sets of genes that could be used to make a diagnosis of CFS on the basis of their expression, we showed previously that the expression levels of 88 "CFS reporter genes", identified using microarrays, could assign individuals to CFS disease or healthy control groups following quantitative PCR of PBMC RNA [7,8].

Microarray analysis has been used frequently to identify groups of genes associated with various diseases, including infectious diseases [11], autoimmune diseases and cancer [12]. In such studies, a microarray dataset featuring tens of thousands of genes is computationally reduced to several hundred genes found to be significantly differentially expressed between healthy and diseased individuals, or between different stages of disease. Computational methods, such as support vector machines (e.g. [13]), artificial neural networks (e.g. [14]) and simple selective naïve Bayes classifiers [15], are able to identify such gene sets for disease classification. These methods require training the underlying statistical models on data representative of diseased and healthy phenotypes in order to make such predictions. Ideally, models are trained on a well-characterised dataset and evaluated using a separate, preferably blinded, test set consisting of new samples from diseased and healthy individuals. The use of different non-array based methods for quantification of gene expression, such as reverse transcription polymerase chain reaction (RT-PCR) is also desirable. On this basis, a gene profile can be formally assessed as a multiplex diagnostic tool.

Here, we have undertaken such an analysis to determine the predictive power of our 'CFS signature genes' identified previously [7,8]. We have assessed the CFS disease predictive genes in the original study data and in a new blinded sample set of CFS disease and healthy control samples. We show, using a variety of methods, that these genes do not identify robustly patients with CFS disease.

Results

CFS class prediction using a 44 gene classifier

To develop a robust CFS diagnostic metric, we used as a training set the mRNA relative quantities (RQ, defined as $2^{-\Delta\Delta CT}$) for the 44 most discriminating reporter genes identified

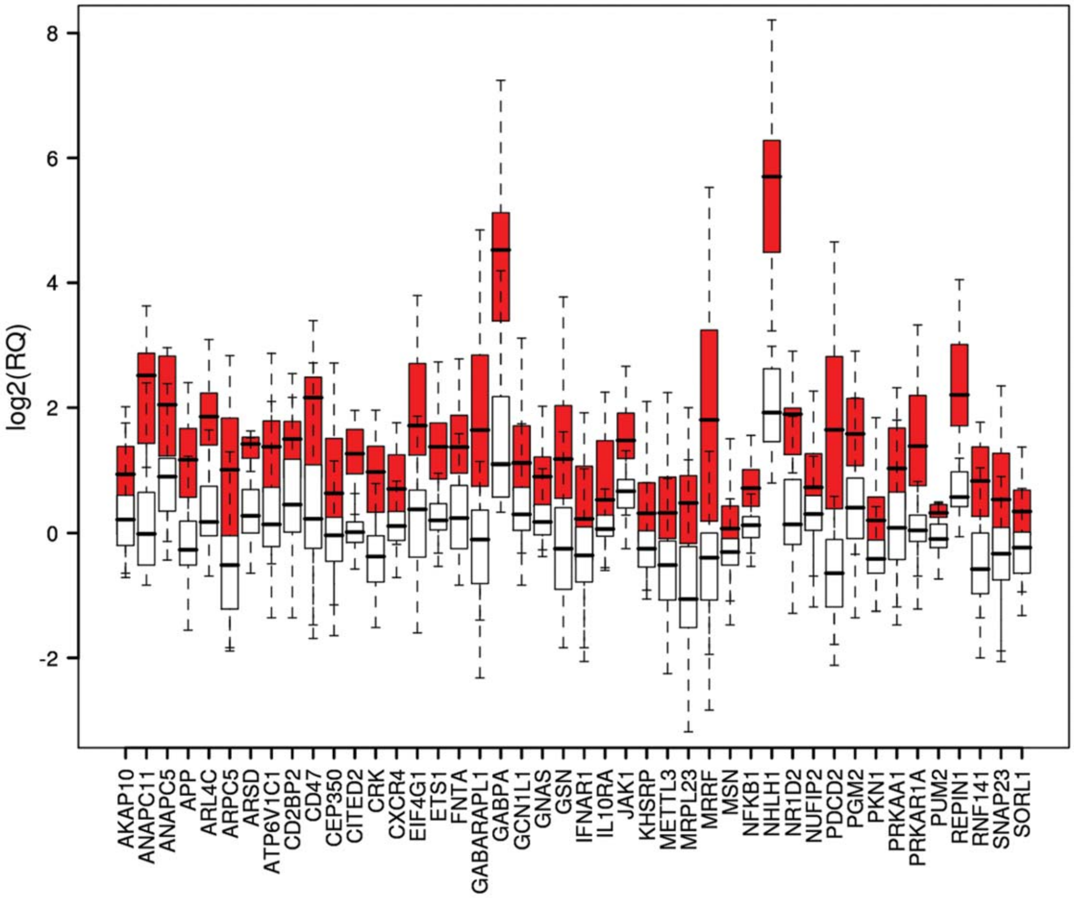

Figure 1. A boxplot illustrating the distribution of $\log_2$(RQ) values for each of the reporter genes in the training set. Distributions for CFS samples are shown in red, healthy samples in white. The boxes represent the inter-quartile range (25–75%) of the data with the median being shown as a solid horizontal line within each box. Whiskers extend to 1.5 times the inter-quartile range. The data is $\log_2$ scaled and outliers omitted for purposes of clarity.
doi:10.1371/journal.pone.0016872.g001

previously by us [7,8] (Table S1). The data were normalised to GAPDH and to a calibrator sample. Initially, we assessed the ability of each of the individual reporter genes to be used as an accurate predictor of CFS disease in this dataset of PBMC gene expression, determined for 56 patients suffering from CFS and 75 healthy volunteers.

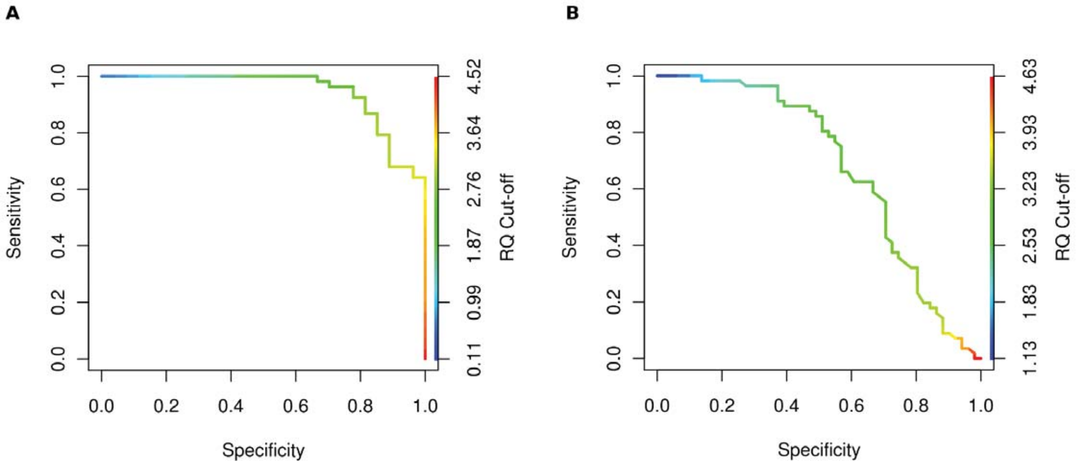

Figure 2. ROC curves obtained when using (A) ARL4C and (B) NUFIP2 as CFS predictors on the training set. The graphs are coloured according to the scale bar on the right hand y-axes to indicate the RQ value cut-offs associated with each pair of sensitivity and specificity values at that particular point in the ROC curve. The AUCs were found to be: (A) 0.94 and (B) 0.67.
doi:10.1371/journal.pone.0016872.g002

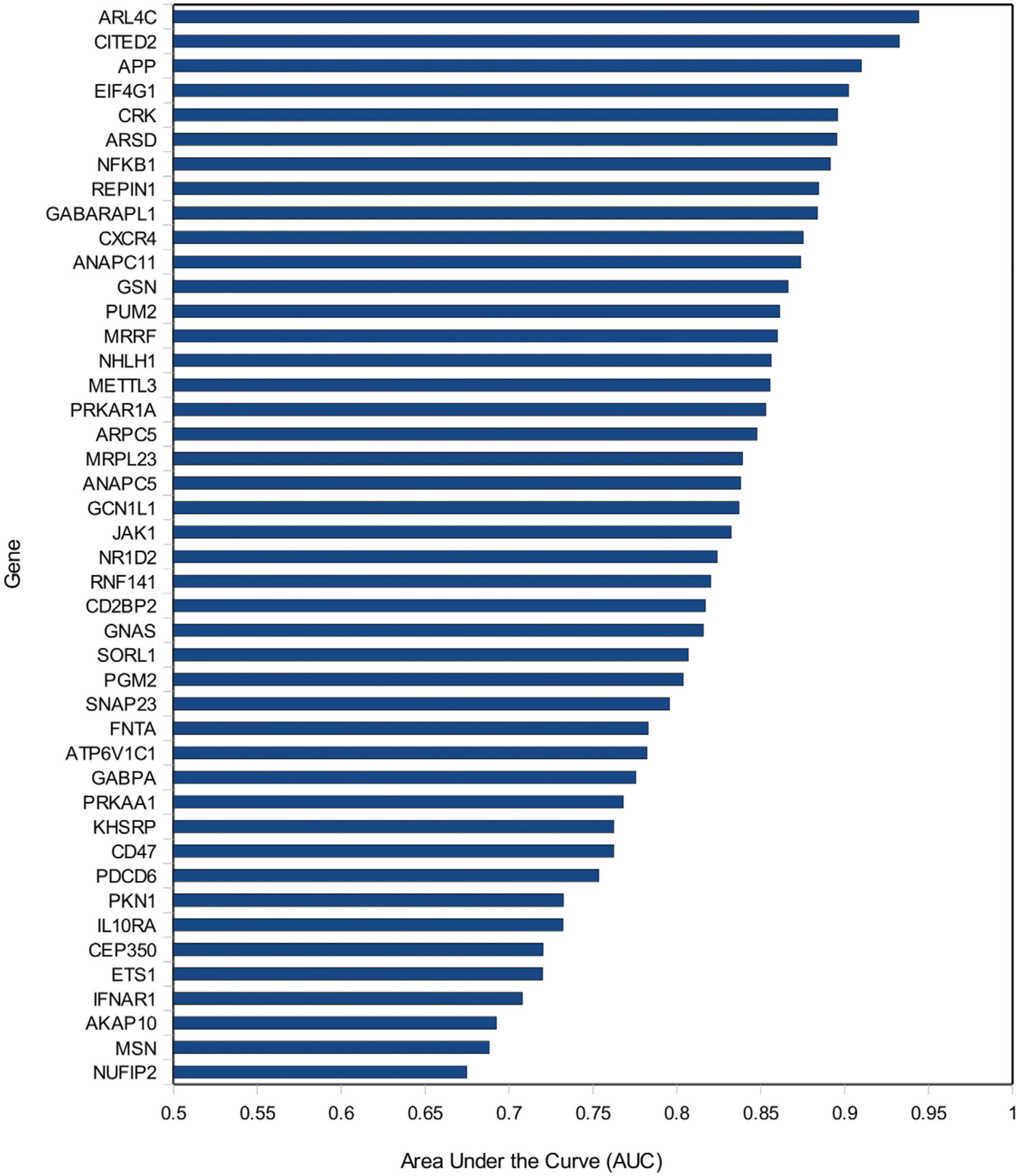

Figure 3. Area under the curve (AUC) values for each of the reporter genes when used as CFS predictors on the training set. All perform better than random (AUC = 0.5).
doi:10.1371/journal.pone.0016872.g003

For each of the 44 reporter genes, we assessed the distribution of RQ values, finding that each of the genes was expressed at higher levels in the CFS samples than controls (Figure 1). In some cases, the RQ values of the reporter genes were clearly separate and the mean RQ differed significantly between CFS and controls, for example ANAPC11. For other classifier genes, although the mean RQ values differed significantly between CFS and control samples, the separation of the individual RQ values was not so distinct, for example MRRF.

For each of the genes we performed receiver operator curve (ROC) analysis, using different RQ value cut-offs to produce ROC curves and a corresponding assessment of sensitivity, specificity, true

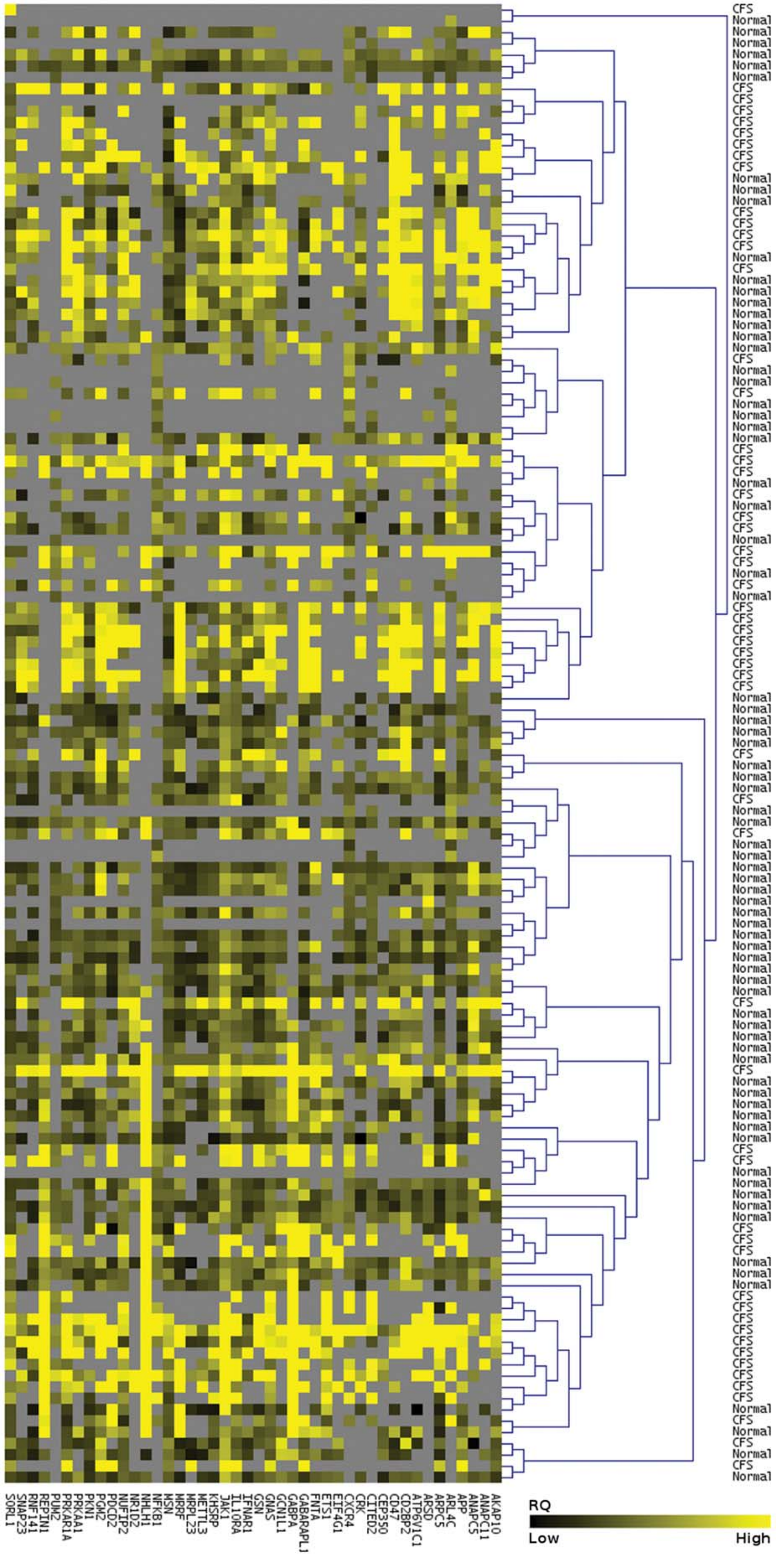
CFS
Normal
Normal
Normal
Normal
Normal
Normal
CFS
CFS
CFS
CFS
CFS
CFS
CFS
CFS
Normal
Normal
Normal
CFS
CFS
CFS
CFS
Normal
CFS
Normal
Normal
Normal
Normal
Normal
Normal
Normal
CFS
Normal
Normal
CFS
Normal
Normal
Normal
Normal
CFS
CFS
CFS
Normal
CFS
Normal
CFS
CFS
Normal
CFS
CFS
Normal
CFS
Normal
CFS
CFS
CFS
CFS
CFS
CFS
CFS
CFS
Normal
Normal
Normal
Normal
Normal
CFS
Normal
Normal
Normal
CFS
Normal
Normal
CFS
Normal
Normal
Normal
Normal
Normal
Normal
Normal
Normal
Normal
Normal
Normal
Normal
Normal
Normal
CFS
Normal
Normal
Normal
Normal
Normal
CFS
Normal
Normal
Normal
Normal
Normal
Normal
CFS
CFS
Normal
Normal
Normal
Normal
Normal
CFS
CFS
CFS
Normal
Normal
Normal
CFS
CFS
CFS
CFS
CFS
CFS
CFS
CFS
CFS
CFS
Normal
CFS
Normal
CFS
Normal
CFS
Normal
SORL1
SNAP23
RNF141
REPIN1
PUM2
PRKAR1A
PRKAA1
PKN1
PGM2
PDCD2
NUFIP2
NR1D2
NHLH1
NFKB1
MSN
MRRF
MRPL23
METTL3
KHSRP
JAK1
IL10RA
IFNAR1
GSN
GNAS
GCN1L1
GABPA
GABARAPL1
FNTA
ETS1
EIF4G1
CXCR4
CRK
CITED2
CEP350
CD47
CD2BP2
ATP6V1C1
ARSD
ARPC5
ARL4C
APP
ANAPC5
ANAPC11
AKAP10
RQ
Low
High

Figure 4. Hierarchical clustering across samples of RQ values from the training set by Euclidean distance with average linkage. Reporter genes are arranged in columns, samples in rows. Missing values are shown in grey.
doi:10.1371/journal.pone.0016872.g004

and false positive rates (TPR, FPR) for each gene. We used the results of the ROC analyses to identify an RQ value cut off for each gene that maximised the true positive, and minimised the false positive, classification of CFS samples. A sample was classified as CFS-positive if its RQ value was greater than the calculated cut-off.

Consistent with the distribution of RQ values for each gene (Figure 1) the area under the curve (AUC) from the ROC analysis for the 44 reporter genes varies substantially, from 0.67 to 0.94 (Figures 2, 3 and Table S2). ARL4C was found to be the best CFS predictor (AUC = 0.94), whilst NUFIP2 performs poorly (AUC = 0.67); thus, NUFIP2 performs only marginally better than a random predictor (AUC = 0.5). Consequently, the individual 44 reporter genes vary considerably in their ability to predict CFS in the training set accurately, although all perform better than a random predictor (Figure 3). Those with high AUCs yield fewer false positives at low RQ cut-offs and fewer false negatives at higher RQ cut-offs.

Using individual genes as classifiers is often problematic due to missing data, where gene expression levels are not available for all patients and individual variation of single gene expression confounds predictive power. Thus, in principle, no single reporter gene would be able to classify CFS accurately across all samples. We therefore explored the data structure of the 44 gene expression set using unsupervised hierarchically clustering of the PCR-derived RQ gene expression levels in the training set using Euclidean distance (Figure 4). Although there is clearly an underlying structure to the data, with many of the CFS and control samples clustering in distinct groups, there is also clearly overlap between the two groups. This is most likely due to missing data and occasionally large RQ outliers skewing the clustering. Consistent with this observation, the 44 reporter genes could not produce a CFS classifier using support vector machines by training on a subset of the training set and assessing the predictive power on a separate subset (data not shown). To minimise the effect of missing gene expression levels in each sample, we filtered the training set to include only samples for which there were at least 22 different RQ gene expression values (i.e. 50% of the reporter genes). However, this did not improve the ability to predict CFS cases using hierarchical or k-means clustering, or SVMs.

Gene Profile Score Metric

Because standard clustering and classification methods were unable to classify the training set, we developed a scoring metric based around a gene-profiling approach. For each sample, a gene profile was generated using a binary classification of CFS or control for each reporter gene: a sample with a RQ value greater than the defined gene specific cut-off corresponding to a 5% false positive rate determined from the earlier ROC analysis was assigned a '1' ("present"), or otherwise a '0' ("absent"). Any missing data values were assigned a '0'. The per gene score is summed over the 44 genes, resulting in a profile score of between 0 and 44 for each sample. Samples with high scores should be over represented for CFS disease whilst samples with low scores all should be healthy controls.

We assessed the ability of this metric to differentiate between CFS and healthy samples in the training set again using ROC analysis on the profile score, obtaining true and false positive rates for a range of profile score cut-offs (Figure 5). This method results in an AUC of 0.95, with 47 CFS (84%) and 72 healthy controls (96%) correctly predicted in the 131 sample training set. This therefore represents a better predictor of CFS than any of the genes used individually (Figure 3).

We applied this gene profile score metric to analyse a blind study dataset, selecting a profile score cut-off of 5, above which samples were classed as CFS disease and equal to or below were classed as healthy controls. At this level of score threshold, the training set gave a FPR of 5%. Under these criteria, our predictor

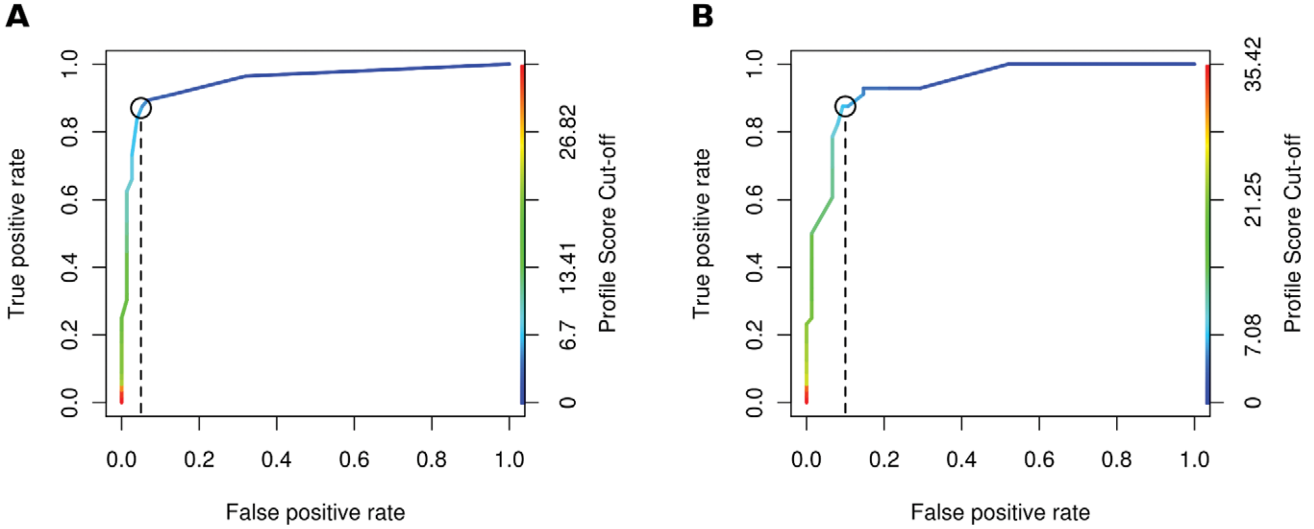

Figure 5. ROC curves for the gene profile score metric using a false positive rate (FPR) on the training set of (A) 5% and (B) 10%. For more detail, see Materials and Methods. The graph is coloured to indicate the gene profile score cut-offs associated with each pair of true and false positive rates at that particular point in the ROC curve. The range of profile cut-off scores are coloured on the right hand y-axes of each graph. The profile score cut-offs used to classify the test set were derived by selecting scores yielding FPR's of (A) 5% and (B) 10%, as indicated by the dashed lines.
doi:10.1371/journal.pone.0016872.g005

Table 1. The performance of the gene profile score metric in classifying samples in the test set.

Method		Positive	Negative
5% FPR	True text	22	51
	False text	10	45
10% FPR	True text	43	37
	False text	24	24

Cut-offs were used on the test set corresponding to false positive rates (FPR) of 5% and 10% on the training set.
doi:10.1371/journal.pone.0016872.t001

assigned from the 128 blinded samples, 32 CFS positive samples, of which 22 were true-positives (69%) and 10 were false-positives (31%) (Table 1). Correspondingly, it predicted 96 healthy samples of which 51 were true negatives (53%) and 45 were false negatives (47%). Overall, this resulted in 73 (58%) correct classifications (TP+TN), based on our gene expression scoring metric. An empirical assessment of the blinded test set based on expert assessment of the RQ values resulted in the correct class prediction of 79 of 128 samples (61%).

To determine if the predictive power that was evident in the training data was lost on the test data due to a subset of poorly performing genes we looked at the number of correctly predicted samples per gene (Figure 6). This showed that except for 3 genes (MRPL23, PKN1 and ARL4C) all other genes had significantly lower predictive power in the test set. Although a number of reporter genes are good CFS predictors in the training set, the majority perform no better than random, 32 of the 44 making more incorrect than correct predictions on the test set (Tables S3, S4). Those that are good CFS predictors in the training set are not necessarily good predictors in the test set: there is no correlation between the number of correct predictions made by each reporter gene individually across the two data sets (Pearson's correlation coefficient = 0.07).

To determine if the calibrator reference sample confounded the RQ-based analysis we investigated the properties of normalized real time ΔCT values. Performing ROC analysis and generating profile scores for the ΔCT values, we found no significant difference in predictive outcome for the 44 reporter genes in terms of their ability to classify CFS and healthy samples, either when used as individual classifiers or by combining them to generate profile scores (Figure S1, Tables S5, S6).

For the training set, mean RQ values are significantly higher in CFS than healthy samples for each of the 44 reporter genes (Welch 2-sample t-test, $P<0.01$). However, no significant difference is observed between the means for any of these genes in the test set (Welch 2-sample t-test, all $P>0.01$). Thus, although the reporter genes could be used (either by profile score or, to a lesser extent, individually) to classify CFS in the original study, they are not robust in a new sample population.

Discussion

Here we have assessed the ability of a proposed 44 gene classifier [8] to discriminate between CFS patients and healthy control individuals. This classifier was able to discriminate correctly between CFS and healthy control samples in 95% of the training samples. However, when assessed on a new, blinded 128 sample test set only 58% of samples were predicted correctly. Importantly, a high number of false-positive predictions were made, with 31% of CFS-positive predictions being from healthy volunteers. In addition, a high number of false negative predictions were made, 57% of the CFS samples being predicted as healthy controls. Therefore, with the methods used here we cannot predict CFS disease based on the analysis of expression of 44 classifier genes in the peripheral blood.

There may be several reasons for the poor performance on the blinded test set for what, as far as the training set is concerned, would otherwise be considered a good predictive metric. Firstly, the low 5% FPR per gene in the development of cut-off scores for the scoring metric may have resulted in over-fitting to the training set. This seems unlikely, however, as similar results were observed based on a more relaxed training set FPR of 10% with the training set still producing a classifying AUC of 0.94 (Figure 5). Although roughly two-thirds (43/67, 64%) of the CFS samples from the blinded test set were classified correctly using a profile score cut-off corresponding to a FPR of 10%, the rate of misclassification was

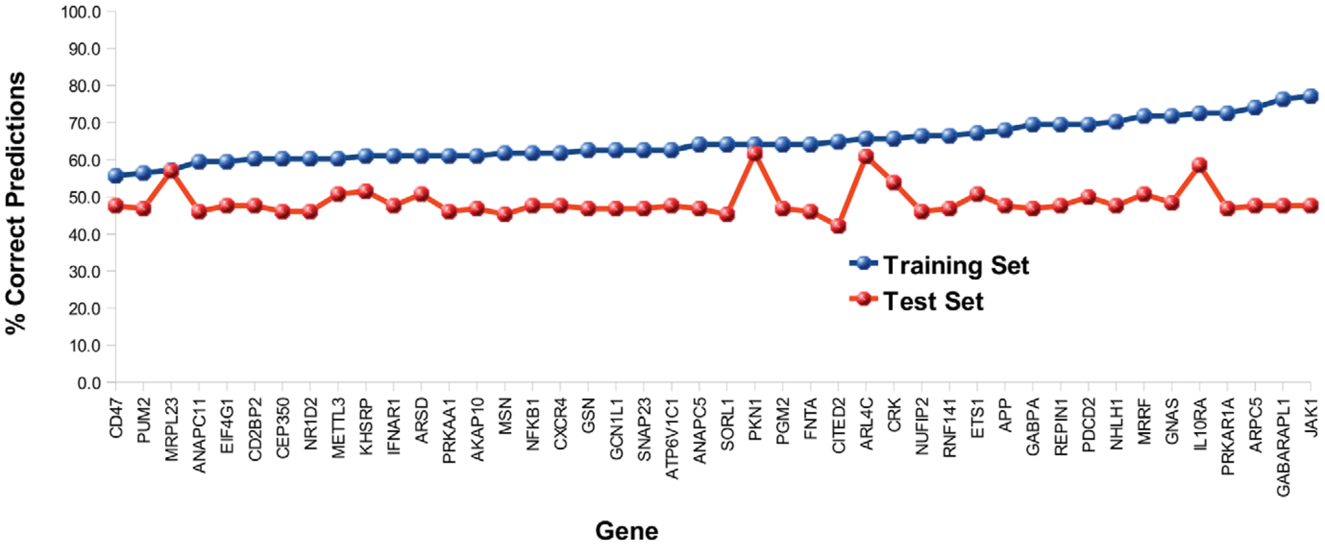

Figure 6. Percentage of predictions made correctly by each of the reporter genes using cut-offs corresponding to a FPR of 5% on the training set (blue) and the test set (red).
doi:10.1371/journal.pone.0016872.g006

higher (24/67, 36%; Table 1). Secondly, there may be fundamental biological differences between the training and test set due to for example, population stratification, age, onset of CFS or other factors which confound both clustering and gene-profiling approaches. In the worst case, the 44 reporter genes may not be representative beyond the small study in which they were identified [7].

The determination of microarray study specific gene classifiers has lead to the proposed identification of various CFS reporter genes, based on microarray analysis of gene expression [3–10]. There is little overlap in the gene sets, a reflection of the difference between studies. Larger CFS microarray studies, together with analysis by comparison with different control groups, may help in identify CFS disease-classifying gene sets. Alternatively, a meta-analysis of existing CFS data sets would provide a valuable extension with the potential to identify gene expression signatures for formal assessment as disease classifiers in new samples, as outlined here.

Materials and Methods

Datasets

Two datasets of quantitative real time RT-PCR values pre-processed as mRNA relative quantities (RQ), defined as $2^{-\Delta\Delta CT}$ for 44 reporter genes were used in this study. The first dataset, was generated from PBMC samples taken from 57 CFS patients and 75 healthy volunteers and had previously been reported [8]. This was used as a training set. The second study was from a blinded study and was used as a test set. This comprised PBMC samples from 64 CFS patients and 64 healthy volunteers. This sample set was blinded and the disease classification was not assessed until the classes were predicted computationally. The data for the blinded study was collected via clinical questionnaires as outlined previously [8]. The patients were provided with paper copies for completion at home and subsequent return by post to the clinical centre. Data from the questionnaires was collated and used for clinical characterisation. The diagnosis of CFS was based on the Fukuda criteria and this diagnosis was made by CFS clinical experts. The blinded study was approved by the Wandsworth Research Ethics Committee, St George's Hospital, London, SW17 0QT. Verbal consent was given by all subjects for their information to be stored and used for this study.

Analysis

Hierarchical and k-means clustering of RQ values was performed using the TIGR MeV software suite [16]. Support vector machines (SVM) were also created using MeV, using the training set to train the SVM and the test set to evaluate the accuracy of SVM classification.

Receiver operator curve (ROC) analysis was performed using the ROCR library [17] within R/BioConductor [18]. Individual reporter genes were used as CFS predictors for the training set, the corresponding sensitivity, specificity, true and false positive rates were determined for all possible RQ cut-offs (from the smallest to largest RQ value). Areas under the curve were determined for each of the reporter genes as a measure of prediction accuracy. The same procedure was used to evaluate the gene profile score metric, outlined below.

Gene profile score metric

A gene profile score metric was generated as follows. Firstly, RQ cut-offs were determined for each reporter gene such that the associated false-positive rate of CFS prediction in the training set, using that gene alone, was 5%. These cut-offs were used to create binary profiles (44 digits in length) for each sample, with a '1' indicating that particular sample had an RQ greater than the cut-off and a 0, otherwise (either a lower RQ or a missing value).

The norm of the profile was then calculated by summing all 44 scores to give the "gene profile score". Thus, if the reporter genes are accurate predictors of CFS, samples from CFS patients would yield a higher profile score than those from healthy volunteers. The gene profile score metric itself was evaluated using ROC analysis using the training set. From this analysis. a profile score cut-off was generated at which level of the metric yielded a 5% false positive rate when classifying the training set. Profile scores were then generated for the blinded test set and samples were classified as "CFS" or "healthy" according to this cut-off.

We assessed the effect of altering the RQ cut-off score such that the false positive rate for both individual reporter genes and for the resultant gene profile score metric was 10% in terms of classifying CFS and healthy samples in the training set.

Supporting Information

Figure S1 ROC curves for CFS classification at a 5% FPR of the training set by gene profile score using ΔCT (red) and RQ (blue) values. Both yield AUCs of 0.95.
(TIFF)

Table S1 The 44 CFS reporter genes used in this study.
(DOC)

Table S2 Area under the curve (AUC) values for each of the reporter genes when used as CFS predictors on the training set.
(DOC)

Table S3 Performance of each of the reporter genes used individually as CFS predictors on the training set.
(DOC)

Table S4 Performance of each of the reporter genes used individually as CFS predictors on the test set.
(DOC)

Table S5 Relative performance of each of the scoring metrics.
(DOC)

Table S6 AUCs for the individual reporter genes used as predictors on the training and test sets, using RQ and ΔCT values.
(DOC)

Author Contributions

Conceived and designed the experiments: DF JK TJH PK. Performed the experiments: DF JK. Analyzed the data: DF. Contributed reagents/materials/analysis tools: DF JK. Wrote the paper: DF JK TJH PK.

References

1. Jason LA, Richman JA, Rademaker AW, Jordan KM, Plioplys AV, et al. (1999) A community-based study of chronic fatigue syndrome. Arch Intern Med 159(18): 2129–2137.
2. Reynolds KJ, Vernon SD, Bouchery E, Reeves WC (2004) The economic impact of chronic fatigue syndrome. Cost Effectiveness and Resource Allocation 2(1): 4.
3. Aspler AL, Bolshin C, Vernon SD, Broderick G (2008) Evidence of inflammatory immune signalling in chronic fatigue syndrome: a pilot study of gene expression in peripheral blood. Behav Brain Funct 4: 44.
4. Fang H, Xie Q, Boneva R, Fostel J, Perkins R, et al. (2006) Gene expression profile exploration of a large dataset on chronic fatigue syndrome. Pharmacogenomics 7(3): 429–440.

5. Fostel J, Boneva R, Lloyd A (2006) Exploration of the gene expression correlates of chronic unexplained fatigue using factor analysis. Pharmacogenomics 7(3): 441–454.
6. Gow JW, Hagan S, Herzyk P, Cannon C, Behan PO, et al. (2009) A gene signature for post-infectious chronic fatigue syndrome. BMC Med Genom 2: 38.
7. Kaushik N, Fear D, Richards SC, McDermott CR, Nuwaysir EF, et al. (2005) Gene expression in peripheral blood mononuclear cells from patients with chronic fatigue syndrome. J Clin Pathol 58: 826–832.
8. Kerr JR, Petty R, Burke B, Gough J, Fear D, et al. (2008) Gene expression subtypes in patients with chronic fatigue syndrome/myalgic encephalomyelitis. J Inf Dis 197: 1171–1184.
9. Presson AP, Sobel EM, Papp JC, Suarez CJ, Whistler T, et al. (2008) Integrated weighted gene co-expression network analysis with an application to chronic fatigue syndrome. BMC Sys Biol 2: 95.
10. Saiki T, Kawai T, Morita K, Ohta M, Saito T, et al. (2007) Identification of marker genes for differential diagnosis of chronic fatigue syndome. Mol Med 14(9–10): 599–607.
11. Jenner RG, Maillard K, Cattini N, Weiss RA, Boshoff C, et al. (2003) Kaposi's sarcoma-associated herpesvirus-infected primary effusion lymphoma has a plasma cell gene expression profile. Proc Natl Acad Sci U S A 100(18): 10399–10404.
12. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, et al. (2002) Gene expression profiling predicts clinical outcome of breast cancer. Nature 415: 530–536.
13. Brown MP, Grundy WN, Lin D, Cristianini N, Sugnet CW, et al. (2000) Knowledge-based analysis of microarray gene expression data by using support vector machines. Proc Natl Acad Sci U S A 97(1): 262–267.
14. Xu Y, Selaru FM, Yin J, Zou TT, Shustova V, et al. (2002) Artificial neural networks and gene filtering distinguish between global gene expression profiles of Barrett's esophagus and esophageal cancer. Cancer Res 62: 3493–3497.
15. Vinciotti V, Tucker A, Kellam P (2006) Robust selection of predictive genes via a simple classifier. Appl Bioinformatics 5(1): 1–11.
16. Saeed A, Bhagabati NK, Braisted J, Liang W, Sharov V, et al. (2006) TM4 microarray software suite. Meth Enzymol 411: 134–93.
17. Sing T, Sander O, Beerenwinkel N, Lengauer T (2005) ROCR: visualizing classifier performance in R. Bioinformatics 21(20): 3940–3941.
18. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5(10): R80.