Genetic Predisposition for Immune System, Hormone, and Metabolic Dysfunction in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: A Pilot Study

Melanie Perez 1, Rajeev Jaundoo 2,3, Kelly Hilton 1, Ana Del Alamo 1,3, Kristina Gemayel 1, Nancy G. Klimas 1,3,4, Travis J. A. Craddock 1,2,3,5, and Lubov Nathanson 1,3

1 Dr. Kiran C. Patel College of Osteopathic Medicine, Nova Southeastern University, Fort Lauderdale, FL, United States, 2 Department of Psychology and Neuroscience, Nova Southeastern University, Fort Lauderdale, FL, United States, 3 Institute for Neuro Immune Medicine, Nova Southeastern University, Fort Lauderdale, FL, United States, 4 Veterans Affairs Medical Center, Miami, FL, United States, 5 Department of Computer Science, Nova Southeastern University, Fort Lauderdale, FL, United States

Introduction: Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) is a multifactorial illness of unknown etiology with considerable social and economic impact. To investigate a putative genetic predisposition to ME/CFS we conducted genome-wide single-nucleotide polymorphism (SNP) analysis to identify possible variants.

Methods: 383 ME/CFS participants underwent DNA testing using the commercial company 23andMe. The deidentified genetic data was then filtered to include only non-synonymous and nonsense SNPs from exons and microRNAs, and SNPs close to splice sites. The frequencies of each SNP were calculated within our cohort and compared to frequencies from the Kaviar reference database. Functional annotation of pathway sets containing SNP genes with high frequency in ME/CFS was performed using over-representation analysis via ConsensusPathDB. Furthermore, these SNPs were also scored using the Combined Annotation Dependent Depletion (CADD) algorithm to gauge their deleteriousness.

Results: 5693 SNPs were found to have at least 10% frequency in at least one cohort (ME/CFS or reference) and at least two-fold absolute difference for ME/CFS. Functional analysis identified the majority of SNPs as related to immune system, hormone, metabolic, and extracellular matrix organization. CADD scoring identified 517 SNPs in these pathways that are among the 10% most deleteriousness substitutions to the human genome.

Keywords: myalgic encephalomyelitis/chronic fatigue syndrome, genome-wide, single-nucleotide polymorphism, immune system, hormone, metabolic
INTRODUCTION

Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) is a complex illness characterized by disabling fatigue, disturbed sleep patterns, pain, and flu-like symptoms. Patients report a high degree of physical disability, and a decreased quality of life with 24% being homebound (1), causing a US economic loss ranging from $9.1 to $51 billion (2). Currently, there are three main sources of diagnosis criteria, the Center for Disease Control (CDC) Empiric (3), Fukuda (4), and Canadian Consensus (5), showing 2.54, 1.0, and 0.10% of the population affected, respectively. This variation highlights the lack of a concrete illness definition. Although research studies have identified various aspects such as immune abnormalities and exposure to toxins relevant to the pathogenesis of ME/CFS (6), ME/CFS is still not yet fully understood. The genetic and environmental pathogenesis of ME/CFS remains unclear. Currently, treatment of ME/CFS is dependent on management of symptomology and improvement on quality of life (6). An improved understanding of the molecular mechanisms affected and dysfunction in the regulatory systems will translate into better diagnostic methods and more targeted approaches to treatment. There are numerous studies suggesting that genes and single nucleotide polymorphisms (SNPs) within those genes might play a role in the development and progression of ME/CFS (7–9). Results of these studies are very interesting and useful, however, one of these studies was focused on mitochondrial DNA (7), and the other two were limited by 80 study subjects (8, 9). The aim of the current study is to increase the size of the ME/CFS cohort and identify the most harmful variants associated with ME/CFS.

MATERIALS AND METHODS

Patient Population

Individuals with ME/CFS were selected through an online English pre-screening questionnaire via the RedCap platform. The study was restricted to adults (18–70 years of age) that endorsed a clinically diagnoses of chronic fatigue syndrome (CFS), post-infection fatigue (IF), or myalgic encephalomyelitis (ME) and endorsed criteria meeting the 1994 CDC definition of CFS (4); four or more of the following symptoms over a minimum of 6 consecutive months and not predating fatigue: sore throat, tender cervical or axillary lymph nodes, muscle pain, multiple joint pain without swelling or redness, headaches of new type, pattern or severity, unrefreshing sleep, post-exertional malaise, and impaired memory or concentration. Furthermore, study subjects were excluded if they had HIV infection, or dementia precluding full participation/consent. Qualified prescreened participants then completed an online consent form describing the study in detail, asking them to accept or decline the opportunity to continue with the study, via the RedCap online platform. Consenting participants then securely uploaded their genotyping data received from 23andMe into a secure database using the RedCap online platform.

Ethics Approval and Consent to Participate

All study subjects signed an informed consent approved by the Institutional Review Board (IRB) of Nova Southeastern University (NSU). Ethics review and approval for data analysis was also obtained by the IRB of NSU.

23andME Genotyping

23andME processes saliva containing DNA that was sent by the study subjects collected with the 23andMe kits according to the supplied instructions. The 23andMe CLIA-certified lab extracted DNA and processed the DNA on a genotyping chip that reads hundreds of thousands of variants in the human genome. Samples were collected starting in July 2016 until August 2018 and processed with 23andMe chip versions 4 (∼570 k SNPs; prior to August 2017) and version 5 (∼640 k SNPs; after August 2017). Genotyping calls were performed by 23andMe. Personalized reports based on well-established scientific and medical research were returned to study subjects and subsequently uploaded to the NSU RedCap online platform.

SNP Filtering and Analysis

All variants received from study participants were annotated using SeattleSeq 138 (10) for Genes, Distance-To-Nearest Splice Site, and microRNAs. Based on the annotation we focused our analysis on only non-synonymous and non-sense SNPs located in the gene's coding regions, near the splice sites and in microRNAs. The frequency of each of these SNPs was calculated in ME/CFS cohort (study participants). We compared these frequencies with the frequencies of the corresponding SNPs from the reference database Kavier [hg19 (GRCh37)] (11). Kavier contains over 162 million SNPs from 35 projects, including dbSNP, 1000Genomes and other and does not include the data from cancer genomes.

For functional analysis we selected SNPs that satisfied following criteria: the frequency at least 10% in either reference or ME/CFS cohort and the ratio in frequencies between the ME/CFS cohort and the reference cohort is more than two in either direction (Supplementary Table 1).

All variants that prevail in ME/CFS cohort were also scored using the Combined Annotation Dependent Depletion algorithm (CADD) (12) (Supplementary Table 2).

Functional Annotation

Functional annotation of SNPs was performed using the ConsensusPathDB (13–15) to provide biological pathway information. Over-representation analysis (13) incorporating the Kyoto Encyclopedia of Genes and Genomes (KEGG) (73.0) (16), Netpath (1.1.2015) (17), the Integrating Network Objects with Hierarchies (INOH) (1.1.2015) (18), Biocarta (2009.05.12) (19), Humanycyc (18.5) (20), Signalink (8.1.2015) (21), Edinburgh human metabolic network (Ehmn) (1.1.2015) (22), Reactome (51) (23), Wikipathways (9.1.2015) (24) and the Pathway Interaction Database (PID) (2014.02.14) (25) pathway sets was used to interpret the functions the identified SNPs may play. Here the significance of the observed overlap between the gene module and the members of known pathways, compared to random expectations, was calculated based on a
RESULTS AND DISCUSSION

Functional analysis of SNPs identified three main clusters of pathways as sharing at least 30% SNP related genes (Figure 1). The first is dominated in size via the pathway Cytokine Signaling in Immune System and includes other immune-related pathways such as interferon signaling, autoimmune responses, and T-cell receptor signaling. This cluster highlights a module of immune-related SNPs.

The second cluster is dominated in size via the Nuclear-Receptors Meta-Pathway and includes hormone related pathways such as steroid hormone, estrogen, and androgen biosynthesis, glucuronidation, and the pregnant x receptor pathway. This cluster highlights modules of hormone-related SNPs.

The final cluster is dominated in size by Pathways in Cancer, however, closer inspection shows many metabolic processes such as enzyme reactions (protein kinase A, calcium and calmodulin signaling), and G proteins signaling which regulate metabolic enzymes, which are all involved in the regulation of glycogen, sugar and lipid metabolism. This cluster highlights a module of metabolism-related SNPs.

While there is an overlap between the metabolic and immune modules, the hormone module remains isolated with main connections only formed via Ovarian Steroidogenesis and the Wnt signaling pathway. Finally, there is a group of loosely connected pathways involved in an extracellular matrix organization.

While this organization highlights the interplay between immune, hormone and metabolic activity underlying ME/CFS, overlay of the location of CADD scores illustrates where the most deleterious effects occur (Figure 1; lower panel).

Of the 11,485 SNPs that passed prefiltering according to the annotations (see Methods), 8,593 SNPs had frequency more than 10% in either reference or ME/CFS cohort. Out of them, 5,693 SNPs had a two-fold difference between ME/CFS and the reference cohorts in either direction (Supplementary Table 1).

SNPs that prevailed in ME/CFS cohort were scored using the CADD algorithm (12). According to the CADD algorithm, C-scores above 10 indicate that these SNPs are predicted to be among the 10% most harmful, and C-scores above 20 indicate the 1% most deleterious substitutions (12). Table 1 shows 50 SNPs that are the most frequent in the ME/CFS cohort and have C-scores above 10.

Of the 50 most frequent deleterious SNPs found in our ME/CFS cohort compared to the reference database (Table 1), 10 were found to have a frequency of 70% or more in the ME/CFS group. This includes CYP2D6, PRRT4, and PRSS56 at a frequency over 90%, C14orf37, ANKDD1B, at over 80%, and GPBAR1, LHB, ADAMTS19, VARS2, and CPLX2 at over 70%.

CYP2D6 (Cytochrome P450 2D6) is primarily expressed in the liver, but also highly expressed in areas of the central nervous system, including the substantia nigra, and is one of the most important enzymes involved in the metabolism of xenobiotics in the body. A significantly higher frequency of polymorphisms CYP2D6 was found in ME/CFS study subjects with Fibromyalgia than in controls and could differentiate these study subjects from study subjects with multiple chemical sensitivity (27). CYP2D6 was found in the xenobiotics metabolism, androgen and estrogen biosynthesis and metabolism, tyrosine metabolism, codeine and morphine metabolism, oxidation by cytochrome P450, metapathyway biotransformation phase I and II, and cytochrome P450—arranged by substrate type pathways all of which belong to the hormone related cluster.

PRSS56 (putative serine protease 56) is a serine protease that has been implicated in human eye development (28) and in the regulation of cerebellum activity of mice in exercise (29). It was not found to be a member of any of the annotated pathways.

GPBAR1 (G Protein-Coupled Bile Acid Receptor) functions as a cell surface receptor for bile acids and participates in the production of intracellular cAMP and activation of a MAP kinase signaling pathway. This receptor plays a big role in the suppression of macrophage functions and regulation of energy homeostasis by bile acids (30). Finding of the deleterious SNP in GPBAR1 (Table 1) is in agreement with the results of the recent study that showed disturbances in bile acid metabolism in ME/CFS study subjects (31). GPBAR1 was not among any of the pathways annotated.

LHB (luteinizing hormone beta polypeptide) is expressed in the pituitary gland and is essential for spermatogenesis and ovulation by stimulating the testes and ovaries to synthesize steroids (32, 33). LHB was found among the GnRH signaling pathway and ovarian steroidogenesis pathway.

ADAMTS19 is a member of the large ADAMTS (a desintegrin-like and metalloprotease with thrombospondin type 1 motif) family of metalloproteases (metal binding enzymes). ADAM proteins are responsible for the proteolytic cleavage of many transmembrane proteins and the release of their extracellular domain. ADAMTS19 is considered as a possible candidate for premature ovarian failure (34). Only the O-linked glycosylation pathway was found to contain ADAMTS19.

VARS2 (valyl-tRNA synthetase 2, mitochondrial) is important for the mitochondrial protein synthesis. Mutations in this gene are associated with cardiomyopathy (35), microcephaly and epilepsy (36), deficiency of the mitochondrial respiratory chain complex I and oxidative phosphorylation deficiency (37). VARS2 was not found among any of the annotated pathways.
FIGURE 1 | Pathway overlap networks. Pathways identified via over-representation analysis (nodes) connected by shared genes with SNPs (edges). Upper panel gives total number of SNPs per pathway and likelihood of annotation (p-value). Lower panel shows pathways most affected by deleterious SNPs.
CPLX2 gene encodes the complexin 2 protein that participates in neurotransmitter release by directly interacting with the neuronal SNARE complex (38). CPLX2 is known to be overexpressed in aging and downregulated by sleep deprivation (39), and this shows a connection of CPLX2 expression to fatigue. CPLX2 was also not found among any of the annotated pathways.

The remaining genes PRRT4, C14orf37, and ANKDD1B are obscure without much literature to support their function and not found among any of the annotated pathways. It was determined that PRRT4 (proline-rich transmembrane protein 4) showed biased expression in adult ovary, lung, adrenals, CNS and whole brain, while C14orf37 showed bias in brain, kidney, and ovary (ncbi.nlm.nih.gov). Little information was found for ANKDD1B (ankyrin repeat and death domain containing 1B).

Although SNPs in MYBPC3 and HLA genes have lower frequencies in ME/CFS cohort (0.19 for MYBPC3 and 0.13-0.44 for various HLA isoforms, respectively), these SNPs could be used for subgrouping of ME/CFS study subjects in larger studies because of their possible association with ME/CFS and fatigue. Multiple deleterious SNPs in HLA genes are in agreement with known impairment of the immune system in ME/CFS (40). Increased frequency of HLA-DQA1 alleles and decreased expression of HLA-DRB1 was found to be associated with HLA (41). MYBPC3 (myosin binding protein C, cardiac) dysfunction is also associated with hypertrophic cardiomyopathy and corresponding fatigue (42).

These results contrast with previous SNP studies in ME/CFS (9, 43–45) which have found statistically significant associations in multiple loci including in neuroendocrine effector and receptor genes (43), TRP ion channels and AChRs (44, 45), and genes regulating the HPA axis (9, 46). This difference is most likely due to a combination of factors such as, (i) differences in array types used between studies, (ii) difference in the methods of analysis, (iii) differences between cohorts and the general heterogeneity of ME/CFS, and (iv) small-effect variants due to the relatively small sample sizes in each of these previous studies, compared to our relatively large cohort.

To date, this is the largest study known using SNP data and its affected pathways in combination with study subjects’ self-reported symptoms. The results generated from our study will enhance the current understanding of ME/CFS and will generate new studies, all of which will lead to a better method for diagnosis and targeted genetic therapy. Replicative larger studies are warranted to improve the reliability of the results.

While these results are novel there are some limitations to the current analysis that are worth noting. First, there is no control over the chip version used by 23andMe for genotyping. This can result in loss of precision in the determination of study subject genotyping signature. Second, this initial pilot analysis was only conducted on SNPs in protein coding regions, miRNA regions, and regions close to splice junctions. SNPs in non-coding regions may be important in the cause of the illness. Finally, this analysis does not include rare variants. Moving forward, future studies based on this on-going collection of study subject information will address these limitations, will increase sample size, and provide more detailed statistical analyses. Building on
this dataset we also aim to correlate these findings with our ongoing research on gene expression (47), miRNA expression and DNA methylation (48).

ETHICS STATEMENT

All subjects signed an informed consent approved by the Institutional Review Board of Nova Southeastern University. Ethics review and approval for data analysis was also obtained by the IRB of Nova Southeastern University.

AUTHOR CONTRIBUTIONS

NK, LN, and TC conceived the study. MP, RJ, TC, and LN analyzed data. NK consulted on ME/CFS symptoms. AD, KH, KG, and MP led the recruitment of study participants.

REFERENCES

1. Pendergrast T, Brown A, Sunquist M, Jantke R, Newton JL, Strand EB, et al. Housebound versus nonhousebound patients with myalgic encephalomyelitis and chronic fatigue syndrome. Chronic Illn. (2016) 12:292–307. doi: 10.1177/1742393516644770
2. Lin J-MS, Resch SC, Brimmer DJ, Johnson A, Kennedy S, Burstein N, et al. The economic impact of chronic fatigue syndrome in Georgia: direct and indirect costs. Cost Eff Resour Alloc. (2011) 9:1. doi: 10.1186/1748-7549-9-1
3. Bruberg KG, Fenhus MS, Larun L, Flottorp S, Malterud K. Case definitions for chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME): a systematic review. BMJ Open. (2014) 4:e003973. doi: 10.1136/bmjopen-2013-003973
4. Fukuda K, Straus SE, Hickie I, Sharpe MC, Dobbins JG, Komaroff A. The chronic fatigue syndrome: a comprehensive approach to its definition and study. Ann Intern Med. (1994) 121:953–9. doi: 10.7326/0003-4819-121-12-199412150-00009
5. Reeves WC, Jones JF, Maloney E, Heim C, Hoaglin DC, Boneva RS, et al. Prevalence of chronic fatigue syndrome in metropolitan, urban, and rural Georgia. Popul Health Metr. (2007) 5:5. doi: 10.1186/1478-7954-5-5
6. Bested AC, Marshall LM. Review of myalgic encephalomyelitis/chronic fatigue syndrome: an evidence-based approach to diagnosis and management by clinicians. Rev Environ Health. (2015) 30:223–49. doi: 10.1515/reveh-2015-0026
7. Billing-Ross P, Keinan A, Gu Z, Hanson MR. Mitochondrial DNA variants correlate with symptoms in myalgic encephalomyelitis/chronic fatigue syndrome. J Transl Med. (2016) 14:19. doi: 10.1186/s12976-016-0771-6
8. Smith AK, Fang H, Whistler T, Unger ER, Rajeevan MS. Convergent genomic studies identify association of GRIK2 and Npas2 with chronic fatigue syndrome. Neuropsychobiology. (2011) 64:183–94. doi: 10.1111/j.1497-2563.2010.00692
9. Schlauch KA, Khaitoullina SE, De Meirleir KL, Rawat S, Peteret J, Rizvano VA, et al. Genome-wide association analysis identifies genetic variations in subjects with myalgic encephalomyelitis/chronic fatigue syndrome. Transl Psychiatry. (2016) 6:e730. doi: 10.1038/tpp.2015.208
10. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature. (2009) 461:272–6. doi: 10.1038/nature08250
11. Glusman G, Caballero J, Mauldin DE, Hood L, Roach JC, KaviarH: an accessible system for testing SNV novelty. Bioinformatics. (2011) 27:3216–7. doi: 10.1093/bioinformatics/btr540
12. Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. (2014) 46:310–5. doi: 10.1038/ng.2892
13. Kamburov A, Wierling C, Lehrach H, Herwig R. ConsensusPathDB—a database for integrating human functional interaction networks. Nucleic Acids Res. (2009) 37:D623–8. doi: 10.1093/nar/gkn698
14. Kamburov A, Pentechev K, Galicka H, Wierling C, Lehrach H, Herwig R. ConsensusPathDB: toward a more complete picture of cell biology. Nucleic Acids Res. (2011) 39:D712–7. doi: 10.1093/nar/gkr1156
15. Kamburov A, Stelzl U, Lehrach H, Herwig R. The ConsensusPathDB interaction database: 2013 update. Nucleic Acids Res. (2013) 41:D793–800. doi: 10.1093/nar/gks1055
16. Kanehisa M, Goto S, Yamanishi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. (2011) 40:A109–14. doi: 10.1093/nar/gkr988
17. Kandasamy K, Mohan SS, Raju R, Keerthikumar S, Kumar GSS, Venugopal AK, et al. NetPath: a public resource of curated signal transduction pathways. Genome Biol. (2010) 11:R3. doi: 10.1186/gb-2010-11-1-r3
18. Yamamoto S, Sakai N, Nakamura H, Fukagawa H, Fukuda K, Takagi T. INOH: ontology-based highly structured database of signal transduction pathways. Database. (2011) 2011:bar052. doi: 10.1093/database/bar052
19. Nishimura D. BioCarta. BioTech Softw Internet Rep. (2001) 2:117–20. doi: 10.1027/152971601750294344
20. Romero P, Wagg J, Green ML, Kaiser D, Kruppenacker M, Karp PD. Computational prediction of human metabolic pathways from the complete human genome. Genome Biol. (2004) 6:R2. doi: 10.1186/gb-2004-6-1-r2
21. Fazekas D, Koltai M, Türei D, Módos D, Pálfy M, Dül Z, et al. SignaLink 2–a signaling pathway resource with multi-layered regulatory networks. BMC Syst Biol. (2013) 7:7. doi: 10.1186/1752-0509-7-7
22. Hao T, Ma H-W, Zhao X-M, Gorynian I. Compartmentalization of the Edinburg human metabolic network. BMC Bioinformatics. (2010) 11:393. doi: 10.1186/1471-2105-11-393
23. Croft D, Mundo AF, Flaw R, Milacic M, Weiser J, Wu G, et al. The reactome pathway knowledgebase. Nucleic Acids Res. (2014) 42:D472–7. doi: 10.1093/nar/gkt1102
24. Kelder T, van Iersel MP, Hanspers K, Kutmon M, Conklin BR, Evelo CT, et al. WikiPathways: building research communities on biological pathways. Nucleic Acids Res. (2012) 40:D1301–7. doi: 10.1093/nar/gkr1074
25. Schaefer CF, Anthony K, Krupa S, Buchhoff J, Day M, Hannay T, et al. PID: the pathway interaction database. Nucleic Acids Res. (2009) 37:D674–9. doi: 10.1093/nar/gkn653
26. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. (2003) 13:2498–504. doi: 10.1101/gr.1239303
27. Caccamo D, Cesareo E, Mariani S, Raskovic D, Ientile R, Curro C, et al. WikiPathways: building research communities on biological pathways. Oxid Med Cell Longev. (2013) 2013:831969. doi: 10.1155/2013/831969
28. Novilaty SR, Khan AO, Aldmahesh MA, Tabbara KM, Al-Amri A, Alkuraray F. Biometric and molecular characterization of clinically diagnosed

ACKNOWLEDGMENTS

We acknowledge the student of Dr. Kiran C. Patel College of Osteopathic Medicine of Nova Southeastern University Christopher Larrimore for his help in managing RedCap database. We would like to thank students of Halmos College of Natural Sciences and Oceanography of Nova Southeastern University Valentina Ramirez, Maria Cash and Pallavi Samudrala for their help with the analysis of data. This study was partially supported by the Health Professions Division grant from Nova Southeastern University awarded to KH.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fped.2019.00206/full#supplementary-material
47. Jeffrey MG, Nathanson L, Aenlle K, Barnes ZM, Baig M, Broderick G, et al. Treatment avenues in myalgic encephalomyelitis/chronic fatigue syndrome: a split­­gender pharmacogenomic study of gene­expression modules. Clin Ther. (2019). doi: 10.1016/j.clinthera.2019.01.011. [Epub ahead of print].

48. Trivedi MS, Oltra E, Sarria L, Rose N, Beljanski V, Fletcher MA, et al. Identification of myalgic encephalomyelitis/chronic fatigue syndrome­associated DNA methylation patterns. PLoS ONE. (2018) 13:e0201066. doi: 10.1371/journal.pone.0201066

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Perez, Jaundoo, Hilton, Del Alamo, Gemayel, Klimas, Craddock and Nathanson. This is an open­access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.