Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) and mitochondria: A review

A Thesis

Presented to the Faculty of the Graduate School

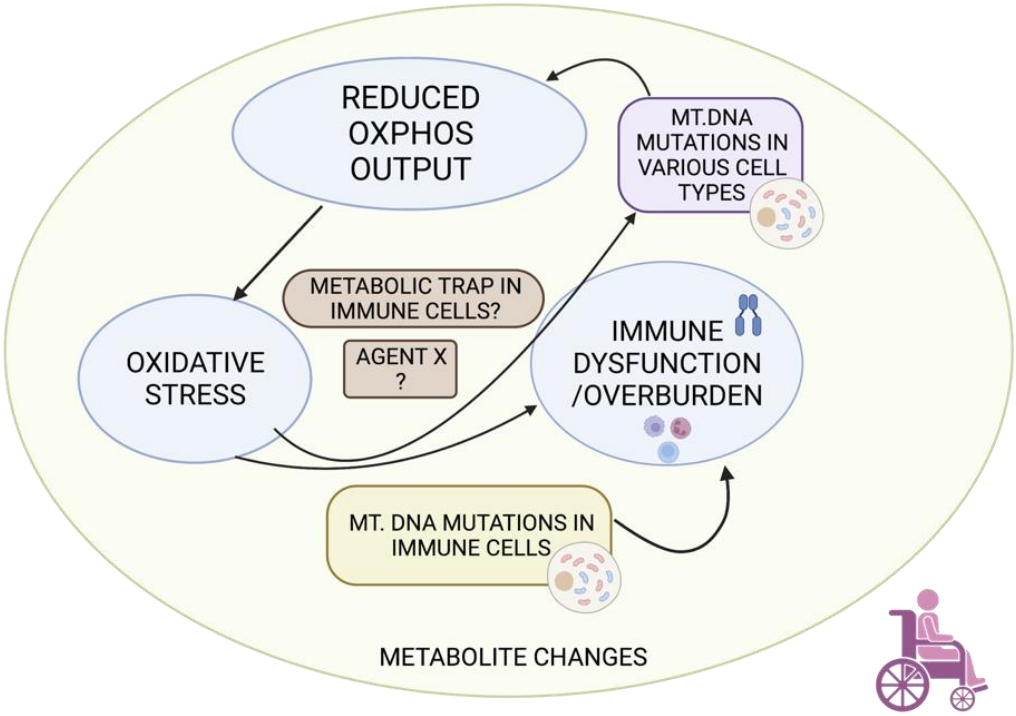
of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

by
Aakarsha Ajit Pandey
May 2022

GRAPHICAL ABSTRACT



Schematic representing a speculation of the physiological processes that might be occurring during ME/CFS and how one process is related to the others.

ABSTRACT

Myalgic Encephalomyelitis/ Chronic Fatigue Syndrome (ME/CFS) is a debilitating disorder that affects about 60 million people worldwide and manifests commonly in the aftermath of a viral infection. Symptoms affect immune function, sleep patterns, and cognition and leave patients severely fatigued after normal or less than normal exertion. Mitochondria are responsible for energy metabolism, cell signaling, and oxidative stress pathways in majority of the tissues in the body. Presented here is a review of studies investigating the role of mitochondrial DNA mutations and oxidative phosphorylation output in immune cells and skeletal muscle cells. Also included are studies that are broad metabolomic investigations of blood plasma of ME/CFS patients, and studies that measure markers of oxidative stress. Immune dysfunction emerges to be playing a key role in ME/CFS pathophysiology as well as oxidative stress. Ultimately the interdependence of these processes with the mitochondria at the center starts to paint a clearer picture of the mechanisms at play in this disease.

Keywords: ME/CFS, OXPHOS, oxidative stress, mtDNA, PBMCs, free radicals

BIOGRAPHICAL SKETCH

Aakarsha Ajit Pandey was born in Riyadh, Saudi Arabia to Indian parents; her father worked as an architect and planner and her mother served as an anesthesiologist at a maternity children's hospital. Her family moved to Gujarat, India in 1996 where she completed primary, secondary, and high school education. Her parents encouraged all round development of their two daughters, Aakarsha and her older sister. Aakarsha was awarded for her academic performance throughout school and participated in extracurricular activities like dance, public speaking, and sports. She developed a keen interest in Biology since a young age and her anesthesiologist mother fueled this passion. Exposure to the world of recombinant DNA technology and biotech research in high school spurred her in the direction of studying molecular biology. She decided to pursue an undergraduate degree majoring in pharmaceutical sciences with the subsequent goal of pursuing post graduate education in molecular and cell biology.

She attended New York University (NYU) as a master's candidate pursuing research in heart and pharyngeal muscle development using *Ciona*– a chordate invertebrate and highly tractable laboratory organism. The close evolutionary relationship between *Ciona* and vertebrates makes them an excellent model organism and Aakarsha was assigned two independent projects that involved implementation of CRISPR/Cas9 in *Ciona*. Thanks to this experience she was hired by the Di Gregorio lab at NYU as a research technician to implement CRISPR/Cas9 gene editing with a goal of targeting notochord specific genes in *Ciona*. Her work including the knockout of the *Leprecan* gene homolog in *Ciona* and characterization of the resulting transgenic embryos was published in a Springer book chapter. Getting into Cornell University's Biochemistry, Molecular and Cell Biology program as a PhD student was a dream come true and she learned a lot from her rotations.

Ultimately, she joined the Gu lab and discovered that she was interested in contributing to the field of ME/CFS research.

Throughout her career she constantly explored her passion for teaching. After completing her undergraduate education, she worked as a Science Communicator on-board the "Science Express"- a Government of India-sponsored exhibition on rails. Here she worked with a team of science communicators to develop interactive tools and techniques that would not only help explain the science exhibits, but also help develop scientific temper among the students who visited the exhibition. The science exhibits focused on the "Biodiversity of India" and the exhibition train traveled to more than sixty Indian cities over a period of seven months. This experience was extremely unique and left a deep impression on Aakarsha with regards to the importance of outreach, science education and environment conservation.

She also taught undergraduate students as a Teaching Assistant (TA) during her time at NYU and Cornell and was awarded the Outstanding Teaching Assistantship award by the Office of Undergraduate Biology, Cornell University. With the goal of giving back and using her voice to help share the magic of biology she participated in outreach programs like Center for talented Youth (CTY), GRASSHOPPR and STEP-UP. She hopes that she continues to make a difference using her gifts and training and that this work contributes a piece to the puzzle of ME/CFS.

To Allison Williams whose story of suffering with ME/CFS as a part of David S. Bell's book impacted me deeply.

To Dr. David S. Bell for writing an engrossing and detailed book about ME/CFS more than three decades ago that is still relevant.

To all those who silently battle an uncharacterized illness and the ignorance because of it.

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Zhenglong Gu for providing me with the opportunity to work in the field of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS). I would like to thank Dr. Xiaoxian Guo for serving as a mentor when I had just joined the lab. I would also like to thank Weiliin Xu in helping me interpret and understand computational analysis aspects of my project and other ongoing projects in the lab. I would like to thank Inna Sirota for always listening and helping me with my concerns and providing feedback on my presentations and writing. Thanks to you, Inna the last few months have been enjoyable. I would like to acknowledge Dr. Chris Fromme who served as my field DGS and supported me throughout the phase of thesis completion, defense scheduling and communicating with committee members. Chris is the most approachable and helpful DGS and is a true student advocate.

I would like to thank my committee members Drs Thomas Fox and Maureen Hanson for taking the time to provide their valuable feedback on my thesis and defense presentation. I would like to thank my friends and peers especially Bhargav Sanketi, Katie Gordon and Saket Bagde for their advice, help and company whenever I needed it. I am grateful to my family and friends including Krunal my husband for his constant effort towards providing the best possible environment to help me succeed. I am grateful to my sister Manasi Shah and friend Sakshi Bhatia for constantly pushing me to complete my tasks and achieve targets. I would like to extend a heartfelt thanks to Neha, Kushal and Aahana Budhwar for providing a comfortable oasis for the last leg of my thesis completion; I couldn't have done this without you. I am grateful to my parents Manisha and Ajit Pandey who inspire me every day by the way they lead their life.

I would like to thank Dr. Darlene Campbell for giving me the opportunity to serve as a T.A. for BIOG1445 and effectively giving me my most enriching academic experience at Cornell. Being a

T.A. for this course breathed life into my daily routine and motivated me to keep giving my best to my research as well. I would also like to thank all the T.A.s who taught the course with me, the students I interacted with and everyone responsible for awarding me the Outstanding Teaching Assistant award. I would like to use this opportunity to thank all the professors with whom I rotated including Drs Natasza Kurpios, Kelly Liu, Gunther Hollopeter and Haiyuan Yu. I learned a lot of important lessons from these rotations and hope to implement these in the future. I would l like to acknowledge my extra-curricular activity partners who helped me enjoy my time at Cornell including members of the Salsa Palante group: Kunal Pattanayak, Michael Ristorucci, Ruju Dani; GRASSHOPPR and STEP-UP program partners: Mariela Nunez Santos and Jumana Badar; Gardening workshop organizer: Sarah Post. I would like to thank Manju Sinha, Ashok Sinha, Suman Singh and Aaditya Singh for supporting me at the very start of this journey. I would like to thank Shiva and Madhav my nephews for always believing in me and finally I would like to acknowledge the Universal Consciousness that has guided me up until this point.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviation	Explanation
ACTH	Adrenocorticotropic Hormone
ADP	Adenosine diphosphate
ANAPC11	Gene that encodes Anaphase-promoting complex subunit 11
ATP	Adenosine triphosphate
BCAA	Branched chain amino acids
BRMS1	Gene that encodes Breast cancer metastasis-suppressor 1 protein
CCC	Canadian Consensus Criteria
CD2BP2	Gene that encodes CD2 antigen cytoplasmic tail-binding protein 2
CD4	Cluster of differentiation 4
CD8	Cluster of differentiation 8
CDC	Center for Disease Control

CE-TOF/MS	Capillary Electrophoresis- Time of Flight/ Mass Spectrometer
CFI	Chronic Fatigue Initiative
CPET	Cardiopulmonary Exercise Testing
CSF	Cerebrospinal fluid
DNA	Deoxyribonucleic acid
EBV	Epstein Bar Virus
EDA	Exploratory Data Analysis
EDTA	Ethylenediamine Tetra Acetic acid
EIF2B4	Gene that encodes Translation initiation factor eIF-2B subunit delta
EIF4G1	Gene that encodes Eukaryotic translation initiation factor 4 gamma 1 protein
EL	Extension- Ligation
ETC	Electron Transport Chain
FABP4	Gene that encodes fatty acid binding protein
FC	Fukuda Criteria
FGF21	Gene that encodes Fibroblast growth factor 21

FL	Follicular lymphoma
GABARAPL1	Gene that encodes Gamma-aminobutyric acid receptor-associated protein-like 1
GSH	Glutathione
GSN	Gene that encodes Gelsolin protein
HEK293	Human Embryonic Kidney 293 cells
HEK293T	HEK293 variant that expresses a temperature-sensitive allele of the SV40 T antigen
HeLa	Henrietta Lacks immortal embryonic cell line
HPLC	High Performance Liquid Chromatography
IBS	Inflammatory Bowel Disease
ICC	International Consensus Criteria
IDO	Indoleamine 2,3-Ddioxygenase
IGHV3-23	Gene that encodes Immunoglobulin heavy variable 3-23
IL-10RA	Gene that encodes Interleukin-10 receptor subunit alpha

IL8	Gene encoding Interleukin 8
IOMC	Institute of Medicine Criteria
KB	Kilo base
KHSRP	Gene that encodes KH-Type Splicing Regulatory Protein
MAIL1	Gene that encodes Main Like -1 protein
MDD	Major Depressive Disorder
ME/CFS	Myalgic Encephalomyelitis/Chronic Fatigue Syndrome
MELAS	Mitochondrial Encephalo-myopathy, Lactic Acidosis, and Stroke-like episodes
MPSS	Massively parallel signature sequencing
MRPL23	Gene that encodes 39S ribosomal protein L23, mitochondrial
MS	Mass Spectrometry
mtDNA	Mitochondrial DNA
NAD	Nicotinamide Adenine Dinucleotide
NF-κB	Gene encoding Nuclear Factor Kappa light chain enhancer of activated B cells

NHL	Non-Hodgkin lymphoma
NK cells	Natural Killer cells
NKCC	Natural Killer Cellular Cytotoxicity
NMR	Nuclear Magnetic Resonance
NTE	Gene that encodes Neuropathy target esterase protein
OCR	Oxygen Consumption Rate
OXPHOS	Oxidative Phosphorylation
PBMCs	Peripheral Blood Mononuclear Cells
PCR	Polymerase Chain Reaction
PCr	Phosphocreatine
PDCD2	Gene that encodes Programmed cell death protein 2
PEM	Post Exertional Malaise
PEX16	Gene that encodes Peroxisomal membrane protein
POLR2G	Gene that encodes DNA-directed RNA polymerase II subunit RPB7 protein

PRKCL1	Gene that encodes Protein Kinase N1 protein
PSNS	Parasympathetic nervous system
qPCR	Quantitative PCR
RNA	Ribonucleic acid
ROC	Receiver Operating Curve
ROS/RONS	Reactive Oxygen Species/Reactive Oxygen or Nitrogen Species
rRNA	Ribosomal RNA
RT- qPCR	Real Time Quantitative PCR
SATET	Sub Arachnoid Threshold Exercise Test
SNPs	Single Nucleotide Polymorphisms
STAMP	Sequencing by Targeted Amplification of Multiplex Probes
TCA	Tricarboxylic acid
TFAM	Mitochondrial Transcription Factor A
TNF ALPHA	Gene encoding Tumor Necrosis Factor alpha
TOF MS	Time of Flight Mass Spectrometry
TOR	Target of Rapamycin
tRNA	Transfer RNA
Tsp	Thermostable polymerase

INTRODUCTION

Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) has affected about 2.5 million Americans and about 60 million people worldwide (O'Neal & Hanson, 2021). The most common symptoms include brain fog, severe headache, flu symptoms, musculoskeletal pain, orthostatic intolerance, and post-exertional malaise (PEM) (Germain et al., 2021). These symptoms allude to neurological, immunological, and metabolic abnormalities that lead the patients to be subjected to a slew of tests that result in mild irregularities that are insufficient for a definite diagnosis (Bell, 1994a; Morris & Maes, 2013). This inflicts a considerable financial and psychological burden on the patients and their families (Bell, 1994a; Morris & Maes, 2013).

ME/CFS is an endemic disease and occurs as sporadic cases or cluster outbreaks (Underhill, 2015). There is a debate about whether the disease dates to a hundred years or is new, and it may, in fact, be an old disease that remained uninvestigated due to more concerning diseases like smallpox and plague (Bell, 1994a). Over the last few decades, numerous outbreaks bearing a resemblance to ME/CFS have been seen in the United States (L.A.-1934; New York State-1950; Washington D.C.-1953; Lake Tahoe-1985; Lyndonville, NY-1985), Australia (Adelaide-1945), the UK (Royal Free Hospital, London-1955; Great Ormond Street, London-1974; Ayrshire, Scotland-1980) and Iceland (Akureyri, Iceland-1948) and many names of the "unknown" disease emerged like "Iceland disease," "Akureyri Disease" and "Royal free disease" (Bell, 1994b). It was called the Epstein Barr Virus (EBV) disease after the Lake Tahoe outbreak in 1985 and was also known as abortive poliomyelitis due to symptom similarity to poliomyelitis (Bell, 1994b). After two massive epidemics in 1988, the CDC set up a committee to arrive at a consensus about the symptoms of the disease-causing the epidemics, and it was named chronic fatigue syndrome. It was recognized as a distinct illness (Bell, 1994b). It continued to be known as myalgic encephalomyelitis in other parts of the world, but the

name Epstein Bar virus Disease and others were completely ruled out. Due to this history, the disease is also called "the disease of a thousand names" (Bell, 1994b).

According to the Center for Disease Control (CDC), there are three main degrees of severity of ME/CFS- mild, moderate, and severe. The severe cases render the patients completely bedridden and, in many cases, suffer extreme sensitivity to light and sound (Chang et al., 2021). The estimated population suffering from ME/CFS in the United States is between 1.7 and 3.38 million, and the worldwide prevalence may be as much as 65 million (Hanson & Germain, 2020; O'Neal & Hanson, 2021; Valdez et al., 2019). About 25 percent of ME/CFS patients are severely affected, indicating that about 850000 Americans live with severely impaired mental and physical agility (Chang et al., 2021).

Over the years, many criteria have been developed to diagnose ME/CFS as a distinct illness, including the 1994 Fukuda Criteria (FC), the 2003 Canadian Consensus Criteria (CCC), the 2011 International Consensus Criteria (ICC), and the 2015 Institute of Medicine Criteria (IOMC) (Holden et al., 2020). All these criteria use symptom clusters to diagnose the disease, and the required or primary symptoms of each standard have been tabulated in Table 1. Post-exertional malaise, unrefreshing sleep, and impairment of physical activity are common among CCC, FC, and IOMC, and post-exertional malaise is shared among all four. Exertion, or the act of spending energy, routinely involves metabolism in muscle cells, brain cells, and immune cells, and mitochondria are the centers of metabolism in all these cells.

The mitochondria, famously referred to as the cell's powerhouse, are involved in cell signaling, immune cell activation, and apoptosis (West et al., 2011). Mitochondria possess their genetic material, which is encompassed within a 16.6 kb circular genome containing 22 tRNA and two rRNA genes used for mitochondrial protein synthesis, as well as genes that encode 13 evolutionarily conserved protein complexes that form a part of the electron transport chain (ETC)

(Figure 1) (Oliveira, 2018). Each cell possesses more than one mitochondrion ranging from a few tens in immune cells to thousands in liver cells. In general, the mitochondrial genome is not identical in all the mitochondria of a cell, generating a population of varied mitochondrial genomes in a single cell- a state called heteroplasmy (N. Lane, 2005; Oliveira, 2018). Homoplasmy means the presence of genetically identical mitochondria and is seen in germ cells (N. Lane, 2005; Oliveira, 2018). As mitochondria are the organelles where oxidative phosphorylation (OXPHOS) occurs, the environment within the mitochondria is highly oxidative. Even though TFAM is known to play a role in coating and to protect mtDNA from damage, the proximity of the electron transport chain predisposes mtDNA to readily acquire mutations in the coding regions (N. Lane, 2005; Starkov, 2008; Wang et al., 2013). This can lead to improper formation of respiratory chain complexes, leading to electron leakage. The free electrons from free radicals can damage mtDNA or proteins, and these free radicals need to be quenched by antioxidants (See section "Oxidative Stress Pathways and ME/CFS") (N. Lane, 2005; Russell et al., 2020). Free radicals play an essential role in signaling and help mitochondria correct their genetic mutations. They can also help eliminate mitochondria whose genomes have many mild mutations and pathogenic mutations (N. Lane, 2005). Oxidative stress occurs when the free radicals are very high in number and is a phenomenon that occurs during aging and many other diseases (Starkov, 2008).

ME/CFS is an illness involving multiple organ systems, leading to neurological symptoms, exhaustion, and "flu-like symptoms" due to immune activation (Morris & Maes, 2014). Since the mitochondria are present in most body tissues and are involved in pathways like immune signaling, oxidative stress, and OXPHOS, many studies have been conducted studying the role of mitochondria in ME/CFS patients vis a vis these many conduits (Holden et al., 2020; Morris & Maes, 2014). This review summarizes and discusses research about aspects of mitochondrial DNA mutations, oxidative stress, OXPHOS output, immune function, metabolites, and viral infection and

their role in ME/CFS. Many of these aspects feed into each other, and their interrelatedness is also discussed.

Table 1. The four criteria used for diagnosis of ME/CFS include only the required "primary" symptoms(Carruthers et al., 2011; Fukuda et al., 1994; Holden et al., 2020).

FUKUDA	CCC	ICC	IOM
Clinically evaluated, unexplained, persistent, or relapsing chronic fatigue that is of new or definite onset (has not been lifelong); is not the result of ongoing exertion; is not substantially alleviated by rest; and results in substantial reduction in previous levels of occupational, educational, social, or personal activities.	Fatigue	Post exertional neuroimmune exhaustion	A substantial reduction in the ability to engage in pre-illness levels of activity.
Four or more of the following symptoms: substantial impairment in short term memory or concentration; sore throat; tender lymph nodes; muscle pain; multi-joint pain without swelling or redness; headaches of a new type, pattern, or severity; unrefreshing sleep; and post-exertional malaise lasting mor than 24 hours	Post-exertional malaise	At least one of the following: gastrointestinal/genitourinary/immune impairment	Post-exertional malaise
	Sleep dysfunction	At least one neurological impairment	Unrefreshing sleep
	Pain	At least one ion transport/energy metabolism impairment	Cognitive impairment or orthostatic intolerance

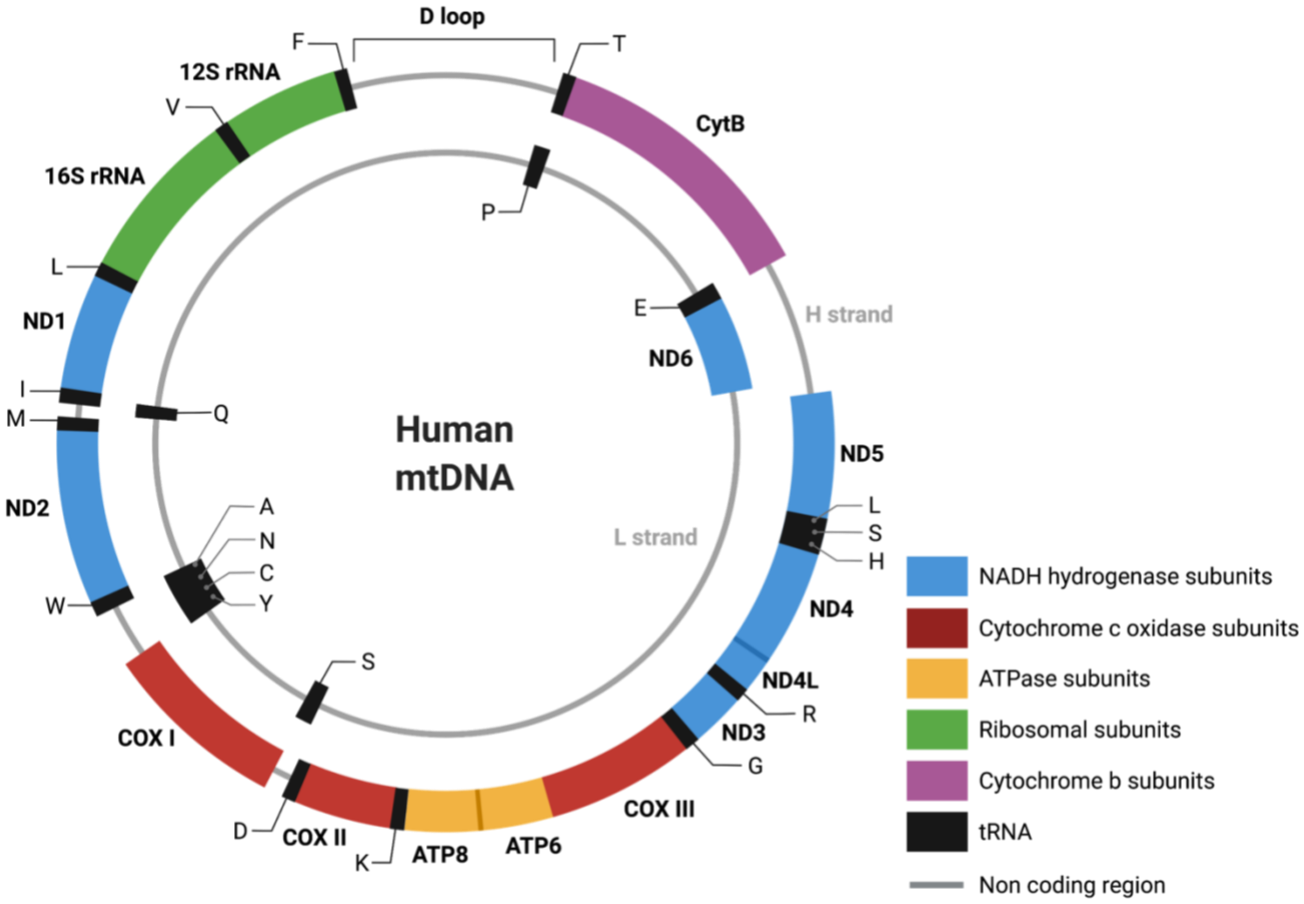


Figure 1. A Schematic representing the 16,569 bp human mtDNA. The mitochondrial genome is a double-stranded circular DNA molecule. 13 protein-encoding genes are shown as different color blocks (blue, maroon, yellow, and purple). The 22 transfer RNA genes are shown in black, and the 2two ribosomal RNA genes are shown as green colored blocks. The non-coding regain has no colored blocks. *Created with BioRender.com*

Role of mitochondrial mutations in ME/CFS

Mitochondrial disorders

Mitochondrial disorders are characterized by a dysfunction in mitochondria and are usually chronic, heterogeneous diseases affecting adults and children. They are predominantly defined by a genetic defect in mitochondrial genes or nuclear genes that cause structural defects in proteins participating in oxidative phosphorylation. These diseases can affect any tissue or organ and show a significant amount of clinical variation due to varying degrees of heteroplasmy in mitochondrial mutations across patients and are hard to diagnose (Gorman et al., 2016; Russell et al., 2020). The first time a human disorder was linked to mitochondrial mutation was in 1988 (Gorman et al., 2016; Wallace et al., 1988), and since then, a considerable amount of research has been conducted in this field, leading to the discovery of about 200 pathogenic mitochondrial mutations before the advent of the 21st century (Howell', 1999). Leigh Syndrome, which presents as neurodevelopment regression, and may lead to seizures, is the most common mitochondrial disease in children. MELAS is another common mitochondrial disease, which presents as stroke-like episodes in adult individuals and may be associated with diabetes mellitus and deafness; there are many more mitochondrial diseases (Gorman et al., 2016).

ME/CFS is characterized by immunological symptoms, fatigue, and PEM, among other symptoms, and as a result, many studies have been conducted to analyze the difference in OXPHOS output between patients and healthy controls using immune cells, skeletal muscle tissue (R. J. M. Lane, Barrett, Taylor, et al., 1998; R. J. M. Lane, Barrett, Woodrow, et al., 1998; Missailidis, Annesley, et al., 2020; Tomas et al., 2017) etcetera; see section "OXPHOS output in ME/CFS." This ongoing research has also led to an interest in sequencing mitochondrial DNA or RNA to determine specific mutations that may be enriched in ME/CFS.

Different approaches can analyze the role of mitochondrial mutations in ME/CFS: one approach is by sequencing mitochondrial DNA mutations and looking for mutations that correlate with a disease condition or disease symptoms. Billing-Ross et al. (2015) sequenced the mitochondrial DNA from 193 cases and 196 age- and gender-matched healthy control samples obtained from the Chronic Fatigue Initiative (CFI). The samples were total genomic DNA extracted from Peripheral Blood Mononuclear cells (PBMCs), and the entire cohort included residents of the United States alone. They used Polymerase Chain Reaction (PCR) first to generate two fragments of about 10.5 and 11 kb in length covering the entire 16.6 kb mitochondrial genome with sufficient overlap, followed by library preparation for sequencing by the Illumina MiSeq platform (Ravi Rupesh Kanchi and Walton, 2018).

They used the sequencing data to analyze the relationship between mitochondrial SNPs, haplogroups, and heteroplasmy levels with two aspects: disease status and symptom status. Single Nucleotide Polymorphisms (SNPs) are mutations that are the most common reason for variation in a population and are single-base substitutions. A Haplotype is a group of SNPs that are usually inherited together. For example, mutations found close together on a single chromosome or ,in the case of mitochondrial DNA, SNPs clustered together (Lee et al., 2017). Haplogroups are groups of specific haplotypes (regions of mitochondrial DNA) that are inherited together and were likely passed down to a specific population by a single ancestor (Lee et al., 2017). Mitochondrial DNA haplogroups have been used for maternal ancestry detection (Översti et al., 2019) and were used in this study in association testing. Specifically, individuals with haplogroup J and U had lower joint pain and bloating distress levels, respectively, and haplogroup H individuals were more susceptible to "feeling dead" after exercise. Symptom clusters were defined based on the Short Form- 36 and DePaul Symptom Questionnaires.

There were eight significant associations of SNPs with specific symptoms like difficulty in performing at work, bloating, and gastrointestinal distress, and it was found that the overall heteroplasmy levels were low. No known pathogenic/ deleterious mutations were detected in ME/CFS patients. This study mirrors the notion rising from metabolomic analyses wherein variation exists among specific upregulated and downregulated metabolites across patients even though the overall presentation/disease state is uniform and supports the need for individualized treatment and monitoring. (See section "Metabolomic studies in ME/CFS patients") (Naviaux et al., 2016).

A more robust mitochondrial sequencing technology called Sequencing by Targeted Amplification of Multiplex Probes (STAMP) was developed to serve as a cost-effective and low error sequencing methodology (Guo et al., 2020). STAMP is both labor and cost-effective and possessed high mapping rates to mtDNA and in-depth coverage of unique reads. The protocol begins with a 20-hour hybridization step wherein 46 Extension Ligation (EL) probe pairs are made to bind the 16.6 kb human circular mtDNA from total extracted genomic DNA. Each probe pair consists of an extension probe with three parts: an extension arm that is complementary in sequence to the targeted mtDNA region, a unique molecular tag, and a standard PCR primer annealing region used for PCR amplification of the captured target. The second step is the gap-filling step, wherein the extension arm of the extension probe is extended and ligated with the ligation arm. This generates a unique read, capturing a region of the mitochondrial genome. Following this, each read is amplified by PCR. Most protocols used for enriching mitochondrial DNA involve many PCR cycles, hence introducing amplification errors, and STAMP allows the minimization of these errors. This is thanks to the unique molecular barcode that allows for the detection of duplicate reads; duplicate reads can be used to generate a consensus read that maps to a specific region of the genome. The third and final step includes amplifying the reads using indexing primers, consisting of

P5 and P7 Illumina adapter sequences. Amplified libraries with unique combinations of P5 and P7 Illumina adapter sequences can be pooled and purified for sequencing by the Illumina platform (Guo et al., 2020).

The experiment shown in Appendix A of this work demonstrates the application of STAMP on ME/CFS patients and control samples used in the study done by Billing-Ross et al. (2016). Billing-Ross et al. (2016) sequenced the mitochondrial genome by generating two 10.5, and 11 kb fragments for each sample converted into libraries using the Nextera XT DNA Sample Preparation Kit (96 samples, FC-131-1096, http://www.illumina.com). Though much longer sequences are more useful in sequence assembly and finishing the assembled genome after sequencing, they have the disadvantage of producing single-pass sequence reads with every 8^{th} or 9^{th} base incorrect. The shorter reads (550-650 bp) generated in STAMP and the ability to wean out duplication errors might lead to the identification of SNPs and haplogroup associations that were missed. And if this is not the case, it could be helpful to validate the original study's data (Billing-Ross et al., 2016). Appendix A shows detailed information regarding the steps followed and data obtained.

Another study carried out genetic association analyses using two cohorts from two different regions: South Africa and the United Kingdom (Venter et al., 2019). Like Billing-Ross et al. (2016), they were interested in looking at the correlation of haplogroups and proven pathogenic mutations with disease status. They analyzed whether 89 moderate and 29 severely affected patients from the UK cohort and 143 moderately affected patients from the South African cohort harbored pathogenic mutations above the threshold required for disease or at the subthreshold level and found that the two mutations detected were present at levels unlikely to cause pathogenicity. Then, they specifically studied mildly deleterious mutations and found that patients had more individuals without such mildly deleterious mutations. When analyzing haplogroups, they discovered that there

was no correlation of any haplogroups with disease status in cohorts from both regions. They advocated the need for more studies, including larger population sizes across country boundaries.

Another approach to understanding the role mitochondrial genetic elements play in ME/CFS is conducting transcriptomic analyses. Vernon et al. (2002) extracted total RNA from PBMCs of 5 ME/CFS patients and 17 healthy controls to determine if the gene expression profile was different between patients and healthy controls. The RNA was labeled and hybridized to 1764 genes of the Atlas Human cDNA expression array (588 genes) and the Atlas Human 1.2 Array II (1176 genes). Eight genes were found to be differentially expressed between age-matched cases and controls and when comparing all ME/CFS samples to the control samples. Among these were CMRF35 antigen, IL-8, and HD protein involved in immunologic functions. This led the authors to believe that there may be a component of immune dysfunction in the pathophysiology of ME/CFS. Powell et al. (2003) analyzed PBMCs from 7 ME/CFS patients and four healthy controls by "differential display" using gel electrophoresis. The patients and control samples were screened for different banding patterns, and genes corresponding to differential expression were cloned and identified. They found that 12 genes were overexpressed in lymphocytes from ME/CFS patients, two of which corresponded to *cathepsin C* and *MAIL1*. Seven of the differentially expressed gene levels were verified by TaqMan quantitative PCR, out of which six were found to be overexpressed in ME/CFS patients. According to the authors, the upregulation in patient samples of *cathepsin C* and *MAIL1* – regulated genes in activated lymphocytes (Porter et al., 2013; Powell et al., 2003)-supports the notion that ME/CFS is a disease characterized by immune system activation. A study used microarrays to profile the expression of 3800 genes of 23 women with ME/CFS. Array data were analyzed by clustering most of the subjects into groups of "gradual onset of ME/CFS" and "sudden onset of ME/CFS." One hundred seventeen genes were differentially expressed, out of which 19

were increased in the gradual onset group, and 98 were found to be decreased in this group. The authors found that 20 genes were downregulated in the gradual onset cluster and were mainly involved in the regulation of glycolysis, glucose and disaccharide metabolism, oxidative phosphorylation, and amino acid biosynthesis. The authors believe that their results support the notion that ME/CFS is a heterogeneous disease (Whistler et al., 2003).

Kaushik et al. (2005) published a pilot study of gene expression using PBMCs of 25 ME/CFS patients and gender and age-matched 25 healthy controls using a single-color microarray representing 9522 human genes. Using a cut-off fold difference of expression <= 1.5 and a P value of 0.001, 35 differentially expressed genes were shortlisted. These were further analyzed using TaqMan real-time PCR in fresh samples, and 15 upregulated genes were identified in ME/CFS patient samples (*PRKCL1, MRPL23, CD2BP2, GSN, NTE, POLR2G, PEX16, EIF2B4, EIF4G1, ANAPC11, PDCD2, KHSRP, BRMS1, GABARAPL1*). *IL-10RA* was found to be downregulated in ME/CFS patients. These results led the author to speculate that T-cell activation and perturbation of neuronal and mitochondrial function might be a feature of ME/CFS (Kaushik et al., 2005).

Next, the same group undertook a comprehensive study wherein gene expression in PBMCs of 25 ME/CFS patients and 50 healthy controls were analyzed. They used a cutoff fold difference of expression of 2.5 or higher and found that 182 genes were differentially expressed. They then performed TaqMan RT-PCR using 55 patient samples and 75 healthy control samples and confirmed that 88 genes (from the previously shortlisted 182 genes) were differentially expressed. Eighty-five of these genes were upregulated in ME/CFS, and 3 were found to be downregulated (Kerr, Petty, et al., 2008). All 16 genes reported in their pilot study (Kaushik et al., 2005) were upregulated with the discrepancy that *IL10RA* was also upregulated in ME/CFS. Clustering the RT PCR values from ME/CFS patients led to the identification of 7 subtypes that differed in clinical

symptoms, severity, and Medical Outcomes Survey Short Form- 36 scores (Lowrie et al., 2003). Subtypes 1 and 7 had the most severe clinical phenotypes, and these were followed by 2, 4, 5, 6, and 3 on the disease severity scale. The functions represented among the differentially expressed genes were a hematologic disease, immunologic disease, cancer, cell death, immune response, and infection (Kerr, Petty, et al., 2008).

Kerr et al. (2008) report that they demonstrated the same gene expression abnormalities along with the existence of subtypes with marked clinical differences. It was pointed out that the presence of ME/CFS subtypes could explain the difficulty of identifying consistent and reproducible abnormalities in ME/CFS patients. They also explained that microarray screening had the shortcoming of not identifying novel genes that were differentially expressed in ME/CFS (Powell et al., 2003). They also screened ME/CFS patients using a combination of Massive Parallel Signature Sequencing (MPSS) using 20 ME/CFS and 20 healthy control blood donors. However, these results did not add to the microarray testing results (Kerr et al., 2008). All the studies performed by this group included patients and healthy control samples from the UK. Bouquet et al. (2019) set out to determine whether CPET could trigger immune dysregulation or virus reactivation. They performed RNA Sequencing analysis on 14 ME/CFS patients and 11 matched sedentary controls. The entire cohort was made to participate in CPET on two consecutive days and followed up to 7 days post-exercise. Although ME/CFS patients had worse symptoms after exercise, transcriptome analysis yielded only six differentially expressed gene candidates compared to controls. And none of the gene candidates were involved in immune function. Also, the number of viral reads was not significantly associated with PEM.

Sweetman et al. (2019) analyzed the transcriptome of PBMC cells from 10 ME/CFS patients age/gender-matched with 10ten healthy control subjects. The patients belonged to a New Zealand

cohort, and they found that 27 gene transcripts were at higher levels (1.5-6-fold) and six transcripts were at lower levels (3-6-fold) in patients compared to healthy controls. While the cohort size is small, the study did find significantly higher transcript levels of genes like *IL8* and *NFKBIA* (validated through quantitative polymerase chain reaction (qPCR)). The RNA libraries were sequenced and data analysis by the Illumina HiSeq platform (https://www.illumina.com/systems/sequencing-platforms/hiseq-2500.html) was done in R using EdgeR. The three top upregulated genes in the study- *IL8, NFKBIA,* and *TNFAIP3*- are involved in TNF-induced NF- ϰB activation, which turns on critical pathways of inflammation, and immune cell proliferation and differentiation (Hayden & Ghosh, 2014). These results support the speculation that ME/CFS patients may have overactive immune and inflammation signaling driven by TNFα ME/CFS patients (Bell, 1994a).

As can be gleaned from the above-reported studies, activated immune response is a feature that is speculated to be a characteristic of ME/CFS patients (Kaushik et al., 2005; Kerr et al., 2008; Kerr, Petty, et al., 2008; Powell et al., 2003; Sweetman et al., 2019; Vernon et al., 2002; Whistler et al., 2003), but not in case of all studies(Bouquet et al., 2019; Kerr et al., 2008). None of the differentially expressed genes are mitochondrial genes however upregulated immune signaling does affect mitochondria and could lead to mtDNA damage (Breda et al., 2019). Also. Some differentially expressed genes include those involved in oxidative phosphorylation and glycolysis regulation. There is a need for more transcriptome/RNA seq analysis studies, including larger ME/CFS cohorts and better mtDNA sequencing methodologies to uncover more meaningful relationships between ME/CFS and mtDNA. Analyzing the role of nuclear DNA mutations that play a direct or indirect role in mitochondrial output is also desirable. A method like STAMP can be applied at the single-cell

level to detect mildly deleterious mutations present in mitochondria at deficient levels and are hence not detected when cells are pooled together.

OXPHOS output in ME/CFS

In ME/CFS, the symptoms of exhaustion and tiredness immediately point towards a lack of the energy currency of our body, ATP, and cast a suspicion on the efficiency of ATP production by mitochondria. Studies may include Magnetic Resonance Spectroscopy to investigate tissue metabolism or the use of technology like the Seahorse extracellular flux analyzer to measure various aspects of mitochondrial energy output in cells.

Studies focused on muscle tissue.

One of the earliest studies investigating OXPHOS output included using NMR spectroscopy to measure the high energy phosphate levels in skeletal muscles (Wong et al., 1992). The study analyzed ^{31}P levels by performing nuclear magnetic resonance (NMR) spectroscopy on the gastrocnemius muscles of 22 ME/CFS patients and 21 control subjects from Alberta, Canada. The ^{31}P levels were compared during rest, graded dynamic exercise to exhaustion and recovery. The authors found that though ME/CFS patients and control subjects had similar skeletal muscle metabolic patterns during training, ME/CFS patients reached exhaustion more rapidly, at which point their intracellular ATP concentrations were also low. The study suggests that there is a defect in oxidative metabolism and acceleration of glycolysis in working skeletal muscles of ME/CFS patients. A more recent study measured Phosphocreatinine (PCr)/ adenosine triphosphate (ATP) ratio and other parameters in cardiac and skeletal muscles from 12 ME/CFS patients and 8 healthy controls. The study showed an association between cardiac muscle bioenergetic abnormality with previously documented skeletal muscle bioenergetic abnormalities in ME/CFS patients (Jones et al., 2010). They also found that cardiac contractility increases in ME/CFS patients on standing and is correlated with the observed cardiac bioenergetic impairment.

As a part of the SATET (Subarachnoid threshold exercise test), many ME/CFS patients show increased lactate levels following short periods of moderate exercise- such patients are called SATET +ve, and those who show normal lactate levels are called SATET -ve (R. J. M. Lane, Barrett, Taylor, et al., 1998). Phosphorous magnetic resonance imaging was used to analyze forearm muscles of 10 SATET +ve patients, 9 SATET -ve patients, and 13 sedentary volunteers. Though resting spectra showed no differences, there was a significantly lower intracellular pH level in the case of SATET +ve patients. ATP synthesis was also significantly lower during recovery of SATET +ve patients (R. J. M. Lane, Barrett, Taylor, et al, 1998). The same group conducted another study wherein they used electron microscopy to examine 105 quadriceps muscle biopsies from ME/CFS patients ,and it was found that patients with abnormal lactate responses to exercise had a significantly lower proportion of mitochondria-rich type 1 muscle fibers (R. J. M. Lane, Barrett, Woodrow, et al., 1998). Another study, including subjects that participated in CPET (cardiopulmonary exercise test), found that there was a lower rate of oxygen uptake by cardiac muscle cells relative to cardiac output at the time of exercise (Vermeulen & Vermeulen van Eck, 2014).

The papers discussed above include the analysis of different types of muscle tissues (cardiac, skeletal, quadricep, etc.), use of various analytical parameters (lactate levels, Phosphocreatine (PCr)/ adenosine triphosphate (ATP) ratio) while subjecting the patients to varying exercise regimens or no exercise. A common finding is that there is bioenergetic impairment in ME/CFS patient muscle tissue and that exercise exacerbates fatigue and exhaustion symptoms in ME/CFS patients.

Studies employing Seahorse Flux Analyzer

The Seahorse Flux analyzer (XF Mito stress test, XF Glycolysis stress test, Seahorse Bioscience, North Billerica, MA, USA) allows rapid and noninvasive measurement of bioenergetics

like basal oxygen consumption, glycolysis rates, ATP production and respiratory capacity in a single experiment (Agilent Technologies, 2018). The studies discussed below all use this technology in various cells/tissues.

Tomas et al. (2017) showed that the basal respiration using Seahorse was much lower in ME/CFS patient PBMC samples than healthy controls (Tomas et al., 2017). One of the most recent studies used immortalized lymphoblasts from 51 ME/CFS patients and 22 age and gender-matched controls. The Seahorse extracellular flux analysis assessed mitochondrial function, and many exciting results, including upregulation of mitochondrial membrane potential, and upregulation of expression of complexes involved in fatty acid metabolism, were shown. The rate of ATP synthesis by Complex V was significantly reduced and hyperactivated TOR Complex I stress signaling was observed. They concluded that respiratory efficiency is lower because of which fatty acid metabolism pathways were upregulated. Hence, when acute energy-demanding situations like exercise or stress arise, the cells would not be able to cope as their fatty acid metabolism pathways would already be activated (Missailidis, Annesley, et al., 2020).

Nguyen et al. (2019) conducted a study with the goal of looking at the metabolic output in resting Natural Killer (NK) cells from ME/CFS patients. Their cohort consisted of 6 ME/CFS patients of age and sex-matched with the same number of healthy controls from Australia. They performed mitochondrial stress tests to measure parameters of mitochondrial function and glycolytic stress tests to measure parameters such as glycolytic reserve, glycolysis, etcetera. They observed a significant reduction of glycolytic reserve in the ME/CFS patient samples and no significant difference in mitochondrial respiration between patient samples and healthy controls. In conclusion, the authors believe that reduced glycolytic reserve in resting NK cells shows that ME/CFS patient NK cells have a reduced ability to deal with increased ATP demand and believe that this may have a role in ME/CFS pathogenesis.

Tomas et al. (2017) have used PBMCs, whereas Nguyen et al. (2019) and Missailidis, Annesley, et al. (2020) have focused their attention on lymphoblasts and NK cells respectively, which are subtypes of PBMCs. The largest cohort of the three studies discussed above reported that maximal respiration was significantly impaired in ME/CFS patients (Tomas et al., 2017). Nguyen et al. (2019) studied 6 ME/CFS patients and equal number of age and gender-matched healthy controls. They found no significant differences between mitochondrial respiration, basal respiration, proton leak, ATP production, maximum respiration, spare respiratory capacity, and coupling efficiency. The cohort studied was relatively small and this could be a contributing factor to the fact that no significant differences in cellular respiration were observed between healthy controls and patients. However, the authors did find a significant reduction of glycolytic reserves in ME/CFS patients. Missailidis et al. (2020) observed that basal oxygen consumption rate (OCR) and the rate of ATP synthesis by Complex V was significantly reduced in ME/CFS (Cohort: 51 patients; 22 healthy controls). Overall, in the two studies wherein the cohort size was larger, an impairment in cellular respiration the case of ME/CFS patients was observed.

Mitochondria and Immune function in ME/CFS

ME/CFS is often described by patients as "the flu that they never recovered from" and the immune system's roll in ME/CFS is evidenced by the symptoms (Siegel et al., 2018). In his 1994 book, "A Doctor's Guide to Chronic Fatigue Syndrome," David S. bell also speculates that the "cytokine theory" that postulates an immune system that has been pushed into overdrive state, is the most promising theory explaining ME/CFS(Bell, 1994a). There has been a huge interest in analyzing various aspects of immune cell metabolism and function and many studies have been undertaken using peripheral blood mononuclear cells (PBMCs) as a whole or isolating specific cell types like T cells, Natural Killer (NK) cells (King, 2020).

Tomas et al. (2017) applied the Seahorse flux analyzer technology to PBMCs isolated from 52 ME/CFS patients and 35 control samples and analyzed the following parameters: basal respiration, ATP production, proton leak, maximal respiration, reserve capacity, non-mitochondrial respiration, and coupling efficiency. They found that maximal respiration was the key parameter that was significantly different between patients and controls with the former being found to occur at lower levels in ME/CFS patients. They concluded that the PBMCs are likely experiencing physiological stress and are unable to fulfill cellular energy demands. Many other studies have used PBMCs for measuring OXPHOS output or to measure metabolites indicative of oxidative stress and these studies have been discussed in those respective sections of this work (Castro-Marrero et al., 2013; Missailidis, Sanislav, et al., 2020). Mitochondrial sequencing (Billing-Ross et al., 2016; Venter et al., 2019) and RNA sequencing studies (Sweetman et al., 2019) described previously also use PBMC cells and further support the theory of dysfunctional immune cell- mediated pathophysiology of ME/CFS.

Mandarano et al. (2020) investigated the OXPHOS output of isolated $CD4^+$ and $CD8^+$ T cells from 53 patients with ME/CFS and 45 healthy controls from Incline Village, Nevada, USA. The Seahorse flux analyzer was used to measure glycolysis rates at rest and after activation. It was found that both $CD4^+$ and $CD8^+$ T cells from patients had reduced glycolysis at rest, and $CD8^+$ T cells had reduced glycolysis following activation. Plasma cytokine analysis was also performed on isolated plasma from these samples. Plasma was collected in EDTA tubes, and 48 cytokines and chemokines were measured using Bio-Plex pro Human Cytokine Screening Panel, 48-plex (Bio-Rad). It was observed that proinflammatory cytokines unexpectedly correlate with T cell hypometabolism. The authors speculate that T cells could have lost responsiveness to pro-inflammatory markers in ME/CFS patients which suggests a strong role of metabolically abnormal immune cells in ME/CFS as do other previous studies (Fletcher et al., 2010; Mandarano et al., 2020; Tomas et al., 2017). Also, Mandarano et al. (2020) tease out the specific role of T cell subtypes in ME/CFS which provides more nuanced and meaningful information with regards to treating the disease and understanding the exact pathophysiology. This study highlights the need for future studies wherein PBMCs are sorted to generate individual cell subtype populations which are then analyzed with respect to aspects of metabolic output and more. Also, worth including in this section is the fact that, NK cell cytotoxicity (NKCC) in ME/CFS patients has been found to be markedly reduced making the same one of the most consistent immune abnormalities observed in ME/CFS patients (Carruthers et al., 2011; Fletcher et al., 2010).

Oxidative Stress Pathways and ME/CFS

Oxidative stress occurs due to increased amounts of reactive oxygen/nitrogen species and is a phenomenon that occurs during aging as well as chronic diseases. Reactive oxygen/nitrogen species contain free radicals -molecules that have unpaired electrons- and at least one oxygen atom/nitrogen atom. Mitochondrial mutations that affect respiratory chain complexes can lead to production of free radicals which can then serve as signals for repair within the mitochondrion (N. Lane, 2005). If the problem is not solved within the mitochondria, free radicals send signals to the nucleus to assist in correcting the mutations. In case the mitochondria get trapped in a vicious cycle wherein free radicals damage mitochondrial DNA causing more mtDNA mutations which further produce more free radicals, then the cell is eliminated by the immune system (N. Lane, 2005). Hence while free radicals are important for signaling, they should not be present at very high levels. Antioxidants like CoenzymeQ10 and Vitamin E scavenge free radicals and prevent cellular damage. An imbalance of free radicals and antioxidants causes oxidative stress (Starkov, 2008)(N. Lane, 2005). ME/CFS symptoms include fatigue, inflammation and immune reactivity and as previously discussed, cellular respiration is found to be somewhat impaired in various tissues(Holden et al., 2020). The study mentioned below show investigations that measure levels of oxidative stress markers like glutathione and Coenzyme Q10 and provide evidence that ME/CFS patients do exhibit oxidative stress (Morris & Maes, 2012, 2013).

Jammes et al. (2012) conducted a study with 43 ME/CFS patients matched with 28 healthy controls and showed increased levels of oxidative stress markers in blood plasma which further increased after exercise. Similar results were obtained from a study much earlier which looked at analyzed whole blood samples in 33 ME/CFS patients' gender and age matched with 27 healthy controls. As a part of this study, markers of oxidative stress like malondialdehyde and 2,3-

diphosphoglycerate were correlated with variations in cognitive symptoms of ME/CFS (Richards et al., 2000).

Castro-Marrero et al. (2013) performed High Performance Liquid Chromatography (HPLC) to measure Coenzyme Q10 levels in PBMCs of fibromyalgia patients (n=20), ME/CFS patients (n=23) and healthy controls (n=15) (all patients were from Spain). Coenzyme Q10, commonly referred to as an antioxidant, carries electrons from Complex 1 to 3 in the ETC and the authors found these levels to be significantly lower in ME/CFS patients as compared to healthy controls and same was the case for fibromyalgia patients. They also found that ATP levels were lower in ME/CFS and fibromyalgia patients and that lipid peroxidation was higher in ME/CFS patients and fibromyalgia patients. They concluded that ME/CFS and fibromyalgia patients had increased signs of oxidative damage and decreased antioxidants (Castro-Marrero et al., 2013).

Another group carried out a neuroimaging study wherein they used ^{1}H Magnetic Resonance Spectroscopic Imaging (MRSI) technologies to measure CerebroSpinal Fluid (CSF) lactate in ME/CFS patients(n=15), Major Depressive Disorder (MDD)(n=15) patients and healthy volunteers (n=13) (Shungu et al., 2012). They also measured levels of cortical glutathione (GSH) as a marker of antioxidant capacity, cerebral blood flow and high energy phosphates. There was an elevation in lactate levels and reduction in GSH levels in both ME/CFS and MDD levels in comparison to healthy controls. They did not detect any differences in high-energy phosphate metabolites in all three groups. This study was a third instalment of brain imaging studies carried out by the same group and they concluded that increased oxidative stress might be a key player in ME/CFS pathophysiology (Mathew et al., 2009; Murrough et al., 2010). Similar studies were conducted with a larger cohort (Natelson, Mao, et al., 2017),including fibromyalgia patients, and similar results were obtained (Natelson, Vu, et al., 2017).

Metabolomic and Proteomic studies in ME/CFS patients

Metabolomic studies

The studies described here are focused mostly on profiling the metabolites in blood plasma from patients and healthy controls and analyzing the differences. Receiver Operating Curve (ROC) analyses have been done in most of the studies described here to estimate the power of using certain metabolite levels for diagnosis of ME/CFS. Various techniques including Liquid chromatography-mass spectrometry (LC-MS), capillary electrophoresis-mass spectrometry (CE-MS) and aptamer interrogation have been used.

In a relatively recent study, blood and urine samples were collected from a cohort of females (34 ME/CFS patients and 25 healthy controls) to perform metabolic profiling using one-dimensional ^{1}H Nuclear Magnetic Resonance spectroscopy and blood glucose was found to be elevated in ME/CFS patients (Armstrong et al., 2015). On the contrary blood lactate, urine pyruvate, and urine alanine were reduced indicating an inhibition of glycolysis that may affect the citric acid cycle by not providing enough substrate for the same. Yamano et al. (2016) performed metabolomic analysis on plasma samples obtained from a cohort of 133 Japanese individuals (67 ME/CFS, 66 Healthy Controls) with the aim of establishing a parameter for diagnosis of ME/CFS. The authors performed a comprehensive analysis using Capillary Electrophoresis – Time of Flight Mass Spectrometry (CE-TOFMS) on two sets of samples from the original cohort: the training data set (47 ME/CFS, 46 Healthy Controls) and the validation data set (20 ME/CFS, 20 Healthy Controls). They identified 31 metabolites that were significantly different between patients and controls. These consisted of metabolic intermediates of the TriCarboxylic Acid cycle (TCA) and the urea cycle. Subsequently, they identified that the combination of ratios of pyruvate/ isocitrate and ornithine/citrulline could be used as index markers to differentiate ME/CFS patients from healthy controls. They developed a multiple logistic regression model using these indices and the training

data set and validated the same using the validation data set. They concluded that the combination of ornithine/citrulline and pyruvate/isocitrate ratios clearly discriminated CFS patients from healthy controls.

A similar study was performed by Naviaux et al. (2016) in which they used liquid chromatography, electrospray ionization and tandem mass spectrometry to target metabolites in plasma of ME/CFS patients and healthy controls. Total number of subjects included was 84 (45 patients, 39 controls) and they, too, set out with the goal of developing a diagnostic parameter for ME/CFS. The authors suggest that ME/CFS patients have a distinct metabolic signature that can be used for diagnosis and that the metabolic profile of ME/CFS patients resembles that of the evolutionarily conserved Dauer state. Out of 63 targeted metabolomic pathways, they found that 20 were abnormal in ME/CFS patients with most of these involving metabolites like sphingolipids and phospholipids. Purine ($p = 0.044$), Cholesterol (0.035) and Riboflavin ($p= 0.005$) were also significantly reduced in ME/CFS patients in comparison to healthy adults. Branched Chain Amino Acid (BCAA) ($p =0.023$) levels were also found to be reduced in comparison to patients, which supports findings of previous studies (Blomstrand et al., 1991).

Maximum percent of reduced metabolites were sphingolipids and glycosphingolipids, which in turn affect fatty acid metabolism. The authors postulated that the body might be switching its metabolic state to that of an inactive survival state in response to some stressor, very similar to the Dauer state (Lant & Storey, 2010). The authors suggested that while evolutionarily the Dauer state serves to support survival by opting for fewer firing pathways and preventing acute damage, in the case of ME/CFS, it might lead to significant impairment in activity and severe fatigue. The study also tries to analyze how the ME/CFS metabolic signature differs among male and female patients; for example: the authors find that more sphingomyelin species in females are present at normal levels in comparison to males. The study supports and strengthens the claim that fatty acid and lipid

metabolism is adversely affected in ME/CFS patients and puts forth the idea that despite the heterogenous nature of ME/CFS, there is a consistent manifestation in terms of metabolite profile, hence making the disease diagnosable and perhaps treatable.

Germain et al. (2017) applied mass spectrometry on blood plasma from ME/CFS patients and age/gender matched controls to analyze differences in metabolic profiles of these subjects. The study showed that there were 74 metabolites that were present at significantly different levels ($P<0.05$) in patients versus healthy controls and they found that 35 of these were retained when they made the statistical criteria more stringent ($P<0.05$, $Q<0.15$). The differentially accumulating metabolites were employed for pathway analysis to understand if there were metabolic pathways that might be affected. It was discovered that lipid metabolism was impacted due to lower accumulation of metabolites like glucose, oxaloacetate, and taurine; membrane phospholipids were also found to be lowered in patients, as well as purines like ADP and ATP.

The same authors sought to investigate the plasma extracted from 32 patients and 19 healthy controls (all females) using the Metabolon®platform (Germain et al., 2018). 832 metabolites covering eight classes or "super pathways" were measured and alpha ketoglutarate was found at significantly different levels in patients as compared to healthy controls. Other such differentially detected metabolites include cofactors and vitamins like Heme and gamma-CEHC; nucleotides like cAMP and the peptide gamma-glutamyl threonine. Alpha- ketoglutarate belongs to the TCA cycle sub-pathway and the fact that TCA cycle has surfaced consistently in ME/CFS metabolomic analyses supports the notion of dysfunction in energy metabolism in ME/CFS. The conclusions from the study also support previous findings from similar sized cohort studies (Armstrong et al., 2015; Germain et al., 2017) but does not support the notion of Dauer state existence as observed by Naviaux et al. (2016). The same group next performed another metabolomic study wherein 52 plasma samples from female subjects were analyzed using Metabolon's mass spectrometry(Germain

et al., 2020). The samples consisted of equal number of control and patient samples and provided data for 1750 blood compounds. The study reported that patient samples showed a decrease in four long-chain acyl lipids (arachidonoylcholine, linoleoylcholine, oleoylcholine, and palmitoylcholine) as compared to healthy controls. These long-chain acyl lipids are known to have a pressor effect with reference to blood pressure and are believed to have a stimulating effect on the parasympathetic nervous system (PSNS) (Marshall, 1911). Cholate- a bile acid which plays an important role fat absorption - levels were also found to be 3 times lower in-patient samples. provide support to the observation that there is a disruption in the acyl choline class of lipids. The reduction in acyl lipids` is in alignment with the authors' previous study and so is the finding that dipeptides are decreased in patients vs healthy controls (Germain et al., 2018). The new finding, they have shown in this study is that three classes of steroids are reduced in ME/CFS patients which include androgenic, progestin and cortico- which can have major repercussions on health of an individual (Germain et al., 2020). The controversial finding includes that the ceramides and sphingomyelins are found to be increased in the case of patients which is contradictory to what was reported by Naviaux et al. (2016), who found the levels to be decreased regardless of gender. Germain et al. (2020) found that when splitting the patient cohort based on gut symptoms, ceramide levels for patients without inflammatory bowel disease (IBS) further increased which is inconsistent with a report published by yet another group (Nagy-Szakal et al., 2018). These inconsistencies according to the authors, point to a possibility that the sphingomyelin and ceramide changes in patients are due to irregular diets and lifestyle abnormalities resulting from living with ME/CFS rather than being disease specific abnormalities.

Hoel et al. 2021 conducted Exploratory Data Analysis (EDA) to compare blood serum of 83 ME/CFS patients with 35 healthy controls. They used the HD4 platform from Metabolon to carry out global metabolomics, lipidomics and hormone measurements. They found that circulating

glycerol levels were elevated which indicated that lipolysis was elevated. They found that there was an alteration in amino acid (BCAAs and tryptophan) usage and that three sphingolipid varieties elevated. They did not find any changes in morning cortisol and ACTH but did detect increased levels of *FABP4* and *FGF21* which points towards systemic metabolic stress (Hoel et al., 2021). Since the decrease in sphingolipids was not replicated in ME/CFS patients of this study and the study by Germain et al. (2020), the Dauer state hypothesis as proposed by Naviaux et al. (2016) is brought into question.

As can be gleaned from the above discussed reports, untargeted metabolomic studies have uncovered metabolic patterns in ME/CFS patients that represent disturbed cellular energetics, although the specifically affected metabolites do vary between studies (Germain et al., 2017, 2018, 2020; Nagy-Szakal et al., 2018; Naviaux et al., 2016). Also, targeted metabolomic analyses further support the notion that fatty acid metabolism and purines like ATP are downregulated and in ME/CFS patients which are both pathways that implicate the energy metabolism by mitochondria (Armstrong et al., 2015; Litwack, 2018; Yamano et al., 2016). A mathematical modelling study proposes a theory known as "metabolic trap hypothesis" to suggest a possible pathophysiological mechanism that may be at play in ME/CFS (Kashi et al., 2019). With the goal of identifying predisposing mutations the authors start by looking for common damaging mutations in genes coding for enzymes involved in energy metabolism. They found four such mutations in IDO2 gene that encodes for enzyme indoleamine 2,3-dioxygenase 2. The IDO1 and IDO2 genes both encode for indoleamine 2,3-dioxygenase enzymes 1 and 2 respectively both of which can cleave the aromatic ring of the least-abundant essential amino acid, L-Trp (L-tryptophan), to produce L-kynurenine (Yeung et al., 2015). IDO1 can be substrate inhibited and there is a critical limit of L-tryptophan levels beyond which IDO1 becomes non-functional. This phenomenon in conjunction with mutated IDO2 enzyme can lead to the cell being "trapped" in a metabolically dysfunctional state wherein L-

kynurenine is not produced, and downstream electron carrier NAD^+ cannot be synthesized (Yeung et al., 2015). In certain cells the IDO system is the only pathway for kynurenine production (dendritic cells and macrophages) and this metabolic trap can cause them to lose the main pathway of de novo NAD+ synthesis, which in turn could affect oxidative phosphorylation. The IDO2 mutation is not causal but could be predisposing amid other deleterious factors that might all together lead to ME/CFS. In conclusion, large scale and small-scale metabolomics studies shed light on many pathways that might have a role to play in ME/CFS and enough such studies would help in piecing together mechanisms that could be useful for diagnosing and treating ME/CFS.

Proteomics studies

A relatively recent mass spectrometry-based study examined the proteomes of PBMCs from ME/CFS patients and controls and found 60 proteins that were significantly differentially expressed. A proportion of the identified proteins were involved in mitochondrial function, OXPHOS, and redox regulation and a significant number of these were implicated in immune inflammatory response, DNA methylation and apoptosis. The authors concluded that their results supported a notion of dysfunction in ATP production which was compensated by pathways upstream of Complex V. The authors suggest that their results points to elevation in oxidative stress in PBMCs (Sweetman et al., 2020).

Milivojevic et al. (2020) obtained plasma samples from 50 ME/CFS patients and 50 controls and performed ultra-performance liquid chromatography-tandem mass spectrometry on the same. The subjects met the 1994 CDC Fukuda and/or Canadian consensus criteria and were from clinics in New York, Utah, Nevada, and Florida. Of the samples mention ed above they used 39 ME/CFS patient samples and 41 control samples to eliminate potential batch effects. They found there was a significant association between ME/CFS and immunoglobulin heavy variable (IGHV) region 3-23/30. In addition, they were able to predict ME/CFS disease state with a high degree of accuracy

(AUC = 0.774 – 0.838) using a panel of proteins selected by 3 different algorithms. Given that IGHV3-23 is one of the most used heavy variable regions in the human immunoglobulin (Ig) repertoire (Brezinschek et al., 1997) and the fact that it has been liked with diseases like non-Hodgkin lymphomas (NHL) and follicular lymphoma (FL) (Berget et al., 2015; García-Álvarez et al., 2019) the authors speculate ME/CFS subjects may have increased IGHV3-23 levels due to antigen driven clonal expansion and suggest that kinase inhibitors of B-cell receptor signaling inhibitors ought to be explored as potential treatment targets.

Yet another recent study uses an aptamer-based approach to characterize the plasma proteome of ME/CFS patients (Germain et al., 2021). Aptamers can be DNA/RNA or derived oligonucleotides or peptides and can be designed specifically to target proteins. An example of a naturally occurring aptamer is ribozyme and in experiments the 3-D structure of folded RNA/DNA or peptides is harnessed to bind target molecules through non-covalent interactions (Kaur & Shorie, 2019). Germain et al. (2021) conducted a pilot study on isolated plasma from 20 ME/CFS patients and 20 age and gender matched controls and found that 19 proteins were present at significantly different levels in patients and controls. The pathways, these differentially accumulating proteins were part of, included immune function and cell-cell communication pathways. Further analysis showed that the Ephrin pathway, which is involved in angiogenesis, axon guidance, epithelial cell migration and immune response, was upregulated in ME/CFS patients because many of the ephrin family proteins were present at higher levels in patients as compared to healthy controls. Some studies discussed above support the notion that immune signaling pathways are implicated in ME/CFS patients (Germain et al., 2021; Milivojevic et al., 2020) and Sweetman et al. (2020) support findings from metabolomic studies regarding impairment of energy generating metabolic pathways. More such studies will help generate a better understanding of the proteomes of ME/CFS patients and how the same interact with other metabolites to precipitate disease.

Discussion

This review looks at various aspects of physiology with mitochondria at the center like oxidative stress, metabolism, and oxidative phosphorylation. The role of mitochondrial DNA mutations and mitochondrial function in immune cells has also been discussed. All papers from the systematic review by Holden et al. (2020) except two (Booth et al., 2012; Light et al., 2013) have been included and several others which are older (Bouquet et al., 2019; Kaushik et al., 2005; Kerr, Petty, et al., 2008; R. J. M. Lane, Barrett, Taylor, et al., 1998; R. J. M. Lane, Barrett, Woodrow, et al., 1998; Porter et al., 2013; Powell et al., 2003; Whistler et al., 2003; Wong et al., 1992) and some latest studies have been included (Germain et al., 2020, 2021; Milivojevic et al., 2020). For many of the studies reviewed, categorization into the separate sections was difficult as many of these physiological features (particularly those of immune activation, oxidative stress, and mitochondrial mutations) are interrelated (N. Lane, 2005; Morris & Maes, 2013).

Mitochondrial DNA heteroplasmy levels would be expected to be detected at minimal levels (N. Lane, 2005) unless single cell profiling is done. However, insights regarding haplogroup associations with certain symptoms can be quite useful in individualized treatment(Billing-Ross et al., 2016). Many studies support the notion, that oxidative phosphorylation and energy metabolism is dysfunctional in ME/CFS patients when analyzing varied tissues (Holt et al., 1988; R. J. M. Lane, Barrett, Taylor, et al., 1998; R. J. M. Lane, Barrett, Woodrow, et al., 1998; Wong et al., 1992). And as more studies report similar results particularly using Seahorse platform in immune cells – the theory of immune dysfunction in ME/CFS continues to get support (Mandarano et al., 2020; Nguyen et al., 2019; Tomas et al., 2017). Finally, metabolic profiling provides many insights due to the high throughput nature of the technologies used; differentially accumulated metabolites that can be linked to pathways of oxidative stress, fatty acid metabolism and energy generation pathways however the specific metabolites that are differentially accumulated vary between studies (Germain et al., 2017,

2018, 2020; Hoel et al., 2021; Naviaux et al., 2016; Starkov, 2008). All these approaches with mathematical modelling analysis can help elucidate target pathways which could be investigated to help reveal one or more central mechanisms at play in ME/CFS.

Viral infection has been a trigger for ME/CFS occurrence in majority patients all over the world and several speculations regarding the existence of an "Agent X" which perhaps prompts the ME/CFS disease state have been proposed since the early 1990s (Bell, 1994b). Among the many ME/CFS outbreaks that have occurred over the last few decades the one in Akureyri, Iceland in 1948 coincided with poliovirus epidemic (Bell, 1994b; O'Neal & Hanson, 2021). The patients who were affected with the so called "Iceland disease" did not have antibodies to polio but when they were immunized with a polio vaccine their antibody responses suggested a previous exposure to polio(Bell, 1994b). This indicates that there could be an "Agent X" belonging to the same family as poliovirus (enterovirus family) that could be a triggering factor in ME/CFS (O'Neal & Hanson, 2021). Any theory that explains ME/CFS would have to include an explanation to the antibody production seen during the 1948 Iceland outbreak.

At the same time, ME/CFS outbreaks have coincided with Epstein Barr Virus outbreaks(Bell, 1994b) and there is a lot of concern regarding the increased prevalence of ME/CFS after the COVID-19 pandemic (Paul et al., 2021; Wood et al., 2021). Perhaps "Agent X" is already present in most individuals but is triggered by other viral infections like poliovirus, Epstein Barr Virus and coronavirus and could hence lead to ME/CFS manifestation. The studies described in this review all contribute important pieces to the puzzle of ME/CFS and ultimately more funding is required to propel many more such studies to be done at a rapid rate to help fill in the blank pieces of ME/CFS etiology.

Literature cited

Armstrong, C. W., Mcgregor, N. R., Lewis, D. P., Butt, H. L., & Gooley, P. R. (2015). *Metabolic profiling reveals anomalous energy metabolism and oxidative stress pathways in Chronic Fatigue Syndrome patients*. http://www.ebi.ac.uk/metabolights

Bell, D. S. (1994a). *The Doctor's Guide to Chronic Fatigue Syndrome* (p.).

Bell, D. S. (1994b). The Historical Perspective. In *The Doctor's Guide to Chronic Fatigue Syndrome* (pp. 181–193).

Berget, E., Molven, A., Løkeland, T., Helgeland, L., & Vintermyr, O. K. (2015). IGHV gene usage and mutational status in follicular lymphoma: Correlations with prognosis and patient age. *Leukemia Research*, *39*(7), 702–708. https://doi.org/10.1016/j.leukres.2015.03.003

Billing-Ross, P., Germain, A., Ye, K., Keinan, A., Gu, Z., & Hanson, M. R. (2016). Mitochondrial DNA variants correlate with symptoms in myalgic encephalomyelitis/chronic fatigue syndrome. *Journal of Translational Medicine*, *14*(1). https://doi.org/10.1186/s12967-016-0771-6

Blomstrand, E., Hassmén, P., Ekblom, B., & Newsholme, E. A. (1991). Administration of branched-chain amino acids during sustained exercise - effects on performance and on plasma concentration of some amino acids. *European Journal of Applied Physiology and Occupational Physiology*, *63*(2). https://doi.org/10.1007/BF00235174

Booth, N. E., Myhill, S., & McLaren-Howard, J. (2012). Mitochondrial dysfunction and the pathophysiology of myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). *International Journal of Clinical and Experimental Medicine*, *5*(3).

Bouquet, J., Li, T., Gardy, J. L., Kang, X., Stevens, S., Stevens, J., VanNess, M., Snell, C., Potts, J., Miller, R. R., Morshed, M., McCabe, M., Parker, S., Uyaguari, M., Tang, P., Steiner, T., Chan, W. S., de Souza, A. M., Mattman, A., … Chiu, C. Y. (2019). Whole blood human transcriptome and virome analysis of ME/CFS patients experiencing post-exertional malaise following cardiopulmonary exercise testing. *PLoS ONE*, *14*(3). https://doi.org/10.1371/journal.pone.0212193

Breda, C. N. de S., Davanzo, G. G., Basso, P. J., Saraiva Câmara, N. O., & Moraes-Vieira, P. M. M. (2019). Mitochondria as central hub of the immune system. In *Redox Biology* (Vol. 26). Elsevier B.V. https://doi.org/10.1016/j.redox.2019.101255

Brezinschek, H.-P., Foster, S. J., Brezinschek, R. I., Dörner, T., Domiati-Saad, R., & Lipsky, P. E. (1997). Analysis of the Human V H Gene Repertoire Differential Effects of Selection and Somatic Hypermutation on Human Peripheral CD5 /IgM and CD5 /IgM B cells. In *J. Clin. Invest* (Vol. 99, Issue 10).

Carruthers, B. M., van de Sande, M. I., de Meirleir, K. L., Klimas, N. G., Broderick, G., Mitchell, T., Staines, D., Powles, A. C. P., Speight, N., Vallings, R., Bateman, L., Baumgarten-Austrheim, B., Bell, D. S., Carlo-Stella, N., Chia, J., Darragh, A., Jo, D., Lewis, D., Light, A. R., … Stevens, S. (2011). Myalgic encephalomyelitis: International Consensus Criteria. In *Journal of Internal Medicine* (Vol. 270, Issue 4). https://doi.org/10.1111/j.1365-2796.2011.02428.x

Castro-Marrero, J., Cordero, M. D., Sáez-Francas, N., Jimenez-Gutierrez, C., Aguilar-Montilla, F. J., Aliste, L., & Alegre-Martin, J. (2013). Could mitochondrial dysfunction be a differentiating marker between chronic fatigue syndrome and fibromyalgia? *Antioxidants and Redox Signaling*, *19*(15), 1855–1860. https://doi.org/10.1089/ars.2013.5346

Chang, C. J., Hung, L. Y., Kogelnik, A. M., Kaufman, D., Aiyar, R. S., Chu, A. M., Wilhelmy, J., Li, P., Tannenbaum, L., Xiao, W., & Davis, R. W. (2021). A comprehensive examination of severely ill me/cfs patients. *Healthcare (Switzerland)*, *9*(10). https://doi.org/10.3390/healthcare9101290

Fletcher, M. A., Zeng, X. R., Maher, K., Levis, S., Hurwitz, B., Antoni, M., Broderick, G., & Klimas, N. G. (2010). Biomarkers in chronic fatigue syndrome: Evaluation of natural killer cell function and dipeptidyl peptidase IV/CD26. *PLoS ONE*, *5*(5). https://doi.org/10.1371/journal.pone.0010817

Fukuda, K., Straus, S. E., Hickie, I., Sharpe, M. C., Dobbins, J. G., & Komaroff, A. (1994). The chronic fatigue syndrome: a comprehensive approach to its definition and study. International Chronic Fatigue Syndrome Study Group [see comments]. *Annals of Internal Medicine*, *121*(12).

García-Álvarez, M., Alonso-Álvarez, S., Prieto-Conde, I., Jiménez, C., Sarasquete, M. E., Chillón, M. C., Medina, A., Balanzategui, A., Maldonado, R., Antón, A., Puig, N., Rodríguez, M., Blanco, O., Tamayo, P., González-Calle, V., Martín, A., García-Sanz, R., González, M., Caballero, M. D., & Alcoceba, M. (2019). Immunoglobulin gene rearrangement IGHV3-48 is a predictive marker of histological transformation into aggressive lymphoma in follicular lymphomas. *Blood Cancer Journal*, *9*(7). https://doi.org/10.1038/s41408-019-0213-9

Germain, A., Barupal, D. K., Levine, S. M., & Hanson, M. R. (2020). Comprehensive circulatory metabolomics in ME/CFS reveals disrupted metabolism of Acyl lipids and steroids. *Metabolites*, *10*(1). https://doi.org/10.3390/metabo10010034

Germain, A., Levine, S. M., & Hanson, M. R. (2021). In-depth analysis of the plasma proteome in ME/CFS exposes disrupted ephrin-eph and immune system signaling. *Proteomes*, *9*(1), 1–19. https://doi.org/10.3390/PROTEOMES9010006

Germain, A., Ruppert, D., Levine, S. M., & Hanson, M. R. (2017). Metabolic profiling of a myalgic encephalomyelitis/chronic fatigue syndrome discovery cohort reveals disturbances in fatty acid and lipid metabolism. *Molecular BioSystems*, *13*(2), 371–379. https://doi.org/10.1039/c6mb00600k

Germain, A., Ruppert, D., Levine, S. M., & Hanson, M. R. (2018). Prospective biomarkers from plasma metabolomics of myalgic encephalomyelitis/chronic fatigue syndrome implicate redox imbalance in disease symptomatology. *Metabolites*, *8*(4). https://doi.org/10.3390/metabo8040090

Gorman, G. S., Chinnery, P. F., DiMauro, S., Hirano, M., Koga, Y., McFarland, R., Suomalainen, A., Thorburn, D. R., Zeviani, M., & Turnbull, D. M. (2016). Mitochondrial diseases. *Nature Reviews Disease Primers*, *2*. https://doi.org/10.1038/nrdp.2016.80

Guo, X., Wang, Y., Zhang, R., & Gu, Z. (2020). STAMP: a multiplex sequencing method for simultaneous evaluation of mitochondrial DNA heteroplasmies and content. *NAR Genomics and Bioinformatics*, *2*(4). https://doi.org/10.1093/nargab/lqaa065

Hanson, M. R., & Germain, A. (2020). Letter to the editor of metabolites. In *Metabolites* (Vol. 10, Issue 5). MDPI AG. https://doi.org/10.3390/metabo10050216

Hayden, M. S., & Ghosh, S. (2014). Regulation of NF-κB by TNF family cytokines. In *Seminars in Immunology* (Vol. 26, Issue 3, pp. 253–266). Academic Press. https://doi.org/10.1016/j.smim.2014.05.004

Hoel, F., Hoel, A., Pettersen, I. K. N., Rekeland, I. G., Risa, K., Alme, K., Sørland, K., Fosså, A., Lien, K., Herder, I., Thürmer, H. L., Gotaas, M. E., Schäfer, C., Berge, R. K., Sommerfelt, K., Marti, H. P., Dahl, O., Mella, O., Fluge, Ø., & Tronstad, K. J. (2021). A map of metabolic phenotypes in patients with myalgic encephalomyelitis/ chronic fatigue syndrome. *JCI Insight*, *6*(16). https://doi.org/10.1172/jci.insight.149217

Holden, S., Maksoud, R., Eaton-Fitch, N., Cabanas, H., Staines, D., & Marshall-Gradisnik, S. (2020). A systematic review of mitochondrial abnormalities in myalgic encephalomyelitis/chronic fatigue syndrome/systemic exertion intolerance disease. In *Journal of Translational Medicine* (Vol. 18, Issue 1). https://doi.org/10.1186/s12967-020-02452-3

Holt, I. J., Harding, A. E., & Morgan-Hughes, J. A. (1988). Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies. *Nature*, *331*(6158). https://doi.org/10.1038/331717a0

Howell', N. (1999). *Answering Questions and Questioning Answers*.

Jones, D. E. J., Hollingsworth, K. G., Taylor, R., Blamire, A. M., & Newton, J. L. (2010). Abnormalities in pH handling by peripheral muscle and potential regulation by the autonomic nervous system in chronic fatigue syndrome. *Journal of Internal Medicine*, *267*(4). https://doi.org/10.1111/j.1365-2796.2009.02160.x

Kashi, A. A., Davis, R. W., & Phair, R. D. (2019). The IDO metabolic trap hypothesis for the etiology of ME/CFS. *Diagnostics*, *9*(3). https://doi.org/10.3390/diagnostics9030082

Kaur, H., & Shorie, M. (2019). Nanomaterial based aptasensors for clinical and environmental diagnostic applications. In *Nanoscale Advances* (Vol. 1, Issue 6, pp. 2123–2138). Royal Society of Chemistry. https://doi.org/10.1039/c9na00153k

Kaushik, N., Fear, D., Richards, S. C. M., McDermott, C. R., Nuwaysir, E. F., Kellam, P., Harrison, T. J., Wilkinson, R. J., Tyrrell, D. A. J., Holgate, S. T., & Kerr, J. R. (2005). Gene expression in peripheral blood mononuclear cells from patients with chronic fatigue syndrome. *Journal of Clinical Pathology*, *58*(8), 826–832. https://doi.org/10.1136/jcp.2005.025718

Kerr, J. R., BCh, M., & St George, Frcp. (2008). *Gene Profi ling of Patients with Chronic Fatigue Syndrome/Myalgic Encephalomyelitis*.

Kerr, J. R., Petty, R., Burke, B., Gough, J., Fear, D., Sinclair, L. I., Mattey, D. L., Richards, S. C. M., Montgomery, J., Baldwin, D. A., Kellam, P., Harrison, T. J., Griffin, G. E., Main, J., Enlander, D., Nutt, D. J., & Holgate, S. T. (2008). Gene expression subtypes in patients with chronic fatigue syndrome/myalgic encephalomyelitis. *Journal of Infectious Diseases*, *197*(8), 1171–1184. https://doi.org/10.1086/533453

King, D. (2020). *PBMCs a powerful research tool*. https://www.allcells.com/pbmcs-a-powerful-research-tool/

Lane, N. (2005). *Power, Sex, Suicide; Mitochondria and the Meaning of Life*.

Lane, R. J. M., Barrett, M. C., Taylor, D. J., Kemp, G. J., & Lodi, R. (1998). *Heterogeneity in chronic fatigue syndrome: evidence from magnetic resonance spectroscopy of muscle*.

Lane, R. J. M., Barrett, M. C., Woodrow, D., Moss, J., Fletcher, R., & Archard, L. C. (1998). Muscle fibre characteristics and lactate responses to exercise in chronic fatigue syndrome. *Journal of Neurology Neurosurgery and Psychiatry*, *64*(3), 362–367. https://doi.org/10.1136/jnnp.64.3.362

Lant, B., & Storey, K. B. (2010). An Overview of Stress Response and Hypometabolic Strategies in Caenor-habditis elegans: Conserved and Contrasting Signals with the Mammalian System. In *Int. J. Biol. Sci* (Vol. 6, Issue 1). http://www.biolsci.org9

Lee, W. T., Sun, X., Tsai, T. S., Johnson, J. L., Gould, J. A., Garama, D. J., Gough, D. J., McKenzie, M., Trounce, I. A., & st. John, J. C. (2017). Mitochondrial DNA haplotypes induce differential patterns of DNA methylation that result in differential chromosomal gene expression patterns. *Cell Death Discovery*, *3*. https://doi.org/10.1038/cddiscovery.2017.62

Light, K. C., Agarwal, N., Iacob, E., White, A. T., Kinney, A. Y., VanHaitsma, T. A., Aizad, H., Hughen, R. W., Bateman, L., & Light, A. R. (2013). Differing leukocyte gene expression profiles associated with fatigue in patients with prostate cancer versus chronic fatigue syndrome. *Psychoneuroendocrinology*, *38*(12), 2983–2995. https://doi.org/10.1016/j.psyneuen.2013.08.008

Litwack, G. (2018). Metabolism of Fat, Carbohydrate, and Nucleic Acids. In *Human Biochemistry*. https://doi.org/10.1016/b978-0-12-383864-3.00014-4

Lowrie, E. G., Curtin, R. B., LePain, N., & Schatell, D. (2003). Medical Outcomes Study Short Form-36: A consistent and powerful predictor of morbidity and mortality in dialysis patients. *American Journal of Kidney Diseases*, *41*(6), 1286–1292. https://doi.org/10.1016/S0272-6386(03)00361-5

Mandarano, A. H., Maya, J., Giloteaux, L., Peterson, D. L., Maynard, M., Gottschalk, C. G., & Hanson, M. R. (2020). Myalgic encephalomyelitis/chronic fatigue syndrome patients exhibit altered T cell

metabolism and cytokine associations. *Journal of Clinical Investigation*, *130*(3), 1491–1505. https://doi.org/10.1172/JCI132185

Marshall, E. K. (1911). *REID HUNT 1870-1948*.

Mathew, S. J., Mao, X., Keegan, K. A., Levine, S. M., Smith, E. L. P., Heier, L. A., Otcheretko, V., Coplan, J. D., & Shungu, D. C. (2009). Ventricular cerebrospinal fluid lactate is increased in chronic fatigue syndrome compared with generalized anxiety disorder: An in vivo 3.0 T 1H MRS imaging study. *NMR in Biomedicine*, *22*(3). https://doi.org/10.1002/nbm.1315

Milivojevic, M., Che, X., Bateman, L., Cheng, A., Garcia, B. A., Hornig, M., Huber, M., Klimas, N. G., Lee, B., Lee, H., Levine, S., Montoya, J. G., Peterson, D. L., Komaroff, A. L., & Lipkin, W. I. (2020). Plasma proteomic profiling suggests an association between antigen driven clonal B cell expansion and ME/CFS. *PLoS ONE*, *15*(7). https://doi.org/10.1371/journal.pone.0236148

Missailidis, D., Annesley, S. J., Allan, C. Y., Sanislav, O., Lidbury, B. A., Lewis, D. P., & Fisher, P. R. (2020). An isolated complex v inefficiency and dysregulated mitochondrial function in immortalized lymphocytes from ME/CFS patients. *International Journal of Molecular Sciences*, *21*(3). https://doi.org/10.3390/ijms21031074

Missailidis, D., Sanislav, O., Allan, C. Y., Annesley, S. J., & Fisher, P. R. (2020). Cell-based blood biomarkers for myalgic encephalomyelitis/chronic fatigue syndrome. *International Journal of Molecular Sciences*, *21*(3). https://doi.org/10.3390/ijms21031142

Morris, G., & Maes, M. (2012). Increased nuclear factor-κB and loss of p53 are key mechanisms in Myalgic Encephalomyelitis/chronic fatigue syndrome (ME/CFS). *Medical Hypotheses*, *79*(5). https://doi.org/10.1016/j.mehy.2012.07.034

Morris, G., & Maes, M. (2013). Myalgic encephalomyelitis/chronic fatigue syndrome and encephalomyelitis disseminata/multiple sclerosis show remarkable levels of similarity in phenomenology and neuroimmune characteristics. *BMC Medicine*, *11*(1). https://doi.org/10.1186/1741-7015-11-205

Morris, G., & Maes, M. (2014). Mitochondrial dysfunctions in Myalgic Encephalomyelitis/chronic fatigue syndrome explained by activated immuno-inflammatory, oxidative and nitrosative stress pathways. In *Metabolic Brain Disease* (Vol. 29, Issue 1, pp. 19–36). Springer New York LLC. https://doi.org/10.1007/s11011-013-9435-x

Murrough, J. W., Mao, X., Collins, K. A., Kelly, C., Andrade, G., Nestadt, P., Levine, S. M., Mathew, S. J., & Shungu, D. C. (2010). Increased ventricular lactate in chronic fatigue syndrome measured by 1H MRS imaging at 3.0 T. II: Comparison with major depressive disorder. *NMR in Biomedicine*, *23*(6). https://doi.org/10.1002/nbm.1512

Nagy-Szakal, D., Barupal, D. K., Lee, B., Che, X., Williams, B. L., Kahn, E. J. R., Ukaigwe, J. E., Bateman, L., Klimas, N. G., Komaroff, A. L., Levine, S., Montoya, J. G., Peterson, D. L., Levin, B., Hornig, M., Fiehn, O., & Lipkin, W. I. (2018). Insights into myalgic encephalomyelitis/chronic fatigue syndrome phenotypes through comprehensive metabolomics. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-28477-9

Natelson, B. H., Mao, X., Stegner, A. J., Lange, G., Vu, D., Blate, M., Kang, G., Soto, E., Kapusuz, T., & Shungu, D. C. (2017). Multimodal and simultaneous assessments of brain and spinal fluid abnormalities in chronic fatigue syndrome and the effects of psychiatric comorbidity. *Journal of the Neurological Sciences*, *375*. https://doi.org/10.1016/j.jns.2017.02.046

Natelson, B. H., Vu, D., Coplan, J. D., Mao, X., Blate, M., Kang, G., Soto, E., Kapusuz, T., & Shungu, D. C. (2017). Elevations of ventricular lactate levels occur in both chronic fatigue syndrome and fibromyalgia. *Fatigue: Biomedicine, Health and Behavior*, *5*(1). https://doi.org/10.1080/21641846.2017.1280114

Naviaux, R. K., Naviaux, J. C., Li, K., Bright, A. T., Alaynick, W. A., Wang, L., Baxter, A., Nathan, N., Anderson, W., & Gordon, E. (2016). Metabolic features of chronic fatigue syndrome. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(37), E5472–E5480. https://doi.org/10.1073/pnas.1607571113

Nguyen, T., Staines, D., Johnston, S., & Marshall-Gradisnik, S. (2019). Reduced glycolytic reserve in isolated natural killer cells from myalgic encephalomyelitis/chronic fatigue syndrome patients: A preliminary investigation. *Asian Pacific Journal of Allergy and Immunology*, *37*(2), 102–108. https://doi.org/10.12932/AP-011117-0188

Oliveira, P. J. (2018). Mitochondrial biology and experimental therapeutics. In *Mitochondrial Biology and Experimental Therapeutics*. https://doi.org/10.1007/978-3-319-73344-9

O'Neal, A. J., & Hanson, M. R. (2021). The Enterovirus Theory of Disease Etiology in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: A Critical Review. In *Frontiers in Medicine* (Vol. 8). https://doi.org/10.3389/fmed.2021.688486

Översti, S., Majander, K., Salmela, E., Salo, K., Arppe, L., Belskiy, S., Etu-Sihvola, H., Laakso, V., Mikkola, E., Pfrengle, S., Putkonen, M., Taavitsainen, J. P., Vuoristo, K., Wessman, A., Sajantila, A., Oinonen, M., Haak, W., Schuenemann, V. J., Krause, J., … Onkamo, P. (2019). Human mitochondrial DNA lineages in Iron-Age Fennoscandia suggest incipient admixture and eastern introduction of farming-related maternal ancestry. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-51045-8

Paul, B. D., Lemle, M. D., Komaroff, A. L., & Snyder, S. H. (2021). Redox imbalance links COVID-19 and myalgic encephalomyelitis/chronic fatigue syndrome. In *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 118, Issue 34). https://doi.org/10.1073/pnas.2024358118

Porter, K., Lin, Y., & Liton, P. B. (2013). Cathepsin B Is Up-Regulated and Mediates Extracellular Matrix Degradation in Trabecular Meshwork Cells Following Phagocytic Challenge. *PLoS ONE*, *8*(7). https://doi.org/10.1371/journal.pone.0068668

Powell, R., Renw, J., Lewith, G., Barclayw, W., Holgate, S., & Almondz, J. (2003). *Identification of novel expressed sequences, up-regulated in the leucocytes of chronic fatigue syndrome patients*. www.appliedbiosystems.com/support/

Ravi Rupesh Kanchi and Walton, K. and K. M. (2018). MiSeq: A Next Generation Sequencing Platform for Genomic Analysis. In J. K. DiStefano (Ed.), *Disease Gene Identification: Methods and Protocols* (pp. 223–232). Springer New York. https://doi.org/10.1007/978-1-4939-7471-9_12

Richards, R. S., Roberts, T. K., McGregor, N. R., Dunstan, R. H., & Butt, H. L. (2000). Blood parameters indicative of oxidative stress are associated with symptom expression in chronic fatigue syndrome. *Redox Report*, *5*(1). https://doi.org/10.1179/rer.2000.5.1.35

Russell, O. M., Gorman, G. S., Lightowlers, R. N., & Turnbull, D. M. (2020). Mitochondrial Diseases: Hope for the Future. In *Cell* (Vol. 181, Issue 1, pp. 168–188). Cell Press. https://doi.org/10.1016/j.cell.2020.02.051

Shungu, D. C., Weiduschat, N., Murrough, J. W., Mao, X., Pillemer, S., Dyke, J. P., Medow, M. S., Natelson, B. H., Stewart, J. M., & Mathew, S. J. (2012). Increased ventricular lactate in chronic fatigue syndrome. III. Relationships to cortical glutathione and clinical symptoms implicate oxidative stress in disorder pathophysiology. *NMR in Biomedicine*, *25*(9), 1073–1087. https://doi.org/10.1002/nbm.2772

Siegel, Z. A., Brown, A., Devendorf, A., Collier, J., & Jason, L. A. (2018). A content analysis of chronic fatigue syndrome and myalgic encephalomyelitis in the news from 1987 to 2013. *Chronic Illness*, *14*(1), 3–12. https://doi.org/10.1177/1742395317703175

Starkov, A. A. (2008). The role of mitochondria in reactive oxygen species metabolism and signaling. *Annals of the New York Academy of Sciences*, *1147*. https://doi.org/10.1196/annals.1427.015

Sweetman, E., Kleffmann, T., Edgar, C., de Lange, M., Vallings, R., & Tate, W. (2020). A SWATH-MS analysis of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome peripheral blood mononuclear cell proteomes reveals mitochondrial dysfunction. *Journal of Translational Medicine*, *18*(1). https://doi.org/10.1186/s12967-020-02533-3

Sweetman, E., Ryan, M., Edgar, C., Mackay, A., Vallings, R., & Tate, W. (2019). Changes in the transcriptome of circulating immune cells of a New Zealand cohort with myalgic encephalomyelitis/chronic fatigue syndrome. *International Journal of Immunopathology and Pharmacology*, *33*. https://doi.org/10.1177/2058738418820402

Technologies, A. (2018). *Seahorse XFe Analyzer Operating Manual*.

Tomas, C., Brown, A., Strassheim, V., Elson, J., Newton, J., & Manning, P. (2017). Cellular bioenergetics is impaired in patients with chronic fatigue syndrome. *PLoS ONE*, *12*(10). https://doi.org/10.1371/journal.pone.0186802

Underhill, R. A. (2015). Myalgic encephalomyelitis, chronic fatigue syndrome: An infectious disease. *Medical Hypotheses*, *85*(6). https://doi.org/10.1016/j.mehy.2015.10.011

Valdez, A. R., Hancock, E. E., Adebayo, S., Kiernicki, D. J., Proskauer, D., Attewell, J. R., Bateman, L., DeMaria, A., Lapp, C. W., Rowe, P. C., & Proskauer, C. (2019). Estimating prevalence, demographics, and costs of ME/CFS using large scale medical claims data and machine learning. *Frontiers in Pediatrics*, *6*(JAN). https://doi.org/10.3389/fped.2018.00412

Venter, M., Tomas, C., Pienaar, I. S., Strassheim, V., Erasmus, E., Ng, W. F., Howell, N., Newton, J. L., van der Westhuizen, F. H., & Elson, J. L. (2019). MtDNA population variation in Myalgic encephalomyelitis/Chronic fatigue syndrome in two populations: a study of mildly deleterious variants. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-39060-1

Vermeulen, R. C. W., & Vermeulen van Eck, I. W. G. (2014). Decreased oxygen extraction during cardiopulmonary exercise test in patients with chronic fatigue syndrome. *Journal of Translational Medicine*, *12*(1). https://doi.org/10.1186/1479-5876-12-20

Vernon, S. D., Unger, E. R., Dimulescu, I. M., Rajeevan, M., & Reeves, W. C. (2002). Utility of the blood for gene expression profiling and biomarker discovery in chronic fatigue syndrome. In *Disease Markers* (Vol. 18). IOS Press.

Wallace, D. C., Singh, G., Lott, M. T., Hodge, J. A., Schurr, T. G., Lezza, A. M. S., Elsas, L. J., & Nikoskelainen, E. K. (1988). Mitochondrial DNA Mutation Associated with Leber's Hereditary Optic Neuropathy. In *New Series* (Vol. 242, Issue 4884).

Wang, Y. E., Marinov, G. K., Wold, B. J., & Chan, D. C. (2013). Genome-Wide Analysis Reveals Coating of the Mitochondrial Genome by TFAM. *PLoS ONE*, *8*(8). https://doi.org/10.1371/journal.pone.0074513

West, A. P., Shadel, G. S., & Ghosh, S. (2011). Mitochondria in innate immune responses. In *Nature Reviews Immunology* (Vol. 11, Issue 6). https://doi.org/10.1038/nri2975

Whistler, T., Unger, E. R., Nisenbaum, R., & Vernon, S. D. (2003). *Integration of gene expression, clinical, and epidemiologic data to characterize Chronic Fatigue Syndrome*. http://www.translational-medicine.com/content/1/1/10

Wong, R., Lopaschuk, G. ;, Gang, Z., Walker, D., Catellier, D., Burton, D., Teo, K., Collins-Nakai, R., & Montague, T. (1992). *Skeletal Muscle Metabolism in the Chronic Fatigue Syndrome* In Vivo Assessment by 31 P Nuclear Magnetic Resonance Spectroscopy*.

Wood, E., Hall, K. H., & Tate, W. (2021). Role of mitochondria, oxidative stress and the response to antioxidants in myalgic encephalomyelitis/chronic fatigue syndrome: A possible approach to SARS-CoV-2 'long-haulers'? In *Chronic Diseases and Translational Medicine* (Vol. 7, Issue 1, pp. 14–26). KeAi Communications Co. https://doi.org/10.1016/j.cdtm.2020.11.002

Yamano, E., Sugimoto, M., Hirayama, A., Kume, S., Yamato, M., Jin, G., Tajima, S., Goda, N., Iwai, K., Fukuda, S., Yamaguti, K., Kuratsune, H., Soga, T., Watanabe, Y., & Kataoka, Y. (2016). Index markers of chronic fatigue syndrome with dysfunction of TCA and urea cycles. *Scientific Reports*, *6*. https://doi.org/10.1038/srep34990

Yeung, A. W. S., Terentis, A. C., King, N. J. C., & Thomas, S. R. (2015). Role of indoleamine 2,3-dioxygenase in health and disease. In *Clinical Science* (Vol. 129, Issue 7, pp. 601–672). Portland Press Ltd. https://doi.org/10.1042/CS20140392

APPENDIX A

Primary experiment:
Aim: To use Sequencing by Targeted Amplification of Multiplex Probes (STAMP) to perform mitochondrial DNA library synthesis for sequencing ME/CFS PBMC patient and healthy control samples.
The goal of the experiment was to use STAMP- an accurate, low error and low-cost multiplex sequencing technology targeted to amplify mitochondrial DNA (Guo et al., 2020)- to sequence the mitochondrial genome of 193 ME/CFS PBMC patient samples, age, and gender-matched with 196 healthy control samples obtained from the Hanson Lab (same cohort as described in (Billing-Ross et al., 2016); see section "Role of mitochondrial mutations in ME/CFS").

Materials and Methods
Sample Prep
The samples were transferred from their original plates to new 96 well plates wherein the concentration of all the samples was adjusted to 50 ng/µl. These plates with normalized concentrations were then used to set up the "reaction plates" for the STAMP protocol to be applied. The layout and sample condition of samples of the reaction plates has been shown in tables A2-A15; each plate has genomic DNA samples isolated from HEK293T cells serving as controls. In reaction plates 1, 2R, 3R, 4 and 5 control genomic DNA samples were aliquoted to H4 and H5 wells (Figure A1,A3,A5,A6, and A7; tables A2,A3,A6,A7,A10,A11,A12,A13,A14 and A15) The reason for this was the fact that the control genomic DNA samples if aliquoted to the last well of any plate would be susceptible to evaporation and loss of the control sample from the plate, as was the case for reaction plates 2 and 3 (Figures A2 and A4). The STAMP protocol was repeated for reaction plates 2 and 3 hence giving rise to reaction plates 2R and 3R (Figures A2-A5).

Special note on Reaction Plate 5:
Reaction plate 5 had all the samples which had not been tested before in row B- from B1 to B12 (PID:163,235,242,270,313,329,332,347,349,358,401,005). Also, it had samples from the corners of each reaction plate (1,2=2R,3=3R, and 4) namely samples from wells A1, A2, A11, A12, H1, H2, H11, H12 in case these were lost due to evaporation. It also contained samples that were replaced due to genomic DNA control samples in the other reaction plates -namely H4 and H5 of reaction plates 1,2R,3R and 4 (Tables A2-A15).
Having all the reaction plates prepared in this manner, the STAMP protocol was applied.

The process involved three main steps:
Step 1: Hybridization
In this step 46 pairs of single strand Extension Ligation (EL) probe pairs are hybridized with the mitochondrial DNA to capture the entire 16.6 kb circular human mitochondrial DNA. At the same time 5 probe pairs that target nuclear DNA are also included (this is useful to quantify the mtDNA copy number across samples). The reaction mixture has a volume of 8µl including Ampligase buffer at 1X volume and test/control genomic DNA template at 150ng. Thermal conditions were as follows:

Cycle Step	Temp	Duration
Denaturation	95 °C	10 min

Hybridization	55 °C	20 hours

Step 2: Gap filling
As the name suggests this step fills the space between the probe pairs to produce a fragment of captured mtDNA having a unique molecular barcode(Guo et al., 2020). The gap filling mix was prepared as follows:

Nuclease free water	3.45 µl
10X Ampligase buffer	0.6 µl
Betaine(4M)	1 µl
(NH4)2SO4 (1M)	0.7 µl
dNTPs (10 mM)	0.05 µl
Ampligase(5U/µl)	0.1 µl
Tsp DNA Polymerase(5U/µl)	0.1 µl

The entire reaction mixture was incubated without including Ampligase, at 25 °C for 1 minute followed by 40 °C to deactivate antibodies present in Tsp DNA Polymerase. A total of 6 µl gap-filling mix was added to the reaction mixture from Step 1. The resulting reaction mixture of 14 µl was incubated at 55°C for another 20 hours for gap filling.

Step 3: Amplification of captured targets.
Next the captured targets were amplified using indexing primers that comprised P5 and P7 Illumina adapter sequences, an 8-nt index sequence, a 13- or 14-nt primer adaptor sequence, and a universal sequence designed at the 3' terminus of EL probe. The tables A2-A15 show the primer pair combination/layout used and tables.

2 µl of the captured product was used to set up a 25 µl reaction as follows:

Component	Volume (1X)
NF water	12.25 µl
5X Phusion HF buffer	5 µl
10mM dNTPs	0.5 µl
Phusion Hot Start II DNA Polymerase	0.25 µl
Captured DNA	2 µl
I7 index primer (5 µM)	2.5 µl
I5 index primer (5 µM)	2.5 µl

The size and integrity of PCR products were visually verified by agarose gel electrophoresis. The gels showcased below show the images of the amplified libraries of a subset of samples from each reaction plate (Figures A1-A7); PCR products of an effective gap-filling reaction using control genomic DNA as a template show a band centered at about 550bp-600bp. Pilot studies using genomic DNA samples were performed to validate approach in experimenter's hands/practice (Figure A8).

Step 4. Library purification

The PCR products resulting from the library amplification step were pooled and purified using Ampure XP magnetic beads. 5µl from each well of any given 96 well reaction plate, were pooled together following which double size selection (Beckman Coulter, Inc.) using the beads was performed. In brief, the beads were first brought to room temperature and 0.25 volume of beads was used to bind >700bp of the PCR products following which the supernatant was transferred to a fresh tube. -Now 0.4 volume of beads was added to the supernatant to bind fragments smaller than 500 bp. Following this the beads were washed with 70 per cent alcohol, dried and treated with 20µl of nuclease free water twice in series to maximize yield. A total of 40µl of pooled library sample was obtained, which was gel purified using Monarch gel purification kit quantified using Qubit® 2.0 Fluorometer (Life Technologies, Inc.). Gel electrophoresis was used to check if the size of bead purified libraries was correct (Figure A9) and concentrations of gel purified libraries are depicted in Table 1. Each pooled library represents one reaction plate so there are a total of 7 library samples representing reaction plates 1, 2, 3, 4, 5, 2R and 3R; all gels were 1 % agarose. Every sample on every plate has a unique combination of primer pairs (Tables A2-A15)
Libraries can be sequenced by Illumina Mi Seq 2 x 125bp paired end sequencing.

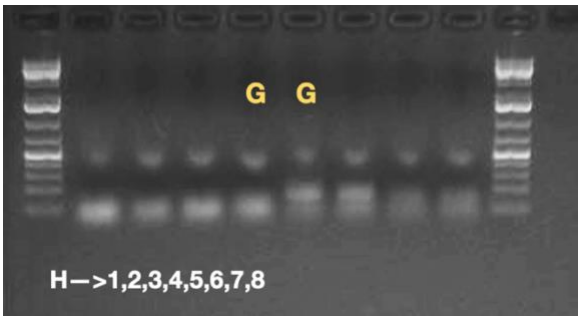


Figure A1. Subset of samples from Reaction plate 1: The gel image showcases the amplified libraries of samples in wells H1 to H8 in reaction plate 1 (Tables A2 and A3). The lanes labelled with G indicate control genomic DNA samples from HEK293T cells. As can be seen the bands shown are of the expected band size for each sample (550bp-600bp). 1 kb plus DNA ladder was used.

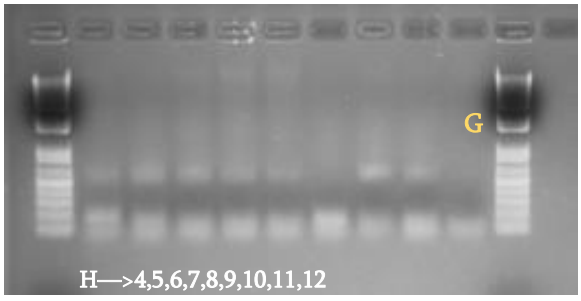


Figure A2. Subset of samples from Reaction plate 2: The gel image showcases the amplified libraries of samples in wells H1 to H12 in reaction plate 2 (Tables A4 and A5). The lane labelled with G indicate control genomic DNA samples from HEK293T cells. As can be seen, the control genomic DNA sample was lost and the STAMP process was redone for this plate (Figure A3, Tables A6 and A7). As can be seen the bands shown are of the expected band size for each sample (550bp-600bp). 1 kb plus DNA ladder was used.

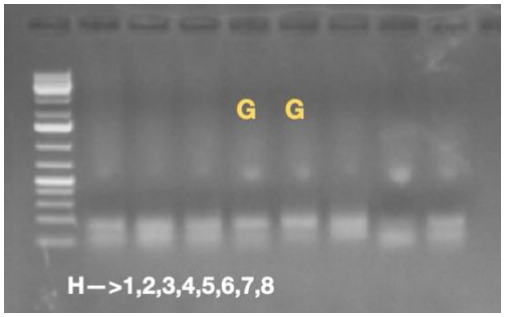


Figure A3. Subset of samples from Reaction plate 2R: The gel image showcases the amplified libraries of samples in wells H1 to H8 in reaction plate 2R (Tables A6 and A7). The lanes labelled with G indicate control genomic DNA samples from HEK293T cells. This includes a repeat of the STAMP procedure for reaction plate 2 (Figure A2, Tables A4 and A5). As can be seen the bands shown are of the expected band size for each sample (550bp-600bp). 1 kb plus DNA ladder was used.

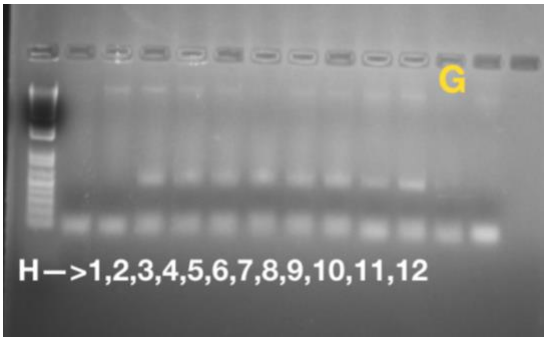


Figure A4. Subset of samples from Reaction plate 3: The gel image showcases the amplified libraries of samples in wells H1 to H12 in reaction plate 3 (Tables A8 and A9). The lanes labelled with G indicate control genomic DNA samples from HEK293T cells. As can be seen the control genomic DNA sample was lost and the STAMP process was redone for this plate (Figure A5, Tables A10 and A11). As can be seen the bands shown are of the expected band size for each sample (550bp-600bp). 1 kb plus DNA ladder was used.

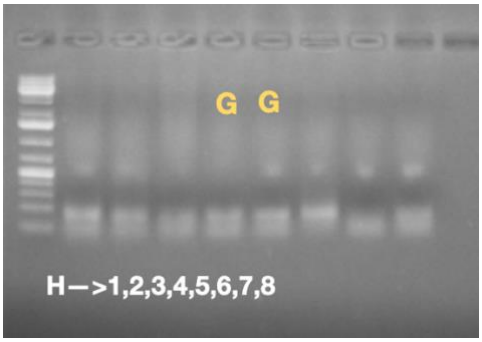


Figure A5. Subset of samples from Reaction plate 3R: The gel image showcases the amplified libraries of samples in wells H1 to H8 in reaction plate 3R (Tables A10 and A11). The lanes labelled with G indicate control genomic DNA samples from HEK293T cells. This includes a repeat of the STAMP procedure for reaction plate 2 (Figure A2, Tables A4 and A5). As can be seen the bands shown are of the expected band size for each sample (550bp-600bp). 1 kb plus DNA ladder was used.

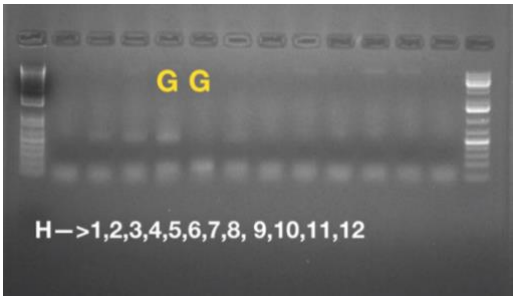


Figure A6. Subset of samples from Reaction plate 4: The gel image showcases the amplified libraries of samples in wells H1 to H12 in reaction plate 4 (Tables A12 and A13). The lanes labelled with G indicate control genomic DNA samples from HEK293T cells. As can be seen the bands shown are of the expected band size for each sample (550bp-600bp). 1 kb plus DNA ladder was used.

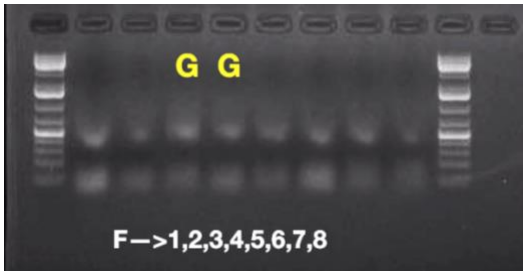


Figure A7. Subset of samples from Reaction plate 5: The gel image showcases the amplified libraries of samples in wells F1 to F8 in reaction plate 5 (Tables A14 and A15). The lanes labelled with G indicate control genomic DNA samples from HEK293T cells. As can be seen the bands shown are of the expected band size for each sample (550bp-600bp). 1 kb plus DNA ladder was used.

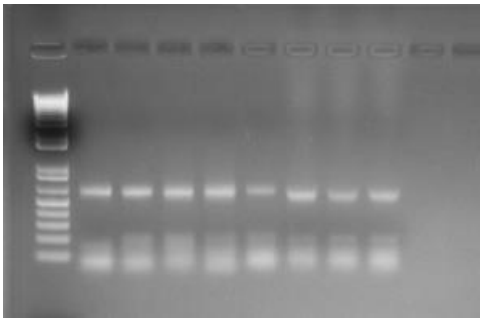


Figure A8. Amplified libraries using control genomic DNA samples as template: The gel image showcases the amplified libraries of a genomic DNA sample extracted from HEK293T cells. Each well represents a reaction prepared individually using the same genomic DNA. As can be seen the bands shown are of the expected band size for each sample (550bp-600bp). I kb plus DNA ladder was used.

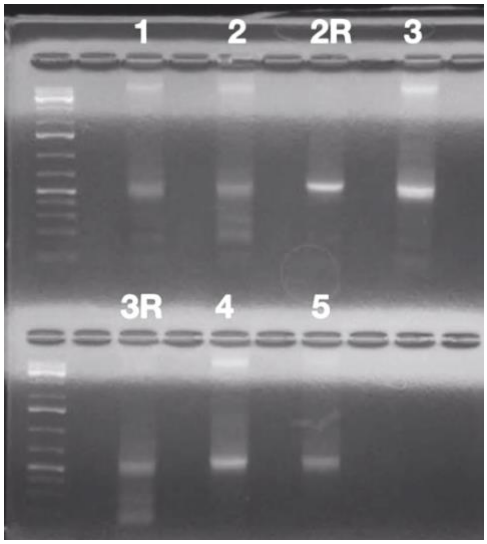


Figure A9. Purified libraries of pooled samples from each Reaction plate. This figure shows the bands for AMPure XP magnetic bead purified libraries. The samples 1, 2, 2R, 3, 3R, 4 and 5 represent pooled and purified library samples from the reaction plates 1, 2, 2R, 3, 3R, 4 and 5 respectively. The bands are of the desired size (550bp-600bp) and were gel purified and then quantified by Qubit® 2.0.

Table A1. Library concentrations using Qubit® 2.0 after gel purification.

Pooled library sample	Concentration reading on Qubit (ng/mL)	Final concentration of library (ng/µl)
Reaction plate 1(all reaction plates have 96 samples)	22.1	4.42
Reaction plate 2	31.3	6.26
Reaction plate 2-repeat	53.7	10.6
Reaction plate 3	79.4	15.88
Reaction plate 3-repeat	87.5	17.5
Reaction plate 4	42.9	8.58
Reaction plate 5(*it has less samples - 57)	25.6	5.12

Table A2. CFI cohort sample layout in the 96 well plate Reaction plate 1. This table depicts a top view of the samples loaded in reaction plate 1 and shows the Illumina P5, P7 primer pair used (from lab stock)

P7 primer →		737	738	739	740	741	742	743	744	745	746	747	748
P5 primer ↓		1	2	3	4	5	6	7	8	9	10	11	12
533	A	006	101	136	164	177	250	293	338	367	394	403	404
534	B	406	410	411	106	402	405	408	409	414	415	421	412
535	C	416	417	419	420	422	423	424	425	426	428	430	431
536	D	434	418	429	432	433	407B	079	354	020	040	012	112
545	E	337	053	081	051	146	254	226	044	302	386	050	343
546	F	022	126	183	198	203	340	348	059	086	120	173	227
547	G	142	239	370	083	127	186	195	259	285	317	384	121
548	H	149	266	286	G.DNA CONTROL	G.DNA CONTROL	030	038	065	080	088	092	125

Table A3. Sample list, template concentration used, Illumina P5, P7 primer pair (from lab stock) used and condition (0=control, 1= patient) in Reaction plate 1.

Well id	PID/ID1	Concentration in. plate (ng/µl)	Condition (0 = control, 1 = patient)	Illumina P5, P7 primer pair (from lab stock)
A1	006	150	0	533,737
A2	101	150	0	533,738
A3	136	150	0	533,739
A4	164	150	1	533,740
A5	177	150	0	533,741
A6	250	150	0	533,742
A7	293	150	1	533,743
A8	338	150	0	533,744
A9	367	150	0	533,745
A10	394	150	0	533,746
A11	403	150	0	533,747
A12	404	150	0	533,748
B1	406	150	0	534,737
B2	410	150	0	534,738
B3	411	150	1	534,739
B4	106	150	1	534,740
B5	402	150	1	534,741
B6	405	150	1	534,742
B7	408	150	1	534,743
B8	409	150	0	534,744
B9	414	150	0	534,745
B10	415	150	1	534,746
B11	421	150	1	534,747
B12	412	150	1	534,748
C1	416	150	0	535,737
C2	417	150	1	535,738
C3	419	150	0	535,739
C4	420	150	0	535,740
C5	422	150	1	535,741
C6	423	150	0	535,742
C7	424	150	1	535,743
C8	425	150	0	535,744
C9	426	150	0	535,745
C10	428	150	0	535,746
C11	430	150	0	535,747
C12	431	150	0	535,748
D1	434	150	1	536,737
D2	418	150	0	536,738
D3	429	150	0	536,739

D4	432	150	0	536,740
D5	433	150	1	536,741
D6	407B	150	1	536,742
D7	079	150	1	536,743
D8	354	150	0	536,744
D9	020	150	1	536,745
D10	040	150	1	536,746
D11	012	150	0	536,747
D12	112	150	1	536,748
E1	337	150	1	545,737
E2	053	150	1	545,738
E3	081	150	1	545,739
E4	051	150	0	545,740
E5	146	150	1	545,741
E6	254	150	1	545,742
E7	226	150	1	545,743
E8	044	150	0	545,744
E9	302	150	1	545,745
E10	386	150	1	545,746
E11	050	150	0	545,747
E12	343	150	1	545,748
F1	022	150	0	546,737
F2	126	150	0	546,738
F3	183	150	1	546,739
F4	198	150	1	546,740
F5	203	150	1	546,741
F6	340	150	0	546,742
F7	348	150	0	546,743
F8	059	150	0	546,744
F9	086	150	1	546,745
F10	120	150	1	546,746
F11	173	150	1	546,747
F12	227	150	1	546,748
G1	142	150	1	547,737
G2	239	150	1	547,738
G3	370	150	1	547,739
G4	083	150	1	547,740
G5	127	150	0	547,741
G6	186	150	1	547,742
G7	195	150	0	547,743
G8	259	150	0	547,744
G9	285	150	1	547,745
G10	317	150	1	547,746
G11	384	150	1	547,747
G12	121	150	1	547,748

H1	149	150	1	548,737
H2	266	150	1	548,738
H3	286	150	1	548,739
H4	G.DNA CONTROL	150	-	548,740
H5	G.DNA CONTROL	150	-	548,741
H6	030	150	1	548,742
H7	038	150	1	548,743
H8	065	150	1	548,744
H9	080	150	1	548,745
H10	088	150	1	548,746
H11	092	150	1	548,747
H12	125	150	0	548,748

Table A4. CFI cohort sample layout in the 96 well plate Reaction plate 2. This table depicts a top view of the samples loaded in reaction plate 2 and shows the Illumina P5, P7 primer pair used (from lab stock)

P7 primer →		701	702	703	704	705	706	707	708	709	710	711	712
P5 primer ↓		1	2	3	4	5	6	7	8	9	10	11	12
525	A	141	159	160	191	228	258	291	327	045	078	137	147
526	B	197	206	264	273	008	023	130	132	148	162	166	160
527	C	202	225	351	369	390	400	018	031	042	064	168	204
528	D	222	048	109	111	143	161	190	200	210	238	241	269
529	E	318	366	376	387	037	098	105	153	192	216	260	267
530	F	294	373	067	103	118	122	185	280	287	288	362	380
531	G	004	039	104	196	205	221	284	305	310	323	344	350
532	H	399	119	155	157	213	256	271	276	282	311	297	G.DNA CONTROL

Table A5. Sample list, template concentration used, Illumina P5, P7 primer pair (from lab stock) used and condition (0=control, 1= patient) in Reaction plate 2.

Well id	PID/ID1	Concentration in. plate (ng/µl)	Condition (0 = control, 1 = patient)	Illumina P5, P7 primer pair (from lab stock)
A1	141	150	1	525,701
A2	159	150	1	525,702
A3	160	150	1	525,703
A4	191	150	1	525,704
A5	228	150	1	525,705
A6	258	150	1	525,706
A7	291	150	1	525,707
A8	327	150	1	525,708
A9	045	150	1	525,709
A10	078	150	1	525,710
A11	137	150	0	525,711
A12	147	150	1	525,712
B1	197	150	1	526,701
B2	206	150	0	526,702
B3	264	150	1	526,703
B4	273	150	1	526,704
B5	008	150	0	526,705
B6	023	150	0	526,706
B7	130	150	0	526,707
B8	132	150	1	526,708
B9	148	150	0	526,709
B10	162	150	1	526,710
B11	166	150	0	526,711
B12	169	150	0	526,712
C1	202	150	0	527,701
C2	225	150	0	527,702
C3	351	150	1	527,703
C4	369	150	0	527,704
C5	390	150	1	527,705
C6	400	150	1	527,706
C7	018	150	0	527,707
C8	031	150	0	527,708
C9	042	150	0	527,709
C10	064	150	0	527,710
C11	168	150	1	527,711
C12	204	150	1	527,712
D1	222	150	1	528,701
D2	048	150	1	528,702
D3	109	150	0	528,703

D4	111	150	0	528,704
D5	143	150	0	528,705
D6	161	150	1	528,706
D7	190	150	1	528,707
D8	200	150	0	528,708
D9	210	150	1	528,709
D10	238	150	1	528,710
D11	241	150	0	528,711
D12	269	150	1	528,712
E1	318	150	0	529,701
E2	366	150	1	529,702
E3	376	150	1	529,703
E4	387	150	1	529,704
E5	037	150	0	529,705
E6	098	150	0	529,706
E7	105	150	0	529,707
E8	153	150	0	529,708
E9	192	150	1	529,709
E10	216	150	0	529,710
E11	260	150	1	529,711
E12	267	150	0	529,712
F1	294	150	1	530,701
F2	373	150	1	530,702
F3	067	150	0	530,703
F4	103	150	1	530,704
F5	118	150	1	530,705
F6	122	150	1	530,706
F7	185	150	1	530,707
F8	280	150	0	530,708
F9	287	150	1	530,709
F10	288	150	0	530,710
F11	362	150	1	530,711
F12	380	150	0	530,712
G1	004	150	0	531,701
G2	039	150	0	531,702
G3	104	150	1	531,703
G4	196	150	0	531,704
G5	205	150	0	531,705
G6	221	150	1	531,706
G7	284	150	1	531,707
G8	305	150	1	531,708
G9	310	150	1	531,709
G10	323	150	0	531,710
G11	344	150	1	531,711
G12	350	150	1	531,712

H1	399	150	0	532,701
H2	119	150	1	532,702
H3	155	150	0	532,703
H4	157	150	0	532,704
H5	213	150	1	532,705
H6	256	150	1	532,706
H7	271	150	0	532,707
H8	276	150	1	532,708
H9	282	150	0	532,709
H10	311	150	0	532,710
H11	297	150	0	532,711
H12	G.DNA CONTROL	150	-	532,712

Table A6. CFI cohort sample layout in the 96 well plate Reaction plate 2R. This table depicts a top view of the samples loaded in reaction plate 2R and shows the Illumina P5, P7 primer pair used (from lab stock)

P7 primer →		713	714	715	716	717	718	719	720	721	722	723	724
P5 primer		1	2	3	4	5	6	7	8	9	10	11	12
513	A	141	159	160	191	228	258	291	327	045	078	137	147
514	B	197	206	264	273	008	023	130	132	148	162	166	160
515	C	202	225	351	369	390	400	018	031	042	064	168	204
516	D	222	048	109	111	143	161	190	200	210	238	241	269
541	E	318	366	376	387	037	098	105	153	192	216	260	267
542	F	294	373	067	103	118	122	185	280	287	288	362	380
543	G	004	039	104	196	205	221	284	305	310	323	344	350
544	H	399	119	155	G.DNA CONTROL	G.DNA CONTROL	256	271	276	282	311	297	300

Table A7. Sample list, template concentration used, Illumina P5, P7 primer pair (from lab stock) used and condition (0=control, 1= patient) in Reaction plate 2R.

Well id	PID/ID1	Concentration in. plate (ng/µl)	Condition (0 = control, 1 = patient)	Illumina P5, P7 primer pair (from lab stock)
A1	141	150	1	513,713
A2	159	150	1	513,714
A3	160	150	1	513,715
A4	191	150	1	513,716
A5	228	150	1	513,717
A6	258	150	1	513,718
A7	291	150	1	513,719
A8	327	150	1	513,720
A9	045	150	1	513,721
A10	078	150	1	513,722
A11	137	150	0	513,723
A12	147	150	1	513,724
B1	197	150	1	514,713
B2	206	150	0	514,714
B3	264	150	1	514,715
B4	273	150	1	514,716
B5	008	150	0	514,717
B6	023	150	0	514,718
B7	130	150	0	514,719
B8	132	150	1	514,720
B9	148	150	0	514,721
B10	162	150	1	514,722
B11	166	150	0	514,723
B12	169	150	0	514,722
C1	202	150	0	515,713
C2	225	150	0	515,714
C3	351	150	1	515,715
C4	369	150	0	515,716
C5	390	150	1	515,717
C6	400	150	1	515,718
C7	018	150	0	515,719
C8	031	150	0	515,720
C9	042	150	0	515,721
C10	064	150	0	515,722
C11	168	150	1	515,723
C12	204	150	1	515,724
D1	222	150	1	516,713
D2	048	150	1	516,714
D3	109	150	0	516,715
D4	111	150	0	516,716

D5	143	150	0	516,717
D6	161	150	1	516,718
D7	190	150	1	516,719
D8	200	150	0	516,720
D9	210	150	1	516,721
D10	238	150	1	516,722
D11	241	150	0	516,723
D12	269	150	1	516,724
E1	318	150	0	541,713
E2	366	150	1	541,714
E3	376	150	1	541,715
E4	387	150	1	541,716
E5	037	150	0	541,717
E6	098	150	0	541,718
E7	105	150	0	541,719
E8	153	150	0	541,720
E9	192	150	1	541,721
E10	216	150	0	541,722
E11	260	150	1	541,723
E12	267	150	0	541,724
F1	294	150	1	542,713
F2	373	150	1	542,714
F3	067	150	0	542,715
F4	103	150	1	542,716
F5	118	150	1	542,717
F6	122	150	1	542,718
F7	185	150	1	542,719
F8	280	150	0	542,720
F9	287	150	1	542,721
F10	288	150	0	542,722
F11	362	150	1	542,723
F12	380	150	0	542,724
G1	004	150	0	543,713
G2	039	150	0	543,714
G3	104	150	1	543,715
G4	196	150	0	543,716
G5	205	150	0	543,717
G6	221	150	1	543,718
G7	284	150	1	543,719
G8	305	150	1	543,720
G9	310	150	1	543,721
G10	323	150	0	543,722
G11	344	150	1	543,723
G12	350	150	1	543,724
H1	399	150	0	544,713

H2	119	150	1	544,714
H3	155	150	0	544,715
H4	G.DNA CONTROL	150	-	544,716
H5	G.DNA CONTROL	150	-	544,717
H6	256	150	1	544,718
H7	271	150	0	544,719
H8	276	150	1	544,720
H9	282	150	0	544,721
H10	311	150	0	544,722
H11	297	150	0	544,723
H12	300	150	1	544,724

Table A8. CFI cohort sample layout in the 96 well plate Reaction plate 3. This table depicts a top view of the samples loaded in reaction plate 3 and shows the Illumina P5, P7 primer pair used (from lab stock)

P7 primer →		701	702	703	704	705	706	707	708	709	710	711	712
P5 primer ↓		1	2	3	4	5	6	7	8	9	10	11	12
525	A	303	331	375	395	029	102	116	123	176	257	283	388
526	B	016	025	056	113	115	167	187	236	279	299	309	382
527	C	087	099	124	154	193	278	306	335	013	015	033	035
528	D	047	075	085	128	150	184	214	274	290	322	393	010
529	E	069	097	114	201	208	237	240	249	304	341	359	365
530	F	379	381	392	024	028	139	218	275	308	334	374	094
531	G	156	181	219	268	292	319	321	372	377	002	014	041
532	H	070	071	138	165	182	188	209	217	244	253	G.DNA CONTROL	397

Table A9. Sample list, template concentration used, Illumina P5, P7 primer pair (from lab stock) used and condition (0=control, 1= patient) in Reaction plate 3.

Well id	PID/ID1	Concentration in. plate (ng/μl)	Condition (0 = control, 1 = patient)	Illumina P5, P7 primer pair (from lab stock)
A1	303	150	0	525,701
A2	331	150	1	525,702
A3	375	150	0	525,703
A4	395	150	???	525,704
A5	029	150	0	525,705
A6	102	150	0	525,706
A7	116	150	1	525,707
A8	123	150	1	525,708
A9	176	150	0	525,709
A10	257	150	1	525,710
A11	283	150	0	525,711
A12	388	150	0	525,712
B1	016	150	0	526,701
B2	025	150	0	526,702
B3	056	150	1	526,703
B4	113	150	1	526,704
B5	115	150	1	526,705
B6	167	150	0	526,706
B7	187	150	1	526,707
B8	236	150	1	526,708
B9	279	150	0	526,709
B10	299	150	1	526,710
B11	309	150	0	526,711
B12	382	150	1	526,712
C1	087	150	???	527,701
C2	099	150	1	527,702
C3	124	150	1	527,703
C4	154	150	0	527,704
C5	193	150	1	527,705
C6	278	150	0	527,706
C7	306	150	0	527,707
C8	335	150	0	527,708
C9	013	150	0	527,709
C10	015	150	1	527,710
C11	033	150	0	527,711
C12	035	150	0	527,712
D1	047	150	0	528,701
D2	075	150	0	528,702
D3	085	150	0	528,703
D4	128	150	1	528,704

D5	150	150	0	528,705
D6	184	150	0	528,706
D7	214	150	???	528,707
D8	274	150	0	528,708
D9	290	150	0	528,709
D10	322	150	0	528,710
D11	393	150	0	528,711
D12	010	150	0	528,712
E1	069	150	1	529,701
E2	097	150	1	529,702
E3	114	150	0	529,703
E4	201	150	1	529,704
E5	208	150	0	529,705
E6	237	150	???	529,706
E7	240	150	0	529,707
E8	249	150	0	529,708
E9	304	150	1	529,709
E10	341	150	1	529,710
E11	359	150	1	529,711
E12	365	150	0	529,712
F1	379	150	0	530,701
F2	381	150	0	530,702
F3	392	150	1	530,703
F4	024	150	1	530,704
F5	028	150	1	530,705
F6	139	150	1	530,706
F7	218	150	0	530,707
F8	275	150	0	530,708
F9	308	150	0	530,709
F10	334	150	1	530,710
F11	374	150	1	530,711
F12	094	150	1	530,712
G1	156	150	1	531,701
G2	181	150	1	531,702
G3	219	150	1	531,703
G4	268	150	1	531,704
G5	292	150	0	531,705
G6	319	150	0	531,706
G7	321	150	1	531,707
G8	372	150	0	531,708
G9	377	150	1	531,709
G10	002	150	0	531,710
G11	014	150	1	531,711
G12	041	150	0	531,712
H1	070	150	0	532,701

H2	071	150	1	532,702
H3	138	150	0	532,703
H4	165	150	1	532,704
H5	182	150	1	532,705
H6	188	150	1	532,706
H7	209	150	1	532,707
H8	217	150	0	532,708
H9	244	150	0	532,709
H10	253	150	1	532,710
H11	G.DNA CONTROL	150	-	532,711
H12	397	150	0	532,712

Table A10. CFI cohort sample layout in the 96 well plate Reaction plate 3R. This table depicts a top view of the samples loaded in reaction plate 3R and shows the Illumina P5, P7 primer pair used (from lab stock)

P7 primer →		713	714	715	716	717	718	719	720	721	722	723	724
P5 primer ↓		1	2	3	4	5	6	7	8	9	10	11	12
541	A	303	331	375	395	029	102	116	123	176	257	283	388
542	B	016	025	056	113	115	167	187	236	279	299	309	382
543	C	087	099	124	154	193	278	306	335	013	015	033	035
544	D	047	075	085	128	150	184	214	274	290	322	393	010
545	E	069	097	114	201	208	237	240	249	304	341	359	365
546	F	379	381	392	024	028	139	218	275	308	334	374	094
547	G	156	181	219	268	292	319	321	372	377	002	014	041
548	H	070	071	138	G.DNA CONTROL	G.DNA CONTROL	188	209	217	244	253	261	397

Table A11. Sample list, template concentration used, Illumina P5, P7 primer pair (from lab stock) used and condition (0=control, 1= patient) in Reaction plate 3R.

Well id	PID/ID1	Concentration in. plate (ng/μl)	Condition (0 = control, 1 = patient)	Illumina P5, P7 primer pair (from lab stock)
A1	303	150	0	541,713
A2	331	150	1	541,714
A3	375	150	0	541,715
A4	395	150	???	541,716
A5	029	150	0	541,717
A6	102	150	0	541,718
A7	116	150	1	541,719
A8	123	150	1	541,720
A9	176	150	0	541,721
A10	257	150	1	541,722
A11	283	150	0	541,723
A12	388	150	0	541,724
B1	016	150	0	542,713
B2	025	150	0	542,714
B3	056	150	1	542,715
B4	113	150	1	542,716
B5	115	150	1	542,717
B6	167	150	0	542,718
B7	187	150	1	542,719
B8	236	150	1	542,720
B9	279	150	0	542,721
B10	299	150	1	542,722
B11	309	150	0	542,723
B12	382	150	1	542,722
C1	087	150	???	543,713
C2	099	150	1	543,714
C3	124	150	1	543,715
C4	154	150	0	543,716
C5	193	150	1	543,717
C6	278	150	0	543,718
C7	306	150	0	543,719
C8	335	150	0	543,720
C9	013	150	0	543,721
C10	015	150	1	543,722
C11	033	150	0	543,723
C12	035	150	0	543,724
D1	047	150	0	544,713
D2	075	150	0	544,714
D3	085	150	0	544,715

D4	128	150	1	544,716
D5	150	150	0	544,717
D6	184	150	0	544,718
D7	214	150	???	544,719
D8	274	150	0	544,720
D9	290	150	0	544,721
D10	322	150	0	544,722
D11	393	150	0	544,723
D12	010	150	0	544,724
E1	069	150	1	545,713
E2	097	150	1	545,714
E3	114	150	0	545,715
E4	201	150	1	545,716
E5	208	150	0	545,717
E6	237	150	???	545,718
E7	240	150	0	545,719
E8	249	150	0	545,720
E9	304	150	1	545,721
E10	341	150	1	545,722
E11	359	150	1	545,723
E12	365	150	0	545,724
F1	379	150	0	546,713
F2	381	150	0	546,714
F3	392	150	1	546,715
F4	024	150	1	546,716
F5	028	150	1	546,717
F6	139	150	1	546,718
F7	218	150	0	546,719
F8	275	150	0	546,720
F9	308	150	0	546,721
F10	334	150	1	546,722
F11	374	150	1	546,723
F12	094	150	1	546,724
G1	156	150	1	547,713
G2	181	150	1	547,714
G3	219	150	1	547,715
G4	268	150	1	547,716
G5	292	150	0	547,717
G6	319	150	0	547,718
G7	321	150	1	547,719
G8	372	150	0	547,720
G9	377	150	1	547,721
G10	002	150	0	547,722
G11	014	150	1	547,723
G12	041	150	0	547,724

H1	070	150	0	548,713
H2	071	150	1	548,714
H3	G.DNA CONTROL	150	-	548,715
H4	G.DNA CONTROL	150	-	548,716
H5	182	150	1	548,717
H6	188	150	1	548,718
H7	209	150	1	548,719
H8	217	150	0	548,720
H9	244	150	0	548,721
H10	253	150	1	548,722
H11	261	150	1	548,723
H12	397	150	0	548,724

Table A12. CFI cohort sample layout in the 96 well plate Reaction plate 4. This table depicts a top view of the samples loaded in reaction plate 4 and shows the Illumina P5, P7 primer pair used (from lab stock)

P7 primer →		701	702	703	704	705	706	707	708	709	710	711	712
P5 primer ↓		1	2	3	4	5	6	7	8	9	10	11	12
501	A	001	009	046	060	096	220	233	245	252	289	346	353
502	B	361	385	007	036	082	110	129	158	172	230	262	281
503	C	315	330	355	368	383	391	011	032	107	152	174	234
504	D	396	398	049	052	063	077	180	189	232	251	333	336
505	E	026	074	089	199	207	224	229	231	243	248	265	352
506	F	058	093	117	135	223	255	277	307	325	326	342	356
507	G	357	363	364	003	017	019	027	034	061	072	084	140
508	H	144	246	247	G.DNA CONTROL	G.DNA CONTROL	312	339	066	090	095	100	134

Table A13. Sample list, template concentration used, Illumina P5, P7 primer pair (from lab stock) used and condition (0=control, 1= patient) in Reaction plate 4.

Well id	PID/ID1	Concentration in. plate (ng/µl)	Condition (0 = control, 1 = patient)	Illumina P5, P7 primer pair (from lab stock)
A1	001	150	1	501,701
A2	009	150	0	501,702
A3	046	150	0	501,703
A4	060	150	1	501,704
A5	096	150	0	501,705
A6	220	150	1	501,706
A7	233	150	0	501,707
A8	245	150	0	501,708
A9	252	150	0	501,709
A10	289	150	1	501,710
A11	346	150	1	501,711
A12	353	150	0	501,712
B1	361	150	0	502,701
B2	385	150	0	502,702
B3	007	150	1	502,703
B4	036	150	0	502,704
B5	082	150	0	502,705
B6	110	150	???	502,706
B7	129	150	1	502,707
B8	158	150	1	502,708
B9	172	150	0	502,709
B10	230	150	0	502,710
B11	262	150	0	502,711
B12	281	150	0	502,712
C1	315	150	0	503,701
C2	330	150	0	503,702
C3	355	150	1	503,703
C4	368	150	1	503,704
C5	383	150	0	503,705
C6	391	150	1	503,706
C7	011	150	???	503,707
C8	032	150	1	503,708
C9	107	150	0	503,709
C10	152	150	0	503,710
C11	174	150	0	503,711
C12	234	150	1	503,712
D1	396	150	0	504,701
D2	398	150	0	504,702
D3	049	150	0	504,703
D4	052	150	1	504,704
D5	063	150	0	504,705

D6	077	150	0	504,706
D7	180	150	1	504,707
D8	189	150	1	504,708
D9	232	150	0	504,709
D10	251	150	0	504,710
D11	333	150	1	504,711
D12	336	150	1	504,712
E1	026	150	1	505,701
E2	074	150	0	505,702
E3	089	150	1	505,703
E4	199	150	1	505,704
E5	207	150	0	505,705
E6	224	150	0	505,706
E7	229	150	1	505,707
E8	231	150	0	505,708
E9	243	150	0	505,709
E10	248	150	0	505,710
E11	265	150	0	505,711
E12	352	150	1	505,712
F1	058	150	0	506,701
F2	093	150	0	506,702
F3	117	150	1	506,703
F4	135	150	0	506,704
F5	223	150	0	506,705
F6	255	150	1	506,706
F7	277	150	0	506,707
F8	307	150	1	506,708
F9	325	150	0	506,709
F10	326	150	1	506,710
F11	342	150	1	506,711
F12	356	150	0	506,712
G1	357	150	1	507,701
G2	363	150	0	507,702
G3	364	150	1	507,703
G4	003	150	1	507,704
G5	017	150	0	507,705
G6	019	150	0	507,706
G7	027	150	0	507,707
G8	034	150	1	507,708
G9	061	150	0	507,709
G10	072	150	0	507,710
G11	084	150	1	507,711
G12	140	150	0	507,712
H1	144	150	1	508,701
H2	246	150	0	508,702

H3	247	150	0	508,703
H4	G.DNA CONTROL	150	-	508,704
H5	G.DNA CONTROL	150	-	508,705
H6	312	150	1	508,706
H7	339	150	1	508,707
H8	066	150	0	508,708
H9	090	150	1	508,709
H10	095	150	0	508,710
H11	100	150	1	508,711
H12	134	150	0	508,712

Table A14. CFI cohort sample layout in the 96 well plate Reaction plate 5. This table depicts a top view of the samples loaded in reaction plate 5 and shows the Illumina P5, P7 primer pair used (from lab stock)

P7 primer →		737	738	739	740	741	742	743	744	745	746	747	748
P5 primer ↓		1	2	3	4	5	6	7	8	9	10	11	12
	A	Blank (H2O)	Blank (H2O)	Blank (H2O)	Blank (H2O)	Blank (H2O)	Blank (H2O)	Blank (H2O)	Blank (H2O)	Blank (H2O)	Blank (H2O)	Blank (H2O)	Blank (H2O)
534	B	163	235	242	270	313	329	332	347	349	358	401	005
535	C	006	101	403	404	149	266	092	125	314	378	141	159
536	D	137	147	399	119	092	125	157	213	303	331	283	388
545	E	070	071	261	397	165	182	001	009	346	353	144	246
546	F	100	134	296	298	G.DNA CONTROL	G.DNA CONTROL	347	Blank (H2O)	Blank (H2O)	Blank (H2O)	Blank (H2O)	Blank (H2O)
	G	Blank (H2O)	Blank (H2O)	Blank (H2O)	Blank (H2O)	Blank (H2O)	Blank (H2O)	Blank (H2O)	Blank (H2O)	Blank (H2O)	Blank (H2O)	Blank (H2O)	Blank (H2O)
	H												

Table A15. Sample list, template concentration used, Illumina P5, P7 primer pair (from lab stock) used and condition (0=control, 1= patient) in Reaction plate 5.

Well id	PID/ID1	Concentration in. plate (ng/µl)	Condition (0 = control, 1 = patient)	Illumina P5, P7 primer pair (from lab stock)
A1	Blank(H2O)	150	-	
A2	Blank(H2O)	150	-	
A3	Blank(H2O)	150	-	
A4	Blank(H2O)	150	-	
A5	Blank(H2O)	150	-	
A6	Blank(H2O)	150	-	
A7	Blank(H2O)	150	-	
A8	Blank(H2O)	150	-	
A9	Blank(H2O)	150	-	
A10	Blank(H2O)	150	-	
A11	Blank(H2O)	150	-	
A12	Blank(H2O)	150	-	
B1	163	150	0	534,737
B2	235	150	0	534,738
B3	242	150	0	534,739
B4	270	150	0	534,740
B5	313	150	1	534,741
B6	329	150	0	534,742
B7	332	150	0	534,743
B8	347	150	0	534,744
B9	349	150	0	534,745
B10	358	150	1	534,746
B11	401	150	0	534,747
B12	005	150	0	534,748
C1	006	150	0	535,737
C2	101	150	0	535,738
C3	403	150	0	535,739
C4	404	150	0	535,740
C5	149	150	1	535,741
C6	266	150	1	535,742
C7	092	150	1	535,743
C8	125	150	0	535,744
C9	314	150	1	535,745
C10	378	150	1	535,746
C11	141	150	1	535,747
C12	159	150	1	535,748
D1	137	150	0	536,737
D2	147	150	1	536,738
D3	399	150	0	536,739

D4	119	150	1	536,740
D5	092	150	1	536,741
D6	125	150	0	536,742
D7	157	150	0	536,743
D8	213	150	1	536,744
D9	303	150	0	536,745
D10	331	150	1	536,746
D11	283	150	0	536,747
D12	388	150	0	536,748
E1	070	150	0	545,737
E2	071	150	1	545,738
E3	261	150	1	545,739
E4	397	150	0	545,740
E5	165	150	1	545,741
E6	182	150	1	545,742
E7	001	150	1	545,743
E8	009	150	0	545,744
E9	346	150	1	545,745
E10	353	150	0	545,746
E11	144	150	1	545,747
E12	246	150	0	545,748
F1	100	150	1	546,737
F2	134	150	0	546,738
F3	296	150	1	546,739
F4	298	150	1	546,740
F5	G.DNA CONTROL	150	-	546,741
F6	G.DNA CONTROL	150	-	546,742
F7	347	150	0	546,743
F8	Blank(H2O)	150	-	546,744
F9	Blank(H2O)	150	-	546,745
F10	Blank(H2O)	150	-	546,746
F11	Blank(H2O)	150	-	546,747
F12	Blank(H2O)	150	-	546,748
G1	Blank(H2O)	150	-	
G2	Blank(H2O)	150	-	
G3	Blank(H2O)	150	-	
G4	Blank(H2O)	150	-	
G5	Blank(H2O)	150	-	
G6	Blank(H2O)	150	-	
G7	Blank(H2O)	150	-	
G8	Blank(H2O)	150	-	
G9	Blank(H2O)	150	-	
G10	Blank(H2O)	150	-	

G11	Blank(H2O)	150	-	
G12	Blank(H2O)	150	-	

APPENDIX B

Timeline of previous lab work.

	Timeline	Projects
	07/24/2019	DNA/RNA isolation • Used HELA cells and HEK293T cells to perform mtDNA and genomic DNA library prep.
	08/02/2019	• mtDNA capture and analyzing capture ratio by qpcr
	11/27/2019	• Extracted RNA and DNA from HEK293T cells with and without blue light stimulation
	07/8/2020- 02/08/2021	• STAMP practice – First started with single samples then 8 samples together using HEK293T and HELA cell extracted control genomic DNA • Sequenced 2 libraries
	03/26/2021- 07/20/2021	Applied STAMP on PBMC samples from CFI cohort

APPENDIX C

ROTATION PROJECTS, TEACHING EXPERIENCE AND OUTREACH

Rotation 1 Principal Investigator: Dr. Haiyuan Yu Fall 2018

- Cultured HEK293T cells to generate protein samples for use in Mass Spectrometry
- Prepared and purified protein samples, followed by running through a Mass Spectrometer

Rotation 2 Principal Investigator: Dr. Jun (Kelly) Liu Spring 2019

Attempted to unravel the role *Syndecan* gene in the Bone Morphogenetic Pathway (BMP) by performing preliminary experiments.

- Generated, successfully a *C. elegans* strain (*ccls4438(cc::gfp); sdn-1(zh20))* which was mutant for the *SynDecaN* gene and expressed *GFP* in coelomocyte cells
    - Set up crosses of appropriate male and hermaphrodite lines; selected best candidates in the F2 generation by scoring for phenotype using fluorescence microscopy; genotyped the candidate strains; and saved the newly generated strains
- Attempted to generate a *C. elegans* strain *(sdn-1(zh20); sma-9(cc604))* which was mutant for the *SynDecaN* gene and SMAII gene
    - Set up crosses of appropriate male and hermaphrodite lines; selected best candidates from 1000 F1 worms as the recombination frequency for the genes was 0.2 %.
- Attempted to generate BMP-Like (*dbl-1)* Over Expression (OE*)* strain through extrachromosomal integration

Rotation 3 Principal Investigator: Dr. Natasza Kurpios Spring 2019

Investigating the role of smooth muscles in long chain fatty acids.

- Dissected mouse embryos at the embryonic stage to obtain small intestine tissue.
- Generated villi tissue sections and performed antibody staining to test antibodies against the *Smoothelin* (SMTN) class of proteins.
- Performed double immuno-staining to detect colocalization between smooth muscle actin (SMA) and 3,4,6 trichlorophenol (H3P)
- Imaged via fluorescence microscopy the mounted slides and used imaging software to score the number of colocalizing H3P loci with SMA

TEACHING EXPERIENCE

Served as a Graduate Teaching Assistant to Dr. Darlene Campbell in "BIOG 1445: Introduction to Comparative Anatomy and Physiology, Individualized Instruction" from Fall 2019-Spring 2021
Responsibilities included:

- Tutoring students individually and in groups, developing and administering oral examinations, teaching an inquiry-based Cardio-pulmonary function lab, and grading student lab reports
- teaching assigned students, the entire course material through virtual Oral test evaluations and virtual tutoring during the COVID-19 pandemic.
- Administering the Cardio-pulmonary function lab, virtually using simulation software during the COVID-19 pandemic.
- Tutoring students through rat dissection demonstrations of different organ systems

Cumulative Summary of teaching evaluations over all four semesters:

	Fall 2019		Spring 2020		Fall 2020		Spring 2021		Overall	
Field	Mean (out of 5)	Number of students	Mean (out of 5)	Number of students	Mean (out of 5)	Number of students	Mean (out of 5)	Number of students	Mean (out of 5)	Number of students
Knowledge of course material	4.92	12	4.80	10	4.71	7	4.88	8	4.83	37
Fairness and impartiality	4.92	12	5.00	9	4.71	7	4.88	8	4.88	36
Clarity of Explanations	4.67	12	4.70	10	4.71	7	4.63	8	4.68	37
Friendliness and Helpfulness	5.00	13	5.00	10	5.00	7	4.88	8	4.97	38

Was awarded the Outstanding Graduate Teaching Assistantship Award based on distinguished performance as a Teaching Assistant in the College of agricultural and life sciences for the 2020-2021 academic year.

LEADERSHIP AND OUTREACH

Teaching Fellow, Science and Technology Entry Program – Upward Partnership (STEP-UP) Summer 2021

Developed and taught a 4-lesson mini-course in inheritance, mendelian genetics, plant reproduction and environmental factors affecting plant growth to high school students belonging to underrepresented communities.

Teaching Fellow, Graduate Student School Outreach Program (GRASSHOPR) Spring 2021

Developed and taught a 4-lesson mini-course in plant reproduction to high school biology students at Southern Cayuga Jr/Sr High in Auburn, NY.

Committee Chair, Appropriations Committee, Cornell Graduate and Professional Students' Association (GPSA)

August 2020-2021

Managed and allocated GPSA funds by drafting a resolution, reviewed the policies and procedures for setting and allocating the Activity Fee and made recommendations.

Editor and Essay Author, Frontiers- Techniques that Changed Biology, Vol. 1, Issue 1, July 2019

MBG Department, Cornell University

Contributed an essay for a pilot magazine publication based on exciting, paradigm-shifting technologies in biology, served as an editor on the student committee responsible for compiling and completing the final magazine draft.