



OpenMedicine Foundation (OMF)

Scientific Advisory Board

Director

Ronald Davis, Ph.D.
*Genome Technology Center
Stanford University*

Paul Berg, PhD
*Nobel Laureate
Molecular Genetics
Stanford University*

Mario Capecchi, Ph.D.
*Nobel Laureate
Genetics & Immunology
University of Utah*

Mark Davis, Ph.D.
*Immunology
Stanford University*

H. Craig Heller, Ph.D.
*Biology & Exercise Physiology
Stanford University*

Andreas Kogelnik, MD, Ph.D.
*Infectious Disease
Open Medicine Institute*

Baldomero Olivera, Ph.D.
*Neurobiology
University of Utah*

Ronald Tompkins, MD, Sc.D.
*Trauma & Metabolism
Harvard Medical School*

James Watson, Ph.D.
*Molecular Genetics
Nobel Laureate
Human Genome Project*

Wenzhong Xiao, Ph.D.
*Computational Genomics
Stanford University
Harvard University*

Foundation Board
Executive Director,
Linda Tannenbaum
Treasurer, Kimberly Hicks
Secretary, Patricia Linsley
Tina Orkin
R.P. Channing Rodgers, MD
Deborah Rose, MD
H. Kenneth Walker, MD, PhD

End ME/CFS Project Biomarker Discovery: Severely Ill Big Data Study

INTRODUCTION

The IOM report on ME/CFS states: “Myalgic encephalomyelitis (ME) and chronic fatigue syndrome (CFS) are serious, debilitating conditions that impose a burden of illness on millions of people in the United States and around the world. Somewhere between 836,000 and 2.5 million Americans are estimated to have these disorders.” The report further states: “Diagnosing ME/CFS in the clinical setting remains a challenge. Patients often struggle with their illness for years before receiving a diagnosis, and an estimated 84 to 91 percent of patients affected by ME/CFS are not yet diagnosed.”

This proposal will conduct a “BIG DATA” analysis on ME/CFS in search of a sensitive and specific molecular biomarker(s). Although many of the symptoms are neurological, molecular biomarker(s) may be found in the blood, saliva, sweat, urine, and feces. Identification of biomarker(s) in these easily assayed fluids can be convenient, inexpensive, and could be conducted on a bed-bound patient.

This BIG DATA set will be released to the scientific community and serve to provide a better understanding of the disease and lead to effective treatment and prevention. It is vital to find a molecular biomarker(s) for the diagnosis of ME/CFS. Switching from a symptom-base to a molecular biomarker(s)-based diagnosis will remove doubt about the diagnosis from the patient, from the medical community and the medical researcher. There are several problems in finding the biomarkers. The disease may be heterogeneous, which will require biomarkers for each disease type. We don’t know where to look. It could be in the DNA, RNA, protein, carbohydrate, or metabolite—or could be from an associated microbe in or on the body. The biomarkers need to be easily retrieved (in a home for the seriously ill) and thus should be in the blood, urine, saliva, stool, and cerebral spinal fluid, sweat or breath.



The biomarkers need to be present at all severities of the disease. It would be helpful for the quantitative biomarkers that their intensity is reflected the severity of the disease. The biomarkers needs to uniquely identify ME/CFS and distinguish from all other diseases (e.g. Fibromyalgia, Lyme, depression, trauma, infection, etc.).

The heterogeneity of the disease may be evaluated by biomarker-profiling patients with identical or nearly identical symptoms. A clustering of a unique set of biomarkers with symptoms may indicate a unique form of the disease. Another method of clustering is to include the immune repertoire DNA sequence. This might include information in the sequence that uniquely identifies the cause of the original stress or infection.

We could also use biomarkers to indicate most likely treatment. Patients could be extensively profiled for biomarkers prior to any treatment. Those patients who respond favorably to treatment could then be retrospectively analyzed for unique biomarker profiles that would prescribe the best treatment options for new patients. The biomarkers need not be a single component or several components of the same type (i.e., several RNA species) but could be a mixture of different components (i.e., several proteins from blood, a metabolite from urine and several DNA alleles). Development of a mixed biomarker set would be greatly facilitated if the search for biomarkers is conducted on the same large clinical sample for both cases and controls.

The biomarkers contained in the DNA could be single nucleotide polymorphisms, structural rearrangements, unique methylation or demethylation, unique binding or unbinding of proteins, and other unique arrangements of DNA. The cellular origin of the DNA could be any cell that gives a unique signal but most likely would be a cell from the blood or other easily accessible source. We must also be aware of the possibility of mosaicism in the origin of the DNA

The use of RNA as a source of biomarkers is likely to be a quantitative determination. High precision and reproducibility is necessary to give the best resolution and accurate diagnoses and should be tested and demonstrated for biomarker discovery. A unique immune cell type is likely to be the best source for the RNA. In the past, a mixture of all cell types in the blood have been used for the search for biomarkers. However, this approach is less likely to give clear biomarkers because the RNA quantity in each cell type is different and the number of each cell type is likely different in each patient. We should focus on Natural Killer (NK) cells because they are usually reduced and/or inactive in ME/CFS/SEID patients and could be a good source of biomarkers or a component of a set of biomarkers. Another approach is analyzing individual cells. This will not require separating individual cell types if we can analyze a very large number of cells.

The use of proteins as a source of biomarkers follows a classical approach. There are numerous antibody methods and other assays that allow easy fast analysis. Some of the newer methods allow extensive multiplexing that might be required for ME/CFS. The

discovery phase could use various mass spectroscopy methods that are now quite advanced. Unique protein modifications could also be used.

Many Physicians and researchers speculate that some microbe is the initiating event of ME/CFS. Although this supposed organism(s) may not continue to be present, we must exhaustively search for them. This can be done by standard microbiome DNA sequencing from all body fluids. We can increase sensitivity if we first disrupt all human cells followed by DNase treatment. The DNase resistant DNA will be from DNA containing particles such as viruses, bacteria, fungus, or parasite.

The recent advances in mass spectroscopy make searching for biomarkers among the small molecule metabolites a feasible approach. The biological source is likely to be blood, urine, saliva, or cerebral spinal fluid, although all bodily fluids should be evaluated.

The physiological state of the patient is likely to have a major impact on revealing suitable biomarkers. Because post exercise malaise is a major phenotype of ME/CFS, this is the state that is most likely to contain unique biomarkers. Working with the most severely affected patients is also likely to give good biomarker signatures although they probably will not be able to enter a state of post exercise malaise unless they are constantly in this state.

Conducting all of these molecular investigations with state-of-the-art methodologies on well-phenotyped patients will be a daunting task. It will require significant resources and considerable coordination and cooperation within the scientific and medical communities.

RESEARCH PLAN

Research Tests

- RNA Gene Expression - PBMCs
- RNA Gene Expression - NKs
- RNA Gene Expression - Monocytes
- RNA Gene Expression - T-Cells
- RNA Gene Expression - B-Cells
- RNA Gene Expression - Macrophages
- RNA Gene Expression - Dendritic Cells
- MicroRNA in Plasma
- Proteomics
- Cell-Free DNA
- Whole Genome Sequencing
- Whole Exome Genome Sequencing
- Mitochondrial Genome Sequencing
- Mitochondrial DNA/Nuclear DNA Radio

HLA DNA Sequencing
WBC density
Autoantibody panel
T-cell Repertoire DNA Sequencing-Stanford
T-cell Repertoire DNA Sequencing-Sequentia
CyTOF Standard Panel
CyTOF Additional Panels
Pan Viral Panel
Serum Cytokines
Ceres Nano Lyme
Microbiome - stool
Metabolomics - stool
Metabolomics - plasma
Metabolomics - urine
Heavy Metals - blood
Heavy Metals - urine
Mycotoxins by Magarray

Clinical Testing

Acylcarnitines
Ammonia
Bartonella extended cx
Biotin (Vitamin B7)
CBC (Complete Blood Count)
Chemistry Comprehensive Panel
Creatine Kinase
Copper
C- Reactive Protein
Cysteine
Folate
FSH (Follicle-Stimulating Hormone)
LH (Luteinizing Hormone)
Estrogen
HbA1C (Hemoglobin A1c)
Homocysteine
IgG Subsets
Lactate
Lyme Serology with reflex Western Blot
Lymphocyte Subsets
MMA (Methylmalonic Acid)

MTHFR Mutations (Methylenetetrahydrofolate)
Natural Killer Cell (Count & Function)
Organic Acids - urine
Pyruvate
Serotonin
Testosterone
Thyroid Panel
Uric Acid
Vasopressin/ADH
Viral Digital PCR
Viral Serologies:
 EBV EA (Epstein Barr Virus EA)
 EBV NA (Epstein Barr Virus NA)
 EBV VCA (Epstein Barr Virus VCA)
 CMV (Cytomegalovirus)
 HHV-6 (Human Herpesvirus 6)
 Parvovirus
 HHV-7 (Human Herpes Virus 7)
 HSV1&2 (Herpes Simplex 1 & 2)
Vitamin B12 (Cobalamin)
Vitamin D25 OH

Other Testing/Monitoring

Extended EEG
Sleep monitoring
Fitbit
Cognitive Testing
Questionnaires - SF36, PROMIS

ANALYSIS

We will conduct an extensive bioinformatics analysis on the data sets. We will search for the best combination to generate a diagnostic biomarker set. Because of the large data sets, false discovery is of serious concern. Therefore, we will evaluate our diagnostic molecular biomarker set on different patients with and without ME/CFS to determine sensitivity and specificity. Another problem is the possibility that ME/CFS is heterogeneous, and no biomarker set can be found. We will then attempt to cluster the patient population into groups and search for a biomarker set in each group. The Immune repertoire of all immune cells may be a useful data set to achieve biological meaningful clustering. It is possible that the repertoire will reflect the initiating event. This approach has been successful for other diseases.

This BIG DATA set will be useful toward understanding the disease and may suggest various treatments. It also might suggest methods of prevention. There is the remote possibility that nothing useful is found. In this case, because we focused on using the best and latest technology and focused on best practices we, can move on. The experiments conducted here will not have to be repeated. Also, because the data will be in the public domain, other investigators can explore other analysis and uses without the expense of doing additional experimentation. If BIG DATA of accessible body fluids does not yield useful results, then we will probably have to investigate the brain, which will be more difficult and expensive.

This proposal was written by OMF Scientific Advisory Board Director, Ronald W. Davis, PhD, and Linda Tannenbaum updated 12-9-15.