



The Gene Teams

BY SUZANNE VERNON, PHD, CENTERS FOR DISEASE CONTROL AND PREVENTION

Teams of scientists tackled extensive biomedical data in the CFS Computational Challenge. Here's a play-by-play of the progress they achieved.

AT-A-GLANCE ►►

- CFS was at the center of cutting-edge scientific analysis when multidisciplinary teams examined a plethora of data from extensive medical testing.
- Each team was able to link at least some of its findings to specific genes that appear to be involved in the illness.

In 1996 the Centers for Disease Control and Prevention (CDC) began a large, population-based study of CFS in Wichita, Kansas. After surveying more than 80,000 residents to identify and follow people with CFS, a clinical study was needed to characterize the illness at a detailed biological level. Between November 2001 and December 2002, 227 people identified with CFS, people with other unexplained chronically fatiguing illnesses and randomly selected people without fatigue underwent two days of comprehensive medical and psychiatric testing, creating one of the most intensive clinical explorations of the illness. When all was said and done, we were left with a gold mine of data but no optimal way to mine the nuggets. So the CDC decided to host a unique scientific challenge we coined “C3” for CFS Computational Challenge.

The Game Plan

Twenty-four investigators from around the world joined investigators from the CDC’s CFS research program for the kickoff of the challenge. What made this group of investigators unique was its international prominence and diverse expertise in mathematics, biology, medicine and computation. From this promising group, we formed multidisciplinary teams so that each one included strong computational, biologic and medical expertise. Every team was given the same data set and six months to integrate this extensive, disparate data in hopes of describing the pathophysiology and identifying biomarkers of CFS. Then we let the games begin!

Team One: Mapping Gene Expression to Symptoms

Members: Drs. Elizabeth R. Unger (CDC), Toni Whistler (CDC), Nancy Klimas (University of Miami), Renee Taylor (University of Illinois at Chicago), R. Cameron Craddock (CDC) and Gordon Broderick (University of Alberta)

This team focused on the expression of 20,000 genes and 117 clinical variables. The first step when dealing with this much data is to reduce it to the fewest bits of information while retaining as much potentially CFS-relevant biologic information as possible. So that's where Team One began.

Initially, based on the 1994 CFS case definition, the team found no unique gene expression pattern to distinguish people with CFS from nonfatigued people. But when study subjects were defined by symptoms (referred to as symptom space)—as measured by the Multidimensional Fatigue Inventory (MFI), the Short Form 36 (SF-36), the CDC Symptom Inventory (CDC SI), the Zung Depression Scale and CANTAB computerized cognitive testing—the team successfully identified 39 distinguishing genes. Four of these 39 genes were also identified in an accompanying paper by team member Toni Whistler, who used a technique called Quantitative Trait Analysis (QTA) to find gene expression activity that correlated with fatigue as measured by the MFI. This directly ties a set of 39 biologic indicators (the genes) to fatigued patients defined by standard clinical instruments, essentially removing the more subjective factors from the equation.

Interestingly, the expression of these 39 genes is important in antioxidant activity, immune system activity and apoptosis (cell death)—factors that could underlie physical complaints common to many people with CFS.

This team also identified the most important of the 117 clinical variables that describe the CFS symptom space, finding 17 significant clinical variables including measures of heart rate variability during sleep, orthostatic stability, endocrine function and medications. Once again, these findings were corroborated in another paper, this time by team member R. Cameron Craddock, who also determined that sleep heart rate variability and medications were associated with fatigue.

Everyone was given the same data set and six months to integrate it in hopes of describing the pathophysiology and identifying biomarkers of CFS.

With important gene expression and clinical factors identified, the team was then able to use this information to distinguish the people with CFS from the nonfatigued people in the study (See Figure 1). It found that people with CFS could be distinguished from those with insufficient symptoms or fatigue (ISF) and from nonfatigued controls. As seen in the figure, each group is arrayed in a spectrum that reflects differences in symptom severity.

Team One's findings are significant in the quest to objectively and biologically identify CFS. The results also raise caution about the discriminating power of gene expression markers—particularly of individual genes with single-point-in-time measures, given the dynamic nature of this illness. Since gene expression is dynamic, it should be used in conjunction with other physiologic and functional measures to increase its effectiveness in biomarker discovery.

Team Two: Searching for CFS Subgroups

Members: Drs. Peter White (Queen Mary School of Medicine), Ute Vollmer-Conna (University of New South Wales), Eric Aslakson (CDC), Sol Efroni (NIH), Liran Carmel (NIH) and Mangalathu Rajeevan (CDC)

This team set out to dissect the heterogeneity of CFS. Members began with the premise that the 1994 CFS case definition was not a valid tool for delineating likely subtypes of fatiguing illness, so they derived their CFS cases and controls empirically.

Hypothesizing that people with CFS suffer from a series of different conditions, all with prominent fatigue, the team's objective was to delineate these subgroups using epidemiologic, clinical and biologic data. Once subgroups were derived, the team would then attempt to identify gene expression and genetic associations. They hoped that by giving both form and content to CFS subgroups, they would prove the hypothesis that CFS was heterogeneous and could propose analytical approaches to improve diagnosis and intervention.

Focusing on 159 women from the clinical study, the team identified 38 clinical and biologic variables to define six subgroups: one "well" group and five groups with prominent fatigue distinguished by various biologic parameters. The fatigue subgroups were characterized as follows:

- Obese with postexertional fatigue, sleep hyponea (reduction of airflow) and sleep disturbance
- Obese with sleep hyponea, low sleep heart rate variability and low 24-hour urinary cortisol
- Multisymptomatic with depression, HPA axis disruption and menopausal indicators

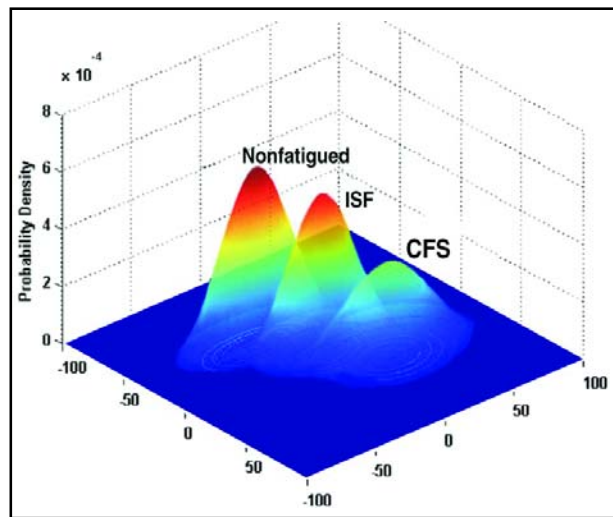


Figure 1: The expression level of 39 genes could distinguish subjects with CFS, insufficient fatigue (ISF) and no fatigue as measured by the MFI, SF36, CDC Symptom Inventory, and cognition and sleep variables.

- Highly symptomatic with depression
- Muscle pain and subjective sleep complaints (also called the interoceptive subgroup since only subjectively reported symptoms and no biologic variables were found)

When the team then compared how people with CFS, according to the 1994 CFS case definition, fit into each of these subgroups, they found that 89% of the nonfatigued (NF) subjects were indeed part of the “well” subgroup and the majority of CFS subjects were in one of the fatigue subgroups.

Next the team wanted to identify gene expressions and genetic associations. While it’s relatively straightforward to identify gene expression correlates between two groups—like sick and healthy—there are few ways to do this when multiple sick groups exist, such as the five fatigued subgroups this team identified.

To address this challenge, team member Liran Carmel developed a computational approach that identified discriminatory gene expression activity that could be applied to any multiclass problem. Using this, the team found that 17 differentially expressed genes could discriminate between the well subjects and the five fatigue subgroups. In fact, one gene (SLC1A6) was expressed 3- to 60-fold greater in all five fatigue subgroups compared with well subjects. While the differential expression of this gene needs to be validated by other methods, the

implications are tantalizing since its function is related to excitatory amino acid transport, an element of central nervous system activity.

Perhaps one of the most exciting outcomes from this team’s work was the finding that genetic variations (known as polymorphisms) were associated with some of the fatigue groups. They analyzed 11 genes related to both hypothalamic pituitary adrenal (HPA) axis function and neurotransmitter systems and found variations in five genes that were significantly associated with three of the fatigue subgroups (See Figure 2). This was the first demonstration that genes important in the function of the HPA axis and the sympathetic nervous system could be used to identify subtypes of CFS.

Team Three: Exploring Extremes

Members: Drs. Andrew Lloyd (University of New South Wales), Roumiana Boneva (CDC), Weida Tong (FDA) and Jennifer Fostel (NIH)

One effective approach for biomarker identification is to compare and contrast the extremes—in other words, the most sick against the healthiest. The third C3 team did just that.

Team Three used responses to the MFI to identify 23 people with the greatest and the least fatigue and the responses to the Zung self-rating depression scale to identify 26 people with the greatest and the least depression. There were 10 people in common among these groups, and these people were excluded, leaving 39 individuals as the basis of this “extreme” analysis. These 39 subjects were then used to explore which of the 20,000 genes in the data set could identify people with CFS. The team found 24 genes that could do this. While these same differentially expressed genes were not found in the other CFS studies, the functions they influence were. Once again, genes important in immune response, ion binding and transport, signal transduction and neuronal activity were shown to differentiate people with CFS.

Then Team Three went back to the remaining people in the data set (leaving out the 39 they used to identify the 24 genes) to see if these 24 genes could again identify the people with CFS. Sure enough, these 24 gene expression markers found most of the people with CFS and distinguished them from the other subjects. The “Extreme Team” had come up with a very similar result as Team One.

This was the first demonstration that genes important in the function of the HPA axis and the sympathetic nervous system could be used to identify subtypes of CFS.

Team Three also attempted to identify gene expression that was specific to the other symptoms of CFS. The team could not unambiguously identify gene expression correlates for any of the symptom dimensions. However, they did identify differential gene expression activity that correlated with blood measures (red and white blood cell counts), biochemical measures (urea, creatinine) and liver function (aspartate and alanine amino transferases).

Team Four: the Allostatic Load Connection

Members: Drs. Elizabeth Maloney (CDC), Ben Goertzel (Biomind LLC), Brian Gurbaxani (CDC) and Jim Jones (CDC)

The fourth team examined the risk of high allostatic load and CFS. Allostatic load is the cumulative wear and tear on the body when it fails to adapt adequately to change, whether it be everyday experiences or nonroutine challenges such as infection and trauma. The body's adaptation to challenge is orchestrated by hormones of the HPA axis and the sympathetic nervous system—both of which seem to be involved in CFS. Since these hormones have receptors throughout the body, increased allostatic load could potentially have an effect on multiple body systems including the immune, endocrine and cardiovascular systems. This was the basis for Team Four's exploration of the subject.

Allostatic load includes measures of metabolic and cardiovascular activity, as well as HPA axis and sympathetic nervous system activity. When the risk of allostatic

load was examined in the study subjects, the team found that the people with CFS, as classified by the 1994 CFS case definition, were almost twice as likely to have a high allostatic load index compared to nonfatigued controls. They also found that certain components of the allostatic load index could predict specific dimensions that define CFS.

For example, the C-reactive protein, norepinephrine, epinephrine and blood components of the allostatic load index specifically predicted increased bodily pain and decreased physical functioning as measured by the SF-36 health survey. Interestingly, however, no component of the allostatic load index predicted any type of fatigue as measured by the MFI.

High allostatic load is associated with increased risk of cardiovascular disease and decline in cognitive and physical function. This was the first demonstration that elevated allostatic load is also associated with CFS, supporting the role of the HPA axis and the sympathetic nervous system in explaining at least some of the pathophysiology of CFS.

The Score?

Everyone wins. The researchers who participated got 14 peer-reviewed papers published in the April 2006 issue of the journal *Pharmacogenomics*. Each team asked different questions of the data and applied unique analytical approaches resulting in novel findings, some common themes and many lessons learned for future CFS research studies. This important accomplishment not only benefits CFS research but establishes a biomedical paradigm demonstrating that data sharing and multidisciplinary research can overcome challenging data integration problems to glean information on some of the most interesting biologic questions.

What does this all mean to the person suffering with CFS? First of all, genes important in the function of the HPA axis and the sympathetic nervous system may make a person more vulnerable to CFS. This genetic vulnerability also appears to be affected by environmental, behavioral and physiologic events experienced over the lifespan. The teams also uncovered some promising genomic biomarkers that we're now actively exploring to improve diagnosis. All of this has implications for targeting therapy and possibly improving interventions based on genomics. ■

Suzanne Vernon, PhD, is a research biologist and human genomics team leader for the Centers for Disease Control and Prevention (CDC). For the past six years, she's directed the molecular epidemiology program in the Chronic Viral Diseases Branch at the CDC.

Figure 2: Team Two identified 38 clinical and biologic variables to define six subgroups: one "well" group and five groups with prominent fatigue distinguished by biologic parameters. Variations in five genes correlated to three of the fatigue subgroups.

