REVIEW

Interpreter of maladies: redescription mining applied to biomedical data analysis

Peter Waltman[1], Alex Pearlman[2] & Bud Mishra[1,2†]

†Author for correspondence
[1]New York University, Courant Institute of Mathematical Sciences, 715 Broadway, New York, NY 10003, USA
Tel.: +1 212 998 3464; Fax: +1 212 998 3484; E-mail: mishra@nyu.edu
[2]New York University, Department of Cell Biology, NYU School of Medicine, New York, NY 10016, USA

Comprehensive, systematic and integrated data-centric statistical approaches to disease modeling can provide powerful frameworks for understanding disease etiology. Here, one such computational framework based on redescription mining in both its incarnations, static and dynamic, is discussed. The static framework provides bioinformatic tools applicable to multifaceted datasets, containing genetic, transcriptomic, proteomic, and clinical data for diseased patients and normal subjects. The dynamic redescription framework provides systems biology tools to model complex sets of regulatory, metabolic and signaling pathways in the initiation and progression of a disease. As an example, the case of chronic fatigue syndrome (CFS) is considered, which has so far remained intractable and unpredictable in its etiology and nosology. The redescription mining approaches can be applied to the Centers for Disease Control and Prevention's Wichita (KS, USA) dataset, integrating transcriptomic, epidemiological and clinical data, and can also be used to study how pathways in the hypothalamic–pituitary–adrenal axis affect CFS patients.

What can be our "responses to a disease thought to be intractable and capricious – that is, a disease not understood – in an era in which medicine's central premise is that all diseases can be cured?" [1]. Our views of such a disease are often multifaceted, metaphorical and, ultimately, mysterious. Unfortunately, as we begin to supplement the existing clinical views of a disease with more disease-related data, details and dimensionality, paradoxically they appear only to exacerbate our confusion and ignorance.

However, looking at those massive multidimensional measurements, we continue to entertain a hope that we will ultimately find the key insights to the disease from the voluminous data and elucidate its pathogenesis. Our hope is further strengthened with the availability of novel biomedical technologies and computational approaches to biomedical data analysis.

A statistically robust strategy for managing multiple views of a disease may be possible through the recently developed methods of redescription mining (RM). A redescription is a shift of vocabulary, or a different way of communicating information about a given subset of data. The goal of redescription mining is to find subsets of data that afford multiple descriptions. By filtering, evaluating, and cross-correlating these multiple redescriptions, we may be able to uncover the core biology of a disease.

Other methods provide similar approaches to data integration: two related techniques currently enjoying some degree of prominence being: information bottleneck (IB) and module network (MN) algorithms. There are many overlapping ideas among these three approaches: RM, IB and MN. Furthermore, it is suspected that they may belong to a common generalized framework.

As a primary example of statistical data-centric approaches to disease modeling, the case of chronic fatigue syndrome (CFS) may be considered. The agreed definitions for this syndrome consist of several easily verifiable clinical criteria: fatigue in such cases must be debilitating; fatigue must be present for 6 months or longer and, finally, CFS can be only diagnosed after ruling out other medical or psychiatric conditions that could cause fatigue. Nonetheless, patients suffering from CFS may vary widely with regard to accompanying symptoms, levels of functional impairment and exclusionary conditions. To date, CFS research has failed to yield any specific diagnostic laboratory abnormalities and, consequently, has even made it doubtful if it represents a single illness. No other measurable biochemical description of the disease has yet emerged, nor is there a correlated redescription of component patho-physiological symptoms of CFS in terms of co-morbid conditions, fatigue level and duration, functional impairment, or a more complex combinatorial formulation that could be composed out of these.

There have been several studies attempting to integrate peripheral blood gene expression results with epidemiological and clinical data to

Keywords: chronic fatigue syndrome, statistical analysis, redescription analysis

future medicine

determine the status of CFS: whether it is a single/unifaceted or heterogeneous/multifaceted disease. Redescription mining approaches could be helpful in cross-correlating clinical data segregated in to multiple facets or modules against the transcriptomic data also segregated to their modules in completely orthogonal manners. Of course, extensions to other viewpoints that can be inferred from genomic data (in terms of polymorphisms, single nucleotide polymorphisms [SNPs] or copy number polymorphisms [CNPs]), static or dynamic transcriptomic data, proteomic data, and clinical data, could further unravel the complex web of interrelationships, presentable through cores of common descriptions and redescriptions.

Fortunately, there is now a considerable amount of data available from the Centers for Disease Control and Prevention's (CDC's) Wichita surveillance study [2], presenting multiple views of CFS as a disease: currently, each patient data consists of clinical data (evaluation of a patient's medical and psychiatric status, stress history, sleep characteristics and cognitive functioning, laboratory test data, e.g., neuroendocrine status, autonomic nervous system function, systemic cytokine profiles, and so on; 227 patients), and transcriptomic data from peripheral blood (e.g., gene expression patterns measured with custom-built single-channel spotted arrays with gold labeling; 177 patients). The original dataset is planned to be augmented with more information: polymorphisms in genes (e.g., SNPs in the coding regions of the hypothalamic–pituitary–adrenal [HPA]-axis-associated genes involved in neurotransmission and immune regulation; 50 patients); and proteomic data (e.g., surface-enhanced laser desorption ionization time-of-flight mass spectrometry [SELDI-TOF] serum data with six fractionations and four assays per patient; 60 patients). The patients were selected by random sampling and were classified as:

- Those meeting the CFS research case definition (CFS)
- Those meeting the CFS research case definition, except that a major depressive disorder with melancholic features was identified (CFS-MMD)
- Those chronically fatigued but not meeting the CFS research case definition because of insufficient number of symptoms or fatigue severity (ISF)
- Those chronically fatigued but with ISF and a major depressive disorder with melancholic features (ISF-MDD)

For a controlled comparison, the Wichita study also selected 'normal' subjects from the same population: nonfatigued controls individually matched to CFS subjects on age, race/ethnicity, sex and body mass index (NF).

Redescription mining, in its simplest form, can be used to identify important atomic propositions from each view and to check if statistically meaningful relationships can be established between atomic propositions taken from two orthogonal views. For instance, one may look for a single gene whose overexpression can be used as a proxy for a closely related clinical criterion that distinguishes a CFS patient from a nonfatigued (NF) normal subject. If so, at this simplest level, each trait could be mapped to a single gene in a typical Mendelian manner. However, such a simple co-association map is unlikely to emerge for a disease as complex as CFS. Perhaps, one should expand the descriptions in each view to more complex formulations in a richer language, and search for one-to-one maps between complex sentences in the resulting extended vocabularies to establish relations among the multiple views. For instance, in the simplest possible extension, one could try to detect association between co-occurrences of multiple clinical criteria to differentially-expressed clusters of genes, and use these complex formulations to differentiate between CFS patients from normal subjects. Ultimately, redescription mining could extend such associations to set theoretic combinations of groups of subjects characterized by multiple clinical criteria, gene expression patterns, polymorphisms and proteomic profiles. Further enrichment can be achieved by combining the experimental data with other available domain knowledge that exists in various ontology and pathway databases, or can be obtained through additional discovery tools.

Redescription mining was originally proposed to analyze multi-'omics' biological data to extract significant relationships, latent in multiple views of a biological process. Ramakrisnan and colleagues also proposed a novel tree-based algorithm (classification and regression trees [CART]wheels) for mining redescriptions, and then applied it to biological problems as a way of generating plausible hypotheses that could be experimentally validated [3]. Intuitively, a redescription is a shift of vocabulary, or a different way of communicating information about a given subset of data. Naturally, redescription mining is ideal for dealing with biological experiments integrating multiple views.

Mathematically speaking, the inputs to redescription mining are the universal set of objects O (e.g., patients and normal subjects) and two sets (X and Y) of subsets of O. The elements of X are the descriptors X_i, and are assumed to form a covering of O:

$$(\bigcup_i X_i = O)$$

Similarly, $\bigcup Y_i = O$. The only requirements of a descriptor are that it be a proper nonempty subset of O, and denote some logical grouping of the underlying objects (for ease of interpretation). The goal of redescription mining is to find equivalence relationships of the form E ~ F that hold at or above a given Jaccard's coefficient, i.e.,

$$\left|E \cap F\right| / \left|E \cup F\right| \geq \theta$$

where E and F are set theoretic expressions involving X_is and Y_is, respectively. For tractability purposes, we may place some restrictions on the length of the allowable set theoretic expressions (but not on their form). Thus, redescription mining involves constructive induction (the task of inventing new features) and exhibits traits of both unsupervised and supervised learning, as noted elsewhere [3]. It is unsupervised because it finds conceptual clusters underlying data, and it can be viewed as supervised because clusters defined using descriptors are given meaningful characterizations (in terms of other descriptors).

In a rather simple illustrative setting, consider the set of all countries in the world. The elements of this set can be described in various ways (e.g., geographical location, political status, scientific capabilities, and economic prosperity). Such features allow us to define various subsets of the given (universal) set, called descriptors. Redescription mining in this setting may discover some nonobvious relationships, by describing a subset in two ways, for instance: 'countries with more than 200 Nobel prize winners' and 'countries with more than 150 billionaires' are two different closely-related descriptions of the same (singleton) set, namely the USA. Such relationships can be mined using techniques from the association rules literature, but the view afforded by redescription mining is much broader in scope, as it also includes set theoretic expressions involving descriptors, for example, 'countries with a defense budget of more than US$30 billion' and 'countries with declared nuclear arsenals' are the same as 'permanent members of UN Security Council' but not 'countries with a history of communism.' Note here that a set intersection on the left and a set difference on the right, from the given descriptors, has been constructed, and a redescription for the three-element set: USA, UK, and France, has been obtained.

To appreciate the power of redescription mining in the context of traditional transcriptomic analysis, gene expression studies in bioinformatics should be considered next. The universal set of genes in a given organism (O) can be studied in many ways, such as functional categorizations, expression level quantification using microarrays, protein interactions, and biological pathway descriptions. Each such methodology provides a different vocabulary to define subsets of O (e.g., 'genes localized in cellular compartment nucleus', 'genes up-expressed twofold or more in heat stress', 'genes encoding for proteins that form the immunoglobin complex', and 'genes involved in glucose biosynthesis'). Instead of following the traditional approach of improvised data mining heuristics to work with each of these vocabularies, redescription mining solves the problem elegantly, since it is able to characterize and analyze the results from any of them.

A naive approach to mining important biological patterns in data would be to first fix the form of the set theoretic expressions and then search within the space of possible instantiations. The more powerful CARTwheel algorithm, developed by Ramakrishnan and colleagues [3], achieves its power by simultaneously constructing set theoretic expressions and searching in the space of possible redescriptions. Such an algorithmic approach could prove enormously useful to tackle the multifaceted datasets incorporating the vast amount of biological information about chronic fatigue syndrome.

However, the Wichita dataset (either in its current form or with planned future extensions), as well as the approach to redescription mining, as described so far, are rather static. There should be a natural apprehension that a complete picture of a disease may not reveal itself through such an instantaneous depiction. As time course gene expression data become available, it would require that the redescription analysis become more flexible in the way it interrelates different components of the data (for example, at different instants). In addition, it would be necessary to extend the description language in which the temporal properties of the biological process could be captured. To fulfill these and other similar needs, a new algorithm, embodied in Gene Ontology Algorithmic Logic and Information Extraction (GOALIE) tool set, has been developed by the New York University (NYU) Bioinformatics group. GOALIE

redescribes numerical gene expression value measurements, sampled over a period of time, into formal temporal logic models of biological processes. It is designed to find extensive uses in the analysis of time course datasets from microarray and other high-throughput biological experiments.

As an example, consider the well known and well studied process of the regulation of the cell cycle in budding yeast. In a traditional diagrammatic representation of a biological process, the mitosis (M) phase is closely followed by cytokinesis and the G_1 phase, during which the cell grows but does not replicate its DNA. There is then a phase of synthesis (S) – DNA replication followed by G_2. Entry to S is carefully controlled where various cellular conditions are checked. If these conditions are not met, then the cell enters a quiescent phase (G_0) and might attempt to continue the cell cycle at a later stage. GOALIE, by examining time course gene expression data [4,5] for budding yeast and by combining the numerical data with qualitative process descriptions in the gene ontology (GO) database, can reconstruct essentially the same diagrammatic representation (formally captured in terms of a Kripke model). GOALIE's representation varies slightly as it splits the G_1 phase into two distinct subphases. It determines through its analysis that since entry to S is carefully controlled, G_1 should be treated in two parts: an early–mid part (G_1 [I]) during which the cell grows in size, and a later part (G_1 [II]) beyond which the cell is committed to undergoing one full cycle. It captures the intuition that G_1 (II) effectively acts as a checkpoint to ensure sufficient availability of nutrients, polypeptide mating factors, and significant growth in cell size.

In general, GOALIE deals with time course data in two logically distinct steps: it first constructs a Kripke model, consisting of labeled states and state transitions, in a manner similar to the diagrammatic representation of cell cycles, and next infers temporal properties that hold true in the Kripke model (and hence also the data), which can be succinctly represented in a propositional temporal logic.

Temporal logics are traditionally defined in terms of Kripke structures M = (V, E, L) [6,7]. Here (V, E) is a directed graph, having the reachable states of the system as vertices and state transitions of the system as directed edges. In the cell-cycle example, there are six states: M, G_1(I), G_1(II), S, G_2 & G_0, with directed edges connecting all except G_0 in one large cycle and separately, G_0 and G_1(I) in another smaller cycle. L is a labeling of the states of the system with properties that hold in each state, and are derived from the auxiliary ontological databases. To obtain a Kripke structure from a reachability graph, one first needs to fix a set of atomic propositions (APs), which denote the properties of individual states. For instance, we can define a proposition p to be 'cell size large enough for division'. Hence, p is not true in states M, G_1(I) and G_0. However, it becomes true in G_1(II). Once we have defined a vocabulary of such propositions, we replace the state symbols (M, G_1[I], and so on) with the set of atomic propositions that are determined to be true in that state. Thus, a Kripke structure can be automatically determined by GOALIE by first extracting the combinatorial graph structure and then labeling the vertices of the graph. The complete algorithm, as formulated in the IB framework, is technically more complex and can be found elsewhere [4,5]. Once a formal Kripke structure has been determined, we can reason about its properties, perform symbolic model checking, and answer queries about pathways. For instance, if we consider the additional propositions, q meaning 'cytokinesis takes place', r meaning 'DNA replication takes place', and s meaning 'cell is in quiescence', we can pose the question 'beginning from when q is true, is there a way to reach a state where r is true, without passing through a state where p is true?' (the answer is no). As another example, 'beginning from when q is true, is there a way to reach a state where r is true without passing through a state where s is true?' (the answer is yes). As is evident, Kripke structures constitute a powerful mechanism to reason about temporal characteristics of biological systems. Also, by changing the underlying vocabulary (atomic propositions labeling the states) we can also interrelate temporal descriptions resulting from different views.

The real power of this extended approach comes mostly from the recently developed efficient model checking algorithms. Upon an already derived Kripke structure model checker imposes a procedure for labeling the possible 'worlds' with more complex temporal formulae by appropriately combining other temporal subformulae that have been shown valid inductively. One can reduce these models to more comprehensible structures by projection and collapsing operations, while maintaining a bisimulation equivalence [6,7], for example, one can answer questions such as: when are two different experimental data sets qualitatively equivalent? Most importantly, one can query this model to see if a

particular biological property holds – one can examine a counter-example to a postulated query when it is falsified, or one may ask for hypothetical properties when certain new properties are speculated to hold true.

These algorithmic tools allow us to not only integrate the static data that have been accumulated from the CFS patients and NF subjects, but also to generate hypotheses about the causes and courses of progression for the disease, and thus, ultimately understand the critical underlying biological processes with detailed time course experiments. Sets of such hypotheses that could be investigated in the context of CFS involve the processes in the HPA axis.

The HPA axis is a classic neuroendocrine system that controls adrenocortical glucocorticoid secretion by the brain. These chemicals have a variety of effects on peripheral tissues, such as prioritizing energy use and distribution toward overcoming the homeostatic challenge posed by stress. Given the speculated connection between stress, the ability to tolerate stress and etiology of CFS, we may focus on understanding the relationship that exists among patient genotypes, lifestyle and environmental conditions, and progression of the processes involved in the HPA axis. However, these biological processes themselves are rather complex and related to each other in a complicated manner, as they affect metabolism, cardiovascular tone and immune reactivity [8,9,10]. The processes have also seemingly contradictory effects on mood and cognition that are controlled through both positive and negative feedback processes. In a normal situation, one suspects that the processes cooperate to provide, not only the ability to respond appropriately to stress, but also to minimize the deleterious effects of an excess of adrenocortical glucocorticoid. By comparing the time course data from CFS patients and NF subjects, a tool such as GOALIE can extract important temporal descriptions that distinguish one group from the other, and see how they are related to polymorphism and clinical data.

It should also be mentioned that many other competing and novel approaches are being actively investigated for the purpose of disease modeling and large-sale data integration. Two notable methods are:

- Graphical models, exemplified by module networks and Bayesian networks
- Information bottlenecks, exemplified by data clustering and compression

In the case of graphical models, the interrelationship among objects in various views are postulated *a priori*, but their exact degrees of statistical dependence are assumed unknown. These dependences are estimated algorithmically from the experimental data. This simpler structure can be further extended by also assuming the existence of hidden modules, whose local structures are then left to the inference engines. The IB theory is essentially a natural generalization of the rate distortion theory that was originally developed in communication engineering to understand design of optimal lossy compression. Using the IB theory, one could imagine a computational approach that attempts to obtain a simple and succinct description (i.e., a lossy compression) in one view, such that its natural mapping to the other related views introduces 'minimal amount of distortion'. Much work remains to be done to create a generalized viewpoint that combines all these notions in one general framework that redescription mining also attempts to achieve. Fortunately, some foundational work in this direction has already been accomplished.

Expert commentary

There are many difficult mathematical and statistical questions that will need to be thought through carefully if such a multidisciplinary approach is to succeed. Below only a few important questions are listed.

The primary among these questions is how can we achieve a better understanding of the nature of experimental noise, number of replicated experiments, number of subjects studied, and their cost–benefit relationships. For instance, a large fraction of the features in the clinical data have to be measured in a subjective manner and cannot yield a reliable numerical value that can be then computationally modeled. The temptation is to introduce new *ad hoc* features based on the researchers' understanding of the disease, which may seriously bias the analysis. All these issues must be addressed through systematic cost-effective improvement of measurement technologies, reproducible protocols and careful selection of the study.

On the computational side, these studies also require better algorithms to model the statistical nature (for example, distributions) of the noise, to reduce the noise and to normalize the data. Although by considering an increased number of features of the disease and multiple views, the chance of capturing the essential variables that will ultimately prove to be directly responsible for the

Highlights

- Chronic fatigue syndrome (CFS) is an illness that affects a large segment of the population with devastating social and economic impacts, but remains nebulous and mysterious in terms of its nosology and etiology.
- Recent advances in biotechnology, bioinformatics, and statistical data analysis have created the hope of obtaining a clearer description of this illness through a data-centric approach.
- The Wichita study data [2], created by the Centers for Disease Control and Prevention (CDC) to study 227 adults from Wichita, KS, USA, provide transcriptomic, epidemiological and clinical data for large subsets of these individuals, who have also been independently classified by CDC into three categories: CFS, nonfatigued (NF) and insufficient symptoms (ISF).
- This review paper describes some recent advances in redescription mining technology that could prove useful in integrating this dataset. Furthermore, other extensions of redescription mining to handle dynamic-datasets (embodied in the Gene Ontology Algorithmic Logic and Information Extraction [GOALIE] toolkit) also enable the understanding of CFS's etiology through the study of hypothalamic–pituitary–adrenal (HPA)-axis gene expression data, obtained by sampling over time.
- However, these statistical approaches suffer from many subtle pitfalls. In order to ensure that these approaches achieve their full potential to generate useful knowledge about CFS, careful interdisciplinary efforts should be targeted to develop tools rigorously and to apply them correctly.

disease is increased, the chance of overfitting the data is also increased. This is especially true if we fail to provide a corresponding increase in the number of data points (e.g., number of patients and normal subjects) to compensate for the increased dimension.

Thus, approaches such as redescription mining must combine with it sound statistical approaches based on shrinkage, dimension reduction, cross-validation, supervised learning, estimation of statistical significance, and so on, if it hopes to generate meaningful hypotheses. Furthermore, the questions of experiment design, cross-validation, hypotheses testing and disease modeling should be naturally viewed as different components of a larger monolithic enterprise. In summary, careful multidisciplinary data-centric approaches have to be designed by paying careful attention to biomedical, biotechnological, bioinformatic, computational, and statistical questions all in one inclusive framework.

Outlook

Based on the preceding discussions, one could develop optimism that the data-centric approach, which has been gaining momentum over the last decade, could eventually deliver powerful tools to tame the capriciousness of a disease such as CFS. It could be possible to settle whether CFS is in fact a heterogeneous disease, grouping together many different related ailments into one all-encompassing syndrome. If so, then the patients can be considered as being segregated into one of several important categories, with each category characterized by a clinical description and many associated redescriptions. Each such redescription in each category can then point to the responsible genotypes and polymorphisms (possibly with other influential environmental factors), diagnostic biomarkers, genes and pathways involved in the disease, and perhaps, even a process level description of the disease. This knowledge could be valuable in screening genomic data to determine subjects that are susceptible to a particular form of CFS and in helping them to modify their lifestyle or to choose a suitable profession. Ultimately, it could even bring CFS under the medicine's very hopeful central premise that 'all diseases can be cured'.

More indirectly, an interdisciplinary approach to CFS will also equip us with the necessary weapons to attack other diseases in similar status. The scientists from different fields and subfields will better understand how to effectively collaborate to generate techniques, technologies and theories to deal with biomedical problems. Several competitions set up by the Centers for Disease Control & Prevention (CDC) and an upcoming competition set up by Critical Assessment of Microarray Data Analysis (CAMDA) to study CFS-related data have been important steps in that direction. Still, there is also the educational question of how to better prepare our 'scientists of the future', who will need to tackle problems such as this much more effectively. At NYU, a bioinformatics course tried to address this issue head-on by designing its syllabi directly around the CFS-related data, available from CAMDA. Instead of CFS being a distraction to the main subject, the modified syllabi naturally organized the genomics, transcriptomic, proteomic and statistical analysis algorithm topics in a more meaningful and motivating manner. The students in the class were divided into five teams, each team dealing with one aspect of the data analysis: for example, clinical data for team 1, SNP data for team 2, gene expression data for team 3, proteomic data for team 4, and the data integration task for team 5. While it is too early to judge if this educational exercise has taken us any further in achieving an improved understanding of CFS, it is nonetheless clear that this group of students are now much better prepared in handling complex data more collaboratively (notwithstanding few bitter contentions and competitions within and across the groups). The course is planned to be repeated in the spring of 2006.

Bibliography

1. Susan Sontag: *Illness as Metaphor*. Farrar, Straus & Giroux, New York, NY, USA (1978)
2. Reeves WC, Wagner D, Nisenbaum R *et al.*: Chronic fatigue syndrome – a clinically empirical approach to its definition and study. *BMC Med.* 3(1), 19 (2005).
3. Ramakrishnan N, Kumar D, Mishra B *et al.*: Turning CARTwheels: an alternating algorithm for mining redescriptions, The Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. *SIGKDD* 266–274 (2004).
4. Antoniotti M, Ramakrishnan N, Mishra B: Reconstructing formal temporal models of cellular events using the go process ontology. Proceedings of the Eighth Annual Bio-Ontologies Meeting, (ISMB'05 Satellite Workshop), Detroit, MI, USA, June 23–24 (2005).
5. Antoniotti M, Ramakrishnan N, Mishra B: GOALIE, a common lisp application to discover kripke models: redescribing biological processes from time-course data. International Lisp Conference, ILC 2005, Stanford University, CA, USA, June 19–22 (2005).
6. Antoniotti M, Policriti A, Ugel N, Mishra B: Model building and model checking for biochemical processes. *Cell Biochem. Biophys.* 38(3), 271–286 (2003).
7. Mishra B, Antoniotti M, Paxia S, Ugel N: Simpathica: a computational systems biology tool within the valis bioinformatics environment. In: *Computational Systems Biology*. Eiles EA. Kriete A (Eds), Elsevier, Amsterdam, The Netherlands (2005).
8. Demitrack MA, Dale JK, Straus SE *et al.*: Evidence for impaired activation of the hypothalamic–pituitary–adrenal axis in patients with chronic fatigue syndrome. *J. Clin. Endocrinol. Metab.* 73, 1224–1234 (1991).
9. Sapolsky RM, Romero LM, Munck AU: How do glucocorticoids influence stress responses? Integrating permissive, suppressive, stimulatory, and preparative actions. *Endocr. Rev.* 21, 55–89 (2000).
10. Vernon SD, Reeves WC: Evaluation of autoantibodies to common and neuronal cell antigens in chronic fatigue syndrome. *J. Autoimmune Dis.* 25(2), 5 (2005).