We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

3

Integrated Analysis of Gene Expression and Genotype Variation Data for Chronic Fatigue Syndrome

Jungsoo Gim[1] and Taesung Park[1,2]
[1]*Intediscplinary program for bioinformatics, Seoul National University,*
[2]*Department of statistics, Seoul National University,*
South Korea

1. Introduction

In the past few years, high throughput technologies, such as gene expression microarrays and genotyping techniques, have provided efficient ways to measure gene expression levels and genotype variation on a genome-wide scale [Schena *et al.,* 1995; Howell *et al.*, 1999]. Various approaches have been proposed to analyse gene expression data and genotype variation data, in order to discover the complex network of biochemical processes of complex diseases such as chronic fatigue syndrome (CFS) [Presson *et al.,* 2008]. In the analysis of gene expression data, for example, the identification of differentially expressed genes between two groups has been of great interest, and various statistical tests have been conducted [Ghazalpour *et al.*, 2008; Brem *et al.*, 2002; Kang *et al.*, 2008]. In analysing genotype variation data, logistic regression has been commonly used to model the relationship between binary clinical outcomes and discrete predictors, such as genotypes [Henshall & Goddard, 1999; Coffey *et al.,* 2004].

Despite the availability of different levels of genome-wide data, most studies have been based on a separate analysis of single-level data to unravel complex biological mechanisms of CFS. Complex diseases such as CFS can be explained at different levels of biological mechanisms, including DNA, gene expression and phenotype levels. While there is a separate mechanism at each level, the mechanisms at different levels are closely related to each other in initiating and influencing CFS. Furthermore, CFS is expected to have complex etiology, which involves the action of many genes in addition to dynamic gene-environment interactions [Lin *et al.*, 2009]. Therefore, separate analyses of single-level data have a limitation in identifying and characterizing genes that are associated with the susceptibility of CFS. The integration of the different types of data (for example, gene expression, genotype variation and clinical outcomes) can provide more comprehensive information related to CFS, hence elucidate complex networks of gene interactions underlying CFS.

In this chapter, we provide an overview of the integrated statistical model (ISM) in order to characterize CFS, which involves integrating genotype variation data and gene expression data. The ISM elucidates the causal relationship between genetic variation, gene expression

level and disease. The ISM consists of two steps. The first step is to determine the causal relationship. Based on the causal relationship determined at the first step, the second step identifies significant gene expression traits of which the effects on disease status or the responses to disease status are modified by the specific genotype variation. By applying the ISM procedure to a CFS dataset, we identified a list of potential causal genes for CFS, and found an evidence for a difference in genetic mechanisms of the etiology between CFS patient and control groups.

Our ISM analyses considering the different levels of data simultaneously, allowed us to elucidate disease susceptibility and differentially expressed genes of genetically different individuals. Some results even showed that integrating genotype and expression data may help the search for new directions for the treatment of CFS that are not being detected by using only one type of data. The integrated analysis provided more information than the two separate analyses of gene expression data and genotype variation data for characterizing CFS that has several possible causes.

2. An overview of Integrated Statistical Model (ISM)

2.1 From genotype to phenotype

In the era of the genome project, the belief came with was that we would answer the questions on how the genes function and how they are related to diseases. The genome project successfully sequenced DNA of various species, including the human. Not only sequencing the genomes, many studies have also identified the gene functions by modifying individual genes in several animals and plants. However, many questions remain unanswered. We still do not know the functions of numerous genes, whether thay are annotated or un-annotated. Especially predicting what genes are associated with disease-related phenotypic variants is of particular interest and still in vague. The problem is complicated, because

i. most phenotypes of medical interest are **complex diseases**, *i.e.,* more than one gene or environmental effect contributes to the phenotype occurrence,
ii. the underlying **molecular mechanism** regulating cellular functions **is complicated,** and
iii. **little genotypic data** (or information) of disease-related phenotypes is available.

High throughput technologies advance for acquiring genome-wide genotyping data of many individuals with and without disease phenotypes. It is of a particular interest to segregate genotypic difference between disease-affected individuals and controls. The variation of genotypes comes from additive and epistatic effects of alleles across multiple genes, resulting in many individuals with phenotypes. Some combinations of genotypic variants result in enhanced traits, whereas other combinations are deleterious to fitness in specific environments. Phenotypic alterations are usually in matters of amount, rather than in the presence or absence of a trait. The field of statistical genetics has developed various methods and tools to map such quantitative traits to regions of chromosomes. These chromosomal regions are known as quantitative trait loci (hereafter QTLs) and are described in terms of the percentage of the variation of a trait that can be attributed to each region.

2.1.1 Quantitative Trait Locus (QTL)

Quantitative traits refer to the characteristics or phenotypes that are quantitative, *i.e.*, vary in degree or continuously, such as height, while dichotomous or discrete traits have two or several characteristic values. A QTL is a specific region of DNA that is associated with these quantitative phenotypic traits. The number of QTLs that explain the variation in the phenotypic trait tells us more about the genetic structure of a specific trait. For example, the research related to QTLs could provide further information about the genes that control human height.

2.1.2 xQTL: Various types of QTL mapping

Microarray technology has elucidated the genetics of gene expression in human populations. It has been less successful to identify genes in underlying diseases by using molecular profiling tools. Since too many genes have been identified to be associated with disease traits, determining and verifying which genes are the true disease-causing genes have been difficult.

Recently, microarray techniques have been combined with genotyping technology to facilitate the identification of key drivers of complex diseases. Figure 1 represents this approach, treating relative transcript abundances as quantitative traits when segregating populations. In this method, chromosomal regions that control the level of expression of a particular gene are mapped as expression quantitative trait loci (eQTL).

This eQTL scheme can be easily extended to other data types, for example, proteome, metabolome and phenome. Figure 2 illustrates this extension: protein expression (pQTL), relative metabolites abundances (mQTL) and phenotype abundances (phQTL).

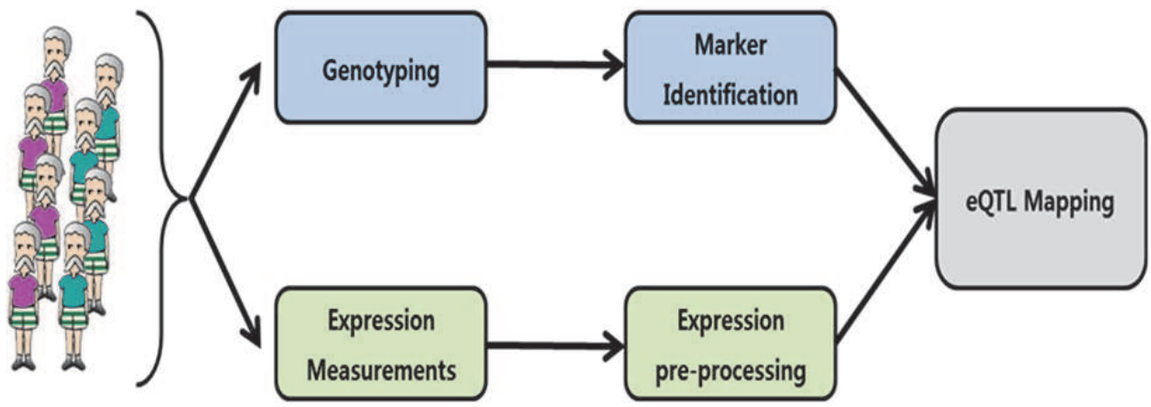


Fig. 1. eQTL pipeline. From disease and normal individuals, genotypes and mRNA expressions are observed.

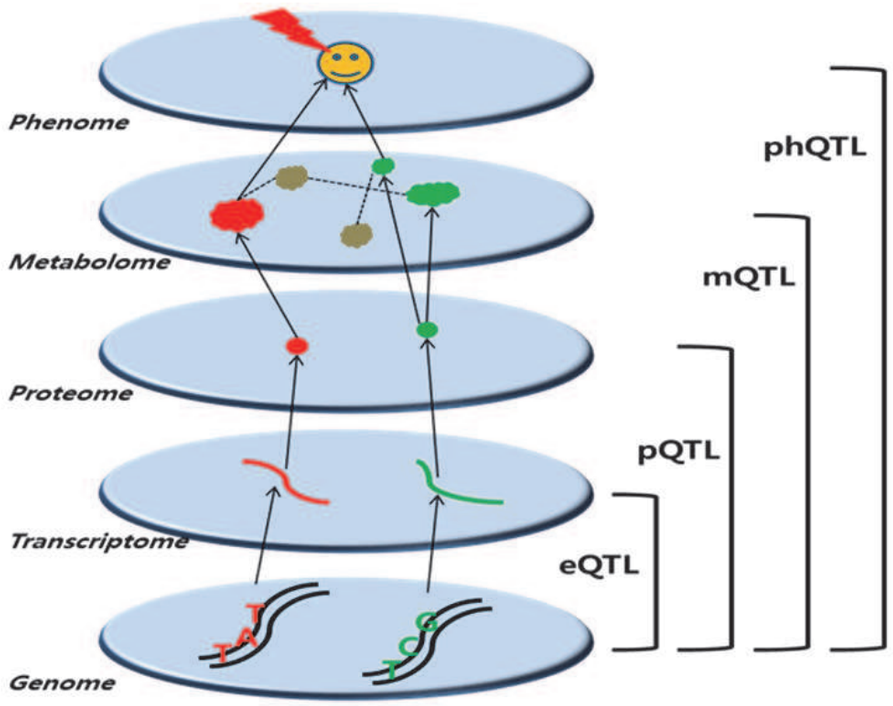


Fig. 2. A schematic representation of extended xQTL analyses.

2.2 Integrative analysis

Fu *et al*. provided the first system-wide evidence for phenotypic buffering in *Arabidopsis* [Fu *et al*., 2009]. Their approach consisted of three steps. Step 1 performed QTL mapping for transcript, protein, metabolite and phenotypic trait data. Then, Step 2 computed significance thresholds for detection of QTL hotspots per level, and finally, Step 3 detected hotspots that appeared across multiple levels. In particular, at Step 2 permutation analysis was used to compute significance thresholds for detecting QTL hotspots. For each of the 250 permutations, all > 40,000 traits were analyzed in order to map QTLs and the most significant marker for each QTL was stored. The number of significnat QTLs were counted over all traits for each marker, and the significant thresholds for hotspot detection per level were derived. For system-wide or multiple level QTL hotspots, Step 3 used the observed QTL hotspots and permutation analysis to compute significance thresholds for detecting QTL hotspots that appeared at multiple levels. Using the results obtained from per-level analysis, the markers per level were ranked from the one with the highest number of traits mapping to it, to the one with the lowest. Then, a rank-product test was performed to find markers that ranked significantly high at multiple levels [Breitling *et al*., 2004]. For each permutated sample, the *p*-value was computed for the rank-product test at each of the 144 markers, and a threshold was derived for hotspot detection by the procedure controlling 5% of the false discovery rate (FDR) [Benjamini & Hochberg, 1995].

Using this approach, 162 recombinant inbred lines (RIL) of *Arabidopsis thaliana* were profiled for variation in transcript, protein and metabolite abundance, and were mapped to QTL for 40,580 of these molecular traits. Only six QTL hotspots were found which underlied variation in 16% of the transcript traits, 25% of the protein traits, 55% of the metabolite traits

and 77% of the phenotypic traits for which QTL could be mapped. QTL for 16%, 25%, 55% of all transcript, protein and metabolite traits with a QTL, respectively, mapped to the same six QTL hot spots, compared to 77% of phenotypic traits. Consequently, screening for mutants at the molecular level would increase the probability of identifying new causal loci that could not be identified from morphological screens [Boerjan & Vuylsteke, 2009].

Using microarrays or massively parallel sequencing it is possible to measure both genetic variation and gene expression at genomic level. Hence, eQTL methods allow for studying the association of all regions in a genome with the expression of all genes. In this sense it is worth re-visiting eQTL in deeper look.

If the genotype at a certain locus is associated with the phenotype of a certain gene, this DNA region might contain a regulator of the target gene expression. It could be any functional nucleotide sequences such as protein-coding regions, microRNAs and *cis*-regulatory DNA motifs. The same individuals of a selected population have to be genotyped and phenotyped first. Based on the genotyped data (*e.g.* SNP), selecting markers that are polymorphic in the study population is in need. Then, at the heart of every eQTL study is the correlation of genotype patterns with expression levels in a genetically diverse population. The simplest mapping strategy is to split the population based on the genotypes at a specific marker and check if the expression levels of a given gene are significantly different between the two groups [Ghazalpour *et al*., 2008; Brem *et al*., 2002; Kang *et al*., 2008].

There have been many approaches to elucidate the variants affecting phenotypes, for example, Lan *et al.* explored correlation of expression profiles across a genetic dimension, namely genotypes segregating in a panel of 60 F_2 mice derived from a cross used to explore diabetes in obese mice. By combining the correlation results with linkage mapping information, they identified regulatory networks, made functinoal predictions for uncharacterized genes, and charaterized novel elements of knwon pathways [Lan *et al.*, 2006]. However, their approach did not provide any information about causality relationships among expression profile, genotype and disease.

The mixture over markers (MOM) model proposed by Kendziorski *et al.* combinds a transcript-based (TB) approach, refering to the repeated application of any single-phenotype mapping method to each mRNA transciprt, and a marker-based (MB) approach, refering to the repeated application, at each marker, of any method for identifying differnetially expressed transciprts [Kendziorski *et al.,* 2006]. They applied two MB approches: an empirical Bayes approach and an approach based on the Student's *t*-test. The MOM model is motivated from the fact that separate tests are conducted for each trascript-marker pair, and each measures evidence that the transcipt maps to that marker relative to evidence that it maps nowhere. Since a trancript can map to any of various marker locations, the evidence that a transcript maps to a particular marker should not be judged relative only to the possibility that it maps nowhere, but rather relative to the possibility that it maps nowhere or to some other markers. This model was proved useful in improving the specificity of eQTL identifications, but used only genotype variation and gene expressino data rather than disease status or trait data.

A gene-set approach based on weighted gene co-expression network analysis (WGCNA) by Presson *et al.* constructs a co-expression network, identifies trait-related modules within the network, uses a trait-related genetic marker to prioritize genes within the module, applies

an integrated gene screening strategy to identify candidate genes and carries out causality testing to verify and/or prioritize results [Presson *et al.,* 2008]. Their work includes steps to identify trait-module association and trait-related genetic marker association, but does not provide the model-based statistical tests.

The step-wise approach proposed by Schadt *et al.* includes i) identifying pair-wise relationships among genotype variation, gene expression, and a complex trait, respectively investigated by identifying QTLs for the complex trait, ii) selecting gene expression traits correlated with the complex trait, iii) detecting eQTL, which overlap the identified QTL, for the selected expression traits; and iv) the likelihood based causality model selection (LCMS) test to identify the causal relationships of the genes detected with overlapping loci [Schadt *et al.,* 2005].

3. Two-step integrative analysis

Schadt *et al.*'s approah has two major limitations. First, although the filtering step is effective in reducing the search-space, it might result in more false negatives than exhaustive search approahes in detecting causal relationships of the genes, espeically when a true causal relationship exists based on the interaction effects among genotype, gene expression and a trait of interest, but any pairwise association is weak. Second, the model does not comprehensively handle the interaction effects, which might cause different disease susceptibility. Therefore Lee *et al*. proposed a two-step integrative approach handling with exhaustive search and interaction effects based on LCMS test [Lee *et al.,* 2009]. In this section we provide a detailed review of the Lee *et al.*'s two-step procedure integrating genotype data, gene expression and clinical data, and thus elucidating mechanisms underlying disease susceptibility and progression [Lee *et al,* 2009].

3.1 Introduction

In figure 3, the two-step procedure is presented to illustrate the integration method based on causal relationship among the three different levels of data. In the first step, the most appropriate causality models are selected to understand the causal relationship among genetic variation, gene expression level, and disease for each gene expression-genetic variation combination. In the second step, significance testing is carried out based on a

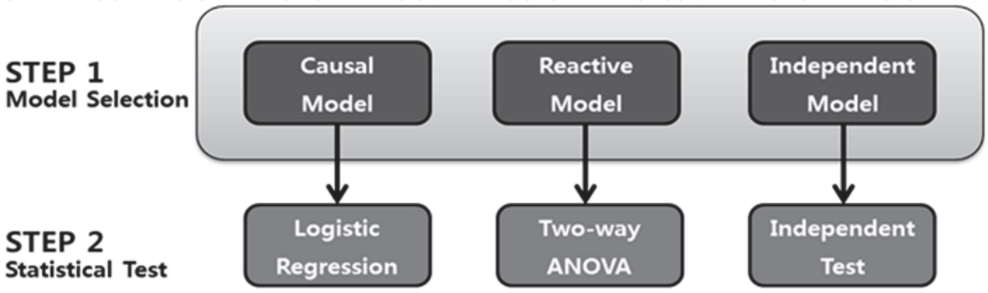


Fig. 3. Two-step procedure illustration of Lee *et al.'s*. In the first step, for each gene expression-genetic variation combination, the most appropriate causality models are selected. Then in the second step, significance test is carried out based on a statistical model for each combination according to the model selected in previous step.

statistical model for each combination, such as logistic regression and a two-way analysis of variance (ANOVA), according to the causality model selected from the first step. Through these tests, gene expression traits whose effects on disease status or responses to disease status are modified by the genotype variation effects.

3.2 The first step: Causality model selection

The possible causal relationships among genetic variation, gene expression level and disease trait, can be summarized as three models. Figure 4 represents three simple models. Causal model assumes the simplest causal relationship with respect to mRNA expression, in which QTL acts on disease through transcript. Reactive model is the model with respect to mRNA expression, in which mRNA expression is modulated by disease. In independent model, QTL at a specific locus acts on these traits independently.

Lee *et al.* assumed that each pair of genetic locus and expressed gene has one of these three simple causal relationships to examine potential relationships among the genotype variation, gene expression level and disease status. In order to find the most possible causal relation, both Lee *et al.* and Schadt *et al.* adapted the likelihood based causality model

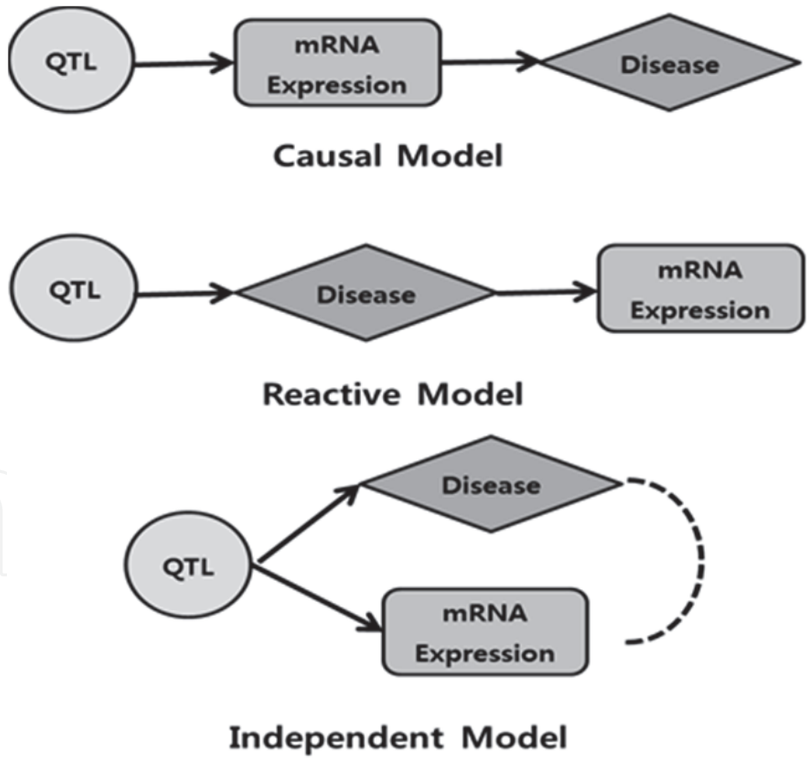


Fig. 4. Three possible causal relationships among genotype variation, mRNA level and complex disease proposed by Schadt *et al*. QTL, mRNA and disease represent any genotype variation like SNP, mRNA expression level of a gene, and complex disease or phenotype of interest, respectively.

selection (LCMS) test, which uses conditional correlation measures for determining the relationships best supported by the data. Unlike multistep procedure of Schadt *et al.*'s method, Lee *et al.* constructed likelihoods associated with each of the causality model and maximized with respect to the model parameters. Then, the best model was selected for each SNP-transcript combination, by choosing the model with the smallest Akaike Information Criterion (AIC) value which can be used to compare different models [Akaike, 1974).

Lee *et al.* and Schadt *et al.* both assumed standard Markov properties for the simple graphs (Fig. 4), the joint probability distributions for the three models are as follows:

- Causal Model: $P(S, R, D) = P(S)P(R|S)P(D|R)$,
- Reactive Model: $P(S, R, D) = P(S)P(D|S)P(R|D)$,
- Independent Model: $P(S, R, D) = P(S)P(R|S)P(D|R, S)$,

where S represents a genotype variation, R gene expression, and D disease status. P(S) is the genotype probability distribution for marker S and is further assumed to be co-dominant. $P(R|S)$ and $P(R|D)$ are the conditional probabilities of R given genotypes S and disease status D, respectively. Lee *et al.* further assumed that the random variable R follows conditional normal distribution, and the random variable D has a binomial distribution. Therefore, in probability $P(D|R)$, the random variable D has a binominal distribution with a success probability that can be modeled by a logistic regression model. $P(D|S)$ is the probability distribution of D conditional on locus S, in which the random variable D also has a binomial distribution. Based on these assumptions, the likelihood of a correspondence to each of the joint probability distributions can be constructed. For each model, the model parameters can be estimated via a standard maximum likelihood method. The best model supported by the data is then chosen based on the AIC, which is commonly used to compare models with different numbers of parameters [Schadt *et al.*, 2005; Lee *et al.* 2009]

3.3 The second step: Statistical test

Step 2 performs statistical tests to determine the significance of the genetic regulatory relationships described in the causality model selected at step 1. The response and independent variables in the statistical models depend on the causality model selected at step 1. These statistical tests can deal with the interaction effects among the three different levels of data and to elucidate differences in disease susceptibility and gene expression pattern across genetically different individuals. The two-step procedure can result in a set of candidate causal and reactive genes, whose expressions affect disease status and respond to disease status under the influence of genotype variation, respectively.

3.3.1 The causal model

In order to investigate gene expression traits whose effects on disease status are modified by genotype variation, the interaction effect of genotype variation and gene expression level on the disease status can be examined using logistic regression below:

$$\log it(\pi) = S + R + S \times R, \quad (1)$$

where π represents the probability of getting the disease; S represents the effect of genotype variation such as SNPs; R represents the effect of gene expression levels; and S × R represents the interaction effect between genotype variation and gene expression level.

3.3.2 The reactive model

For investigating gene expression traits whose responses to disease status are affected by genotype variations, one can fit the following two-way ANOVA model with the interaction between genotype variation and disease groups:

$$R = S + D + S \times D, \tag{2}$$

where S represents the effect of genotype variation; D represents the effect of disease groups; and S × D represents the interaction effect between genotype variation and disease groups.

3.3.3 The independent model

When the independent model is selected at step 1, the effect of genotype variation on each of gene expression and disease can be investigated separately. First, the logistic regression is employed to detect genotypic markers linked to disease loci:

$$\log it(\pi) = S. \tag{3}$$

Next, it is possible to identify genotypic markers that regulate gene expression levels, based on the one-way ANOVA model where the dependent variable is R and the independent variable is S.

In step 2, significant associations among genotype variation, gene expression and disease status are declared via statistical tests for all possible pairs of gene expression-genotype variation. Due to the large number of tests, the multiple-testing problem needs to be addressed. In order to adjust this multiplicity, Lee *et al.* used a step-up procedure controlling false discovery rate (FDR) [Benjamini & Hochberg, 1995].

4. Application

Lee *et al.* applied their two-step procedure to chronic fatigue syndrome (CFS) data to elucidate a list of potential causal genes of CFS. In this section, we provide the application of two-step procedure of Lee *et al.*'s

4.1 Chronic Fatigue Syndrome (CFS) dataset

Chronic fatigue syndrome (CFS) is a debilitating illness lacking consistent anatomic lesions and eluding conventional laboratory diagnosis. CFS has no confirmatory physical signs or laboratory abnormalities, and its etiology and pathophysiology are unknown. This disease characterized by chronic fatigue, lasting at least 6 months, which is accompanied by symptoms such as impairment in short-term memory or concentration, sore throat, tender lymph nodes, and muscle pain. The Centers for Disease Control and Prevention (CDC) Chronic Fatigue Syndrome Research Group produced the dataset including gene expression

of 177 subjects, proteomic of 60 subjects, single nucleotide polymorphism (SNP) of 50 subjects, and clinical data of 227 subjects. All the data set is available on the following web site (*http://www.camda.duke.edu/camda06/datasets/index.html*).

According to severity of symptoms, the patients were originally classified into five groups of CFS. Lee *et al.*'s study, however, only consider three groups of total 101 subjects: 46 subjects meeting the CFS research case definition (CFS), 19 subjects meeting the CFS research case definition and having 'a major depressive disorder with melancholic features' (CFS-MDD/m), and 36 subjects who show no fatigue (NF).

This CFS dataset has been analysed by many research groups for identifying molecular markers and elucidating pathophysiology of CFS, for finding two differentially expressed genes related with fatigue and depression, respectively, for discriminating classes of unexplained chronic fatigue based on differential gene expressions, and for examining the relationship between CFS and allostatic load based on the clinical dataset. In the CFS dataset, the expression levels of 20,160 genes were assessed from peripheral blood mononuclear cells, via custom-printed single-channel oligonucleotide chips. Quantile normalization was conducted on the gene expression data which were pre-processed by the original CDC research group. For genotype data, the whole blood DNA was extracted and specific areas of the genes of interest were amplified by PCR.

For illustration, we summarized the analyses results from the multi-step procedure of Schadt *et al.* and the two-step approach of Lee *et al.* The detailed description of the results is provided in Lee *et al.* [Lee *et al.*, 2009].

4.2 Results

4.2.1 Multi-step procedure by Schadt *et al.*

The multi-step procedure proposed by Schadt *et al.* was applied to the same datasets for the purpose of comparison. First, a gene expression analysis was carried out to detect differentially expressed genes across clinical outcomes. Only a few genes were identified as differentially expressed (Table 1A) by three commonly used approaches such as the t-test, significance analysis of microarray (SAM) [Tusher *et al.*, 2001] and the Bayesian regression approach [Baldi & Long, 2001]. Second, genotype variation data and clinical outcomes were analyzed via logistic regression to detect the susceptibility genes of disease. Out of all 41 markers tested, nine markers were detected with significant genotype effect on initiation of CFS at a 5% significance level, while only four markers were detected with 5% FDR [Benjamini & Hochberg, 1995] (Table 1B). Interestingly, different sets of susceptible genes were identified as having statistically significant association with CFS and CFS-MDD/m. From the CFS vs. NF comparison, the seven markers in the NR3C1 gene were identified as significant markers linked to CFS. On the other hand, the CFS-MDD/m vs. NF comparison revealed the two significant markers in the COMT gene. Finally, for each of the differentially expressed genes across clinical outcomes, eQTL were searched at each of the markers that were identified at the second step, via one-way ANOVA of genotype variation and gene expression data. No significant association between gene expression level and genotype variation was found for any genotype–gene expression combination at a 5% significant level.

Dataset	*t*-test	SAM test	Bayesian model		
A. Number of genes with significant change in expression levels over different disease status, which were detected via *t*-test, SAM test and Bayesian model.					
CFS vs. NG	1	2	0		
CFS-MDD/m vs. NF	1	1	0		
Gene	**SNP[a]**	**Chromosome**	**Position(Mb)[b]**	**CFS vs. NF**	**CFS-MDD/m vs. NF[c]**
B Significant genotype variation linked to disease loci, which were detected via logistic regression					
NR3C1[e]	rs2918419	5	142.641	0.0104	0.3950
	rs1866388	5	142.702	0.0010[f]	0.0472
	rs860458	5	142.739	0.0104	0.3950
	rs852977	5	146.642	0.0035[f]	0.1878
	rs6196	5	146.660	0.0208	0.6423
	rs6188	5	146.667	0.0027[f]	0.0396
	rs258750	5	146.674	0.0035[f]	0.1009
COMT[g]	rs933271	22	18.311	0.0649	0.0025
	rs5993882	22	18.317	0.4306	0.0114

Table 1. parallel analyses for respective association of gene expression and genotype variation with disease status (by courtesy of the authors) [Lee *et al.,* 2009]

As multiple filtering steps is Schadt et al.'s procedure, the separate analyses were conducted respectively on two datasets, CFS vs. NF groups and CFS-MDD/m vs. NF groups. Bold numbers indicate p-values < 0.05.

[a] NCBI dbSNP Build number is 125 using Human Genome Build 35.1

[b] Position of SNP on chromosome.1

[c] p-value from logistic regression with CFS vs. NF data.

[d] p-value from logistic regression with CFS-MDD/m vs. NF data.

[e] Glucocorticoid receptor located at 5q34.

[f] Significant at the 5% false discovery rate (FDR).

[g] Catechol-O-methyltransferase located at 22q11.1.

In other words, no significant results were detected for both datasets from the Schadt *et al.*'s multi-step method.

4.2.2 Two-step integrative analysis

Lee *et al.* analyzed each combination of 20,160 genes and 41 SNPs with their two-step integrative analysis on two datasets, CFS vs. NF groups and CFS-MDD/m vs. NF groups. For each gene–SNP combination, the best causal relationship was detected via the causality model selection at step 1. In comparing CFS with NF groups, the reactive model was selected for ~70% of 20,160 genes on average, for all nine markers within two known CFS-related genes, such as NR3C1 and COMT (Table 2). However, in comparing CFS-MDD/m with NF groups, the causal model was selected for nearly 70% genes for three markers in the NR3C1 gene. This different tendency in the model selection results between CFS and CFS-MDD/m would imply different genetic mechanisms of CFS and CFS-MDD/m.

At step 2, each gene–SNP combination data was analyzed based on one of the three statistical models, corresponding to the detected causal relationship. For all seven SNPs within NR3C1, significant causal relationships with gene expression levels were detected for either or both datasets (Table 2). Three SNPs (rs258750, rs6188 and rs852977) showed significant relationships with expression levels of a large number of genes, and can be candidates for genetic modulators of CFS- related regulatory pathways.

Gene	SNP	CFS vs. NF			CFS-MDD/m vs. NF		
		Causal[a]	Reactive[b]	Independent[c]	Causal[a]	Reactive[b]	Independent[c]
NR3C1[d]	Rs2918419	0 (639)	7 (16,215)	0 (3306)	8 (13,955)	3 (5912)	0 (293)
	Rs1866388	0 (165)	0 (16,872)	0 (3123)	15 (4136)	71 (15,976)	0 (48)
	Rs860458	0 (639)	7 (16,215)	0 (3306)	8 (13,955)	3 (5912)	0 (293)
	Rs852977	0 (230)	0 (17,001)	0 (2929)	**120 (9760)**	**73 (10,139)**	**0 (261)**
	Rs6196	0 (604)	2 (15,037)	0 (4519)	0 (16,278)	0 (2013)	0 (1869)
	Rs6188	0 (171)	7 (16,970)	1 (3019)	**52 (2939)**	**217 (17,074)**	**0 (147)**
	Rs258750	**0 (242)**	**0 (16,279)**	**105 (3639)**	0 (2769)	14 (12,590)	0 (4801)
COMT[e]	Rs933271	0 (1943)	0 (15,156)	0 (3061)	0 (169)	0 (16,872)	0 (3119)
	Rs5993882	0 (1022)	0 (14,380)	0 (4758)	0 (547)	0 (17,333)	0 (2280)

Table 2. Two-step integration based on causality model selection. (by courtesy of the authors) [Lee *et al.*, 2009]

The integrative analyses were conducted respectively on two datasets, CFS vs. NF groups and CFS-MDD/m vs. NF groups. Note that the results are presented only for nine SNPs within two known CFS-related genes (NR3C1 and COMT). For each combination of 20,160 genes and 41 SNPs, the best causal relationship was detected via causal model selection at step 1. Numbers in parenthesis indicate the numbers of genes having each causal relationship with each SNP and disease status. At step 2, each gene-SNP combination data was analyzed based on one of the three statistical models, corresponding to the detected causal relationship. Outside parenthesis, we present the numbers of significant genes identified by the corresponding statistical models. Three SNPs, each of which involves significant causal relationships with expression levels of more than 100 genes, are marked in bold.

[a] Logistic regression was conducted to identify genes whose expressions have interaction effect with genotype variation on disease status.

[b] Two-way ANOVA was conducted to identify genes whose expressions are affected by interaction between genotype variation and disease status.

[c] Independent test was conducted to identify genes whose expressions differ according to SNP genotypes.

[d] Gluccorticoid receptor located at 5q34.

[e] Catechol-O-methltransferase located at 22q11.1.

Next, pathway enrichment analyses were performed for these three SNPs, and the results are given in the next section. In comparing CFS with NF groups, for the rs258750 marker, 105 genes were identified with differential expression across genotypes with 5% FDR from the independent test. This result is supported by the evidence of the neuroendocrine regulation of immunity, because the gene expression data were obtained from a mononuclear cell, and the role of glucocorticoid receptor (NR3C1) gene is to regulate the level of glucocorticoid.

In the integrated analysis for comparing CFS-MDD/m with NF groups, for the rs6188 marker in the NR3C1 gene, 52 genes showed significant interaction effects with the rs6188 marker on disease status CFS-MDD/m from the logistic regression model. Also, the two-way ANOVA models yielded 217 candidate reactive genes, on which there are significant interaction effects between disease status and genotypes. Note that these candidate genes, especially reactive genes, could not be detected by Schadt *et al.*'s method. The Lee *et al.*'s two-step integration method revealed the causal association among gene expression level, genotype and disease status in depth. Candidate causal/reactive genes were detected also for rs852977 in the NR3C1 gene. However, the candidate gene set for the rs852977 is very similar to that for the rs6188, with slight differences in causality structure. This similarity would be due to a strong linkage between the two SNPs.

4.2.3 Pathway enrichment analysis

In comparing CFS with NF groups, Lee *et al.* further conducted a pathway enrichment analysis for 105 genes that were identified to have a significant relationship with the rs258750 marker from the independent test at step 2. The pathway classification showed that nine different pathways were associated with the rs258750 marker at the 5% significance level (Table 3). Out of nine pathways, four were enriched with genes involved in regulation of transcription, translation or mRNA processing, and three are related with immune system.

For comparing CFS-MDD/m with NF groups, pathway enrichment analyses were conducted on the genes that were identified to have a significant relationship with the rs6188 and/or rs852977 markers at step 2. Because of the linkage between the two SNPs, the results were similar (Tables 4 and 5), and the results was given only for the rs6188. While seven different pathways were detected at the 5% significance level for the 52 candidate causal genes, eleven different pathways were detected for the 217 candidate reactive genes (Table 4). In addition, two other pathways, whose p-values were slightly larger than the 5% significance level, are listed.

In pathway enrichment analyses of the candidate causal genes, the steroid biosynthesis pathway appears to have a direct causal effect on the disease status, CFS-MDD/m, through an integrative action of the rs6188 marker within the NR3C1 gene. The two significantly enriched biological pathways (i.e., 'IL-2 Receptor Beta Chain in T cell Activation', and 'HIV-1 Nef: negative effector of FAS and TNF') are all related to the immune system. On the other hand, the pathway enrichment analysis of the candidate reactive genes showed that several pathways related to lipid metabolism or biosynthesis, such as eicosanoid and fatty acid, appear to be important for responding to CFS-MDD/m. Furthermore, other pathways associated with neuron physiology and neurotransmitters appear to respond to CFS-MDD/m.

Pathway[a]	Model[b]	Source[c]	Nodes[d]	Gene ID [e]	Gene name
Galactose metabolism	Indepe-ndent	KEGG	2/22	B4GALT2 MGAM	UDP-Gal:betaGlcNAc beta 1,4-galactosyltransferase, polypeptide2 Maltase-glucoamylase
Basic mechanisms of SUMOylation	Indepe-ndent	BioCarta	1/4	SUMO3	SMT3 suppressor of mif two 3 homolog 3
Internal ribosome entry pathway	Indepe-ndent	BioCarta	1/8	EIF4E	Eukaryotic translation initiation factor 4E
Neutrophil and its surface molecules	Indepe-ndent	BioCarta	1/8	ITGB2	Integrin, beta 2
Alternative complement pathwy	Indepe-ndent	BioCarta	1/9	CFB	Complement factor B
Mechanisms of protein import into the nucleus	Indepe-ndent	BioCarta	1/9	NUP62	Nucleoporin 62kDa
Polyadenylation of mRNA	Indepe-ndent	BioCarta	1/9	PABP2	Poly(A) binding protein II
B Lymphocyte cell surface molecules	Indepe-ndent	BioCarta	1/9	ITGB2	Integrin, beta 2
Adhesion molecules on lymphocyte	Indepe-ndent	BioCarta	1/9	ITGB2	Integrin, beta 2

Table 3. Significant regulated pathways for SNP rs258750 (by courtesy of the authors) [Lee *et al.*, 2009]

Pathway enrichment analysis was conducted using 105 candidate independent genes, which were identified for rs258750. Significant biological pathways were detected via Fisher's exact test at a 5% significance level. Pathways are listed in order of significance e.g., most significant pathway presents at the top.

[a] Name of biological pathway selected by Fisher's exact test.

[b] Causality models selected ay step1.

[c] Source of pathway

[d] The number of candidate causal/reactive genes associated with pathway/the number of all genes associated with pathway.

[e] Gene ID of candidate genes associated with pathway

Pathway[a]	Model[b]	Source[c]	Nodes[d]	Gene ID [e]	Gene name
Electron transport chain	Causal	GenMapp	2/105	COX11 COX6A1	Cytochrome c oxidase subunit11 Cytochrome c oxidase subunit Via polypeptide 1
Steroid biosynthesis	Causal	GenMapp	1/9	F13B	Coagulation factor XIII,B polypeptide
Blood clotting cascade	Causal	GenMapp	1/19	F13B	Coagulation factor XIII,B polypeptide
FAS signaling pathway(CD95)	Causal	BioCarta	1/30	CFLAR	CASP8 and FADD-like apoptosis regulator
Induction of apoptosis through DR3 and DR4/5 Death Receptor	Causal	BioCarta	1/32	CFLAR	CASP8 and FADD-like apoptosis regulator
IL-2 receptor beta chain in T cell activation	Causal	BioCarta	1/35	CFLAR	CASP8 and FADD-like apoptosis regulator
HIV-1 Nef:negative effector of FAS and TNF	Causal	BioCarta	1/57	CFLAR	CASP8 and FADD-like apoptosis regulator
Agrin in postsynaptic differentiation	Reactive	BioCarta	3/39	UTRN DVL1 ARHGEF6	Utrophin Dishevelled,dsh homolog1 Rac/Cdc42 guanine nucleotide exchange factor(GEF)6
Cell cycle	Reactive	GenMapp	4/87	CDC14A E2F2 CDC20	CDC14 cell division cycle 20homolog E2F transcription factor2 CDC20 cell division cycle 20homolog
Eicosanoid metabolism	Reactive	BioCarta	2/20	PTGES EPHX1	Prostaglandin E synthase Epoxide hydrolase
Biosyntheisis of cysteine	Reactive	BioCarta	1/2	CBS	Cystathionine-beta-synthase
Biosyntheisis of threonine and methionine	Reactive	BioCarta	1/2	CBS	Cystathionine-beta-synthase

Pathway[a]	Model[b]	Source[c]	Nodes[d]	Gene ID [e]	Gene name
Inactivation of Gsk3 by AKT causes accumulation of β-catenin in alveolar macrophages	Reactive	BioCarta	2/25	MYD88 DVL1	Myeloid differentiation primary response gene (88) Disheveled, dsh homolog 1
Fatty acid metabolism	Reactive	KEGG	3/57	HADHB	Hydroxyacyl-Coenzyme A dehydrogenase/3-ketoacyl-Coenzyme A thiolase/enoyl-Coenzyme A hydraatase, beta subunit
Bile acid biosynthesis	Reactive	KEGG	2/26	ADH6	Alcohol dehydrogenase 6 (class V)
Catabolic pathways for methionine,isoleucine,threonine And valine	Reactive	BioCarta	1/4	CBS	Cystathionie-beta-synthase
Basic mechanisms of SUMOylation	Reactive	BioCarta	1/4	SMT3H1	SMT3 suppressor of mif two 3 homolog 3
ALK in cardiac myocytes	Reactive	BioCarta	2/34	DLV1 CHRD	Chordin
Taurine and hypotaurine metabolism[f]	Reactive	KEGG	1/5	GAD1	Glutamate decarboxylase 1
Biosynthesis of neurotransmitters[f]	Reactive	BioCarta	1/6	GAD1	Glutamate decarboxylase 1

Table 4. Significant regulated pathways for SNP rs6188 (by courtesy of the authors) [Lee *et al.,* 2009]

Pathway enrichment analysis was conducted using 52 candidate causal genes and 217 candidate reactive genes, which were identified for rs6188. Significant biological pathways were detected via Fisher's exact test at a 5% significance level. Pathways are listed in order of significance within each of causality models, e.g., most significant pathway presents at the top.
[a] Name of biological pathway selected by Fisher's exact test.
[b] Causality models selected ay step1.
[c] Source of pathway
[d] The number of candidate causal/reactive genes associated with pathway/the number of all genes associated with pathway.
[e] Gene ID of candidate genes associated with pathway.
[f] Pathways with *p*-value that is slightly larger than 0.05.

Pathway[a]	Model[b]	Source[c]	Nodes[d]	Gene ID [e]	Gene name
Agrin in postsynaptic differentiation	Causal	BioCarta	2/39	DMD DVL1	Dystrophin Dishevelled, dsh homolog 1
Steroid biosynthesis	Causal	Gen MAPP	1/9	F13B	Coagulation factor XIII, B polypeptide
Nucleotide GPCRs	Causal	Gen MAPP	1/10	P2RY4	Pyrimidinergic receptor P2Y, G-protein coupled 4
RNA polymerase III transcription	Causal	BioCarta	1/8	GTF3C1	General transcription factor IIIC, polypeptide 1, alpha 220kDa
Blood clotting cascade	Causal	Gen MAPP	1/19	F13B	Coagulation factor XIII, B polypeptide
Bile acid biosynthesis	Causal	KEGG	1/26	ADH6	Alcohol dehydrogenase 6
Tyrosine metabolism	Causal	KEGG	1/37	ADH6	Alcohol dehydrogenase 6I
Inactivation of Gsk3 by AKT causes accumulation of b-catenin in alverolar macrophages	Reactive	BioCarta	1/25	MYD88 DVL1	Myeloid differentiation primary response gene (88) Dishevelled, dsh homolog 1
ALK in cardiac myocytes	Reactive	BioCarta	1/34	DVL1 CHRD	Dishevelled, dsh homolog 1 Chordin
Biosynthesis of neurotransmitter	Reactive	BioCarta	1/6	GAD1	Glutamate decarboxylase 1
Taurine and hypotaurine metabolism	Reactive	KEGG	1/5	GAD1	Glutamate decarboxylase 1
Electron transport chain	Reactive	Gen MAPP	2/105	COX11 COX6A1	Cytochrome c oxidase subunit 11 Cytochrome c oxidase subunit Vla polypeptide 1

Table 5. Significant regulated pathways for SNP rs852977 (by courtesy of the authors) [Lee *et al.*, 2009]

Pathway enrichment analysis was conducted using 120 candidate causal genes, which were identified for rs852977. Significant biological pathways were detected via Fisher's exact test at a 5% significance level. Pathways are listed in order of significance within each of causality model, e.g., most significant pathway presents at the top.

[a] Name of biological pathway selected by Fisher's exact test.

[b] Causality models selected ay step1.

[c] Source of pathway

[d] The number of candidate causal/reactive genes associated with pathway/the number of all genes associated with pathway.

[e] Gene ID of candidate genes associated with pathway.

5. Discussion

The two-step procedure can integrate gene expression data, genotype variation data and clinical data, and identify the genetic mechanism of a complex disease. We described three different statistical tests based on the two-step procedure proposed by Lee *et al.*. For purposes of comparison, two different CFS related datasets were analyzed via the multi-step procedure proposed by Schadt *et al.*. In these specific datasets, no significant results were detected from the multistep method of Schadt *et al.*, while the method of Lee *et al.* enabled us to identify many statistically significant causal relationships, some of which were biologically supported by pathway enrichment analyses. These results demonstrated that the two-step method based on an exhaustive search investigation would provide more power.

Furthermore, the two-step approach provided some interesting results. First, CFS groups and CFS-MDD/m groups would appear to have different genotypes and gene expression profiles even though they had the common characteristic of chronic fatigue. In particular, CFS has major susceptibility markers within the NR3C1 gene, and CFSMDD/m seems to have major susceptibility markers within the catechol-O-methyltransferase (COMT) gene, though they are not statistically significant after FDR correction (Table 1B). The NR3C1 gene regulates the level of glucocorticoid which is the end product of the hypothalamic-pituitary-adrenal (HPA) whereas COMT catalyzes the transfer of a methyl group from S-adenosylmethionine to catecholamines, which is the principal end product of the sympathetic nervous system (SNS), of which the role is maintaining stress-related homeostasis [Elenkov *et al.,* 2000]. The different major susceptibility gene may be related with to the provoking of MDD/m.

Second, polymorphisms in the glucocorticoid receptor NR3C1 gene act on CFS and CFS-MDD/m differently. The polymorphisms (rs258750) within NR3C1 have significant effects on CFS, and the 105 gene expression levels independently. However, in the integrated analysis for comparing CFS-MDD/m and NF groups, polymorphisms within the NR3C1 gene affect the CFS-MDD/m and several gene expression levels differently. For example, the 217 genes are differentially expressed according to the rs6188 marker genotype within NR3C1 and disease status, even though polymorphisms within NR3C1 have no direct significant effects after FDR correction on the CFS-MDD/m. In addition, the 52 genes also regulate the CFS-MDD/m, through integrated action with the rs6188 marker. The different action of the NR3C1 gene on gene expression level and disease may be an outcome of other factors, such as environmental effects or polymorphisms of the COMT gene. The catecholamines which are regulated by the COMT gene, have been often been regarded as immunosuppressive [Elenkov *et al.,* 2000].

Two pathway enrichment analyses for the 52 candidate causal genes and 217 candidate reactive genes indicated that our approach can recover plausible regulatory mechanisms of CFS-MDD/m by comparing CFS-MDD/m and NF groups. From the comparison, we noticed that the pathways related to the immune system and steroid may have causal effect on disease state through an integrative action of the NR3C1 gene. Both the NR3C1 gene that regulates the level of glucocorticoid, and the steroid that includes corticosteroids are known to regulate the immune function [Webster *et al.,* 2002]. A number of studies have found many irregularities in the immune systems in CFS patients [Natelson *et al.,* 2002]. This

suggested that an important cause of CFS-MDD/m would be the immune system dysfunction, regulated by the neuroendocrine system, which rs6188 in the NR3C1 gene seems influence. Another potential implication of this comparison is that the CFS-MDD/m status and genetic polymorphisms can jointly induce different activation and expression of several lipid related metabolisms, neuron physiology differentiation, and neurotransmitters. Our results are supported by the known relationship between eicosanoid or fatty acid and CFS [Grey & Martinovic, 1994; Puri, 2007; Puri *et al.,* 2004; Liu *et al.,* 2003].

However, since fatigue is a core symptom in major depressive disorder [Pae *et al.,* 2007], CFS-MDD/m patients might have fatigue due to the depression rather than unexplained causes, and hence the significant results may be related to a 'major depression disorder with melancholic features' rather than chronic fatigue. For example, the excessive hypothalamus-pituitary-adrenal (HPA) axis responses, of which the end products are glucocorticoids, are known to be hallmarks of depression [Pariante & Miller, 2001; Holsboer, 2000; Pariante, 2004]. Also, the major depression can be associated with the immune activation, dysfunction of neurotransmitters at synapse [Neumeister *et al.,* 2004; Sanacora *et al.,* 2004; Maes & Meltzer, 1995], and essential fatty acids [Van Strater & Bouvy, 2006].

The integrative analyses considering the interaction effect among different levels of data could elucidate different disease susceptibility and differentially expressed genes of genetically different individuals. Some results showed that integrating genotype and expression data may help the search for new directions for the treatment of common human diseases that are not being detected using only one type of data. The integrated analysis provided more information than the two separate analyses of gene expression data and genotype variation data for characterizing CFS that has several possible causes.

In conclusion, the two-step approach to the integration of heterogeneous data sets can be generally applied to other studies in which gene expression data, genotype variation data and clinical data are available, and it can be very useful as the importance of integrated data analysis has been increasing. The two-step approach can also be extended to datasets containing other type of data, such as protein data rather than clinical data. The two-step approach can be readily applicable to quantitative traits rather than binary clinical outcome traits, by employing linear regression analysis. Also, it can be easily applied to genome-wide association studies, and can handle environmental factors, such as age and sex, by treating these factors as covariates in the regression model. Furthermore, the two-step approach can be extended to the gene-set approach, the module based approach or co-expression network as Presson *et al*. [Presson *et al.,* 2008] and Chen *et al*. [Chen *et al.,* 2008] did.

However, there are some limitations to the two-step method. First, the causality models are too simple to represent true mechanisms, which would be more complicated due to possible interactions between causal-reactive genes [Schadt *et al.,* 2005]. Further considerations for more complicated models are necessary in order to identify the genetic mechanism of complex diseases. Second, the two-step approach may need large computing although it is applicable to genome-wide studies because it is not limited in the scale of data. Another limitation would be a misclassification problem in that the proposed method relies on the LCMS. The current two-step approach does not use FDR procedure to account for the model misclassification problem. In fact, FDR procedure was employed only in the second step, not in the first step for the model selection procedure that chooses the model with the minimum

AIC among the three causal models. While anticipating the problem, we still employed the LCMS process because it showed good power for detecting true models in the simulation evaluated by Schadt *et al*. The two-step approach can be extended to account for the errors caused by the model misclassification in the first step. For example, we can test for the difference in the AIC values of three causality models, because the chance for model misclassification would be high when the difference between the smallest AIC value from the selected model and those from the other models is not large. A permutation-based nonparametric test might be developed for this testing. We think it requires a further study to control simultaneously two types of errors: causality model selection, and significant maker-gene pair identification.

6. Acknowledgment

The work was supported by the National Research Foundation (KRF-2008-313-C00086)

7. References

Akaike, H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control.* Vol.19, pp.716-723

Baldi, P. & Long, A.D. (2001) A Bayesian Framework for the Analysis of Microarray Expression Data : regularized *t*-test and statistical inferences of gene changes. *Bioinformatics.* Vol.17, pp.509-519

Benjamini, Y. & Hochberg, Y. (1995). Controlling the False Discovery Rate – a pratical and powerful approach to multiple testing. *Journal of the Royal Statistical Society.* Vol.B57. pp.289-300

Boerjan, W. & Vuylsteke, M. (2009). Integrative genetical genomics in Arabidopsis. *Nature Genetics.* Vol.41, No.2, pp.144-145

Breitling, R. ; Armengaud, P. ; Amtmann, A. & Herzyk, P. (2004). Rank products : a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Letters*. Vol.573, pp.83-92

Brem, R.B. ; Yvert, G.l ; Clinton, R. & Kruglyak, L. (2002). Genetic Dissection of Transcriptional Regulation in Budding Yeast. *Science.* Vol.296, pp.297-392

Chen, Y. ; Zhu, J. ; Lum, P.Y. ; Yang, X . ; Pinto, S. ; MacNeil, D.J. ; Zhang, C. ; Lamb, J. ; Edwards, S. ; Sieberts, S.K. ; Leonardson, A. ; Castellini, L.W. ; Wang, S. ; Champy, M.F. ; Zhang, B. ; Emilsson, V. ; Doss, S. ; Ghazalpour, A. ; Horvath, S. ; Drake, T.A. ; Lusis, A.J. & Schadt E.E. (2008). Variations in DNA Elucidate Molecular Networks That Cause Disease. *Nature*. Vol.452. pp.429-435

Coffey, C.S. ; Hebert, P.R. ; Ritchie, M.D. ; Krumholz, H.M. ; Gaziano, J.M. ; Ridker, P.M. ; Brown, N.J. ; Vaughan D.E. & Moore, J.H. (2004). An Application of Conditional Logistic Regression and Multifactor Dimensionality Reduction for Detecting Gene-Gene Interactions on Risk of Myocardial Infaction : The Importance of Model Validation. *BMC Bioinformatics.* Vol.5:49

Elenkov, J. ; Wilder, R.L. ; Chrousos, G.P. & Vizi, E.S. The sympathetic Nerve – an integrative interface between tow supersystems : the brain and the immune system. (2000). *Pharmacological Review.* Vol.52. pp. 595-638

Fu, J.; Keurentijes, J.J.B. ; Bouwmeester, H. ; America, T. ; Verstappen, F.W.A . ; Ward, J.L. ; Beale, M.H. ; Vos, R.C.H. ; Dijkstra, M. ; Scheltema, R.A. ; Johannes, F. ; Koornneef,

M. ; Vreugdenhil, D. ; Breitling, R. & Jansen, R.C. (2009). System-wide Molecular Evidence for Phenotypic Buffering in Arabidopsis, *Natrue Genetics,* Vol.41, No.2, pp. 166-167

Ghazalpour, A. ; Doss, S. ; Kang, H. ; Farber, C. ; Wen P.Z. ; Brozell, A. ; Castellanos, R. ; Eskin, E. ; Smith, D.J. ; Drake, T.A. & Lusis, A.J. (2008). High-Resolution Mapping of Gene Expression Using Association in an Outbred Mouse Stock. *PloS Genetics.* Vol.4, e100149.

Gray, J.B. & Martinovic, A.M. (1994) Eicosanoids and Essential Fatty-acid Modulation in Chornic Disease and the Chronic Fatigue Syndrome. *Med.Hypotheses.* Vol.43 pp.31-42

Henshall, J.M. ; Goddard, M.E. (1999) Multiple-Trait Mapping of Quantitative Trait Loci After Selective Genotyping Using Logistic Regression. *Genetics*. Vol.151, pp.885-894

Holsboer, F. (2000). The Corticosteroid Receptor Hypothesis of Depression. *Neuropsychopharmacology.* Vol.23. pp.477-501

Howell, W.M, ; Jobs, M. ; Gyllensten, U. & Brookes A.J. (1999). A New Method for Scoring Single Nucleotide Polymorphisms. *Nature Biotechnology.* Vol.17. pp.87-88

Jansen, R.C. (2009). Genetical Genomics Tutorial. Available from http://www.ipk-gatersleben.de/Internet/Forschung/Doktorandenprogramm/StudentBoard/GuestSpeakers/Jansen_2009.pdf

Kang, H.M. ; Zaitlen, N.A. ; Wade, C.M. ; Kirby, A. ; Heckerman, D. ; Daly, M.J. & Eksin, E. (2008) Efficient Control of Population Structure in Model Organism Association Mapping. *Genetics.* Vol.178, pp.1709-1723

Kendziorski, C.M. ; Chen, M. ; Yuan, M. & Attie, L.A.D. (2006) Statistical Methods for Expression Quantitative Trait Loci (eQTL) mapping. *Biometrics*. Vol.62. pp.19-27

Lan, H. ; Chen, M. ; Flowers, J.B. ; Yandell, B.S. ; Stapleton, D.S. ; Mata, C.M. ; Mui, E.T. ; Flowers, M.T. ; Chueler, K.L. ; Manly, K.F. ; Williams, R.W. ; Kendziorski, C. & Attie, A.D. (2006). Combined Expression Trait Correlations and Expression Quantitative Trait Locus Mapping. *PloS Genetics*. Vol.2, e6

Lee, E.; Cho, S.; Kim, K. & Park T. (2009). An Integrated Approach to Infer Associations Among Genes Expression, Genotype Variation, and Disease. *Genomics*. Vol.94, pp.269-277

Lin, E & Hsu S. (2009). A Bayesian Approach to Gene-Gene and Gene-Environment Interactions in Chronic Fatigue Syndrome. *Pharmacogenomics.* Vol.10. pp.35-42

Liu, Z. ; Wang, D. ; Xue, Q. ; Chen, K. ; Bai, X. & Chang, L. (2003) Determinination of Fatty Acid Levels in Erythrocyte Membranes of Patients with Chronic Fatigue Syndrome. *Nutr. Neurosci.* Vol.6. pp.389-392

Maes, M. & Meltzer, H.Y. (1995) *The Serotonin Hypothesis of Major Depression.* Raven Press, New York

Michaelson, J.J. ; Loguercio, S. & Beyer, A. (2009). Detection and Interpretation of Expression QuantitativeTrait Loci (eQTL). *Methods.* Vol.48. pp.265-276

Natelson, B.H. ; Haghighi, M.H. & Ponzio, N.M. Evidence for the Presence of Immune Dysfunction in Chornic Fatigue Syndrome. (2002). *Clin. Diagn. Lab. Immunol.* Vol.9. pp.747-752

Neumeister, A.; Young, T. & Stastny, J. (2004). Implications of Genetic Research on the Role of the Serotonin in Depression: emphasis on the setotonin type 1(A) receptor and the serotonin transporter. *Psychopharmacology.* Vol.174. pp.512-524

Pae, C.U.; Lim, H.K.; Han, C.; Patkar, A.A.; Steffens, D.C.; Masand, P.S. & Lee, C. (2007) Fatigue as a Core Symptom in Major Depressive Disorder: overview and the role of bupropion. *Exper Rev. Neurotherapeutics*. Vol.7. pp.1251-1263

Pariante, C.M. (2004). Glucocorticoid Receptor Function in Vitro in Patients with Major Depression. *Stress.* Vol.7. pp.209-219

Pariante, C.M. & Miller, A.H. (2001). Glucocorticoid Recpetors in Major Depression: relevance to pathophysiology and treatment. *Biol. Phychiatry*. Vol.49. pp.391-404

Presson, A.P. ; Sobel, E.M. ; Papp, J.C. ; Suarez, C.J. ; Whistler, T. ; Rajeevan, M.S. ; Vernon, S.D. & Horvath, S. (2008). Integrated Weighted Gene Co-expression Network Analysis with an Application to Chronic Fatigue Syndrome. *BMC Systems Biology*. Vol.2

Puri, B.K. Long-chain Polyunsaturated Fatty Acids and the Pathophysiology of Myalgic Encephalomyelitis (chronic farigue syndrome). (2007). *Jounral of Clinical Pathology*. Vol.60. pp.122-124

Puri, B.K. ; Holmes, J. & Hamilton, G. (2004) Eicosapentaenoic Acid-rich Essential Fatty Acid Supplementation in Chronic Fatigue Syndrome Associated with Symptom Remission and Structural Brain Changes. *Int. J. Clin. Pract.* Vol.58. pp.297-299

Sanacora, G. ; Gueorguieva, R. ; Epperson, C.N. ; Wu, Y.T., Appel, M. ; Rothman, D.I. ; Krystal, J.H. & Mason, G.F. (2004). Subtype-specific Alterations of Gamma-Aminobutyric Acid and Glutamate in Patients with Major Depression. *Arch. Gen. Psychiatry*. Vol.61. pp.705-713

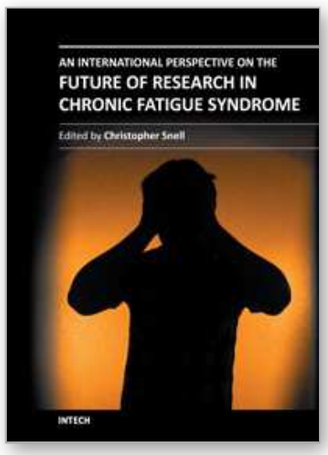
Schadt, E.E. ; Lamb, J. ; Yang, X. ; Zhu, J. ; Edwards, S. ; GuhaTahkurta, D. ; Sieberts, S.K. ; Monks, S. ; Reitman, M. ; Zhang, C. ; Lum, P.Y. ; Leonardson, A. ; Thieringer, R. ; Metzger, J.M. ; Yang, L. ; Castle, J. ; Zhu, H. ; Kash, S.F. ; Drake, T.A. ; Sachs, A. & Lusis A.J. (2005). An Integrative Genomics Approach to Infer Causal Associations Between Gene Expression and Disease. *Nature Genetics.* Vol.37, No.7, pp.710-717

Schena, M. ; Shalon D. ; Davis, R.W. & Brown P.O. (1995) Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. *Science.* Vol.270, No.5235, pp.467-470

Tusher, G. ; Ribshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. (2001). *Proc. Natl. Acad, Sci.* Vol.98, pp.5116-5121

Van Strater, A.C.P. & Bouby, P.F. (2006). Omega-3 Fatty Acids and Mood Disorders. *Am. J. Psychiatr.* Vol.163. p.2018

Webster, J.I. ; Tonelli, L. & Sternberg, E.M. (2002) Neuroendocrine Regulation of Immunity. *Annu. Rev. Immunol.* Vol.20. pp.125-163



An International Perspective on the Future of Research in Chronic Fatigue Syndrome

Edited by Dr. Christopher R. Snell

ISBN 978-953-51-0072-0
Hard cover, 104 pages
Publisher InTech
Published online 15, February, 2012
Published in print edition February, 2012

While the chapters in this book are a long way from solving the enigma that is CFS, they do represent important attempts to understand this complex and perplexing disease. A common theme in them all is CFS as a multisystem disease with the possibility of more than one cause and influenced by a variety of interacting factors. Further, they acknowledge the reality of CFS for persons with this disease and the importance of finding causes, treatments and ultimately a cure. As advanced biomedical research techniques are increasingly applied to the study of CFS, it is surely only a matter of time before biomarkers are identified, etiologies understood, and remedies devised.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen