Transcriptome analysis of patients with Chronic Fatigue Syndrome

Hanna Gräns, MSc

Stockholm 2005

From the Division of Clinical Bacteriology
Department of Laboratory Medicine
Karolinska Institutet, Stockholm, Sweden

Transcriptome analysis of patients with Chronic Fatigue Syndrome

Hanna Gräns, MSc

Stockholm 2005

Printed by Repro Print AB, Stockholm
ISBN 91-7140-448-1

To my dear family

ABSTRACT

Fatigue is a central component of many diseases and illnesses. Fatigue of unknown etiology and pathophysiology lasting more than six months, together with at least four out of eight specified symptoms, is termed chronic fatigue syndrome (CFS). Several causes have been suggested for the illness, including immune dysfunction, stress, sleep disturbances and infectious agents. CFS diagnosis is currently based on self-reported symptoms. The lack of physical abnormalities and laboratory tests makes the diagnosis harder. Identification of biological illness markers would contribute to increase insight into the pathophysiology of the illness and facilitate diagnosis. Powerful methods for transcript analysis have been developed during the past decade. Microarray technology and real-time PCR are two methods commonly used to identify genes involved in disease. The identified genes are disease markers, which may be used for diagnostic purposes.

Researchers involved with microarray experiments need standardization to facilitate comparisons between studies and laboratories. We show here that different RNA extraction methods can yield comparable results. Even so, only one method should be used in any one study and the ambition should be to use identical conditions for each and every experiment. CFS is not characterized by any diseased tissue, and this raises the question of what is a representative sample. The hypothesis has been raised that peripheral blood cells function as indicators for different biological processes going on throughout the human body. We show here that genes involved in psycho-neuroendocrine-immune (PNI) communication can be studied using peripheral blood mononuclear cells (PBMCs). The PBMC sample can be used to study diseases, such as CFS, with unknown pathophysiology and etiology.

The individual transcript expression variability in PBMCs is small and differences in gene activity due to abnormalities caused by illness or disease are larger. We expected to find only small gene expression differences, if any, between CFS patients and healthy controls. In our transcript expression studies we observed only a few differentially expressed genes. We found reduced levels of estrogen receptor β (ERβ) in CFS patients compared to healthy controls using real-time PCR. Three genes were identified using microarray technology with significant expression differences: CD83, NRK1 and BOLA1. The differences were only found between a subgroup of CFS patients, female patients with no previous infection and gradual illness onset, compared with healthy female controls. We verified the results with real-time PCR. The results indicate the need for subgrouping of the heterogeneous group of patients with fatiguing illness in search for pathogenic mechanisms.

In conclusion, the difference in gene expression could contribute to some of the symptoms observed in CFS. Further studies to investigate the protein levels and cellular effects will be required to determine whether any of these genes are involved in CFS pathology. The differences in transcript expression levels could also simply be a marker for changed functions of other cellular components that are involved in CFS. In this case, the altered levels could contribute to diagnostic criteria, they may form a surrogate marker, or they may provide an entry point to identifying potential disease-causing candidate molecules for further study.

LIST OF PUBLICATIONS

This thesis is based on the following papers, which in the text will be referred to by their corresponding Roman numbers:

I. **Ojaniemi H**, Evengård B, Lee DR, Unger ER and Vernon SD
Impact of RNA extraction from limited samples on microarray results
BioTechniques, 35(5):968-973, 2003

II. Nicholson AC, Unger ER, Mangalathu R, **Ojaniemi H** and Vernon SD
Exploration of neuroendocrine and immune gene expression in peripheral blood mononuclear cells
Molecular Brain Research, 129(1-2):193-197, 2004

III. **Gräns H**, Nilsson P and Evengård B
Gene expression profiling in the Chronic Fatigue Syndrome
Journal of Internal Medicine, 258(4):388-390, 2005

IV. **Gräns H**, Evengård B and Nilsson P
Transcriptome analysis of PBMCs from patients with Chronic Fatigue Syndrome
Manuscript

V. **Gräns H**, Nilsson M, Gustafsson J-Å, Dahlman-Wright K and Evengård B
Reduced levels of ERβ mRNA in Swedish patients with Chronic Fatigue Syndrome
Submitted

TABLE OF CONTENTS

INTRODUCTION

Chronic Fatigue Syndrome

Fatigue is a common complaint among the general population in developed countries [1]. Studies show that 5-20% of the population suffers from disabling fatigue [2] and that it is two to three times more common among women than men [1].

There are a number of terms for fatigue, including weariness, tiredness, exhaustion, lethargy and lack of energy. Fatigue can be defined in many ways. One way to define it is "the individual perception of emotional and/or physical incapacitation". Another way is to describe the feeling "the inability to proceed", independently of whether it deals with physical or intellectual work [3]. Fatigue following hard work or lack of sleep is considered as normal fatigue and has an important protective role [3]; both the body and mind need time for recovery. Fatigue is a symptom that often accompanies disease, and it can have different characteristics depending on the type of disease and on individual variation.

Fatigue is often of major importance for a patient because it affects the quality of life profoundly [2]. The medical profession, however, infrequently recognizes it as a major symptom [2]. Diseases connected with fatigue include asthma, arthritis, emphysema, low blood pressure [1], cancer, autoimmune disease, diabetes, hypothyroidism, hypoadrenalism, sleep disorders, multiple sclerosis [2], anaemia, many psychiatric disorders, and illnesses such as chronic fatigue syndrome (CFS) [1, 2]. A disease has a known medical cause, while an illness lacks medical explanation.

What is CFS?

The first time CFS was described in the medical literature was in the middle of the 19th century, although some sources argue that a similar illness was described in the 17th century [4]. Charles Beard, an American neurologist, described an illness called neurasthenia during the 19th century, which resembled the illness now known as CFS [4]. Several other diagnostic labels have been used for CFS, including epidemic neuromyasthenia, Icelandic disease, Royal Free disease, chronic mononucleosis [4], post-viral fatigue syndrome, myalgic encephalomyelitis [1, 4] and chronic Epstein-Barr virus infection [1].

Most people now use the name "chronic fatigue syndrome" and the diagnosis is based on a case definition together with symptoms reported by the patient. The CDC international case definition is based on the consensus of a group of international specialists aid down in 1994 [5], and is the one most widely used. Two more definitions exist, a British [6] and an Australian [7] variant. The three case definitions are compared in Table 1. Ambiguities regarding the CDC-1994 definition have recently

received attention [8, 9]. Topics that have been subjects for debate are the inconsistent case identification, the exclusionary criteria and comorbid conditions [9].

According to the CDC-1994 criteria, the fatigue must be of unknown etiology and pathophysiology and it must have lasted for more than six months, accompanied by at least four out of eight specified symptoms [5]. The symptoms are: impairment of cognition and memory, recurring sore throat, tender lymph nodes, mild muscle pain, joint ache, headaches of a new type, unrefreshing sleep and post-exertional malaise [5].

Table 1: Comparison of three different case definitions for CFS [10].

	CDC-1994	**British**	**Australian**
Minimum duration	6 months	6 months	6 months
Functional impairment	Substantial	Disabling	Substantial
Cognitive or neuro-psychiatric symptoms	May be present	Mental fatigue required	Required
Other symptoms	Four required	Not specified	Not specified
New onset	Required	Required	Not required
Medical exclusions	Clinically important	Known physical causes	Known physical causes
Psychiatric exclusions	Melancholic depression, substance abuse, bipolar disorder, psychosis, eating disorder	Psychosis, bipolar disorder, eating disorder, organic brain disease	Psychosis, bipolar disorder, substance abuse, eating disorder

The CFS patient group has a heterogeneous symptom profile. It is not clear whether the patient cohort consists of one single entity or several entities. This makes the diagnosis more difficult, and subgroups presenting different clinical symptoms may require different treatments. Many studies have attempted to subgroup CFS patients according to clinical symptoms, for example, but no stratification strategy has so far proven to be consistently superior [11]. Study of a heterogeneous CFS patient group can obscure differences between CFS patient subgroups, which emphasises the need for subgrouping of patients [11]. Inconsistent results comparing different CFS studies may also arise from the heterogeneity of the patient populations with different compositions of subgroups in different studies [11].

The diagnosis of CFS would become more credible both within the medical field and within the general public if a biological test could be developed.

How common is CFS?

The prevalence of CFS that is measured depends on the population studied and the case definition used. The CFS or CFS-like illness prevalence rates for five different studies [12-16] are presented in Table 2 with information about country and the type of population studied. Two large community studies have been performed in the US [12, 15]. Both of these studies included medical investigation following a telephone interview. In two Nordic population-based studies [13, 14] using questionnaires, the higher prevalence rates for CFS-like illness compared to CFS are probably due to the absence of a clinical evaluation. Wessely *et al.* applied all three case definitions presented in Table 1 to their primary care patient cohort. The result was prevalence rates ranging from 1.4% to 2.6% with the highest percentage for the CDC-1994 case definition and the lowest for the Australian definition [16].

A predominance of women is seen in all of the community-based studies with 2-4 times higher prevalence rates for women compared to men [12-15]. A study of patients in a Swedish tertiary clinic revealed a preponderance of women over men (70% women and 30% men) [17]. As mentioned before more women than men in the general population complain about fatigue [1].

Table 2: Prevalence of CFS or CFS-like illness in studies using the CDC-1994 criteria [5].

Study	Country	Setting	Prevalence	Ratio ♀/♂
Jason *et al.* [12]	US	Community	0.42%	2.6
Reyes *et al.* [15]	US	Community	0.24%	4.5
Wessely *et al.* [16]	UK	Primary care	0.50%	No data
Lindal *et al.* [14]	Iceland	Community	1.40%*	3.5
Evengard *et al.* [13]	Sweden	Community	2.36%*	4.3

* No medical examination = CFS-like illness

What is the cause of CFS?

The etiology of CFS is poorly understood. Alternative theories about biological, psychological and psychosocial causes including immune dysfunction, hypothalamic-pituitary-adrenal (HPA) axis abnormalities, stress, sleep disturbances and infectious agents, have been suggested. Predisposing, triggering and maintaining factors may all influence the illness process.

The impact of a number of different immunological factors on CFS has been evaluated. Abnormalities in both natural killer cell and T-cell function with elevated T-cell activation have been reported [18]. Interestingly, markers for chronic immune activation have been observed, indicating a constant exposure to an antigen, either foreign or self-antigen [18]. Imbalanced cytokine levels have also been observed [18].

Stress is usually harmful for the organism, but moderate stress can improve the immune system function. Too much stress, however, will reduce the capacity of the immune system resulting in an increased risk for infection. The HPA system, which regulates cortisol secretion, plays a central role in the stress response. Secretion of corticotropin releasing factor from the hypothalamus cause the pituitary glands to release adrenocorticotropic hormone (ACTH) into the blood stream. ACTH acts as a stimulator of cortisol secretion in the adrenal cortex. The cortisol level is regulated by a feedback system that involves receptors in hypothalamus.

Reduced levels of cortisol have been observed in both blood and urine among CFS patients [18]. Abnormally low cortisol levels have recently been suggested as a possible maintaining factor for CFS rather than a triggering factor [19]. The evidence for this is the absence of HPA axis disturbance before the onset of CFS and during the early stages of the illness, with the presence of reduced cortisol levels in a later phase of the illness [19]. Research about deficiencies in both the neuroendocrine and immune communication systems are ongoing and more information should be available in the near future.

Out of the eight specified symptoms in the CFS international case definition [5], unrefreshing sleep is the most prevalent symptom, according to a population-based study [12]. Observation of sleep disturbances such as impaired alpha rhythm within the non-rapid eye-movement sleep and disturbed sleep initiation and maintenance have been reported for CFS patients [18].

Microorganisms have been suggested as triggering factors for CFS. Many CFS patients describe an infectious-like illness onset; they recover from the infection but remain in a fatigued state. Furthermore, several outbreaks of what later have been categorized as CFS have occurred. Two outbreaks in hospital environments have been described, in Los Angeles County Hospital in 1934 and in the Royal Free Hospital in London in 1955 [18]. In a small village called Akureyri on Iceland an outbreak

occurred in 1948, and the latest described outbreak was in Lake Tahoe, on the boarder between California and Nevada in the US, in the late 1980s [18].

The occurrence of different microorganisms, both bacteria and viruses, has been investigated and listed in Table 3. So far no clear relation between any of the microorganisms on the list and CFS have been reported.

Table 3: Microorganisms of interest for CFS [18].

Type	Name
Bacterium	*Borrelia burgdorferi*
	Chlamydia pneumoniae
	Mycoplasma species
Virus	Influenza virus
	Epstein-Barr virus
	Cytomegalovirus
	Human herpesvirus type 6 and 7
	Varicella zoster virus
	Borna disease virus
	Enterovirus

Psychological factors such as the impairment of information processing, impaired cognition, complex information processing speed and efficiency may affect CFS, as may psychosocial factors such as stressful life events, occupational stress, and stress related to personal factors.

How is CFS treated?

CFS causes high costs for both affected individuals and society, as well as productivity losses [8]. The annual productivity losses in the US are estimated to be approximately $9.1 billion ($20,000 per CFS patient), and are about the size of those for immune and nervous system diseases, digestive diseases and skin disorders [8]. The low number of patients reaching full recovery and the lack of effective treatments are two important issues. Studies report a full recovery rate of 0-48% with a median rate of 7% [20]. Psychiatric disorders, the belief in a physical cause for CFS, and low symptom control are connected with poorer outcome [20]. The two most promising treatments for CFS patients are cognitive behaviour therapy (CBT) and graded exercise therapy (GET), but neither of these works for the entire patient population.

In CBT the focus is on how the patient thinks about his/her ill health (cognition) and the way the patient deals with it. CBT as treatment for CFS often includes cognitive restructuring of unhelpful beliefs and assumptions, planned activity and rest, gradual increase of activity and sleep routines [20]. GET includes a gradual increase of physical activity, often an individually designed exercise program with walking [20].

Several other treatments have been studied, but none with any significant degree of success (Table 4).

Table 4: Examples of treatments tested for CFS patients [20].

Category	Treatment
Pharmacological	Antidepressants
	Corticosteroids
	Nicotinamide adenine dinucleotide
Immunological	Immunoglobin
	Interferon
	Staphylococcus toxoid
Other	Nutritional supplements
	Massage therapy

What is a representative clinical sample to study CFS biology?

The CFS diagnosis is based on self-reported symptoms. The absence of physical signs and laboratory tests make the diagnosis more difficult. Unfortunately, this leads to scepticism towards the illness both in the general public and in the medical profession. There is no clearly identifiable diseased tissue in CFS, and this raises the question of what biological tissue should be examined. The answer depends on the believed cause for the illness. One common hypothesis is that peripheral blood cells can serve as indicators for abnormal processes going on throughout the human body [21-26]. According to this theory, investigation of peripheral blood cells can facilitate insight for illness processes in different parts of the body. The individual variation in gene activity in human blood cells is small, which makes the sample suitable for the study of disease in general [27]. Natelson *et al.* believe that CFS originates in brain dysfunction, and have used spinal fluid as a sample [28].

Transcriptome analysis

The end of 20th century and the beginning of 21st century could be called the "omic"-era of molecular biology and biotechnology. The genome consists of deoxyribonucleic acid (DNA) molecules, which contain hereditary information. Genomics is the study of genes and their function. When a gene is activated, it is used as a template to synthesise a messenger ribonucleic acid (mRNA) copy of the gene in a process called transcription. Transcriptomics is the study of RNA. In the next step, the mRNA molecule is transported from the nucleus into the cytosol where it is used as a template for protein production in a process called translation. Proteomics is the study of proteins and their function. The central dogma describes the flow of genetic information from DNA to gene product, and is illustrated in Figure 1.

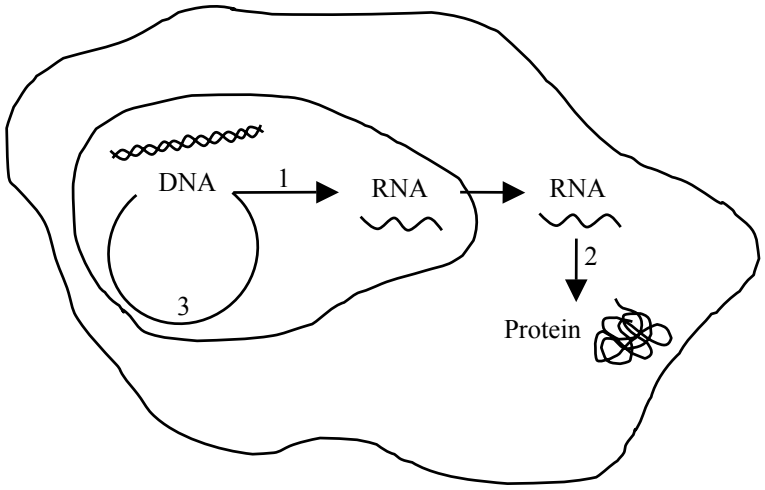


Figure 1: The central dogma of molecular biology: DNA is transcribed to RNA (1) and the RNA is used as template for translation of proteins (2). During the cell division the entire genome is copied in a process called replication (3).

A gene consists of two types of regions, the coding regions, exons, and the non-coding regions, introns. The functional part of a gene consists of exons. The introns are removed in a process called splicing. Messenger RNA molecules are copies of the functional part of a gene and thus consist of only exon regions. The transcriptome is the set of all mRNAs, also called transcripts, generated from genes activated in biological processes in an organism. A definition of transcriptomics is the study of mRNA expression levels and mRNA expression profiling. Transcriptomics aims at determining gene activity levels by determining the amount of mRNA expression.

Transcript profiling techniques have made it possible to study biological processes in a novel way and reveal new information about gene activity levels, the interplay between genes and gene regulation. This area of research is called functional genomics.

Studying protein levels may be a more profitable approach than studying mRNA expression. The mRNA molecules play a crucial and important role, but it is mostly the protein that causes the final effect. There is not necessarily any correlation between mRNA and protein levels. Proteins form an inhomogeneous group of molecules and different proteins require different environmental conditions. Compared to the experimental procedures for evaluation of the transcriptome, the possibilities to study the proteome are much more limited. There is no "protein PCR" [29] and efficient labelling of the entire proteome is much harder. All these facts limit the use of proteins so far, but it is an active field of research and the advent of new powerful methods is expected.

Techniques for transcriptome analysis

A number of techniques for transcript profiling have been developed in recent years. Different methods are suitable for different applications. Some methods are high-throughput methods that screen hundreds to thousands of genes in a single experiment, while other methods look at a single gene. Similarly, some methods determine absolute levels while others determine relative abundances.

Microarray technology is the most commonly used high-capacity technique for transcriptome analysis today [30, 31], and will be described in more detail in the next section. Other methods for analysing large numbers of transcripts are expressed sequence tag (EST) sequencing, serial analysis of gene expression (SAGE) [32] and massively parallel signature sequencing (MPSS) [33]. Unlike microarray technology, where only relative quantification can be performed, these three methods allow absolute quantification. Methods used for the analysis of expression of single genes include Northern blotting [34], reverse transcription polymerase chain reaction (RT-PCR) [35, 36], and real-time PCR [37]. Northern blotting and real-time PCR are widely used for the verification of microarray technology results.

Microarray technology

The levels of mRNA expression of hundreds to thousands of genes, entire genomes, can be studied in one single experiment using microarray technology. The first publications exploring this new technology came in the mid 1990s [30, 31] and reported a very promising technique. Researchers believed that this was the solution for finding answers to several biological questions regarding gene activity. Recently, however, issues have been discussed regarding the data analysis, statistical problems, and difficulties with reproducibility between laboratories. This is not uncommon with new methods, and the microarray technology needs further development. The data analysis part is an especially rapidly evolving area.

The microarray technology has found many useful applications including biomarker discovery for diagnostic purpose, drug discovery, understanding of complex biological systems and toxic assessment.

A microarray consists of DNA fragments representing unique genes attached at high density to a glass or plastic slide. Each spot or feature on a microarray consists of one type of fragment complementary to the labelled mRNA copy from a unique gene. There are two types of microarray, differing mainly in the fragment design. The complimentary DNA (cDNA) microarray has PCR-amplified fragments representing a long part of the gene sequence attached to the surface, while the oligonucleotide microarray has shorter unique sequences of the same size, generally 25-80 bases long, attached to the surface. The oligonucleotide microarrays can be divided into short, 25 bases and long, 40-80 bases. Genes differ in length, which means that the fragments on the cDNA microarray also differ in length.

Cross-hybridisation of unrelated sequences is a problem for short oligonucleotide microarrays [38]. The more stringent hybridisation conditions for cDNA microarrays prevent cross-hybridisation, but highly similar genes may still bind the same fragment [38]. Long oligonucleotide microarrays have higher specificities than short oligonucleotide microarrays, and they can distinguish between splice variants [38].

The first microarrays consisted of a couple of hundred to one thousand features [30, 31], while the arrays available today cover tens of thousands of unique genes. Two main approaches for a microarray experiment exist, the one-colour system (one sample is hybridised to each microarray) and the two-colour system (two samples are hybridised to each microarray). Affymetrix was one of the earliest companies and developed a patented single-colour technique that is widely used [30]. A large number of companies now sell microarrays geared at both eukaryotes and prokaryotes.

One disadvantage with this technology is the high costs, both the microarrays and the reagents are expensive. Prices have decreased with time and development, but it is still difficult for many research groups, especially in the academic world, to perform extensive studies using commercial microarrays. One way of lowering the expense has been to manufacture the microarrays in-house. Another important limitation is the restricted possibility to compare only relative mRNA expression.

Experimental details

Figure 2 shows schematically the steps in a microarray experiment. Hegde et al. [39] and Freeman et al. [40] have written useful summaries of the microarray technique.

A microarray experiment includes the extraction of RNA from cells, either mRNA or total RNA. Pure RNA samples are of great importance for the following steps to work properly [39]. In the early days of the microarray era, 1.5 µg of mRNA or 50-100 µg of total RNA was used [39]. Development has reduced the sample amount required,

but one common limitation when working with biological samples is still the lack of sufficient RNA material. Amplification or enhancer methods can solve the problem. Synthesis of amplified antisense RNA (aRNA) before labelling is one option [41, 42]. There is no risk of introducing bias into the results, if a proper protocol is used, and several methods proven to be unbiased are commercially available.

The RNA extraction step is common for all microarray sample preparation protocols. The second step is the synthesis of copies of the mRNA molecules during incorporation of labelled molecules. This can either be done in a reverse transcription (RT) synthesis to produce labelled cDNA [39] or by synthesis of labelled aRNA following one round of amplification [30]. The latter procedure is used by the Affymetrix chips. More information about the RT synthesis can be found below, in the section describing real-time PCR. Incorporation of labelled molecules enables sample detection in a later step. Equal labelling efficiencies between individual samples and different dyes (two-colour) is crucial to get a true representation of what is actually present in the cell sample.

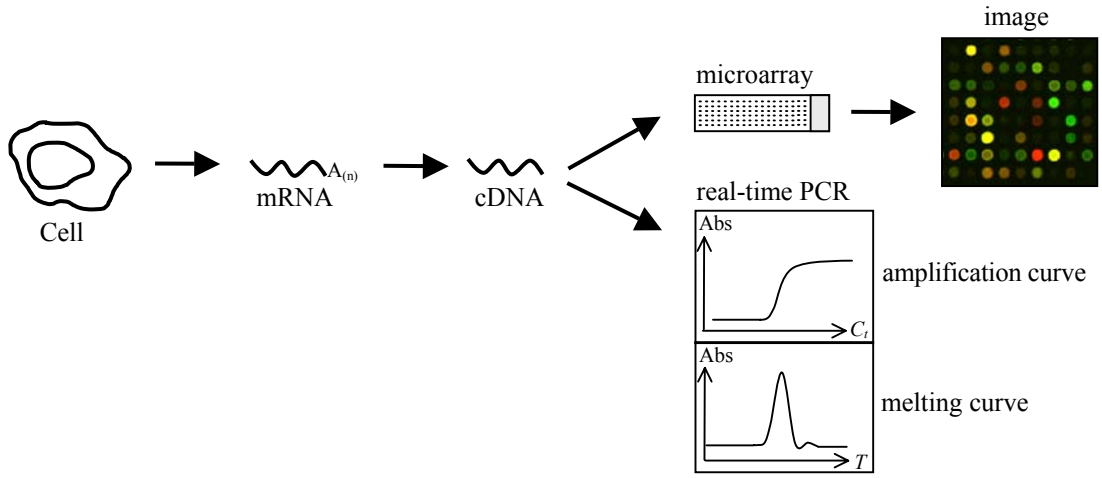


Figure 2: Outline of the experimental steps included in a microarray and real-time PCR experiment. Both technologies include RNA extraction and cDNA synthesis.

Fluorescent labelling dyes (e.g. Cyanine 3 (Cy3) and Cyanine 5 (Cy5)) together with a laser scanner are most commonly used, with photosensitivity as the main disadvantage [39]. Alternative labelling methods exist, such as the resonance light scattering (RLS) system [43] where gold and silver particles are employed for detection with a white light scanner. Other dyes such as fluorescein and X-rhodamine may be used. The RLS system has a 50 times higher sensitivity compared to fluorescent systems [44], which allows the use of smaller RNA samples, and it is not sensitive to photobleaching.

Labelled samples are hybridised to microarrays by Watson-Crick base pairing in an one-colour or a two-colour fashion. A hybridisation reaction with high specificity and

low background signal is desirable [39]. Prehybridisation including bovine serum albumin, for example, is used to block active sites on the microarray to prevent non-specific binding of labelled sample, and unbound DNA is washed away before the sample is added [39].

A hybridisation often lasts for 12-18 hours. When it is finished, unbound sample is washed away and the cleaned microarray is scanned using an appropriate scanner.

Experimental design

Careful design is important in a microarray study in order to be able to answer the intended biological question. Some studies evaluate differences between one or several predefined classes, e.g. patients and controls or different disease states, to find biological differences or to obtain preliminary information about clinical prognosis. Other studies aim at discovering new subclasses within a disease. The goal of the study determines the design of the experiment.

Biological and technical variations are a cause of concern in all experiments. All living organism species have individual variation. In an inbred strain these differences often are smaller [45], but are still present to some degree. Biological variation is due to genetic and/or environmental factors. In all of the steps of an experiment, technical variability is introduced. The biological variability is often larger than the technical variability, and it is therefore often better to perform biological replicate experiments than technical repeats [45], if both cannot be performed. It may be useful in some cases to pool samples, if the material is in limited supply. Pooling of individual samples reduces the biological variability, but does not affect the technical component [46].

In one-colour microarray studies the hybridisation design is fairly simple, one sample is hybridised to each microarray. When it comes to two-colour microarray studies the design is a bit more complicated and there are several options (Figure 3).

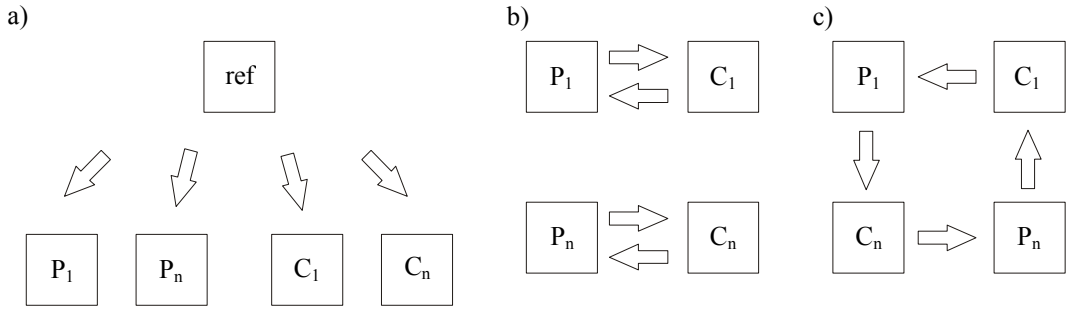


Figure 3: Schematic description of three microarray study designs, a) indirect comparison design, b) dye swap design and c) loop design.

The first decision is between direct and indirect comparison. In direct comparison, two samples, such as from paired patient and control, are hybridised to the same microarray. In indirect comparison each sample is hybridised together with a common

reference sample (Figure 3a). The indirect design is a popular design for two-colour microarray studies (Figure 3a). In this design, all samples are labelled with the same dye and a pool of reference sample labelled with the other dye is created [45, 46]. Two advantages with this design are that all comparisons are made with equal efficiency [46] and the high flexibility in grouping of samples allows the comparison of any subgroup with any other subgroup [45]. This makes cluster analysis possible [45], which is described in the Data analysis section. The indirect design is not as sensitive as other designs for the loss of microarray experiments [45]. The main disadvantage is that half of the hybridisations are used for one sample, the common reference, which gives no biological contribution to the study [45, 46]. Some argue that the design increases the technical reproducibility [46].

An important characteristic of the common reference is the expression of most genes present in the sample under study [45]. This can be achieved by either pooling sample material [46] or by using a mixture of mRNA from several cell types [45, 46]. The same batch of reference should be used for all the hybridisations in one study to avoid the introduction of unnecessary variation [45].

Two designs using direct comparison are the dye swap design (Figure 3b) and the loop design (Figure 3c). Both of these require fewer microarrays to analyse the same number of samples than the indirect strategy [45, 46]. In the dye swap design, samples are paired, and each sample is labelled with both dyes (Figure 3b) [46]. The loop design can be an alternative to the indirect design, but large loops (more than 10 experiments) may be inefficient, and the design is sensitive to hybridisation failures [46].

Data analysis

The experimental work is the easiest and least time-consuming part in a microarray study, analysing the data takes a great deal of time. Further, new methods are continually being developed, with new approaches to improving analysis rapidly evolving.

Raw intensity signals are extracted from scanned images using image analysis software. There are numerous software packages available, but they all include the same basic steps. A grid is created to localize the spots. It is important at this stage to keep track of which spot belongs to which gene among the thousands of features. The spot boundaries are determined and the foreground and background signal intensities are calculated. The background signal intensity is subtracted from the foreground signal intensity to generate the raw data signal. The signal intensities scale ranges from 0 to 65,535. Signal intensities are generally expressed in base 2 logarithm; the data is compressed to make it easier to see trends in the data.

The next step, known as “pre-processing” is to exclude signal intensity features of poor quality. First, a cut-off criterion for what is considered to be a positive signal is set, often background plus two standard deviations [47]. During the pre-processing

proper spot shape, spot size and smooth signal intensity across the feature, etc. are checked. A graph known as "MA-plot" is a useful tool for evaluating the quality of an experiment during [48]. This graph (Figure 4) is simply a plot of the log ratio of the signal intensity (M) on the y-axis against the mean signal intensity of the two channels on the x-axis. One unit on the base 2 logarithmic scales equals a two-fold change in ratio or signal intensity for both down-regulated and up-regulated genes [49]. The plot can be used to detect spot artefacts and intensity-dependent patterns [48, 50]. Other methods to identify spatial patterns exist [51].

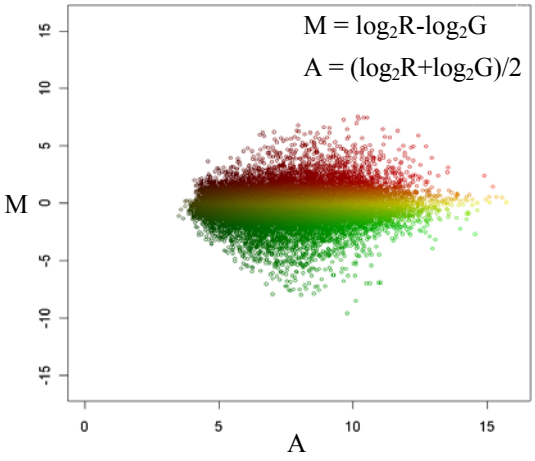


Figure 4: MA-plot of a microarray experiment where M is the log bas 2 ratio and A the log base 2 mean intensity signal of the two channels.

Proper pre-processing leaves only good quality features, but there remains random and systematic variation in the data [50]. These variations are due to such factors as starting amount of RNA [39, 40, 52], uneven labelling [39, 40, 50-53], hybridisation [40, 50], detection efficiency [39, 50-53] and spatial dye effects [51, 53]. Normalization is a process used to correct for the systematic variation. Numerous normalization methods exist, including global normalization utilizing all features on the microarray, normalization with housekeeping genes, use of internal controls, and non-linear normalization taking spatial differences into consideration [47, 51]. The method in which spatial effects are considered is called "locally weighted scatter plot smoothing" (LOWESS or LOESS), and is believed to be one of the best methods available today [47]. Microarray experiments can be normalized both between (Figure 5) experiments and within an experiment [53].

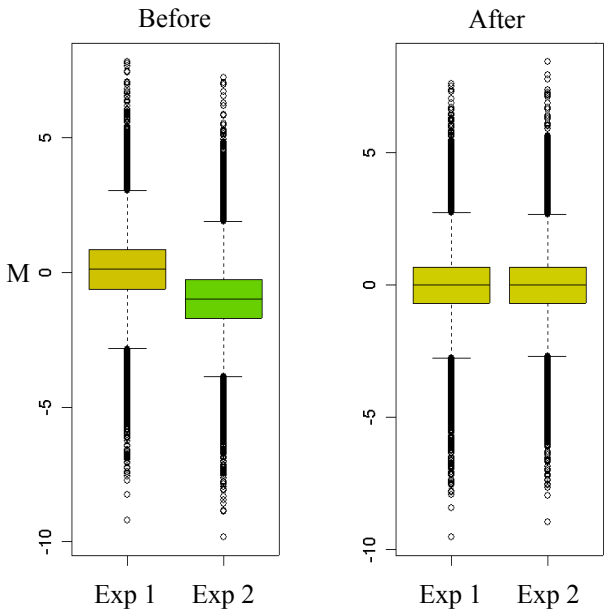


Figure 5: Boxplots of the log ratio of signal intensity (M) for two microarray experiments (Exp 1 and 2) before and after normalization using a LOESS method.

Identification of differential mRNA expression due to biology and not due to the

systemic and random variability is a challenge. The definition of a differentially expressed gene and how to identify it has varied during the years. In the earliest studies a fold change cut-off value in signal intensity was used [31, 54, 55]. Most studies used a two-fold up-regulation or down-regulation as cut-off for differential expression [39]. This criterion is not based on statistics and lacks a measurement of confidence [56].

The process of identification of differentially expressed genes has veered towards statistical tests, both parametric and non-parametric tests. A common limiting factor in order to achieve significance in the statistics is the number of samples/microarrays. The high number of data points, tens of thousand of genes per microarray, has raised statistical issues that need to be considered, such as the problem with false positives.

Use of Student's *t*-test applied on every gene is not suitable because of the assumption of equal variance for all genes, which is not true [56]. Several variants of the *t*-test have been developed that take the non-homogenous error variance into consideration. Some examples of modified *t*-tests are significance analysis of microarrays (SAM) [57], the regularized *t*-test [58] and Bayesian (B) statistics [59]. The tests create ranking lists of genes. The genes most likely to be differentially expressed are given the highest scores and end up on the top of the list. The result can be graphically presented in a volcano plot (Figure 6), in which the signal intensity, M value, is plotted against the ranking score. The interesting genes are found in the upper corners of the graph. They have high-ranking scores with small variance and large log ratios. It is not easy to decide the location on the ranking list at which the cut-off for differential mRNA expression should be set.

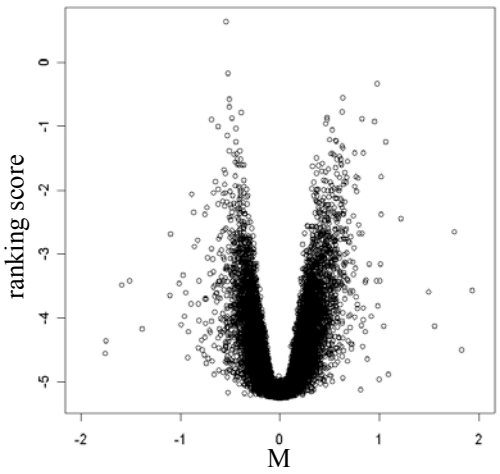


Figure 6: The volcano plot is commonly used to visualize results from statistical tests. The M value (x-axis) is plotted against the ranking score (y-axis).

Other statistical tests that have been used for microarray analysis are the analysis of variance (ANOVA) when several conditions are present, the Mann-Whitney U-test and Wilcoxon's matched pairs signed rank test [56].

The problem with false positives, genes erroneously identified as differentially expressed, has already been mentioned as an important issue during microarray analysis. A second problem is that of false negatives, where differentially expressed genes are missed. Approaches to control the number of false positives are the stringent family-wise type I error rate (FWER) [56] and the less stringent false-discovery rate (FDR) [56, 60]. A too stringent criterion may miss differentially expressed genes, and for this reason it is often preferable to use a less stringent criterion.

In the first reported microarray studies no validation experiments to confirm the microarray results were performed. With knowledge about the problem with the false positives, proper verification of the identified differentially expressed genes using another method became necessary. Different methods, including real-time PCR [61-63], Northern blotting [63, 64], *in situ* hybridisation, ribonuclease protection assay and immunohistochemistry with tissue microarrays [63], have been used for the verification of mRNA levels. The validation experiment is generally performed starting from the same RNA as that used for the microarray experiments. An alternative study approach is to perform microarray experiments on one part of the study cohort to identify significant mRNA expression differences, and real-time PCR verification using the entire population. This design is effective and a higher number of samples can be used at less expense. Various statistics can also be used for validation [64].

Bioinformatics and computational biology are used for putting the results into a biological context. Clustering algorithms arrange experiments together in groups according to similarities in the patterns of mRNA expression. The expression pattern can reveal functional groups of genes regulated in a similar fashion and involved in a certain biological pathway. Clustering methods can be divided into supervised and unsupervised methods [65], of which only the unsupervised will be discussed here. In unsupervised clustering no previous knowledge about the samples is required, only the mRNA expression data. Hierarchical clustering is an often-used method to study microarray data [52, 65]. Two other popular clustering methods are self-organizing maps and k-means clustering [52]. The clustering result is visualized in a dendrogram where the length of the branch describes the degree of similarity [65].

The clustering analysis divides genes into functional groups, but it does not give any information about the biological function of the genes. The Gene Ontology (GO) project aims at categorizing all known genes according to their biological process, molecular function and cellular compartment by searching different biological databases [66]. One gene can be associated with several terms in all three categories. Classification of interesting genes by GO terms can identify biological areas important for the disease under study. Another method for obtaining information concerning biological relevancies is to search the genes whose biological pathways are already known. Such genes are catalogued in databases such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) [67] and ResNet (Ariadne Genomics) database.

The need for standardization of the microarray technology has lately been in focus as problems with reproducibility have been noticed. Two different laboratories studying the same biological question may get different results concerning the up-regulation or down-regulation of mRNA expression with differing experimental protocols and data-analysis approaches [47]. Today most journals require that information about the microarray study is submitted to a database or public repository such as Array Express [68], the Gene Expression Omnibus [69], and the CIBEX database [70] before

publication. The data is organized according to the Minimum Information About a Microarray Experiment (MIAME) proposed by the Microarray Gene Expression Data (MGED) Society (www.mged.org) [71]. MIAME consists of six parts: experimental design, array design, samples, hybridisations, measurements and normalization controls [71]. Recording and reporting of microarray studies in a similar fashion will facilitate the comparison of results between research groups and allow re-analysis of data from another group.

Real-time polymerase chain reaction

The polymerase chain reaction (PCR) is an efficient way of amplifying specific DNA sequences. The method was developed by Kary Mullis in 1988 [72] and includes three main steps: denaturation of the double-stranded DNA, annealing of sequence-specific primers to the DNA, and synthesis of the new DNA strand. Each three-step cycle leads to an exponential amplification in the number of DNA fragments.

Quantitative real-time PCR is the most sensitive technique for measuring mRNA expression levels and is often considered as the gold standard [73]. It is possible to detect one single copy of a specific fragment in a sample, and to differentiate between nearly identical transcripts [74]. The specificity and reproducibility are higher than those of other transcript expression methods like microarray technology and SAGE, but the throughput is lower [73]. Real-time PCR is best suited for the evaluation of a few genes in a high number of samples [73].

Real-time PCR is used for mutation detection and allele detection, diagnostic purposes and the detection of splice variants, etc. A common approach for transcript profiling studies is to use microarray technology for high throughput screening to identify candidate genes, which can be studied in greater detail using real-time PCR.

Several different instruments are available for real-time PCR, including the LightCycler system (Roche) and the ABI system (Applied Biosystems). The method is fast and accurate with extremely high sensitivity [74] and no additional assays such as sequencing are required, which saves both time and reagents [75]. One disadvantage is that the real-time PCR equipment is expensive, as are the reagents required.

Experimental details

The experiment consists of two different reactions: the RT synthesis, in which RNA is copied to cDNA, followed by PCR amplification (Figure 2). The reactions can either be performed in one step, including both RT and PCR in the same tube, or each reaction can take place in a separate tube. The one-step approach minimizes the experimental variation and the risk for DNA contamination [74] compared to the two-step experiment. Separation of the reactions will, on the other hand, allow the use of the same cDNA sample on several occasions, for evaluation of the same gene several times or for completely different genes [74].

The enzyme used for RT is usually either the avian myeloblastosis virus reverse transcriptase (AMV-RT) or the Moloney murine leukemia virus reverse transcriptase (MMLV-RT) [75]. The MMLV-RT is better for the synthesis of full-length cDNA fragments, while AMV-RT is less sensitive to RNA secondary structure [75]. Selection of primer type needs careful consideration depending on the type of study. Gene-specific primers generate lower background but are limited to only one specific gene [75]. Random primers and oligo-dT primers facilitate the analysis of many different genes from one RT-reaction [75].

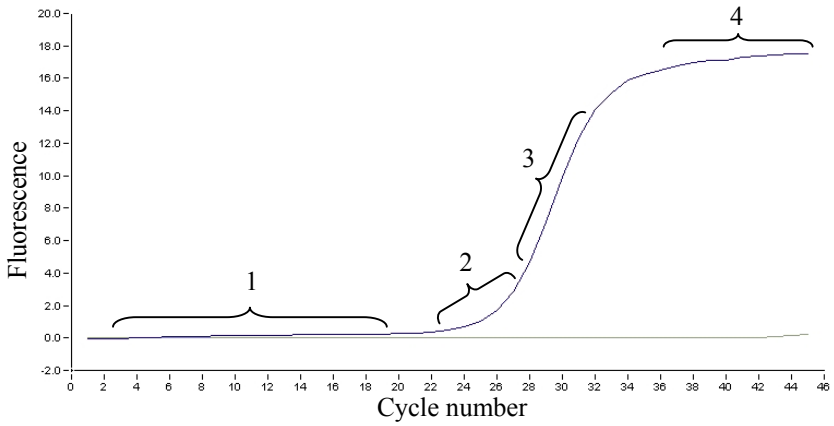


Figure 7: Illustration of the different phases in a PCR amplification with the PCR cycle number represented on the x-axis and the fluorescence on the y-axis: 1) The linear ground phase, 2) the early exponential phase, 3) the exponential phase and 4) the plateau phase.

A PCR can be divided into four phases: 1) the linear ground phase, 2) early exponential phase, 3) exponential phase and 4) the plateau phase [74] (Figure 7). The exponential amplification rate can be described by Equation 1:

$$N_n = N_0 \times (1+E)^n \qquad (1)$$

where N_n is the number of DNA molecules after n cycles, N_0 is the number of DNA molecules before the PCR, and E is the amplification efficiency, which ranges from 0 to an optimum of 1 [73, 75]. Small changes in E accumulate with the number of cycles and can greatly impact the final result. A number of parameters such as the length of the amplified DNA fragment, primer design, reagent concentrations (Mg^{2+} and nucleotide concentration) and PCR settings affect the efficiency [73]. Enzymes with both RNA polymerase (RT) and DNA polymerase (PCR) activity exist, but they are less sensitive than two-enzyme systems [75].

The amount of amplification product is monitored in "real-time" at each PCR cycle during the entire run by detection of the amount of fluorescent light emitted [73] (Figure 7). The increase of fluorescent light is directly correlated to the increase of amplification product [73]. The threshold cycle (C_t), or crossing point (Cp), is the cycle number during which the fluorescence reaches a level significantly above background, which happens during the exponential amplification phase [74, 76]. A common cut-off

is ten times the standard deviation of background [74]. The C_t value is used to calculate the mRNA expression level and the quantification will not be affected by depletion of reagents [75]. The reproducibility decreases with increasing C_t values [75].

Experimental design

Several monitoring systems with approximately equal detection sensitivity are in use. The simplest detection system is fluorescently labelled DNA-binding dyes, such as SYBR Green [77], which non-specifically bind to the double-stranded PCR product. The dyes have high flexibility and can be used to monitor many different genes, although only one gene at a time [74]. Determination of specificity is performed by dissociation curve analysis [74]. The fluorescence is monitored as a function of temperature to create a dissociation or melting curve (Figure 8) [78]. At a certain temperature, the double-stranded PCR product dissociates and releases its incorporated SYBR Green molecules. This leads to a sudden drop in fluorescence, which is seen as a peak (Figure 8). The two-step experimental approach is preferably used with the SYBR Green detection system [74].

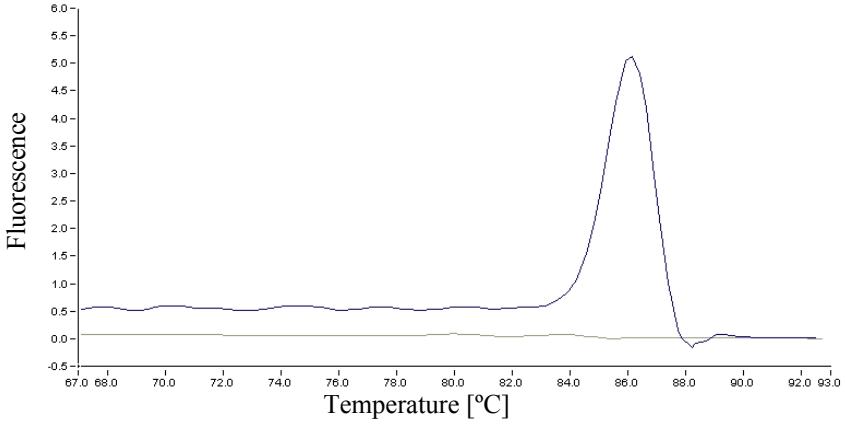


Figure 8: The melting curve is used for quality control of the product formed in the SYBR Green analysis. The melting temperature (x-axis) is plotted against the fluorescence (y-axis).

Hydrolysis probes like the TaqMan™ system (Applied Biosystems) have higher specificity due to the use of a sequence-specific probe. A fluorescent molecule is attached in one end of the probe and a quencher molecule is attached in the other end. As long as the probe is intact no fluorescent light is emitted. During the elongation step of the amplification reaction the DNA polymerase enzyme degrades the annealed probe. The fluorescent molecule is released and when it is no longer in close proximity of the quencher molecule fluorescent light is emitted [73, 75]. A number of other detection systems are also available [74].

Both primer pair and probe require careful design. A number of free and commercial software packages are available to assist in the design process. The amplified fragment should not be too long because short fragments are generally amplified with a higher

efficiency than longer fragments, and short fragments may work satisfactorily even under suboptimal conditions [75].

Primers in a primer pair should bind to different exons to facilitate the exclusion of DNA contamination [75]; this is especially important for the SYBR Green system. The probe should span an exon-exon boundary [75], whereby only correctly spliced transcript will yield amplification product. The most favourable length for a primer is usually 15-20 bases [75], and the optimal probe length varies with the detection system. The melting temperature, which affects the annealing of primer/probe to the sequence, must be considered, as must the G/C content, and secondary structure, etc. [75].

A number of sources introduce variability into a real-time PCR experiment. PCR inhibitors can be carried over from the cDNA synthesis; the quality and concentration of the RNA may vary; as may the performance of the RT reaction, the PCR efficiency, and the biological sample itself [74]. Inclusion of different kinds of controls in the real-time PCR experiment decreases some sources of variability. One way to check for the presence of contaminating DNA is to include a negative RT control in the PCR [74]. The negative RT control is a sample from the RT reaction with all reagents added except for the RT enzyme; no product can be formed. It is also possible to control for undesirable DNA by designing the primer such that it allow amplification only of correctly spliced transcripts.

Individual variation in the amount of starting RNA and the efficiency of the RT synthesis can be controlled by an internal standard in a process often known as "normalization". The internal standard is a gene present in the sample that has equal expression level in all tissues, and at all times regardless of how the sample is treated. So far no such control has been identified. Often one or more housekeeping genes are used, such as β-actin, glyceraldehydes-3-phosphate-dehydrogenase (GAPDH) and ribosomal RNA (rRNA) for the normalization [75]. The housekeeping genes are, however, not always as perfect internal standards as it was first believed [74, 75]. The internal standard and the sample are amplified in separate reactions. Attempts to normalize transcript levels have also been carried out by correlating transcript levels to the total RNA concentration and using the average normalization factor of several housekeeping genes [74]. It is difficult to use total RNA correlation due to the fact that the amount of total RNA may vary depending on cellular state, the quality of the RNA may vary depending on unknown factors, as may the RT reaction efficiency. None of these variations are considered [74].

Contamination of any kind can easily destroy a real-time PCR experiment, by causing false products or by inhibiting the amplification process. A no-template control, including all PCR reagents with no sample added, should always be included in every real-time PCR run to guarantee pure reagents. External controls, for the control of amplification inhibition, will not be discussed here.

Differences in mRNA expression of transcripts with low expression levels are more difficult to detect than expression of transcripts with high expression levels [75].

Data analysis

Real-time PCR is used for both absolute and relative quantification. For absolute quantitation, the C_t values for a serially diluted standard with known concentrations are measured to create a standard curve. The linear relationship between C_t value and the sample amount is used to calculate the concentration of unknown samples [74]. Both DNA and RNA standards can be used to create the standard curve [74].

For relative quantification, a normalization sample is used, based on one or several housekeeping genes. A number of different mathematical approaches can be used to calculate the relative expression levels [74]. Approximately equal amplification efficiency is important for appropriate relative quantitation [74]. In the standard curve method for relative quantification, a standard curve is created for each gene of interest. The concentration of each sample is estimated and the expression level is correlated to the normalization sample. The comparative C_t method expresses the difference in mRNA expression as the relative difference between the expression of the unknown sample and that of the normalization sample [74]. Other methods are the Pfaffl model, Q-Gene, and the amplification plot method [74]. Absolute quantification is more labour-intensive than relative quantification, but it is preferred if experiments are run on different days or in separate laboratories [74].

CFS and transcriptome analysis

Gene expression profiling techniques are valuable tools for evaluation of unfamiliar biological processes involved in diseases and illnesses with unknown etiology. Two different methods have been used for transcript expression level analysis in CFS research: microarray technology [22-24, 26] and differential display PCR (DD-PCR) [21, 25]. All of the studies have profiled peripheral blood cells.

The first CFS mRNA expression study used filter arrays, one of the earliest array types, to compare five female CFS patients with seventeen healthy female controls [22]. Out of 1,764 genes represented on the array, seven were identified as differentially expressed using the nonparametric Wilcoxon test. Several of the genes had an immunological function, which indicated some kind of immune system dysfunction.

Impaired immune system function and reduced T-cell activation was reported in another study comparing 25 CFS patients (16 females and nine males) with 25 healthy controls [26]. Involvement in mitochondrial function and neuronal perturbation was also observed among the 35 genes (out of 9,522 genes) identified as significantly expressed [26]. This study used real-time PCR for seventeen of the patients and a new group of healthy controls for the verification of results. Abnormalities in the metabolic pathways [23] were suggested in a microarray study of female CFS patients identified in a population study [15]. Out of 3,800 genes, 117 genes were indicated as differentially expressed [23]. All these studies used a single-colour system [22, 23, 26].

A DD-PCR study compared seven CFS patients (two female and five male) with an infectious illness onset with four healthy controls, and found indications of subtle changes in the immune system [25].

Post-exertional malaise is one of the eight symptoms listed in the CFS case definition [5]. Two studies, one using microarrays [24] and one using DD-PCR [21], have compared CFS patients and healthy controls before and 24 hours after exercise. The idea with DD-PCR is to search for differences in the PCR banding pattern [79]. Interesting bands are excised from the gel and genes identified by sequencing [79]. The DD-PCR study compared one female CFS patient with a female control, and found genes with differences in expression levels involved in defence and immune system functions before exercise [21]. In the microarray exercise study, 3,800 genes for five female CFS patients and five healthy female controls were compared. Twenty-one genes were identified as differentially expressed, and exercise had greater effect on ion transport and ion activity differences.

Most of the transcript expression profiling studies of CFS reported so far have reported some kind of immune system dysfunction [21, 22, 25, 26], although none of

the genes identified as differentially expressed is in common to any two studies. Genes involved in T-cell function have been reported in several of the studies [22, 25, 26].

Estrogen receptors

Estrogen receptors (ERs) are involved in regulation of the steroid hormone estrogen. The hormone, which is present in both females and males, exerts its many functions by binding to the ERs [80, 81]. Estrogen is involved in a variety of physiological processes including sexual development and the reproductive cycle [80]. The ERs are ligand-activated transcription factors that belong to the nuclear receptor superfamily [80]. There are two types of estrogen receptor, α and β, which have unique and overlapping roles [81]. For ERβ there exists a human splice variant, ERβcx, which differs at the C-terminal end of the protein [82]. ERs are often important in diseases for which the prevalences in men and women are unequal, such as breast cancer, autoimmune disease and osteoporosis [83]. Both ERα and ERβ are expressed from different promoters giving rise to transcripts with differing 5'-untranslated regions [84-87]. Six different untranslated first exons, two untranslated second exons and seven promoters have been identified for ERα in humans [88]. Two alternative first exons have been identified for ERβ, known as 0N and 0K [85], and these alternatives are tissue specific [85, 87].

CFS and estrogen

Estrogen treatment improves status of several cognitive functions [89]. In one study of premenopausal female CFS patients, 22 out of 28 patients reported improved health status following estradiol and cyclic progestin treatment. One fourth of the patients had estrogen deficiency [90]. Female CFS patients often report improved health during pregnancy, when estrogen levels are naturally increased, with a relapse to severe depression following birth [89]. Women with CFS have more problems with their reproductive health than healthy women [91].

AIMS OF THE PRESENT STUDY

The overall aim of this thesis was to evaluate transcript expression levels in peripheral blood mononuclear cells (PBMCs) from chronic fatigue syndrome (CFS) patients and healthy controls using microarray technology and real-time PCR.

The specific aims of this thesis were to evaluate:

- ✧ the impact of technical parameters such as the RNA extraction method and amount of cDNA used for hybridisation.
- ✧ the use of PBMCs to assess expression levels of genes involved in psycho-neuroendocrine-immune (PNI) communication.
- ✧ the eventual existence of differentially expressed genes in PBMCs comparing CFS patients with healthy controls.
- ✧ if there are any differences in the mRNA expression levels of estrogen receptors (ERs) in PBMCs in CFS patients and healthy controls.

MATERIALS AND METHODS

Subjects

Peripheral blood mononuclear cell samples

Peripheral blood mononuclear cells (PBMCs) from randomly selected individuals were used in Papers I and II.

CFS patients

Two different CFS patient cohorts were included in Papers III-IV and V. Five patients participated in both cohorts. All patients fulfilled the CDC-1994 international case definition of CFS [5] and were examined and diagnosed at a clinic for infectious diseases at Karolinska University Hospital, Huddinge in Stockholm (Sweden). Both cohorts were designed to have similar proportions of men and women as typically seen in a tertiary clinic (70% women and 30% men) [17].

The first cohort (Papers III-IV) consists of 20 patients with an average age of 37.9 years (26-60 years). The CFS patients were classified according to the ICD-10 system (infectious or non-infectious illness onset) and the illness onset type (sudden or gradual). Eleven patients (10 females and 1 male) had an infectious illness onset and nine patients (5 females and 4 males) a non-infectious onset. Fifteen of the patients (11 females and 4 males) reported a gradual illness onset, while nine patients (7 females and 2 males) reported a sudden onset. The median illness duration was 3 ¼ years (1-27 years).

The second patient group (Paper V) consists of 30 individuals with an average age of 40.5 years (26-54 years). Epidemiological data was available for 24 of 30 patients. Thirteen patients (9 females and 4 males) reported an infectious illness onset and eleven patients (9 females and 2 males) a non-infectious onset. Eleven of the patients (9 females and 2 males) had a gradual illness onset, while nine patients (6 females and 3 males) reported a sudden onset. The median illness duration for the second cohort was 4 ¼ years (1.5-25 years).

Healthy controls

Healthy voluntary age-matched and sex-matched controls were used in Papers III-V. Fourteen healthy controls were included in Papers III-IV with an average age of 37.8 years (25-57 years). In Paper V, 36 controls with an average age of 43.9 years (26-65 years) participated.

Sample preparation

Whole blood was collected in citrate (Papers I-II) or heparin (Papers III-V) tubes. Immediately following the blood draw, PBMCs were isolated. Cells were washed and counted followed by storage in liquid nitrogen (Papers I-IV) or in cell lysis solution at -80ºC (Paper V) until used for total RNA extraction.

Two different RNA extraction methods were compared in Paper I. The extraction method referred to as single-step method in Paper I was used in the rest of the studies. Due to limited RNA supply in Papers III-IV, the mRNA was amplified to yield aRNA. RNA samples were stored at –80ºC. Complimentary cDNA was synthesized from extracted total RNA (Papers I, II and V) or amplified aRNA (Papers III-IV). A MMLV-RT enzyme with either a combination of oligo-$(dT)_{12-18}$ primer and random primers (Paper I-II) or only random primers (Papers III-V) was used. Samples planned for use in microarray experiments were labelled with biotin (Paper I-II) / fluorescein or Cy3/Cy5 (Papers III-IV). Complementary DNA samples were stored at –20ºC until used for analysis.

Transcriptome analysis

Microarray technology

The one sample-one microarray approach was used in Paper I-II and the two samples-one microarray approach was used in Papers III-IV. Two detection systems with different detection strategies, the RLS system [43] and the fluorescent system, were used. Microarray experiments were performed with both commercially available oligonucleotide microarrays, representing 10,000 unique human transcripts (Papers I-II) and in-house manufactured cDNA microarrays with approximately 30,000 spots representing 19,000 unique human transcripts (Paper III-IV).

Thirty-six hybridisations (12 individuals) were performed in Paper II, three microarrays representing 10,000 genes each for each study participant. Indirect study design with 30 biological replicates was used in Papers III-IV. Automated hybridisation was used in Papers I-II and manual hybridisation was used in Papers III-IV.

In experiments generating unpublished data, which will be discussed later, the RLS two-colour system was used. Samples were prepared as in Paper I and hybridised manually according to manufacturer's instructions. The microarray data was analysed as in Papers III-IV.

Image analysis was performed using either ArrayVision™ (Papers I-II) or GenePix® Pro 5.1 (Papers III-IV) software. Local background subtraction was used in all of the studies. Many different analysis approaches for microarray data exist, and in this thesis two different strategies representing two different research groups have been used. The

Centers for Disease Control and Prevention (CDC) Microarray database (MADB) was used for pre-processing of the data in Papers I-II. The data was normalized to the 75th percentile, and for some parts un-normalized data was used. The Pearson correlation coefficient was used to compare the similarity between experiments and the Lin's concordance coefficient for the reproducibility of the method.

In Papers III-IV pre-processing of microarray data and statistical computing were performed using the academic R-software [92]. Data was filtrated to remove bad spots and leave reliable good quality spots. Pearson correlation coefficient was used as a quality measure for hybridisations. LOESS print-tip normalization was used, which takes uneven dye effects depending on spot intensity and spatial spot position into consideration [51]. Genes qualified for statistical testing required ratio values for at least half of the experiments. B statistics was used to create a ranking list of the genes most likely to be differentially expressed [59]. Genes with small expression differences easily disappear among the tens of thousands of genes with no difference. To solve this problem genes with an absolute ratio (M value) larger than 0.3 were used for statistical calculations. Results were visually presented using volcano plots. Differential mRNA expression was verified using real-time PCR and sequencing.

Multi Experimental Viewer [93] was used for hierarchical clustering with complete alignment and the Pearson correlation as distance measure. Proteins coded by genes with high ranking-scores in the B tests were used for protein pathway analysis using PathwayAssist [94].

Real-time PCR

Two different real-time PCR systems were used in this thesis: In Papers III-IV the LightCycler 2.0 system (Roche) was used for validation of microarray results, and in Paper V the ABI system (ABI 7500 and 7700, Applied Biosystems) was utilized. Both double-stranded DNA-binding dye SYBR Green (Papers III-V) and TaqMan™ system (Paper V) were used. 18S rRNA and GAPDH were used for normalization in all real time PCR experiments.

Samples were run in duplicate (Papers III-IV) or triplicate and quadruplicate (Paper V). All samples for one gene were run on the same day and no template controls were included in each run.

Standard curve analysis (Papers III-V) and comparative analysis (Paper V) methods were used. Data calculations and statistics were performed in Excel. Two-sided Student's *t* test with unequal variance was used (Papers III-V).

RESULTS & DISCUSSION

Microarray technology platforms

Two microarray platforms with completely different detection systems have been used: the RLS system, which uses gold and/or silver particles with white light (Papers I-II), and a fluorophore system with laser detection (Papers III-IV).

Clinical samples are often small and must be used wisely. Standardization and optimisation of a technology like the microarray technology is therefore important. Properties such as RNA quality, labelling efficiency, sample concentration and hybridisation affect the final microarray result [39, 95].

In Paper I, two different total RNA extraction methods, referred to as the single-step extraction method [96] and the glass filter extraction method [97], were compared both with respect to the quality of the RNA obtained and with respect to the reliability of the hybridisation. The quality of the total RNA is important because partially degraded or poorly processed sample may not truly represent the transcript profile [39].

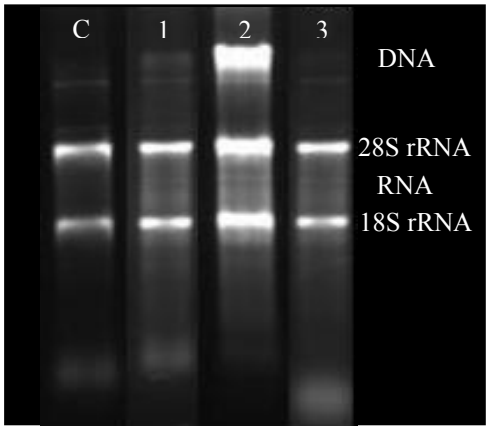


Figure 9: Total RNA extracted with: (1) the single-step method, (2) the glass-filter extraction method before DNase treatment and (3) after DNase treatment. C=control.

Both extraction methods yielded good quality total RNA, although the filterbased method frequently left contaminating DNA (Figure 9). Although the residual DNA did not seem to affect this cDNA labelling synthesis, the presence of DNA would have a serious impact on validation experiments involving a PCR step. The total RNA extracted with the filter-based method was treated with DNase before the RT synthesis.

Variability in final hybridisation results was evaluated for several concentrations of labelled cDNA, starting with total RNA obtained by both extraction methods. Technical replicate experiments for the glass filter extraction method showed more discrepancies for both cDNA concentrations tested than the single-step extraction method, which may be due to the DNase treatment. Differences were small and the two RNA extraction methods yielded comparable final hybridisation results. The same method should be used throughout the study in order to minimize variability. The amount of cDNA that was hybridised to the microarray affected the dynamic range. A greater dynamic range for abundant genes was observed for lower cDNA concentrations, while the dynamic range was greater for genes with low expression when more cDNA was used.

Paper I describes experiments carried out using automated hybridisation, while the other three papers describe experiments that used manual hybridisation. Automated hybridisation reduces human handling and provides superior and efficient control of hybridisation properties such as temperature and degree of mixing. The main disadvantage is that automation stations are expensive. Manual hybridisation requires less expensive equipment that is generally already present in a laboratory. Mixing is not possible and the procedure is more time-consuming.

The RLS system is significantly more sensitive than fluorescent systems, and enables the use of lower sample amounts [43]. One microgram of total RNA with a regular RT reaction is enough for one hybridisation with the RLS system. Other systems recommend higher amounts, about 5-20 µg of total RNA, if no amplification reaction is used. Another advantage is that it is not prone to photodegradation [43].

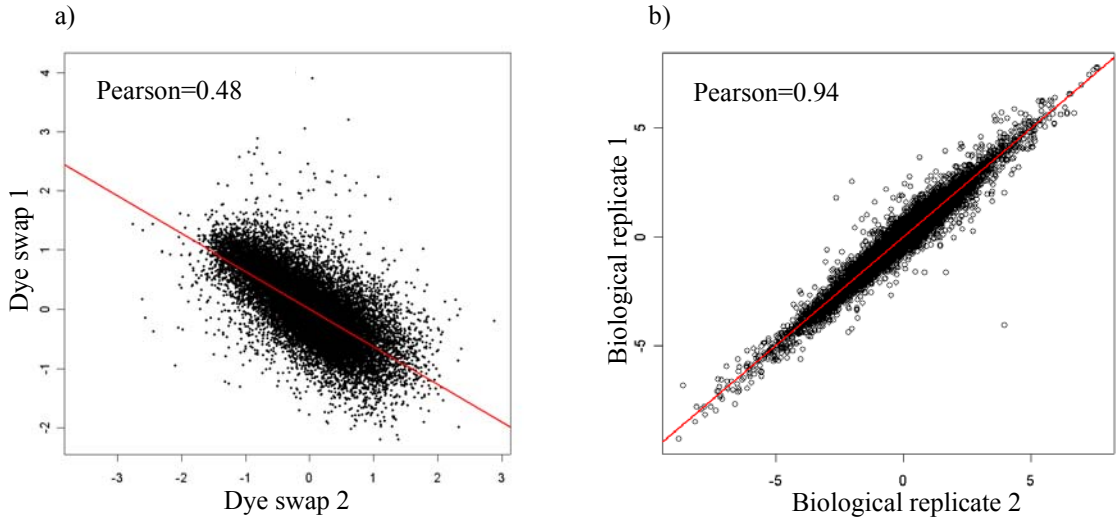


Figure 10: Correlation plots of M values with the Pearson correlation coefficient for: a) two RLS two-colour dye swap experiments and b) two biological replicate experiments from Papers III-IV.

Transcript profiles of the same Swedish CFS patients and healthy controls used in Papers III-IV were compared with the two-colour RLS system. Dye swap design with age-matched and sex-matched pairs was used. The experimental procedure was successful and everything looked promising. RLS data had not previously been analysed with the R software and the data pre-processing required substantially more optimisation than the fluorescent system. The beginning of the data analysis part went really well, but we soon started to run into problems. Hierarchical clustering of all experiments based on M values showed that the samples for which the dye swap correction had been carried out were systematically different from those for which the correction had not been carried out. Further difficulties arose with the discovery of negative correlation between any two technical dye swap experiments (Figure 10a), and almost identical mRNA expression levels for all the silver labellings and all the gold labellings as two separate groups. Unfortunately no sense could be made out of the data. No explanation has been found, although several experts have looked into the

problem. A direct experimental design was used in the RLS study. An indirect study design would not have required dye swap, and it is possible that the problems would have been avoided.

Comparison of the fluorescent and the RLS systems revealed greater dynamic range of the M values for the fluorescent system (Figure 10a & 10b) with lower maximum signal intensities for the RLS system. This is probably due to methodological differences, and cannot be explained by the lower starting amount of total RNA. Signal intensities across the entire dynamic range were observed in one-colour RLS experiments starting out with the same amount of total RNA as for the two-colour experiments. A different hybridisation protocol was used for the one-colour experiments. This protocol differed mainly in that automated hybridisation and mixing were used, and this may explain the higher signal intensities. It is not necessarily the case, however, that the use of the entire dynamic range of intensity (0-65,535) is better than the use of a narrower dynamic range. The lower sample amount requirement makes the RLS system an interesting candidate for the analysis of clinical samples. Each part of the experiment introduces systematic variability and fewer experimental steps lead to less variation. However, most research groups have well-established fluorescent protocols. The problem with limited sample amount can in this case be circumvented by amplification of either the mRNA (used in Papers III-IV) or the signal intensity. An unbiased mRNA amplification reaction is required in order to obtain a true representation of the mRNA expression levels [41, 42].

The Pearson correlation coefficients were in the same range in Paper I as they were in Papers III-IV, although $\log_2$ base signal intensities were used in Paper I and $\log_2$ base signal intensity ratios were used in Papers III-IV. The negative correlation observed in the RLS project was obvious, but the correlation coefficients were lower than those of the other studies.

The microarray research field includes many different approaches with respect to both laboratory procedures and data analysis. With proper experimental design, alternative methods can be used with success to yield good quality data. Consistency is important to decrease experimental variability. The same properties should be used across an entire study. Standardization of the microarray field has started with the MIAME project, which aims for uniform presentation of microarray data [71]. Efforts will make it easier to compare studies between different laboratories in the future but the numerous ways of performing a microarray experiment will still be a limiting factor.

Use of PBMCs in transcriptome analysis

The problems with an appropriate sample for the study of CFS have been described in the Introduction. The hypothesis has been posed that peripheral blood cells serve as indicators for abnormal processes throughout the human body. Paper II describes investigations into the presence of expression of genes involved in psychological, neuroendocrine and immune responses in peripheral blood. It has been hypothesized that all three of these areas, psycho-neuroendocrine-immune (PNI) processes, play a potential role in CFS.

The brain plays a central role in many of the PNI pathways. Due to difficulties in the availability of brain material, *post mortem* tissue or brain-derived cell lines have been used. The availability of *post mortem* tissue is still limited, with accompanying quality issues, and the correlation between cell lines and *in vivo* function is not clear. New ways to study PNI communication are desirable.

Peripheral blood cells circulate throughout the human body, and leukocytes are able to cross the blood-brain barrier. This, and the relatively non-invasiveness and easy accessibility, make peripheral blood mononuclear cells (PBMCs) an interesting option for the study of PNI communication. The individual variability in PBMC mRNA expression is small and the differences are mainly due to age and sex [27]. Variation in mRNA expression levels in human blood cells due to disease is larger than individual variations [27]. Peripheral blood cells have been used to investigate diseases like systemic lupus erythematosus [98, 99], multiple sclerosis [99], sickle cell disease [100], and in cases with no known lesion [101].

Paper II describes the creation of a comprehensive database of 1,622 annotated PNI genes by soliciting molecular biologists, immunologists, endocrinologists, neurologists and psychiatrists, and by reviewing articles from interesting areas. Sixteen percent of the genes were involved in the nervous system, 20% in the endocrine system and 38% in the immune system. The remaining 26% were genes taking part in more than one of the systems, or they were important because of their regulatory properties.

Expression of genes in the PNI database was assessed by querying peripheral blood-specific databases generated from EST data, and by microarray analysis of 30,000 human genes using PBMC samples. Of the 1,622 genes in the database, 566 genes were in common with the EST database, and half of these 566 genes were involved in the immune system. Seventy-nine percent of the genes in the database were represented on the microarray, and 60% of these genes were detected in PBMCs. The proportion of gene function groups among genes positive for hybridisation was similar to that of the PNI database. In total, 1,058 genes (65%) in the PNI database were detected in PBMCs. Several neural and endocrine genes were expressed in the peripheral blood,

including hormone receptors, a hormone responsive transcription factor, and neurotransmitter receptors.

Table 5: The categories and distribution of PNI genes in the three databases.

System	Category	PNI Database	Microarray Database	EST Database
Endocrine	Hormone metabolism	81	33	17
	Hormone receptor	94	43	12
	Hormones	45	22	1
	Regulated by hormones	29	15	11
	Regulates hormone activity	55	20	25
	Regulates hormone expression	19	12	6
Immune	Apoptosis	44	17	30
	Complement component	30	18	8
	Cytokine/chemokine receptors	90	44	38
	Cytokines/chemokines	108	57	31
	Immune: MHC/HLA	22	4	20
	Other immune function	287	123	147
	Regulated by cytokines	9	5	4
	Regulates cytokine activity	22	10	8
	T-cell activation	6	0	3
Nervous	Amyloid processing	18	12	7
	Neurotransmitter	19	12	0
	Neurotransmitter metabolism	33	16	10
	Neurotransmitter receptor	101	44	3
	Other neural function	37	19	3
	Regulated by neurotransmitters	2	1	1
	Regulates neurotransmitter activity	51	29	10
	Regulates neurotransmitter expression	2	2	0
Other	Circadian	7	4	4
	Growth factor	27	13	5
	Growth factor receptor	13	5	2
	Heat shock	20	8	11
	Homeostasis & small molecule transport	37	18	6
	Other	18	10	10
	Other neuroendocrine function	34	20	12
	Protease inhibitor	9	3	4
	Regulation of cell growth	63	28	18
	Signal transduction	76	40	41
	Stress response	10	4	9
	Transcription factor	100	50	46
	Unknown function	4	3	3
	Total	1622	764	566

The predominance of genes involved in the immune system was expected, as the immunological function of PBMCs has been well characterized. A predominance of genes with immune system function was seen in both the microarray and blood EST database. During the microarray analysis, more genes from the nervous and endocrine categories were identified than had been anticipated. This indicates that blood contains a lot of information about biological processes in other parts of the body, and can help to elucidate the communication between the brain and the body. It is possible that analyzing mRNA expression of PNI genes will contribute to the classification and understanding of disease states that have eluded conventional diagnostic approaches.

Transcriptome analysis of CFS

Papers III-V describe transcription analysis of Swedish CFS patients and controls using both microarray technology (Papers III-IV) and real-time PCR (Papers III-V). With the microarray technique, mRNA expression levels for tens of thousands of genes have been evaluated, while real-time PCR has been used to study a few selected genes. Small transcript expression differences, if any, were expected to be found between the entire group of CFS patients and healthy controls. No indication of any differential transcript expression was found between the entire groups in Papers III-IV, no genes had B values above 0 (Figure 11a). We did see significant differential mRNA expression for ERβ (Paper V), which was not present on the microarrays used in Papers III-IV.

Differential mRNA expression identified for a female patient subgroup

The individual variability in PBMC mRNA expression is small, with the largest differences between different blood components [27]. The PBMC sample is a more homogenous sample, consisting of B-lymphocytes and T-lymphocytes, and individual variability was also small. When the entire study cohort in Papers III-IV was subjected to B test and cluster analysis, female and male CFS patients compared with female and male healthy controls, the mRNA expression differences were larger between the sexes than they were between the patient and the control group. A hierarchical clustering applied to all samples, and all genes clustered all women into one large group differentiated from men. However, comparing only female patients with female controls did identify genes with higher probabilities of being differentially expressed (higher B values). Further analysis was performed using only female patients and controls.

It is not clear whether CFS is one single entity or a group, nor is it clear whether males and females suffer from the same or different entities. It is important to study the cohort as one big group and by subgroups according to epidemiological variables like sex and illness characteristics. In order to elucidate all possible new insights into the relatively unknown pathophysiology and etiology of CFS, an unbiased and complete

scheme of pairwise group comparisons based on epidemiological data was performed to identify differentially expressed genes. Female patients were divided into the following subgroups: according to the ICD-10 classification system, illness onset type, illness duration, and number of symptoms (CFS patients in Materials and Methods).

Statistical tests comparing female CFS patients with no previous documented infection (n=10) with female controls (n=12) and female patients with a gradual illness onset (n=9) with female controls identified eight overlapping genes with high ranking scores. Comparing the female patients with both absence of previous documented infection and gradual illness onset (n=8) with healthy controls yielded seven of the eight overlapping genes on top of the ranking list, indicating possible significant mRNA expression differences (Figure 11b). Seven of the eight genes had B scores above 1 and one gene had a B above 0, and all the genes had *p*-values below 0.001 using Student's *t*-test. These eight genes were selected for further investigation.

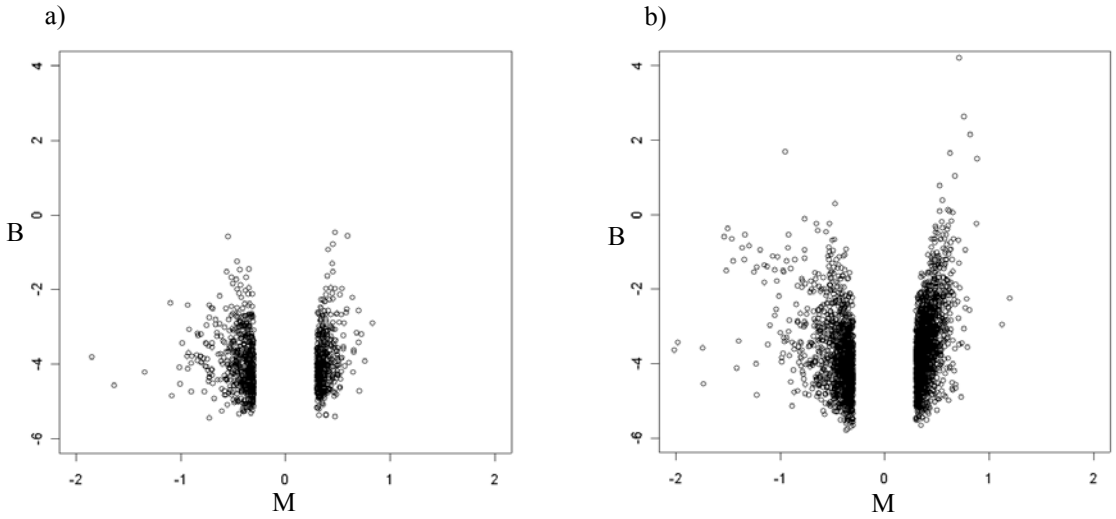
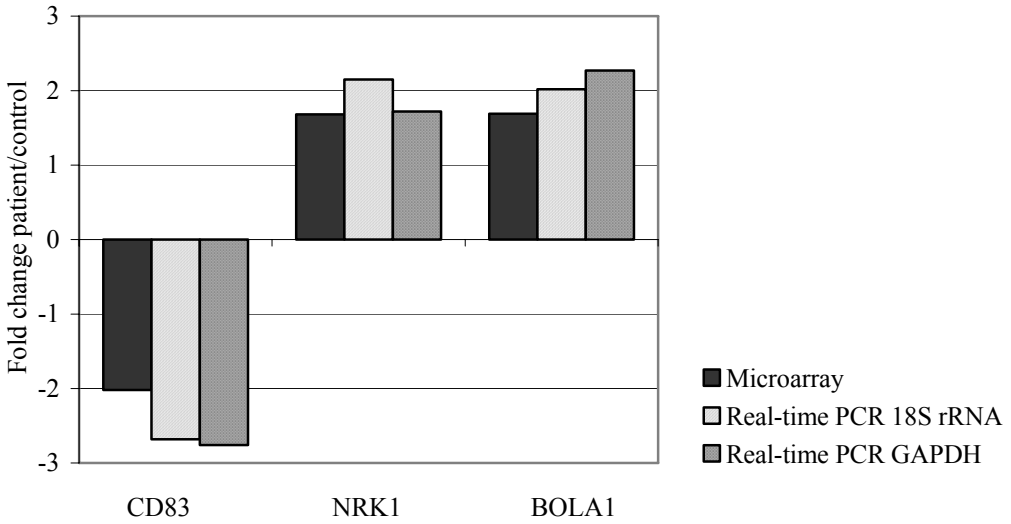


Figure 11: Illustration of B test results for: a) CFS patient versus controls both females and males and b) female patients with gradual and non-infectious illness onset versus female controls.

It was possible to verify significant differences in gene activity for three out of the five genes with known identity (CD83, NRK1 and BOLA1), identified with microarrays, using real-time PCR (Figure 12). Significance was achieved using both 18S rRNA and GAPDH for normalization. For SYNC1, there was a trend of differential expression, but no significance was achieved (p=0.06). Sequencing the fifth gene, WDR47, showed that it was not similar to any known gene. Hierarchical clustering of the subgroups of female CFS patients and female healthy controls using the five genes differentiated most of the patients from controls (Figure 13).

The human glycoprotein CD83 has a molecular weight of 45 kDa and belongs to the Ig superfamily [102, 103]. Western blot analysis of the CD83 protein identified a weak band from a protein of the correct size that is suspected to be CD83. The CD83 protein functions as a maturation marker for dendritic cells (DCs) and is also present in

activated lymphocytes [104]. The full role of the protein in the immune system has not been fully elucidated. Down-regulation of CD83 due to infection by several different viruses has been observed [105]. It has been suggested that this is a viral mechanism to escape host-specific immune responses [105]. Inhibition of the stimulatory function of DCs leads to impaired antiviral T-cell responses [105]. Down-regulation of CD83 mRNA levels in the CFS patient subgroup may lead to lower expression levels of the protein, which may, in turn lead to disturbed T-cell activity. Protein and mRNA levels are not, however, always correlated. Differential mRNA expression levels for other genes involved in T-cell activation has been reported in the other studies of CFS tr anscript expression [22, 25, 26].



***p*-value**			
Microarray	0.00002	0.00007	0.00005
18S rRNA	0.008	0.0002	0.008
GAPDH	0.002	0.03	0.005

Figure 12: Fold change differences for significantly expressed genes comparing female CFS patients with no previous infection and gradual illness onset with healthy female controls. Microarray technology and real-time PCR with 18S rRNA or GAPDH for normalization was used.

A number of psychiatric and medical treatments have been tested for CFS patients with varying success [20]. Nicotinamide adenine dinucleotide has been tested for the treatment of CFS. NRK1 is up-regulated in the CFS patient subgroup. This enzyme is involved in the synthesis of nicotine amide dinucleotide (NAD^+). The function of the third gene, BOLA1, is not known.

Whistler *et al.* have identified genes, mainly involved in metabolic pathways, that distinguish between female patients with sudden and gradual illness onset [23]. We found an overlap of the highest ranked genes in pairwise comparison of female healthy controls with female patients with gradual illness onset and without previous infection. Most of the patients in this study cohort with no previous documented infection reported gradual illness onset. We did not find any metabolic abnormalities. Other

studies of CFS have either looked at female patients without subgrouping [22], or both sexes without subgrouping [25, 26].

Previous microarray studies of CFS have found a larger number of differentially expressed genes than we have found here. The agreement between the different studies is not good, however, although some categories of biological processes are recurrent. These categories include immune responses and T-cell activation [22, 25, 26]. In all the transcript profiling studies, including this one, different microarray platforms, microarrays representing different genes with some genes in common, different analysis approaches, and different stringencies in the definition of differentially expressed genes have been used, which makes them difficult to compare.

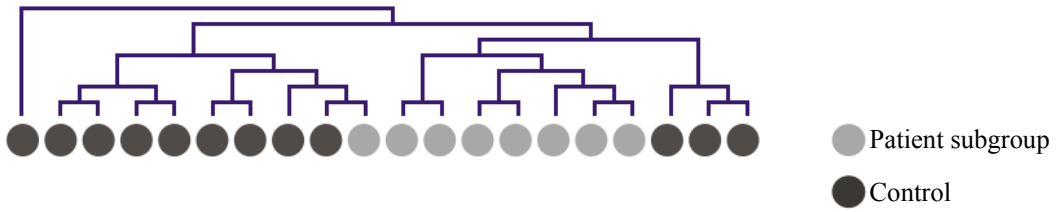


Figure 13: Hierarchical clustering of a subgroup of female CFS patients (with no previous infection and gradual onset) and healthy female controls using transcript expression results for CD83, NRK1, BOLA1, SYNC1 and WDR47.

Statistical issues

Small but significant mRNA expression differences generated in a microarray study may be difficult to detect among the tens of thousands of genes showing little or no variation between the compared groups. A selection criterion for a certain degree of variation between the two groups to be compared can be used to solve this problem. Only genes that pass the selection process are used for further statistical calculations. A suitable threshold must be determined for each study. We tested several criteria and chose a "moderately stringent" criterion. With too stringent variability criteria it is easy to miss small differences of biological importance.

The decision of which cut-off criterion to use for the definition of a differentially expressed gene is not straightforward either. Too stringent a criterion will only identify genes with large changes between the compared groups. Subtle changes in mRNA expression can have significant biological effects, and with too high a stringency these would be impossible to detect. The main disadvantage with a looser criterion is the increase of false positives, which is a well-known problem in microarray data analysis. The stringency requirement differs depending on the type of study. We have used a less stringent criterion because little is known and small changes, if any, are expected. It is of utmost importance to verify the results using a different technique such as real-time PCR.

Indication of differential mRNA expression between patient subgroups

We have seen indications of transcriptional differences between several of the other female patient subgroups. The differences were not statistically significant. This may be due to the low number of samples in some of the patient subgroups. Indications of gene expression differences were observed between female patients with a non-infectious (n=10) and infectious illness onset (n=3), and between patients with gradual illness onset (n=9) compared with sudden onset (n=4), as well as between healthy controls and patients with previous infection and sudden illness onset. No transcriptional differences were observed for the duration of CFS illness or between varying numbers of fulfilled symptoms (4-5 symptoms (n=9) compared to 7-8 symptoms (n=6)).

Hierarchical clustering of all female samples using the 51 most highly ranked genes in the various pairwise comparisons (no infection versus infection, sudden versus gradual illness onset and the patient subgroups versus healthy controls) gathered most of the patients into two clusters (Figure 14). A similar analysis including both female and male patients still differentiated most of the female patients from the controls, but the male samples did not follow the same pattern. This could indicate that men show similar symptoms but suffer from a different illness than women.

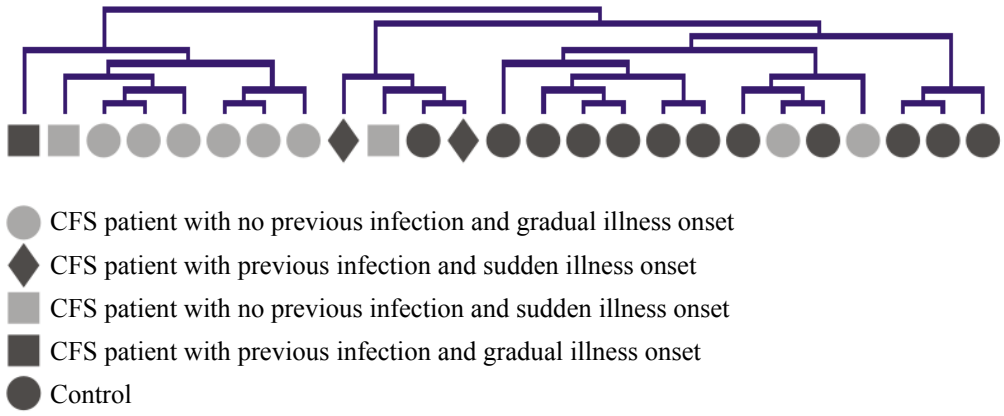


Figure 14: Hierarchical clustering of female CFS patient subgroups and female healthy controls using the 51 top ranked genes in the various pair wise comparisons.

One limitation with this study was the small number of patients in some of the CFS subgroups. Further studies including a larger study cohort with more patients in each of the subgroups and more men are desirable. The transcript level differences also need to be correlated to protein levels.

Reduced estrogen receptor β levels in patients

We have observed significantly lower mRNA expression for ERβ in the CFS patient group compared to the healthy control group (Paper V) (Figure 15). This difference was reproducible when the same RNA was used for an independent cDNA synthesis, and was seen when samples were normalized to both 18S rRNA and GAPDH. All patient subgroups, except for patients with long illness duration (p=0.12), had significantly lower ERβ mRNA levels than healthy controls (p = 0.00002-0.004). Differences were also significant when looking at both sexes separately ($p_{females}$= < 1×10^{-5} and p_{males} < 0.02).

Although the levels of ERβ mRNA are low in the studied cells, such low levels of mRNA may be of biological importance. Furthermore, since the actual target tissue(s) for CFS are unknown, it is possible that these tissues express far higher amounts of ERβ mRNA while maintaining the differential expression observed in this study.

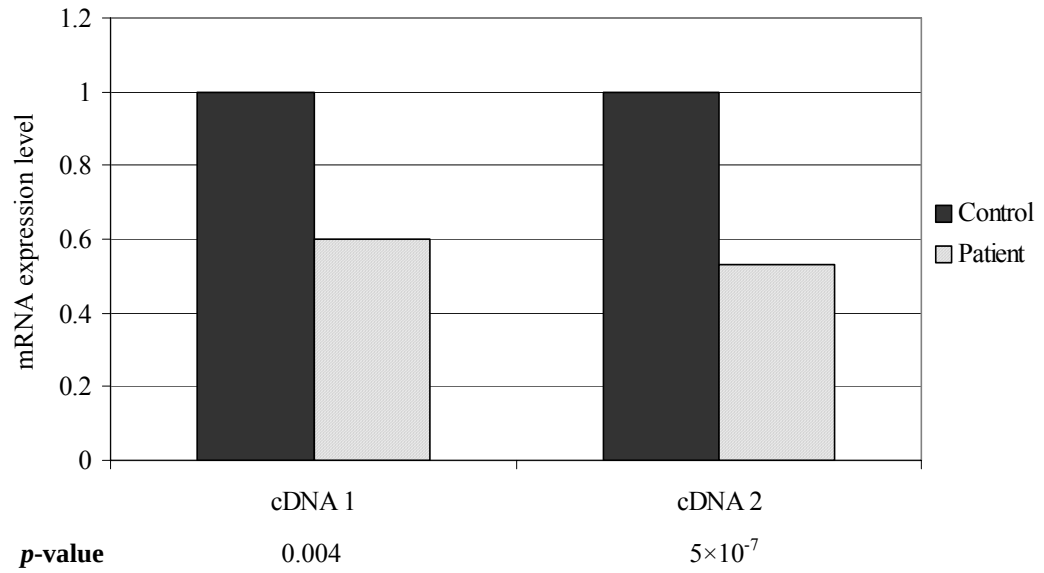


Figure 15: Control and CFS patient mean mRNA expression values for ERβ from two independent cDNA synthesis using 18S rRNA for normalization and comparative analysis.

ERα and ERβcx mRNA expression levels were stable within the study cohort. No mRNA expression level differences were found between the patient group and the control group using either 18S rRNA ($p_{ER\alpha}$ = 0.58 and $p_{ER\beta cx}$ = 0.87) or GAPDH ($p_{ER\alpha}$ = 0.80 and $p_{ER\beta cx}$ = 0.53) for normalization. There were no differences in ERα mRNA levels between any of the patient subgroups and healthy controls. The expression level of ERβcx mRNA was lower in patients with shorter illness duration compared to the patient subgroup with longer duration, irrespective of normalization gene ($p_{18S\ rRNA}$ = 0.003 and p_{GAPDH} = 0.02). However, the groups and the differences

between the groups are small, so the significance of this finding is at present unclear. There were no differences between the other patient subgroups compared to controls.

ERα was represented on the microarray used in Papers III-IV. There was no differential expression for ERα in the microarray study, which is in agreement with the results in Paper V. Neither ERβ nor ERβcx was present on the microarray.

A recent study by Piehl *et al*. has shown that ERα and ERβ mRNA are present in fractionated T-lymphocytes and B-lymphocytes [106]. However, the assay used in the study does not discriminate between ERβ and ERβcx [106]. In our study, the ERβcx mRNA and ERα expression levels were about 100 times higher than the ERβ mRNA expression levels. Concentrations of ERα and ERβcx were in the femtogram range. The PBMC samples contain both T-lymphocytes and B-lymphocytes and it would be interesting to see if the ratios between the ERβ splice variants differ between the fractions previously studied.

ERβ wild type (wt) and ERβcx are different splice variants that use different final exons and hence, have different C-terminal amino acids and 3' UTRs. It is unclear what regulates ERβwt/ERβcx ratios. One possibility is differential promoter usage, so that transcripts originating from different promoters display differential splicing. There are at least two promoters, 0N and 0K for the ERβ gene. To assay if differences in ERβwt but not in ERβcx mRNA levels, observed in CFS patients compared to controls, are due to a specific effect on one of the ERβ promoters, we assayed expression from the 0K and 0N promoters.

Hirata *et al.* have shown by genomic analysis that 0N is coupled to ERβwt and 0K to ERβcx in testis [85]. However, we found only expression from the 0N promoter in PBMCs, despite the fact that ERβcx appears to be the predominant transcript. One possible explanation for this is that the promoter 0N regulates both ERβwt and ERβcx in PBMCs. Tissue-specific promoter usage has been reported [85, 87]. Expression from the 0N promoter did not differ between CFS patients and controls. The latter observation is not surprising if we take the absence of expression from the 0K promoter as an indication that both ERβwt and ERβcx are expressed from the 0N promoter. In this scenario, the much higher expression of ERβcx will obscure differences in ERβwt expression.

Consistent with our results, promoter 0N is used in peripheral leukocytes [85]. Our results do not support promoter usage as a means by which ERβ/ERβcx ratios are regulated. However, the presence of additional ERβ promoters must be considered in this context.

Future perspectives on molecular biotechnology in diagnosis of CFS

Our results support the hypothesis of a heterogeneous CFS cohort. The differences in mRNA expression that are described in Papers III-IV were only identified by comparing female patient subgroups with female healthy controls. It is not clear whether similar differences exist among male patients. A larger male population is needed in order to study this. The ERβ mRNA expression levels were lower for both sexes separately and in all but one CFS patient subgroup compared with healthy controls. It is not clear whether CFS is one single illness or a group of illnesses accompanied by similar symptoms. Differences between CFS patient subgroups can be obscured when looking at the entire patient group, emphasising the need for subgrouping of patients [11]. Difficulties in replication of CFS studies may indicate a heterogeneous patient population with a different mix of subgroups in different studies [11].

The case definition of CFS requires at least four out of eight specified symptoms to be present for diagnosis [5]. One study has shown arbitrariness in this requirement [107]. There was no difference in gene expression comparing patients with varying number of symptoms in any of our studies either. The relevance of the quantity of symptoms needs to be discussed further. Some kind of biological test would simplify diagnosis and give the diagnosis greater creditability.

The differences in mRNA expression between CFS patients, or patient subgroups, and healthy controls that we have observed may contribute to some of the symptoms of CFS. Further studies to investigate protein levels and cellular effects will be required to determine whether any of the genes whose expression is changed are involved in CFS pathology. It is also possible that mRNA level differences are simply markers for changed functions of other cellular components, which are involved in CFS. Altered levels of transcript expression levels could in this case contribute to a diagnostic criterion, may function as a surrogate marker, and/or provide an entry point for the identification of interesting, potentially disease-causing molecules for further study.

Much work remains to be done before it is possible to diagnose CFS with a biological marker. Both microarray technology and real-time PCR are tools well-suited for molecular analysis, and both are potential methods for use in CFS diagnosis. Real-time PCR is already used for diagnosis and assays using microarray technology for diagnostic purposes are under development.

GENERAL SUMMARY AND CONCLUSIONS

The genomic era has seen the introduction of new powerful analytical methods, such as the microarray technology, in functional genomics. It is highly important that the technologies used for research and diagnostic purposes are properly optimised and standardised. Two other issues that have received attention recently are microarray reproducibility and the need for standardized ways of presenting microarray data.

The work presented here has focused on the effect of several technical parameters such as RNA extraction method and the amount of cDNA used for hybridisation. The aim should be to minimize variability by using the same conditions for all experiments. Different extraction methods can yield comparable results, although only one method should be used throughout an entire study. The dynamic range of the signal intensity is affected by the amount of hybridised cDNA, and for this reason similar cDNA concentrations should be used for all samples.

The choice of the most suitable sample to study a certain biological problem is not always straightforward. The location of the lesion may be inaccessible or unknown in certain diseases. It has been suggested that peripheral blood cells serve as indicators for different biological processes going on throughout the body. We have shown that genes involved in psycho-neuroendocrine-immune (PNI) communication can be studied by studying peripheral blood mononuclear cells (PBMCs). The PBMC sample is one option for the study of diseases with unknown pathophysiology and etiology, such as CFS.

The individual variability in PBMC mRNA expression is small, which increases the possibility of identifying genes related to the disease or illness under study. In our transcriptome expression study, we observed only a few differentially expressed genes in a subgroup of the patients. Furthermore, ERβ levels are lower in CFS patients than they are in healthy controls. Differences in the microarray study could only be detected by creating subgroups of the CFS patients, which indicates the need for subgrouping of patients with fatiguing illness in the search for pathogenic mechanisms. Identification of small mRNA expression differences requires a larger number of samples than the number required for finding large differences.

The difference in mRNA expression observed in the studies described here may contribute to some of the symptoms observed in CFS. Further studies will be required into the protein levels and cellular effects in order to determine whether any of the genes whose expression is changed are involved in CFS pathology. The differences in mRNA expression levels may simply be a marker for changed functions of other cellular components, which are involved in CFS. In this case, the altered levels could contribute to a diagnostic criterion, may function as a surrogate marker, and/or provide an entry point for identifying interesting potentially disease-causing molecules for further study.

POPULAR SCIENCE SUMMARY IN SWEDISH

I dagens samhälle är det vanligt att människor känner sig trötta. Studier har visat att 5-20% av befolkningen i i-länderna lider av en trötthet som leder till problem i det vardagliga livet. Det finns flera synonymer som beskriver olika sorters trötthet, t.ex. sömnighet, dåsighet, matthet, kraftlöshet och utmattning.

Patienter upplever ofta trötthet som ett påtagligt symptom då det inverkar på hela livskvaliteten. Det är dock vanligt att den medicinska expertisen bortser från detta symptom. Några sjukdomar som sammankopplas med trötthet är astma, lågt blodtryck, cancer, diabetes, multipel skleros samt kroniskt trötthetssyndrom (KTS).

KTS är ett tillstånd där både orsaken till sjukdom och sjukdomsbilden idag är så gott som okända. För att få diagnosen KTS får det varken finnas någon annan medicinsk eller psykologisk förklaring till trötttheten. Trötttheten måste ha varat i minst sex månader och patienten måste uppvisa minst fyra av totalt åtta specificerade symptom. De åtta symptomen är: försämrad kognition och försämrat minne, återkommande halsont, ömma lymfkörtlar, mild muskelvärk, ledvärk, ny typ av huvudvärk, sömn utan återhämtning samt obehag efter kraftansträngning.

Enligt befolkningsstudier utförda i USA är den allmänna förekomsten av KTS 0,2-0,4% med en övervägande del kvinnor (70-80%). Teorierna om vad som orsakar KTS är många. Några förslag som lagts fram är störningar i immunförsvaret, sömnstörningar och infektioner. Idag saknas det bra behandling för KTS-patienter. Kognitiv beteendeterapi och gradvis ökad fysisk aktivitet är de två metoder som hittills har visat sig fungera bäst men de hjälper långt ifrån alla patienter.

Då det inte finns någon känd sjuklig vävnad för KTS är det svårt att veta vilket provmaterial som bör användas. Vid studier av sjukdomar som cancer ter det sig självklart att använda sig av tumörvävnad. En del forskare tror att hjärnan är inblandad i KTS. Naturligtvis är det svårt att få ett hjärnprov från en levande människa. Det närmaste man kan komma är ett prov från benmärgsvätska, vilket är en smärtsam upplevelse. Önskvärt vore att med en enklare provtagning få information om sjukdomsbilden. En teori är att blodceller kan fungera som ”patrullerande vakter” och ge information om anormala biologiska processer som pågår på olika platser i kroppen.

Målet med detta projekt har varit att identifiera gener i blodceller som spelar en viktig roll för tillståndet KTS. Dessa gener skulle kunna ge ökad information om orsaken till KTS och vara kandidater för att användning vid framtida diagnostisering. För genaktivitetsstudierna har huvudsakligen två olika tekniker använts, mikromatristeknik respektive realtids-PCR.

Generna utgör arvsmassan och styr så att en människa utvecklas till en människa och inte till något annat. Molekylen som bygger upp det genetiska materialet kallas

för deoxyribonukleinsyra (DNA) och återfinns i alla levande celler. Då cellen behöver använda en gen startar en komplicerad process som sker med hjälp av en hel arsenal av olika molekyler. I ett första steg skapas det en kopia av den aktuella genen som kallas för budbärar-ribonukleinsyra (mRNA). Genkopian används i sin tur som mall för att bygga ihop ett protein. Proteinet är genens slutprodukt och en grundläggande byggsten för allt liv.

Genaktivitetsstudier går ut på att mäta koncentrationen av de genkopior, mRNA-molekyler, som finns i en cell. Mängden mRNA-molekyler som uppmäts i ett prov avspeglar vilka gener som används av en cell. På detta sätt kan man jämföra vilka gener som används av olika celler och hur mycket de används. Det viktigaste för att ett experiment av detta slag skall lyckas är att det ursprungliga provmaterialet, mRNA-molekylerna, är av god kvalitet samt att metoden är pålitlig och väl utvecklad.

Den huvudsakliga skillnaden mellan mikromatristeknik och realtids-PCR är antalet gener som går att undersöka per experiment. Med mikromatristeknik är det möjligt att samtidigt mäta aktivitetsgraden för tiotusentals gener i ett prov, att jämföra med en gen per experiment för realtids-PCR. Fördelarna med realtids-PCR är att det är en noggrannare metod och att man kan bestämma exakta koncentrationsnivåer. Med mikromatristekniken går det bara att jämföra genaktiviteten i två prov, alltså endast en relativ jämförelse.

Realtids-PCR och mikromatristeknik är metoder som lämpar sig väl för användning vid diagnostik. Realtids-PCR används redan idag för att t.ex. diagnostisera olika infektionssjukdomar. Det krävs mer utveckling av mikromatristekniken innan den är tillräckligt tillförlitlig för att användas vid diagnostisering.

Det är viktigt att använda optimerade analysmetoder för att i slutändan få resultat av hög kvalitet. I den första studien har vi utvärderat två olika metoder för att få ut mRNA från celler med avseende på provkvalitet och slutligt mikromatrisresultat. Båda metoderna gav provmaterial av jämförbart god kvalitet med snarlika slutliga mikromatrisresultat. Dock visade det sig att den ena metoden fick med delar av cellens DNA tillsammans med provet, vilket inte är önskvärt. Oönskat DNA kan förstöras med ett visst protein så att man bara får kvar det önskade mRNA:t i slutändan.

Nästa steg i studien var att undersöka om blodceller är ett lämpligt prov för att studera biologiska processer som har med hjärna, immunförsvar, hormoner samt samspel mellan dessa områden att göra. Tidigare studier visar att genaktiviteten i blodceller hos olika individer är mycket lika. De största skillnaderna beror på kön och ålder. I och med att det är så pass liten variation i genaktiviteten mellan olika personer så är chansen att upptäcka skillnader som uppstått till följd av sjukdom relativt god. I den andra studien har vi visat att det är möjligt att studera gener som är inblandade i de ovan nämnda biologiska processerna. I ett första steg är därför blodceller ett

lämpligt prov att använda om det inte finns någon känd sjuklig vävnad eller då det är svårt att få tag på den sjuka vävnaden.

Vid jämförelse av genaktiviteter hos svenska patienter som har fått diagnosen KTS och friska ålders- och könsmatchade kontroller har fyra gener med statistiskt signifikanta skillnader identifierats. En gen, östrogenreceptor β, uppvisar skillnad mellan hela patient- och hela kontrollgruppen (både män och kvinnor) medan de tre andra generna har identifierats vid jämförelse mellan en undergrupp av kvinnliga patienter och kvinnliga kontroller. Dessa kvinnliga patienter har haft ett gradvist insjuknande och de har inte haft någon tidigare dokumenterad infektion. Andra skillnader har hittats i liknande studier men överensstämmelsen mellan alla genaktivitetsstudier är dock begränsad. Störningar i immunförsvaret återkommer som ett intressant funktionsområde.

De tre generna som hittats med hjälp av mikromatristeknik heter CD83, NRK1 respektive BOLA1. CD83 finns i lägre grad hos patienterna jämfört med kontrollerna och spelar en viktig roll i immunförsvaret. Genaktiviteten hos de två andra generna är högre hos patienterna jämfört med kontrollerna. NRK1 är involverad i produktionen av ett protein som förkortas NAD+. NAD+ har i sin tur betydelse för en mängd olika reaktioner i ämnesomsättningen. Funktionen för BOLA1 är fortfarande okänd.

I studien har även indikationer på att det finns skillnader mellan andra undergrupper av kvinnliga patienter och kvinnliga friska kontroller framkommit. Dessa har dock inte visat sig vara statistiskt signifikanta, vilket troligtvis beror på ett för litet antal patienter. Det har inte heller varit möjligt att dra några slutsatser om manliga patienterna då det fanns för få manliga studiedeltagare.

KTS patienter uppvisar en lägre nivå av östrogenreceptor β jämfört med friska kontroller. Östrogen är ett hormon som finns hos både män och kvinnor. Hormonet är viktigt för flertalet biologiska processer så som exempelvis sexuell utveckling och förökning. För att östrogenet ska påverka en reaktion måste det först binda till specifika receptorer som kallas östrogenreceptorer. Det finns flera olika varianter av dessa receptorer varav β är en. Östrogenreceptorerna har visat sig spela en viktig roll för sjukdomar med ojämn könsfördelning så som till exempel bröstcancer och benskörhet.

Det går bara att spekulera i vad dessa gener kan ha för betydelse för tillståndet KTS. Skillnaderna som har upptäckts i våra studier skulle kunna bidra till de symptom som man ser hos KTS-patienterna. Genaktivitetsskillnaderna kan möjligtvis också vara varningssignaler för att en förändring har skett på en annan plats i kroppen. Om så är fallet skulle de kunna användas som indikatorer för att allt inte står helt rätt till och vara kandidater för användning vid framtida diagnostik. Dit är vägen dock fortfarande lång och än så länge är forskningen på KTS i ett tidigt stadium. Det kommer att krävas vidare forskning för att komma till insikt om dessa fyra gener har någon biologisk betydelse för sjukdomstillståndet.

ABBREVIATIONS

A	"addition", $\log_2$ mean signal intensity
AMV-RT	avian myeloblastosis viral RT enzyme
ACTH	adrenocorticotropic hormone
aRNA	antisense RNA
B	Bayesian
CBT	cognitive behaviour therapy
CDC	Center for Disease Control and Prevention
CDC MADB	CDC Microarray database
cDNA	complementary DNA
CFS	chronic fatigue syndrome
Cp	crossing point
C_t	threshold cycle
Cy3/Cy5	cyanine3/cyanine5
DC	dendritic cell
DD-PCR	differential display PCR
DNA	deoxyribonucleic acid
E	amplification efficiency
ER	estrogen receptor
EST	expressed sequence tag
FWER	family-wise error I rate
FDR	false discovery rate
GAPDH	glyseraldehyde-3-phosphate dehydrogenase
GET	graded exercise therapy
GO	gene ontology
HLA	human leukocyte antigen complex
HPA	hypothalamic-pituitary-adrenal
KEGG	Kyoto encyclopedia of genes and genomes
KTS	kroniskt trötthetssyndrom
LOWESS/LOESS	locally weighted scatter plot smooth
M	"minus", $\log_2$ ratio of signal intensities
MGED	microarray gene expression data
MHC	major histocompability complex
MIAME	minimum information about a microarray experiment
MMLV-RT	Moloney murine leukaemia virus RT enzyme
MPSS	massively parallell signature sequencing
mRNA	messenger RNA
NAD	nicotine amide dinucleotide
PBMC	peripheral blood mononuclear cell
PCR	polymerase chain reaction
PNI	psycho-neuroendocrine-immune
RLS	resonance light scattering
RNA	ribonucleic acid
rRNA	ribosomal RNA
RT	reverse transcription
SAGE	serial analysis of gene expression
SAM	significance analysis of microarray
wt	wild type

ACKNOWLEDGEMENTS

As I start to write this section of the thesis I get hit by the feeling that now I am really getting towards the end of this project. I will enter a new phase of my life, which feels a bit scary but also exciting. During the time of this doctoral thesis I have experienced a whole lot, both personally and work wise. After only a few months of preparation work in Sweden I was heading towards the States and Atlanta. The only thing I knew about Atlanta was that the city organized the Olympic games in 1996. About one year later I returned with both wonderful and pretty tough experiences behind me. I continued working hard at home with some adversities but mainly success. The work is now almost completed and I want to thank everyone who has helped me during this time and especially:

My main-supervisor **Birgitta Evengård** for giving me the opportunity to join this project and believing in my work. You have taught me a lot about how science works and prepared me for my future working life.

My co-supervisor **Peter Nilsson** for giving me the opportunity to join this project and believing in my work. For supporting me in the microarray project with great enthusiasm and keeping me in good mood despite the "loss" of many months of work due to the RLS ghost.

Carl Erik Nord, chairman for the Division of Clinical Bacteriology, for supporting me in my research and giving wise advice about how to handle tricky situations.

Suzanne Vernon, my boss at CDC, for giving me the chance to a wonderful experience, although some parts were tough. Thanks for guidance in the research world and supporting and believing in my work. **William Reeves** for the opportunity to visit CDC and for repelling and paddling events. **Elizabeth Unger**, **Daisy Lee**, **Irina Dimulescu**, **Toni Whistler**, **Kelly Bowen**, **Raajevan Mangalathu** for supporting me and helping me in my work. I also would like to thank everyone else in the CDC group and especially **David Swan** for the fantastic helicopter ride. One of my dreams came true.

All the volunteers who have donated samples and the people how have helped me with withdrawing blood samples. Without you this thesis would not have been possible.

My co-authors and collaborators **Maria Nilsson**, **Karin Dahlman-Wright** and **Jan-Åke Gustafsson** for good work and fruitful discussions. Thanks Maria for introducing me to real-time PCR.

The other professors at the Division of Clinical Bacteriology: **Andrej Weintraub**, **Bengt Wretlind**, **Gunnar Sandström**, **Jan-Ingmar Flock** and **Charlotta Edlund** for showing interest in my PhD student seminars.

Benjamin Edvinsson for being a very nice fellow PhD student and colleague. Thanks for our scientific and everyday life discussions, for always supporting and helping me with smaller and bigger things and teaching me the LightCycler instrument.

My fellow PhD students at the Division of Clinical Bacteriology, both present and former: **Hanna Billström**, **Andreas Jacks**, **Hadi Abd**, **Cristina Oprica**, **Sohel Islam**, **Amir Saeed**, **Sonja Löfmark**, **Tara Ali**, **Axana Haggar**, **Margareta Flock**, **Anita Forsberg**, **Anna Rennermalm**, **Samuel Vilchez**, **Erick Amaya**, **Oonagh Shannon**, **Nguyen Vu Trung**, **Susanna Falklind Jerkérus**, **Bodil Lund**, **Herin Oh** and **Nagwa El Amin** for practical help, funny events and discussions.

Elisabeth Wahlund for good collaboration in sample collection for the CFS biobank. We had fun but very intense days collecting samples in Lycksele. It was a really interesting experience.

Gudrun Rafi, Kerstin Bergman, **Lena Eriksson**, **Monica Sörensson, Eva Sillerström**, **Ingegerd Löfving Arvholm**, **Karin Olsson**, **Åsa Sullivan**, **Märit Karls**, **Ann-Chatrin Palmgren**, **Ann-Katrine Björnstjerna**, **Annica Nordvall Bodell**, for all the practical help in- and outside the lab and for all lunch discussions.

Oonagh Shannon and **Paula Cunnea** for helping me with the CD83 Western blot experiments.

Emilie Hultin for being a good friend, for the relaxing lunch chats at PhiPhi Island filled with science and life discussions and help with practical things in the lab at KTH.

All the people at the Department of Molecular Biotechnology at KTH who have helped me in any way. Especially thanks to **Malin Eklund**, an excellent exam supervisor, for introducing me to science and **Per-Åke Nygren** for introducing me to my supervisors. **Valtteri Wirta**, **Marcus Gry Björklund** and **Johan Lindberg** for trying to help with the RLS ghost. **Cecilia Laurell** and **Annelie Waldén** for help with microarray stuff.

Linda Eriksson for being such a nice friend. You are supporting and always being there for me and I really enjoy your company. I love our long, long talks about everything and nothing (we know David and Jim are not always pleased) and your sons are so cute.

Cindy Bierl for being there for me during the weird but fun time in Atlanta. Without your lovely friendship I would never have made it. Thanks for introducing me to Cold Stone Creamery and French vanilla ice cream mixed with raspberries, our chats, hiking trips, dinners, movies, shopping trips and laughter.

Eva Jönsson and **Mia Emanuelsson** for being such good friends and always there for me whenever I needed help. **Hanna Sjölund** and **Jessica Sköld** for your friendship. I really enjoyed all the fun we had, chats, dinners, movies, Viking Club events and shopping trips. It is true that Swedish women need to have control in order to be comfortable.

The rest of my friends and relatives for being a part of my life in different ways. No one mentioned no one forgotten.

I would like to thank my "new" family, the **Gräns** family, who have made my family almost three times as large. **Ingrid** and **Karl-Erik** for always supporting me and having an interest in how it is going. **Erik**, **Pernilla**, **Tomas**, **Netta**, **Magnus** and **Therese** for being a nice family. **Lovisa**, **Evelina** and **Frida** for making me relax and forget the work for a moment.

The **Ojaniemi** family: My mother **Vuokko** and father **Unto** for all the support and help during my entire life. I would never have got this far if it was not for you. My little brother **Kalle** for being who you are, although, I wish you sometimes could do things the way I want you to do. I love the three of you a lot.

My dear **David,** for being my companion for life and always supporting me, both through wonderful and hard times. We have just managed the first ten-year period together. The one-year time apart on separate sides of the Atlantic Ocean has been the toughest part so far, but we made it together and came out of it even stronger. I want to have many more years together with you filled with fun experiences and building our own family. I love you more than anything else and I will never stop loving you.

REFERENCES

1. Lewis G and Wessely S, *The epidemiology of fatigue: more questions than answers*, J Epidemiol Community Health, 46(2):92-97, 1992
2. Sharpe M and Wilks D, *Fatigue*, Bmj, 325(7362):480-483, 2002
3. Böttiger L, *Trötthet - en oförmåga att fortsätta*, Läkartidningen, 64(10):984-992, 1967
4. Evengard B, Schacterle RS and Komaroff AL, *Chronic fatigue syndrome: new insights and old ignorance*, J Intern Med, 246(5):455-469, 1999
5. Fukuda K, Straus SE, Hickie I, Sharpe MC, Dobbins JG and Komaroff A, *The chronic fatigue syndrome: a comprehensive approach to its definition and study. International Chronic Fatigue Syndrome Study Group*, Ann Intern Med, 121(12):953-959, 1994
6. Sharpe MC, Archard LC, Banatvala JE, Borysiewicz LK, Clare AW, David A, Edwards RHT *et al.*, *A report - chronic fatigue syndrome: guidlines for research*, J R Soc Med, 84:118-121, 1991
7. Lloyd AR, Wakefield D, Boughton C and Dwyer J, *What is myalgic encephalomyelitis?*, Lancet, 1(8597):1286-1287, 1988
8. Reynolds KJ, Vernon SD, Bouchery E and Reeves WC, *The economic impact of chronic fatigue syndrome*, Cost Eff Resour Alloc, 2(1):4, 2004
9. Reeves WC, Lloyd A, Vernon SD, Klimas N, Jason LA, Bleijenberg G, Evengard B *et al.*, *Identification of ambiguities in the 1994 chronic fatigue syndrome research case definition and recommendations for resolution*, BMC Health Serv Res, 3(1):25, 2003
10. Ranjith G, *Epidemiology of chronic fatigue syndrome*, Occup Med (Lond), 55(1):13-19, 2005
11. Jason LA, Corradi K, Torres-Harding S, Taylor RR and King C, *Chronic fatigue syndrome: the need for subtypes*, Neuropsychol Rev, 15(1):29-58, 2005
12. Jason LA, Richman JA, Rademaker AW, Jordan KM, Plioplys AV, Taylor RR, McCready W *et al.*, *A community-based study of chronic fatigue syndrome*, Arch Intern Med, 159(18):2129-2137, 1999
13. Evengard B, Jacks A, Pedersen NL and Sullivan PF, *The epidemiology of chronic fatigue in the Swedish Twin Registry*, Psychol Med, 35:1317-1326, 2005
14. Lindal E, Stefansson JG and Bergmann S, *The prevalence of chronic fatigue syndrome in Iceland - a national comparison by gender drawing on four different criteria*, Nord J Psychiatry, 56(4):273-277, 2002
15. Reyes M, Nisenbaum R, Hoaglin DC, Unger ER, Emmons C, Randall B, Stewart JA *et al.*, *Prevalence and incidence of chronic fatigue syndrome in Wichita, Kansas*, Arch Intern Med, 163(13):1530-1536, 2003
16. Wessely S, Chalder T, Hirsch S, Wallace P and Wright D, *The prevalence and morbidity of chronic fatigue and chronic fatigue syndrome: a prospective primary care study*, Am J Public Health, 87(9):1449-1455, 1997
17. Evengard B, Jonzon E, Sandberg A, Theorell T and Lindh G, *Differences between patients with chronic fatigue syndrome and with chronic fatigue at an infectious disease clinic in Stockholm, Sweden*, Psychiatry Clin Neurosci, 57(4):361-368, 2003
18. Evengard B and Klimas N, *Chronic fatigue syndrome: probable pathogenesis and possible treatments*, Drugs, 62(17):2433-2446, 2002
19. Cleare AJ, *The HPA axis and the genesis of chronic fatigue syndrome*, Trends Endocrinol Metab, 15(2):55-59, 2004
20. Cairns R and Hotopf M, *A systematic review describing the prognosis of chronic fatigue syndrome*, Occup Med (Lond), 55(1):20-31, 2005
21. Steinau M, Unger ER, Vernon SD, Jones JF and Rajeevan MS, *Differential-display PCR of peripheral blood for biomarker discovery in chronic fatigue syndrome*, J Mol Med, 82(11):750-755, 2004
22. Vernon SD, Unger ER, Dimulescu IM, Rajeevan M and Reeves WC, *Utility of the blood for gene expression profiling and biomarker discovery in chronic fatigue syndrome*, Dis Markers, 18(4):193-199, 2002

23. Whistler T, Unger ER, Nisenbaum R and Vernon SD, *Integration of gene expression, clinical, and epidemiologic data to characterize Chronic Fatigue Syndrome*, J Transl Med, 1(1):10, 2003
24. Whistler T, Jones JF, Unger ER and Vernon SD, *Exercise responsive genes measured in peripheral blood of women with chronic fatigue syndrome and matched control subjects*, BMC Physiol, 5(1):5, 2005
25. Powell R, Ren J, Lewith G, Barclay W, Holgate S and Almond J, *Identification of novel expressed sequences, up-regulated in the leucocytes of chronic fatigue syndrome patients*, Clin Exp Allergy, 33(10):1450-1456, 2003
26. Kaushik N, Fear D, Richards SC, McDermott CR, Nuwaysir EF, Kellam P, Harrison TJ *et al.*, *Gene expression in peripheral blood mononuclear cells from patients with chronic fatigue syndrome*, J Clin Pathol, 58(8):826-832, 2005
27. Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC, Relman DA and Brown PO, *Individuality and variation in gene expression patterns in human blood*, Proc Natl Acad Sci U S A, 100(4):1896-1901, 2003
28. Natelson BH, Weaver SA, Tseng CL and Ottenweller JE, *Spinal fluid abnormalities in patients with chronic fatigue syndrome*, Clin Diagn Lab Immunol, 12(1):52-55, 2005
29. Espina V, Woodhouse EC, Wulfkuhle J, Asmussen HD, Petricoin EF, 3rd and Liotta LA, *Protein microarray detection strategies: focus on direct detection technologies*, J Immunol Methods, 290(1-2):121-133, 2004
30. Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M *et al.*, *Expression monitoring by hybridization to high-density oligonucleotide arrays*, Nat Biotechnol, 14(13):1675-1680, 1996
31. Schena M, Shalon D, Davis RW and Brown PO, *Quantitative monitoring of gene expression patterns with a complementary DNA microarray*, Science, 270(5235):467-470, 1995
32. Velculescu VE, Zhang L, Vogelstein B and Kinzler KW, *Serial analysis of gene expression*, Science, 270(5235):484-487, 1995
33. Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S *et al.*, *Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays*, Nat Biotechnol, 18(6):630-634, 2000
34. Alwine JC, Kemp DJ and Stark GR, *Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes*, Proc Natl Acad Sci U S A, 74(12):5350-5354, 1977
35. Becker-Andre M and Hahlbrock K, *Absolute mRNA quantification using the polymerase chain reaction (PCR). A novel approach by a PCR aided transcript titration assay (PATTY)*, Nucleic Acids Res, 17(22):9437-9446, 1989
36. Wang AM, Doyle MV and Mark DF, *Quantitation of mRNA by the polymerase chain reaction*, Proc Natl Acad Sci U S A, 86(24):9717-9721, 1989
37. Heid CA, Stevens J, Livak KJ and Williams PM, *Real time quantitative PCR*, Genome Res, 6(10):986-994, 1996
38. Barrett JC and Kawasaki ES, *Microarrays: the use of oligonucleotides and cDNA for the analysis of gene expression*, Drug Discov Today, 8(3):134-141, 2003
39. Hegde P, Qi R, Abernathy K, Gay C, Dharap S, Gaspard R, Hughes JE *et al.*, *A concise guide to cDNA microarray analysis*, Biotechniques, 29(3):548-556 , 2000
40. Freeman WM, Robertson DJ and Vrana KE, *Fundamentals of DNA hybridization arrays for gene expression analysis*, Biotechniques, 29(5):1042-1055, 2000
41. Puskas LG, Zvara A, Hackler L, Jr. and Van Hummelen P, *RNA amplification results in reproducible microarray data with slight ratio bias*, Biotechniques, 32(6):1330-1340, 2002
42. Zhao H, Hastie T, Whitfield ML, Borresen-Dale AL and Jeffrey SS, *Optimization and evaluation of T7 based RNA linear amplification protocols for cDNA microarray analysis*, BMC Genomics, 3(1):31, 2002

43. Bao P, Frutos AG, Greef C, Lahiri J, Muller U, Peterson TC, Warden L *et al.*, *High-sensitivity detection of DNA hybridization on microarrays using resonance light scattering*, Anal Chem, 74(8):1792-1797, 2002
44. Francois P, Bento M, Vaudaux P and Schrenzel J, *Comparison of fluorescence and resonance light scattering for highly sensitive microarray detection of bacterial pathogens*, J Microbiol Methods, 55(3):755-762, 2003
45. Simon R, Radmacher MD and Dobbin K, *Design of studies using DNA microarrays*, Genet Epidemiol, 23(1):21-36, 2002
46. Churchill GA, *Fundamentals of experimental design for cDNA microarrays*, Nat Genet, 32 Suppl:490-495, 2002
47. Leung YF and Cavalieri D, *Fundamentals of cDNA microarray data analysis*, Trends Genet, 19(11):649-659, 2003
48. Dudoit S, Yang YH, Callow MJ and Speed TP, *Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments*, Statistica Sinica, 12:111-139, 2002
49. Quackenbush J, *Microarray data normalization and transformation*, Nat Genet, 32 Suppl:496-501, 2002
50. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J and Speed TP, *Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation*, Nucleic Acids Res, 30(4):e15, 2002
51. Smyth GK and Speed T, *Normalization of cDNA microarray data*, Methods, 31(4):265-273, 2003
52. Quackenbush J, *Computational analysis of microarray data*, Nat Rev Genet, 2(6):418-427, 2001
53. Smyth GK, Yang YH and Speed T, *Statistical issues in cDNA microarray data analysis*, Methods Mol Biol, 224:111-136, 2003
54. Schena M, Shalon D, Heller R, Chai A, Brown PO and Davis RW, *Parallel human genome analysis: microarray-based expression monitoring of 1000 genes*, Proc Natl Acad Sci U S A, 93(20):10614-10619, 1996
55. DeRisi JL, Iyer VR and Brown PO, *Exploring the metabolic and genetic control of gene expression on a genomic scale*, Science, 278(5338):680-686, 1997
56. Cui X and Churchill GA, *Statistical tests for differential expression in cDNA microarray experiments*, Genome Biol, 4(4):210, 2003
57. Tusher VG, Tibshirani R and Chu G, *Significance analysis of microarrays applied to the ionizing radiation response*, Proc Natl Acad Sci U S A, 98(9):5116-5121, 2001
58. Baldi P and Long AD, *A Bayesian framework for the analysis of microarray expression data: regularized t -test and statistical inferences of gene changes*, Bioinformatics, 17(6):509-519, 2001
59. Lönnstedt I and Speed T, *Replicated microarray data*, Statistica Sinica, 12:31-46, 2002
60. Reiner A, Yekutieli D and Benjamini Y, *Identifying differentially expressed genes using false discovery rate controlling procedures*, Bioinformatics, 19(3):368-375, 2003
61. Rajeevan MS, Vernon SD, Taysavang N and Unger ER, *Validation of array-based gene expression profiles by real-time (kinetic) RT-PCR*, J Mol Diagn, 3(1):26-31, 2001
62. Rajeevan MS, Ranamukhaarachchi DG, Vernon SD and Unger ER, *Use of real-time quantitative PCR to validate the results of cDNA array and differential display PCR technologies*, Methods, 25(4):443-451, 2001
63. Chuaqui RF, Bonner RF, Best CJ, Gillespie JW, Flaig MJ, Hewitt SM, Phillips JL *et al.*, *Post-analysis follow-up and validation of microarray experiments*, Nat Genet, 32 Suppl:509-514, 2002
64. Nadon R and Shoemaker J, *Statistical issues with microarrays: processing and analysis*, Trends Genet, 18(5):265-271, 2002
65. Eisen MB, Spellman PT, Brown PO and Botstein D, *Cluster analysis and display of genome-wide expression patterns*, Proc Natl Acad Sci U S A, 95(25):14863-14868, 1998
66. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP *et al.*, *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium*, Nat Genet, 25(1):25-29, 2000

67. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H and Kanehisa M, *KEGG: Kyoto Encyclopedia of Genes and Genomes*, Nucleic Acids Res, 27(1):29-34, 1999
68. Brazma A, Parkinson H, Sarkans U, Shojatalab M, Vilo J, Abeygunawardena N, Holloway E *et al.*, *ArrayExpress--a public repository for microarray gene expression data at the EBI*, Nucleic Acids Res, 31(1):68-71, 2003
69. Edgar R, Domrachev M and Lash AE, *Gene Expression Omnibus: NCBI gene expression and hybridization array data repository*, Nucleic Acids Res, 30(1):207-210, 2002
70. Ikeo K, Ishi-i J, Tamura T, Gojobori T and Tateno Y, *CIBEX: center for information biology gene expression database*, C R Biol, 326(10-11):1079-1082, 2003
71. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J *et al.*, *Minimum information about a microarray experiment (MIAME)-toward standards for microarray data*, Nat Genet, 29(4):365-371, 2001
72. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB *et al.*, *Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase*, Science, 239(4839):487-491, 1988
73. Ding C and Cantor CR, *Quantitative analysis of nucleic acids--the last few years of progress*, J Biochem Mol Biol, 37(1):1-10, 2004
74. Wong ML and Medrano JF, *Real-time PCR for mRNA quantitation*, Biotechniques, 39(1):75-85, 2005
75. Bustin SA, *Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays*, J Mol Endocrinol, 25(2):169-193, 2000
76. Gibson UE, Heid CA and Williams PM, *A novel method for real time quantitative RT-PCR*, Genome Res, 6(10):995-1001, 1996
77. Morrison TB, Weis JJ and Wittwer CT, *Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification*, Biotechniques, 24(6):954-962, 1998
78. Ririe KM, Rasmussen RP and Wittwer CT, *Product differentiation by analysis of DNA melting curves during the polymerase chain reaction*, Anal Biochem, 245(2):154-160, 1997
79. Liang P, *A decade of differential display*, Biotechniques, 33(2):338-346, 2002
80. Pettersson K and Gustafsson JA, *Role of estrogen receptor beta in estrogen action*, Annu Rev Physiol, 63:165-192, 2001
81. Matthews J and Gustafsson JA, *Estrogen signaling: a subtle balance between ER alpha and ER beta*, Mol Interv, 3(5):281-292, 2003
82. Ogawa S, Inoue S, Watanabe T, Orimo A, Hosoi T, Ouchi Y and Muramatsu M, *Molecular cloning and characterization of human estrogen receptor betacx: a potential inhibitor ofestrogen action in human*, Nucleic Acids Res, 26(15):3505-3512, 1998
83. Nilsson M, Dahlman-Wright K and Gustafsson JA, *Nuclear receptors in disease: the oestrogen receptors*, Essays Biochem, 40:157-167, 2004
84. Osterlund MK, Grandien K, Keller E and Hurd YL, *The human brain has distinct regional expression patterns of estrogen receptor alpha mRNA isoforms derived from alternative promoters*, J Neurochem, 75(4):1390-1397, 2000
85. Hirata S, Shoda T, Kato J and Hoshi K, *The multiple untranslated first exons system of the human estrogen receptor beta (ER beta) gene*, J Steroid Biochem Mol Biol, 78(1):33-40, 2001
86. Flouriot G, Griffin C, Kenealy M, Sonntag-Buck V and Gannon F, *Differentially expressed messenger RNA isoforms of the human estrogen receptor-alpha gene are generated by alternative splicing and promoter usage*, Mol Endocrinol, 12(12):1939-1954, 1998
87. Zhao C, Lam EW, Sunters A, Enmark E, De Bella MT, Coombes RC, Gustafsson JA *et al.*, *Expression of estrogen receptor beta isoforms in normal breast epithelial cells and breast cancer: regulation by methylation*, Oncogene, 22(48):7600-7606, 2003
88. Kos M, Reid G, Denger S and Gannon F, *Minireview: genomic organization of the human ERalpha gene promoter region*, Mol Endocrinol, 15(12):2057-2063, 2001
89. Panay N and Studd JW, *The psychotherapeutic effects of estrogens*, Gynecol Endocrinol, 12(5):353-365, 1998

90. Studd J and Panay N, *Chronic fatigue syndrome*, Lancet, 348(9038):1384, 1996

91. Harlow BL, Signorello LB, Hall JE, Dailey C and Komaroff AL, *Reproductive correlates of chronic fatigue syndrome*, Am J Med, 105(3A):94S-99S, 1998

92. R Development Core Team: *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2004

93. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J *et al.*, *TM4: a free, open-source system for microarray data management and analysis*, Biotechniques, 34(2):374-378, 2003

94. Nikitin A, Egorov S, Daraselia N and Mazo I, *Pathway studio--the analysis and navigation of molecular networks*, Bioinformatics, 19(16):2155-2157, 2003

95. Vernon SD, Unger ER, Rajeevan M, Dimulescu IM, Nisenbaum R and Campbell CE, *Reproducibility of alternative probe synthesis approaches for gene expression profiling with arrays*, J Mol Diagn, 2(3):124-127, 2000

96. Chomczynski P and Sacchi N, *Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction*, Anal Biochem, 162(1):156-159, 1987

97. Gadbois DM, Salo WL, Ann DK, Downing SW and Carlson DM, *The preparation of poly(A)+mRNA from the hagfish slime gland*, Prep Biochem, 18(1):67-76, 1988

98. Rus V, Chen H, Zernetkina V, Magder LS, Mathai S, Hochberg MC and Via CS, *Gene expression profiling in peripheral blood mononuclear cells from lupus patients with active and inactive disease*, Clin Immunol, 112(3):231-234, 2004

99. Mandel M, Gurevich M, Pauzner R, Kaminski N and Achiron A, *Autoimmunity gene expression portrait: specific signature that intersects or differentiates between multiple sclerosis and systemic lupus erythematosus*, Clin Exp Immunol, 138(1):164-170, 2004

100. Jison ML, Munson PJ, Barb JJ, Suffredini AF, Talwar S, Logun C, Raghavachari N *et al.*, *Blood mononuclear cell gene expression profiles characterize the oxidant, hemolytic, and inflammatory stress of sickle cell disease*, Blood, 104(1):270-280, 2004

101. Heller RA, Schena M, Chai A, Shalon D, Bedilion T, Gilmore J, Woolley DE *et al.*, *Discovery and analysis of inflammatory disease-related genes using cDNA microarrays*, Proc Natl Acad Sci U S A, 94(6):2150-2155, 1997

102. Kozlow EJ, Wilson GL, Fox CH and Kehrl JH, *Subtractive cDNA cloning of a novel member of the Ig gene superfamily expressed at high levels in activated B lymphocytes*, Blood, 81(2):454-461, 1993

103. Zhou LJ, Schwarting R, Smith HM and Tedder TF, *A novel cell-surface molecule expressed by human interdigitating reticulum cells, Langerhans cells, and activated lymphocytes is a new member of the Ig superfamily*, J Immunol, 149(2):735-742, 1992

104. Hock BD, Kato M, McKenzie JL and Hart DN, *A soluble form of CD83 is released from activated dendritic cells and B lymphocytes, and is detectable in normal human sera*, Int Immunol, 13(7):959-967, 2001

105. Lechmann M, Zinser E, Golka A and Steinkasserer A, *Role of CD83 in the immunomodulation of dendritic cells*, Int Arch Allergy Immunol, 129(2):113-118, 2002

106. Phiel KL, Henderson RA, Adelman SJ and Elloso MM, *Differential estrogen receptor gene expression in human peripheral blood mononuclear cell populations*, Immunol Lett, 97(1):107-113, 2005

107. Sullivan PF, Pedersen NL, Jacks A and Evengård B, *Chronic fatigue in a population sample: definitions and heterogeneity*, Psychol Med, 35:1337-1348, 2005