University of Cape Town
Faculty of Science
Department of Molecular and Cell Biology

RESEARCH REPORT
Submitted in fulfilment of the requirements for the degree of
Master of Science (Molecular Biology)
December 2008

Peripheral blood mononuclear cells as non-invasive diagnostic indicators of stress-associated neural states

Johannes Hendrik van Heerden

Supervisors: Prof. Nicola Illing (MCB) and Prof. Dan Stein (Psychiatry)

PLAGIARISM DECLARATION

1. I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is my own.
2. Each contribution to, and quotation in this research report from the work(s) of other people has been contributed, and has been cited and referenced.
3. This research report is my own work and has not previously in its entirety or in part been submitted at any other university for another degree.
4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

Signature: _____

Date: _____

Peripheral blood mononuclear cells as non-invasive diagnostic indicators of stress-associated neural states

Johannes Hendrik van Heerden

VHRJOH008

RESEARCH REPORT

Submitted in fulfilment of the requirements for the degree of

Master of Science (Molecular Biology)

December 2008

- TABLE OF CONTENTS -

- LIST OF FIGURES AND TABLES -

- ABBREVIATIONS -

%OAE	% Open arm entries
%pDIP	% Protected head dips
%pSAP	% Protected stretch attend postures
Ach	Acetylcholine
ACTH	Adrenocorticotropic hormone
Adcy8	Adenylate cyclase 8
Adm	Adrenomedullin
ANS	Autonomic nervous system
aRNA	Antisense RNA
BBB	Blood-brain-barrier
CAT	Close arm time
Cck	Cholecystokinin
CD	Cushing's disease
CE	Centre entries
CED	Centre entry delay
cGR	Centre grooms
Chrna2	Nicotinic cholinergic receptor alpha 2
CNS	Central nervous system
Cort	Cortistatin
cRear	Centre rears
CRH	Corticotrophin releasing hormone
cSAP	Centre stretch attend postures
CT	Centre time
DA	Dopamine
DE	Differentially expressed
E	Epineprhine
Edg5	Endothelial differentiation, sphingolipid G-protein-coupled receptor, 5
EPM	Elevated plus maze
GABA	γ-aminobutyric acid
GH	Growth hormone
GO	Gene ontology
GR	Glucocorticoid receptor
Hcrt	Hypocretin (orexin) neuropeptide precursor
Hic	Hippocampus
HPA	Hypothalamic-pituitary-adrenal
HPAA	HPA axis
Htr3a	5-Hydroxytryptamine (serotonin) receptor 3A
Hyp	Hypothalamus
IFNγ	Interferon-γ
Ig	Immunoglobulin
IL	Interleukin
IVT	In vitro transcription

KNN	K-nearest neighbour
LG	Licking and grooming
LH	Luteinizing hormone
Mch	Melaning concentrating hormone
M-CSF	Macrophage-colony stimulating factor
MS	Maternal separation
NE	Norepinphrine
NEI	Neuro-endocrine-immune
NGFI-A	Nerve-growth-factor-inducible factor a
NK	Natural killer
Nosip	Nitric oxide synthase interacting protein
Npvf	Neuropeptide VF precursor
NPY	Neuropeptide Y
OAT	Open arm time
OF	Open field
OR / *Olfr*	Olfactory receptor
Oxt	Oxytocin
P2ry4	Purinergic receptor, P2Y, G-protein coupled, 4
PBMC	Peripheral blood mononuclear cell
pFC	Prefrontal cortex
PGE2	Prostaglandin E2
pGR	Peripheral grooms
Pmch	Pro-melanin-concentrating hormone
PMIs	Post-mortem intervals
PND	Postnatal day
PNS	Peripheral nervous system
pRear	Peripheral rears
pRet	Peripheral returns
PRL	Prolactin
pSAP	Peripheral stretch attend postures
PSNS	Parasympathetic nervous system
Ptger1	Prostaglandin E receptor 1 (subtype EP1)
RE	Residual error
RIN	RNA integrity number
S.D.	Standard deviation
SAM	Significance of microarrays
SCF	Stem cell factor
SH	Simulated handling
SNS	Sympathetic nervous system
SP	Substance P
Sst	Somatostatin
Stim1	Stromal interaction molecule 1
SVM	Support vector machine
tAE	Total arm entries
TBS	Tricine-buffered-saline

tDIP	Total head dips
tGr	Total grooms
Thpo	Thrombopoietin precursor
TMEV	Tigr MultiExperiment Viewer
TNFα	Tumor necrosis factor-α
tRear	Total rears
tSAP	Total stretch attend postures

- ACKNOWLEDGEMENTS -

I would like to thank Prof. Nicola Illing for her supervision and guidance through a challenging project. Secondly, my sincere thanks go to my co-supervisor Prof. Dan Stein for the conception of this project and his valuable intellectual support during the course of it. A big thank you to Mandy Mason and Dorit Hockman for providing some much needed extra hands during sample collections. Additionally, I am greatly appreciative of time invested by Prof. Vivienne Russel, providing not only intellectual input, but also physical assistance during sample collections. Thanks also go to Joaquin Dopazo and his Bioinformatics group in Valencia, Spain, whom I had the privilege to spend three months with. Specifically, I would like to acknowledge the indispensible help provided David Montaner and Ana Conessa during my stay in Valencia. I would also like to sincerely thank Anè Korff for her specialist and invaluable input in the scoring of mouse behaviours.

Without funding this project would not have been possible and I would like to recognise the National Research Foundation (NRF) and the Cross-University Brain Behaviour Initiative (CUBBI) for the investments made by them.

Last, but not least, I would like to thank my parents for their support and continued encouragement, without which I would not be where I am today.

- ABSTRACT -

Insights into the functional integration of regulatory exchanges between neuro-, endocrine- and immune-systems in health and disease, combined with advances in high-throughput molecular technologies, have provided a foundation for recent interests in the use of peripheral tissue targets as non-invasive diagnostic indicators of neuropsychiatric state. Researchers have demonstrated the ability to predict psychopathological states from human peripheral immune tissue transcriptional profiles, using microarrays. Although evidence in support of such an approach as a viable diagnostic avenue within psychiatric settings is accumulating; it remains to be demonstrated, in an animal model, that transcriptional changes in peripheral tissue targets are paralleled by specific gene expression changes in neural tissues.

A mouse maternal separation model, with aetiological constructs relevant to stress- and anxiety disorders, was used to test the hypothesis that gene expression profiles of circulating peripheral blood mononuclear cells (PBMCs) could be used to predict stress-related psychological states. Importantly, a multi-dimensional strategy involving measures of (1) stress- and anxiety-related behaviours (2) stress-induced corticosterone responses and (3) microarray gene expression profiles of three brain regions (prefrontal cortex, hippocampus and hypothalamus), provided substantial evidence of stress-related differences between maternally separated (MS) and control animals (SH). Within this context, PBMC microarray data highlighted a set of 50 genes which provided sufficient information to predict the treatment status of individual samples with 95% accuracy. Moreover, the results demonstrated that stress-related transcriptome differences in PBMC populations were paralleled by stress-related gene expression differences in neural target tissues.

- CHAPTER 1 -

Stress and whole body homeostasis: Emerging perspectives on the molecular basis of psychiatric disorders

1.1 INTRODUCTION

The prevalence of neuropsychiatric disorders are assuming pandemic proportions (Avissar and Schreiber, 2002; The World Health Report 2001) and are projected to become some of the most important contributors to the world's disease burden within the next two decades (Mathers and Loncar, 2006; Murray and Lopez, 1996). In South Africa, neuropsychiatric disorders are second only to HIV/AIDS in its contribution to the country's disease burden (Bradshaw *et al.*, 2003). There is a growing need for new strategies, both social and clinical; to counter the burgeoning impact neuropsychiatric disorders are having on society.

In recent decades, significant technological advances have led to major shifts in our perceptions regarding the integrated and systemic nature of biological systems. Concomitantly our understanding of the pathophysiology of complex diseases, including neuropsychiatric disorders, has changed dramatically from traditional views. Classically, the nervous and immune systems were viewed as two operationally independent but complimentary regulatory mechanisms, both functioning to maintain the integrity of an organism when challenged by stress, where stress is loosely defined as a state of threatened homeostasis (Pacák and Palkovits, 2001). The nervous system was known to regulate homeostasis by means of an intimate connection with the endocrine system (Solomon *et al.*, 2002), functionally characterized as the neuro-endocrine axis; its role being the regulation of metabolism and behaviour via a cascade of effector molecules known as hormones.

The immune system was thought to be solely involved in the regulation of homeostatic disruptions caused by inflammations and antigenic invasions. It is now understood that interactions between neural, endocrine and immune environments are intimately integrated into an intricate bi-directional network referred to as the neuro-endocrine-immune (NEI) axis, constituting one of the principal homeostatic mechanisms of vertebrates (Wrona, 2006; Dantzer, 2004a; Melmed, 2001; Rabin, 1999; Felten *et al.*, 1991). Dysregulation of components within this network has been implicated in a diverse range of diseases and illnesses (Mašek *et al.*, 2003; Melmed, 2001; Rabin, 1999). Contemporary psychiatric paradigms have taken cognisance of these insights, redefining the scope of potential diagnostic and treatment strategies.

Our understanding of the implications of regulatory exchanges, within this elaborate network, is still far from complete. Nevertheless, an informative picture is starting to emerge, providing a contextual framework for the interpretation of homeostatic disruptions within psychiatric contexts. Existing perspectives can be summarised by focussing on three important themes: (1) The current body of evidence, in general terms, which support the existence of an integrated neuro-endocrine-immune axis, (2) the nature and physiological effects of these interactions, specifically in the context of the classic (hypothalamus activated) stress response and (3) the clinical implications of functional interactions within this axis.

1.2 THE NEURO-ENDOCRINE-IMMUNE AXIS

1.2.1 Cognitive evaluation of stressors and the initiation of a stress response

A brief consideration of stress and its impact on neurophysiological arousal is warranted before discussing the functional nature of the neuro-endocrine-immune

axis. It is important to mention the role that higher-level cognitive functions, those involved in subjective evaluation, play in the interpretation of a specific stimulus as a stressor. The subjective evaluation of a stimulus and its interpretation as being stressful, ultimately determines the specific downstream biochemical and physiological responses associated with stress (Jordaan and Jordaan, 2000; Rabin, 1999).

The focus of this review is on the functional nature, anatomically and physiologically, of stress responses, once initiated. The general stress response is characterized by behavioral, biochemical and physiological changes and has its origin in the central nervous system (CNS). It has been noted that two primary mechanisms exist by which the stress response can be propagated from the CNS to other tissues. Firstly, by means of neuroendocrine cascades, primarily involving the hypothalamic-pituitary-adrenal (HPA) axis; the second mechanism involves the autonomic nervous system (ANS). Important to note, is the central role of the hypothalamus in the activation of both these routes (Jordaan and Jordaan, 2000). The hypothalamus is intimately linked to the pituitary gland, by means of the pituitary stalk, and a number of other brain structures, including the amygdala, the hippocampus, the limbic system and autonomic projections, all of which have been functionally implicated in stress responses (Solomon *et al.*, 2002; Jordaan and Jordaan, 2000; Rabin, 1999). Neurophysiological arousal of the hypothalamus, in response to stressors, leads to the activation of (1) the HPA axis (HPAA) and (2) the ANS (Fig. 1.1) (Shepherd *et al.*, 2005; Jordaan and Jordaan, 2000). In general terms, both these routes function to effect metabolic changes that facilitate the maintenance of homeostasis in response to stimuli which are perceived as being stressful.

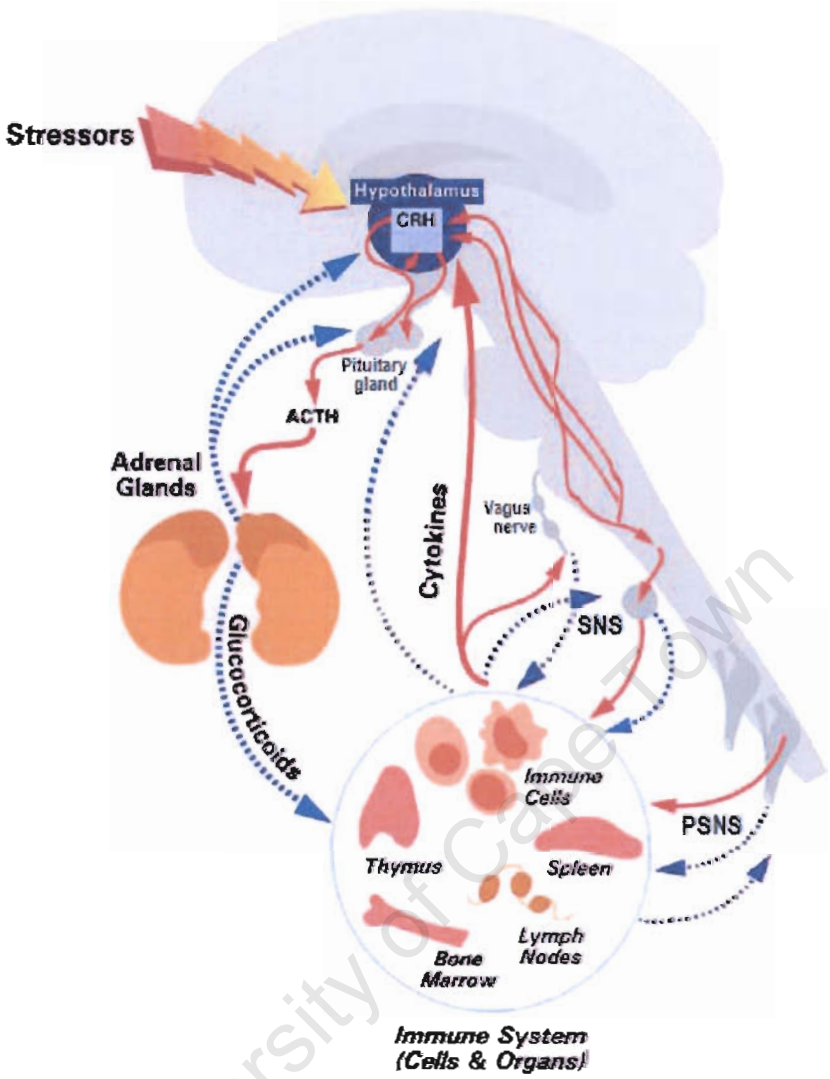

Figure 1.1: Schematic representation of the functional and anatomical relationships between the principal components of the neuro-endocrine-immune axis. The CNS and immune system are bi-directionally connected via both the sympathetic nervous system (SNS) and parasympathetic nervous system (PSNS) components of the ANS, in addition to the humoral routes characteristic of the neuroendocrine cascade. Also indicated is the Vagus nerve (Vagus n.), an important component of the PSNS. Dotted lines represent negative regulatory pathways; solid lines represent positive regulatory pathways (Image adapted from Marques-Deak *et al.*, 2005).

1.2.2 Common physiological and biochemical features within the neuro-endocrine-immune axis

The existence of a functional interactive axis, between the CNS and the immune system, was suspected from the earliest days of formal scientific enquiry, with

several observers linking psychopathological states to disease or illness; concepts which, traditionally, were associated solely with immune function, or rather dysfunction. One of the earliest recorded suggestions of a functional link between the nervous system and the immune system was that made by the Greek physician Galen, who is noted to have declared that melancholic women were more likely to develop cancer (Rabin, 1999). Substantiation of a functional link between the nervous and immune systems was provided by more recent discoveries revealing common features and mechanisms present in the neural, endocrine and immune environments. Anatomical, physiological and pharmacological evidence, provide the underpinnings for the current NEI axis model. Reciprocal exchanges, between the components of this amalgamated homeostatic mechanism, are only possible if the individual systems (neural, endocrine and immune) "speak" a common biochemical language. It has been demonstrated, experimentally, that this is indeed the case, with (1) receptors for cytokines, neuropeptides and neurotransmitters having been demonstrated on immune, endocrine and neural cells and (2) immune and neuro-endrocrine products having been shown to be present in lymphoid, endocrine and neural tissues.

1.2.2.1 Expression of shared receptors

1.2.2.1.1 Receptors for hormones, neuropeptides and neurotransmitters on immunocompetent cells

The existence of hormonal receptors on immune cells was demonstrated as early as 1978 when cytoplasmic glucocoritcoid receptors (GR) were identified in both monocyte and macrophage cell populations (Werb *et al.*, 1978). Since then, receptors for several other hormones, neurotransmitters and neuropeptides,

including insulin, dopamine (DA), prolactin (PRL), growth hormone (GH), estradiol, testosterone, β-adrenergic agents (catecholamines – epinephrine and norepinephrine), acetylcholine (Ach), endorphins, enkephalins, substance P (SP), neuropeptide Y (NPY), somatostatin (Sst) and vasointestinal peptide (VIP) have all been shown to be present in or on immunocompetent cells (Wrona, 2006; Pozo and Delgado, 2004; Rabin, 1999; Besedovksy and Del Rey, 1996). Besedovksy and Del Rey (1996) point out that (1) receptors for neuro-endocrine ligands appear to be differentially expressed on different types of immune cells and (2) the affinity and associated activity of a specific receptor for a given neuro-endocrine agent may change in response to other metabolic events (e.g. T-lymphocyte activation). These latter features are hallmarks of a highly adaptable environment and are essential for the effective regulation of immune-responses to dynamic changes in the neuro-endocrine milieu.

1.2.2.1.2 Receptors for immune-derived products on neural and endocrine cells

Receptors for immune-derived chemical messengers, collectively referred to as cytokines, have been demonstrated on endocrine organs and selected areas of the CNS. Interleukin (IL) receptors have been shown to be present in several endocrine tissues, including the pituitary, adrenal, thyroid, pancreas, testis, and ovaries. The most comprehensively characterized cytokine receptor is that for interleukin-1 (IL-1), which has been shown to be present in all endocrine tissues, except the adrenal glands. Several other studies have illustrated receptors or binding sites for IL-2 and IL-6 on the pituitary gland (Rabin, 1999; Besedovksy and Del Rey, 1996). Cytokine receptors in the CNS have also been extensively documented, with receptors for many of the common cytokines, including IL-1, IL-2, IL-4, IL-6, tumor necrosis factor-

α (TNFα), interferon-γ (IFNγ), macrophage-colony stimulating factor (M-CSF) and stem cell factor (SCF), having been described (Wrona, 2006; Rabin, 1999; Besedovksy and Del Rey, 1996; Dantzer *et al.*, 1993). Again the most commonly characterized receptors are those for IL-1, which have been shown to be present in high concentrations in the hippocampus (Besedovksy and Del Rey, 1996) and the hypothalamus (Dantzer *et al.*, 1993) and at lower levels in the cortex (Rothwell, 1999). In addition to IL-1, receptors for IL-6, IL-10 and TNFα have been well characterized and have been demonstrated in various brain regions including the cortex hypothalamus, hippocampus and the pituitary (Ward *et al.*, 2001; Barkhudaryan and Dunn, 1999; Nadeau and Rivest, 1999; Rabin, 1999; Kobayashi *et al.*, 1997). There is, however, some disagreement amongst researchers regarding the nature of cytokine receptor expression. Some authors have found cytokine receptors to be expressed constitutively, whilst others report their presence only after immune agitation; an ambiguity remaining to be clarified (Besedovksy and Del Rey, 1996).

The reciprocal expression of receptors for hormones, neurotransmitters, neuropeptides and cytokines, by cells of the neuro-endocrine-immune axis, provide a rational basis for potential bi-directional communications within the axis. However, a meaningful functional dialogue, involving shared receptors, can only be facilitated if immune and neuro-endocrine messenger agents coexist in lymphoid, endocrine and neural tissues. Studies demonstrating the presence of hormones, neuropeptides and neurotransmitters in the lymphoid tissues and conversely, the presence of immune-derived products in neural-endocrine structures, serve to substantiate, further, the existence of functional bi-directional exchanges, regulatory in nature, between the neural, endocrine and immune environments.

1.2.2.2 Common chemical messengers

1.2.2.2.1 Neural and endocrine messengers in lymphoid tissues

Lymphoid organs and associated cells are exposed to neuro-endocrine agents, including hormones, neurotransmitters and neuropeptides, generally by means of the circulatory system and specifically as a result of extensive autonomic innervations of both primary and secondary lymphoid tissues (Wrona, 2006; Shepherd *et al.*, 2005; Rabin, 1999; Besedovsky and Del Rey, 1996; Felten and Felten, 1991). Hormones are generally released into interstitial fluids or blood, by endocrine tissues, and are transported to target cells or organs by means of the circulatory system. In contrast, innervation of lymphoid organs by autonomic nerves form the basis for neurotransmitter and neuropeptide mediated regulatory exchanges. Sympathetic innervation by noradrenergic fibers has been extensively characterized in all lymphoid tissues, with some studies having demonstrated synaptic connections with individual lymphocytes (Rabin, 1999). Less substantially characterized is the nature and role of cholinergic parasympathetic innervation, principally via the vagus nerve. The intimate anatomical relationship between the lymphoid organs and the autonomic nervous system provide the functional means for exposure of lymphoid tissues to neurotransmitters and neuropeptides (Wrona, 2006; Shepherd *et al.*, 2005; Besedovsky and Del Rey, 1996). In addition, various studies have shown that cells of the immune system can, endogenously, synthesize and secrete significant amounts of key neuromodulatory agents. These include: (1) hormones - luteinizing hormone (LH), PRL, GH, corticotrophin releasing hormone (CRH) and ACTH; (2) neuropeptides – enkephalins, endorphins and brain-derived neurotrophic factor (BDNF); (3) neurotransmitters – norepinephrine (NE), epinephrine (E) and Ach. Examples include the constitutive secretion of BDNF by peripheral blood monocytes

and the secretion of GH by lymphocytes (Wrona, 2006). The endogenous synthesis of these neuro-endocrine associated agents, reasonably suggests that the immune system can readily participate in communicative exchanges associated with neuro-endocrine environments.

1.2.2.2.2 Cytokines in endocrine and neural tissues

As with hormones, cytokines can be distributed to key organs or tissues via the circulatory system where they can exert their influence as humoral signals. This is the primary means by which endocrine tissues are exposed to immune-derived products. In addition, several studies have shown cytokines to be present in endocrine glands (Besedovksy and Del Rey, 1996). IL-6 presents the best studied example of endogenous cytokine expression by these tissues (Besedovksy and Del Rey, 1996; Spangelo *et al.,* 1990). It has been illustrated that pituitary cells of mice and rats secrete this cytokine spontaneously (Besedovksy and Del Rey, 1996). In addition to IL-6, IL-1, IL-8 and TNFα, have all been shown to be present in specific pituitary tissues. Adrenal glands, specifically, the zona glomerulosa cells, have been shown to secrete IL-6. TNFα mRNA has been characterized in pancreatic islet cells. The gonads, both the testis and the ovaries, have been shown to secrete cytokines, including, IL-1 and TNFα amongst others (Besedovsky and Del Rey, 1996).

Nervous system tissues are, as with the endocrine organs, exposed to cytokines secreted by neural cells themselves, in addition to the circulating cytokines produced by immunocompetent cells. Immune-derived cytokine access, via humoral routes, to CNS tissues is generally prevented by the blood-brain-barrier (BBB). Studies have, however, shown that circulating cytokines can enter the brain through areas with a poorly developed BBB; alternatively, active transport of these agents

across the BBB has been validated as a common mechanism by which immune-derived products gain entry to the CNS (Wrona, 2006). Cytokine interaction with the autonomic branch of the peripheral nervous system (PNS) is less problematic, as many of the afferent and efferent fibers are in direct contact with lymphoid tissues (Wrona, 2006; Shepherd *et al.*, 2005; Rabin, 1999). Endogenous expression of cytokines, by the CNS, was first shown in astrocytes and microglial cells. Since then, several cytokines, including IL-1, IL-2, IL-3, IL-6, IL-8, IL-12 and IFNγ have been shown to be constitutively present in the brain (Besedovsky and Del Rey, 1996). Studies showing the precise localization of these cytokines are lacking and need to be further investigated. Cytokine production by peripheral neural tissues is sparsely characterized, with only a handful of studies having reported the presence of endogenous cytokine-like or cytokine agents. IL-1 presents the best characterized example of endogenous cytokine expression by these tissues; its presence has been demonstrated in cultured sympathetic ganglia (Besedovsky and Del Rey, 1996).

The preceding summary presented a diluted overview of the anatomical and physiological evidence in support of a functional neuro-endocrine-immune regulatory network. The functionality of this network is based on a mutual biochemical language which is facilitated by the presence of common receptors and chemical messengers. The nature and the physiological effects of specific interactions within this network are wide ranging, effecting adaptations in both physiological processes and complex behaviours.

1.2.3 Physiological effects of regulatory exchanges within the neuro-endocrine-immune axis

1.2.3.1 Experimental approaches to functional analyses

It is pertinent to briefly consider the experimental techniques used to elucidate the nature of and levels at which interactions take place within the axis. The functions of key structures or molecules are classically inferred from results obtained by specific and focused experimental techniques, most often using rodent models or in vitro cell culture based assays. Immune, endocrine and non-cortical brain structures and associated functions have been highly conserved across humans, rodents and primates, which make functional inferences possible (Hovatta and Barlow, 2008; Cryan and Holmes, 2005; Huang *et al.*, 2004; Tecott, 2003). Contextual consideration of results is critical, as many classic experimental approaches are selectively focused on one or very few interactions at a time; levels of interactions within the axis are enormously complex and inferred relationships between processes or structures are often considerably oversimplified. Traditionally, two approaches, both involving direct disruption of key structures or processes, were favoured by investigators. The first approach involves the artificial introduction of key regulatory molecules, agonists and antagonists, usually parenterally, which results in the disruption of signalling cascades. By conducting biochemical assays before and after exogenous introduction, the possible roles of specific molecules can be inferentially analysed. The second approach entails the mechanical disruption of principal anatomical structures or physiological relationships within the axis, by direct or chemical surgical means. Selective ablation of neural structures, endocrine glands and lymphoid organs, has permitted researchers to establish the nature of functional interactions between various key structural features within the axis

(Wrona, 2006; Mašek, 2003; Rabin, 1999; Besedovsky and Del Rey, 1996). These traditional approaches have provided the foundations for contemporary insights into the panoptic physiological effects of regulatory exchanges within this NEI axis.

More recently, with the development of gene-targeting techniques, researchers have been able to create specific and stable genetic models, e.g. knockout mice strains, targeting specific genes for use in various functional neuro-endocrine-immune investigations (Cryan and Holmes, 2005). In addition, with an emerging understanding of epigenetic effects on the geno- and phenotype of organisms, researchers have demonstrated the indirect manipulation of key regulatory systems, during development, to create specific experimental models involving neuro-endocrine components (Weaver *et al.*, 2004; Francis *et al.*, 2002). Over the last few years, microarray technologies have made possible the simultaneous inference of mRNA transcript abundance for 1000's of genes in specific tissues. This technique has allowed for efficient exploration of the presence and inference of possible, previously uncharacterized, functional associations between a large numbers of gene products (Mirnics *et al.*, 2006; Wilson *et al.*, 2004; Slonim, 2002).

1.2.3.2 Effects of hormones, neuropeptides and neurotransmitters on the immune system

1.2.3.2.1 Endocrine effects on immune function

Hormonal interactions with the immune system have been well characterized and can lead to either the depression or stimulation of immune responses (Pruett, 2001). Factors influencing the nature of the response include the kind and dose of hormone

and the timing of administration (Wrona, 2006). Endocrine outflow from the HPAA is the principal and best understood mechanism by which the CNS can influence the immune system (Wrona, 2006; Marques-Deak *et al.*, 2005; Padgett and Glaser, 2003; Besedovsky and Del Rey, 1996). Two other neuroendocrine cascades have been implicated in immuno-modulatory processes, these include: the hypothalamic-pituitary-thyroid (HPT) axis and the hypothalamic-pituitary-gonadal (HPG) axis (Marques-Deak *et al.*, 2005). Focus is, however, directed at the immunomodulatory role of the HPAA and its associated hormones.

The Glucocorticoids (cortisol in humans and corticosterone in rodents) are generally considered the main effector molecules of HPAA driven stress responses (Wrona, 2006; Marques-Deak *et al.*, 2005). These molecules have been implicated in the regulation of a wide variety of immune functions, including: cell trafficking, migration, maturation and differentiation (Marques-Deak *et al.*, 2005). The presence of glucocorticoid receptors in various immunocompetent cells (de Kloet *et al.*, 2007; Bartholome *et al.*, 2004; Miller *et al.*, 1998; Werb *et al.*, 1978) facilitates functional interactions with these molecules. Glucocorticoids have been shown to suppress, enhance and generally modulate the immune response. Originally implicated in the stimulatory regulation of immune function (Selye, 1956), the immunosuppressive effects of glucocorticoids has been stressed. It is now generally well established that glucocorticoid driven immunomodulations are complex in nature and can result in a wide range of outcomes (Wrona, 2006; Marques-Deak *et al.*, 2005; Munck, 2005; Mašek *et al.*, 2003; Padgett and Glaser, 2003; Dhabhar, 2002; Pruett, 2001; Rabin, 1999; Besedovsky and Del Rey, 1996). Specifically, glucocorticoids have been shown to play an important role in susceptibility and resistance to autoimmune, inflammatory, infectious and allergic diseases. An important aspect of the functional

effect of glucocorticoid immunomodulation is the nature and extent of a stressor (Marques-Deak *et al.*, 2005; Melmed, 2001; Rabin, 1999). Studies have shown that sustained activation of the HPA axis, and associated maintenance of elevated glucocorticoid levels, as typically occurs in response to chronic stressors, results in a general inhibitory immunomodulatory effect. In contrast, the transient elevation of glucocorticoids associated with acute stress, have been shown to enhance certain immune responses (Marques-Deak *et al.*, 2005; Mašek *et al.*, 2003; Dhabhar, 2002). It is thought that these differences reflect important evolutionary adaptations to different types of stressors (Dhabhar, 2002). Glucocorticoids have been shown to directly effect transcription of various "immuno-genes". They are thought to diffuse freely into the cytoplasm of certain immunocompetent cells, where they bind to and activate cytoplasmic glucocorticoid receptors, forming heterodimeric DNA binding complexes that can modulate the transcription of specific genes (Adcock and Caramori, 2001). Regulation of transcription is often achieved through interactions between the activated glucocorticoid receptor and other transcription factors, such as nuclear factor-$_{K}$B (NF-$_{K}$B) and activator protein-1 (AP-1), both known regulators of inflammatory gene expression (Adcock *et al.*, 2004). Transcriptional regulation of many proinflammatory genes, by glucocorticoids, has been demonstrated, specifically for genes encoding cytokines, adhesion molecules and chemoattractants. A well characterized effect of glucocorticoids on the immune system is the induction of a functional shift from cellular to humoral immunity. This shift comes about as a result of glucocorticoid's enhancing effect on the production of T-helper 2 cell cytokines, IL-4 an IL-10 (which enhance humoral immunity, e.g. stimulating antibody production), and their inhibitory effect on the production of T-helper 1 cell cytokines, IL-1 and TNFα (which enhances cellular immunity, i.e. B- and T-lymphocyte proliferation) (Marques-Deak *et al.*, 2005; Rabin, 1999).

ACTH is produced by the anterior pituitary, and its principal effect is the stimulation of glucocorticoid secretion by the adrenal cortex (Solomon, 2002). In addition, ACTH has been reported to exert direct inhibitory effects on immune functions, independently from its role in the stimulation of glucocorticoid output (Mašek *et al.*, 2003; Rabin, 1999; Besedovsky and Del Rey, 1996). ACTH has been demonstrated to function as a potent inhibitor of antibody production. It is speculated that this inhibition is as a result of its effect on T- rather than B-Lymphocyte function (Mašek *et al.*, 2003). Other hormones produced by the anterior pituitary, and implicated in immunomodulation, include: GH and PRL (Mašek *et al.*, 2003; Rabin, 1999; Besedovsky and Del Rey, 1996). GH, in general, participates in the regulation of growth of body, tissue metabolism and tissue repair. In addition, its role in the regulation of maturation and function of the immune system has been described. Deficiencies of GH have been implicated in depressed T-cell function, natural killer cell (NK) activity and antibody responses. Exogenous administration of GH has been shown to restore, to some degree, many of these immune functions (Mašek *et al.*, 2003; Rabin, 1999). Excessive production of GH is also associated with immune dysregylation, and is implicated in a condition known as acromegaly, which is characterized by severely elevated levels of phagocytic cells (Rabin, 1999). PRL is traditionally known as a hormone fundamental to mammalian reproduction and lactation. PRL has been generally associated with the stimulation of both humoral (antibody) and cellular immune responses. The exact nature of PRL-mediated immune responses is unclear as this hormone has been shown to interact with an extensive array of other hormones and neurotransmitters (Mašek *et al.*, 2003; Rabin, 1999).

Secretion of CRH by the hypothalamus, in response to the stress induced activation of this structure, is directly responsible for the release of ACTH, amongst other hormones, from the anterior pituitary (Fig. 1.2). In addition to its regulatory role within the HPA axis, CRH has loosely been implicated in the regulation of macrophage function. Webster and De Souza (1988) demonstrated the presence of high affinity binding sites for CRH on splenic macrophages. Other studies have described the presence of CRH and its mRNA in human peripheral blood cells, mouse T-lymphocytes (Karalis *et al.*, 1997) and rat spleen and thymus tissues (Baigent and Lowry, 2000). In vitro studies have suggested direct immunomodulatory involvement, but findings have been contradictory. CRH has been shown to have inhibitory, stimulatory or no immunomodulatory effects (Salas *et al.*, 1997). Described effects include: the regulation of both T- and B-lymphocyte proliferation, the stimulation of IL-1, IL-2 and IL-6 secretion, enhanced chemotaxis and NK cell mediated lysis (Karalis *et al.*, 1997). A CRH-ACTH immunomodulatory axis was originally hypothesized from observations that showed CRH induced stimulation of leukocyte derived ACTH and β-endorphins (Smith *et al.*, 1986). In vivo findings have been more difficult to interpret. Central and peripheral administration of CRH has been shown to produce different immunomodulatory outcomes, with both pro- and anti-inflammatory effects having been described (Karalis *et al.*, 1997). This supports the idea that, in vivo, the immunomodulatory effects of CRH on immune function seems to be exerted through interactions with other hormones (Wrona, 2006).

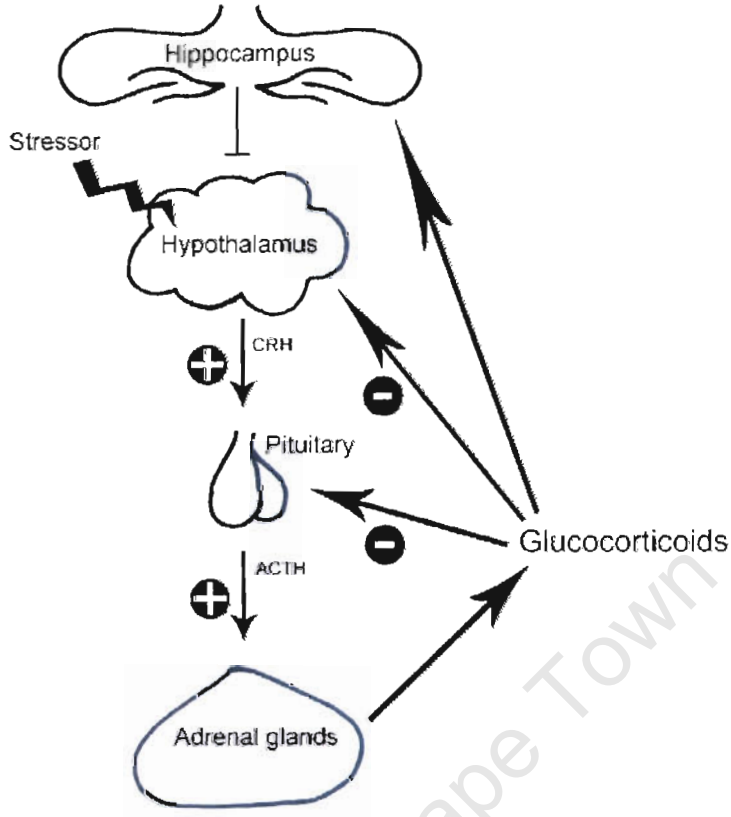

Figure 1.2: Schematic representation of HPA axis feedback mechanisms. The HPA axis is activated in response to a stressor, resulting in the release of CRH from the paraventricular nucleus of the hypothalamus. In turn, CRH induce the release of ACTH, from the anterior pituitary, into the blood stream. The adrenal glands respond to ACTH stimulation by releasing glucocorticoids. Endogenous expression of glucocorticoid receptors, by components of this axis, facilitate a negative feedback mechanism through which secreted glucocorticoids modulate the stress-associated activity of upstream components; effectively inhibiting the prolonged activation of this axis in response to stress. Additionally indicated, is an inhibitory role of the hippocampus on stress-associated hypothalamic activation, also in response to glucocorticoid binding.

The wide ranging immunomodulatory effects of hormones are often difficult to interpret, with many studies arriving at seemingly contradictory findings. In vitro approaches are challenged by several inherent complexities associated with hormonal cascades, including combinatorial effects of different hormones and self-regulatory mechanisms. It is well known that many of the hormonal cascades of vertebrates are characterized by complex feedback mechanisms. The HPAA cascade is no exception (Solomon *et al.*, 2002). This implies that hormones native

to this cascade can exert an indirect influence on immune function by affecting the overall activity of its own cascade. A case in point is the regulatory effect that glucocorticoids have on the overall activity of the HPAA. Glucocorticoids have been demonstrated to, in fact, feedback into the HPAA, affecting the activity of the pituitary and the hypothalamus; its general effect is to inhibit the prolonged activation of these structures, which effectively results in an inhibition of glucocorticoid activity itself (Fig. 1.2) (Marques-Deak *et al.*, 2005; Meany and Szyf, 2005).

1.2.3.2.2 Autonomic regulation of immune function

Traditionally, the autonomic nervous system was thought to effect physiological changes in response to stress exclusively by what has become known as the fight-or-flight response (Jordaan and Jordaan, 2000). This response is characterized by the activation of the sympathetic nervous system (SNS), which leads to the stimulation of target organs by NE, the main effector molecule of the SNS. This has a catabolic effect, mobilizing energy, which serve to prepare the body for action. Characteristics of this response include: Dry eyes and dilated pupils, a dry mouth, sweaty palms, strong rapid heartbeat, increased blood flow to muscles, increase adrenal activity. The parasympathetic nervous system (PSNS) functions to counter the effects of SNS activation, once a stress has been eliminated (Jordaan and Jordaan, 2000). A significant involvement in immunomodulation, by the autonomic nervous system, has been established. However, the extent to which to the sympathetic and parasympathetic branches participate remain controversial (Nance and Sanders, 2007).

A role for the SNS in immune regulation has been well established, with immunosuppressive and immunostimulatory mechanisms having been demonstrated

(Moynihan *et al.*, 2004; Elenkov *et al.*, 2000). The SNS can effect a wide range of immunomodulations, mainly via the actions of catecholamines, including cell proliferation, cytokine and antibody production, lytic activity and migration (Nance and Sanders, 2007; Wrona, 2006; Padget and Glaser, 2003). The catecholamines, E and NE, mediate their effects on immune tissues through interactions with adrenergic receptors. These receptors have been characterized on various immunocompetent cells (Padgett and Glaser, 2003). Although the effects of NE and E on immune regulation are similar, NE has been more extensively studied as it is directly released from sympathetic nerve terminals in response to SNS activation (Elenkov *et al.*, 2000). Both NE and E have been shown to affect the production of leukocytes by direct interaction with β-adrenergic receptors, found on cells of the lymphoid organs. In addition, E has been implicated in the inhibition of complement activation and macrophage-mediated lysis of certain immune-compromised cells (Wrona, 2006). NE has been shown to stimulate the production of several cytokines, including IL-6, IL-8 and IL-10, by lymphocytes (Elenkov *et al.*, 2000). On the other hand, a profound inhibitory effect on NK cell populations, by NE, has been demonstrated. This inhibitory effect involves complex feedback mechanisms, mediated by the SNS, between the lymphoid organs and the CNS (Elenkov *et al.*, 2000). NE induced modification of NK cell receptor ligation efficiency to target cells has been described. This action has a dramatic effect on target-induced activation of cytotoxic NK cell mechanisms. NE-associated inhibition of cytokine secretion by NK cells, which is necessary for NK maturation and differentiation has also been characterized (Wrona, 2006). Chemical sympathectomy, via the administration of 6-hydroxydopamine, has been shown to lead to a general increase in B-lymphocyte activity associated within the spleen (Besedovsky and Del Rey, 1996) suggesting a general inhibitory effect by the SNS on the adaptive immune system. Other studies

involving the 6-hydroxydopamine ablation of sympathetic nerve tracts have, in contrast, demonstrated severe impairments in various parameters of immune function, including a general reduction in T-lymphocyte responses (Mašek *et al.*, 2003). These and other conflicting findings have made the characterization of general SNS-mediated immunomodulatory mechanisms difficult.

Circulating DA, a catecholamine neurotransmitter, has also been associated with SNS-mediated effects on the immune system (Wrona, 2006; Elenkov *et al.*, 2000). It has been suggested that the effect of circulating DA on immune function is in part mediated by the noradrenergic axon terminals of the SNS. The mechanism involves the uptake of circulating DA by axon terminals, which converts it into a NE intermediary that can be released as either NE or DA (Elenkov *et al.*, 2000). Elevated physiological concentrations of DA have been implicated in the inhibition of CD4+ and CD8+ proliferation and cytotoxicity. DA in combination with NE was found to increase lymphocyte activation (Wrona, 2006). In addition to catecholamine exchanges with the lymphoid tissues, several lines of evidence have implicated neuropeptides in immunomodulation. Many important neuropeptides are localized in the nerve terminals of autonomic nerve fibers innervating primary and secondary lymphoid organs, including: SP, Sst, VIP, NPY (Wrona, 2006; Rabin, 1999). Direct effects of these neuropeptides on certain populations of immunocompetent cells have been described. SP was shown to enhance B-lymphocyte production of immunoglobulin (Ig) M and A, but not of IgG (Besedovsky and Del Rey, 1996). Furthermore, SP enhances lymphocyte migration to inflammatory sites and promotes phagocytosis and chemotaxis by these cells (Wrona, 2006; Rabin, 1999). SP has been implicated in the induction of immunoglobulin production by B-lymphocytes as well as increased cytokine release by T-lymphocytes and mononuclear phagocytic

cells (Rabin, 1999). The effects of Sst and VIP on immune function have been shown to be dependent on several parameters, including: effective concentrations, target cell identity and receptor densities on target cells. Both peptides had modulatory effects on B- and T-lymphocyte function (Rabin, 1999; Delgado, 2003; Wrona, 2006). A general immunosuppressive role for NPY has emerged, with a well characterised inhibitory effect of this neuropeptide on NK cell activity (Rabin, 1999). Isolated studies have reported a stimulatory effect of NPY. Elitsur *et al.*, (1994) reported the NPY stimulated increase of gastrointestinal mucosal lymphocyte mitosis. It is interesting to note that the co-localization of NE and NPY has been described (Wrona, 2006; Shepherd *et al.*, 2005; Stjernquist *et al.*, 1987) with observations characterizing the concomitant release of these compounds from sympathetic nerve terminals (Elenkov *et al.*, 2000).

The role of the PSNS in immunomodulation is not as substantially characterized as its functional counterpart, the SNS, and is still considered controversial (Nance and Sanders, 2007). In recent years a few groups have presented findings which infer a possible immunomodulatory role for efferent parasympathetic branches (Czura and Tracey, 2005; Zimring *et al.*, 2005; Pavlov and Tracey, 2004; Kawashima and Fujii, 2003; Bernik *et al.,* 2002; Borovikova *et al.*, 2000). The main effector molecule of the PSNS is Ach. This molecule is a short-acting neurotransmitter and its physiological action generally counters the effect of sympathetic arousal (Solomon *et al.*, 2002). Parasympathetic exchanges, via the efferent vagus nerve, with immune tissues are primarily implicated in the suppression of acute inflammation, now referred to as inflammatory reflexes (Tracey, 2007; Tracey, 2002). An immunomodulatory role was inferred from in vitro observations, which showed that Ach effected the cellular deactivation and inhibition

of pro-inflammatory cytokine (TNFα, IL-1, IL-6 and IL-18) release (Borovikova *et al.*, 2000). Additionally, an in vivo inhibitory effect of Ach on TNFα was shown both in the liver and serum of rats and is suspected to play a role in the prevention of endotoxic shock (Borovikova *et al.*, 2000). Other studies have shown an immunostimulatory effect of Ach on the development of cytolytic T-lymphocytes (Wrona, 2006).

The above summary highlights the many ways by which hormones, neurotransmitters and neuropeptides can affect specific features and cell populations of the immune system. The behaviour of immune effector cells and molecules can be regulated at several levels; ranging from general processes, e.g. intermediate metabolism, to very specific and specialized features, such as selective stimulation of specific immunocompetent cell proliferations (Besedovsky and Del Rey, 1996). An important aspect of these neuro-endocrine and autonomic-neural immunomodulations, is the multidimensionality of these regulatory processes. The outcome of regulatory interactions is a consequence of the simultaneous binding of several neuro-endocrine and autonomic-neural agents. It is reasonable to suspect that a specific complement of these agents can produce unique outcomes, determined not only by the relative activity of the upstream systems but also influenced by processes endogenous to the immune system. The determinate effect, of many of the above mentioned regulatory agents, is therefore difficult to gage, as the complexity of specific and potential interactions necessitates a generally inferential approach. It should be noted, however, that many studies yield results that validate those obtained by others, thereby providing a reliable general model of the nature and extent of regulatory exchanges with the immune system.

1.2.3.3 Effects of cytokines on endocrine and neural components

Maintenance of homeostasis by neuro-endocrine-immune mechanisms requires bi-directional communications between all systems. Uncontrolled activation or deactivation of the immune system, by the neuro-endocrine effectors already mentioned, will lead to severe metabolic disruptions. It is crucial that the immune system communicate its functional state to the CNS by means of complementary neural and humoral routes. Cytokine communications with the CNS has a direct impact on the regulation of the immune system itself; in addition, characteristic behavioural changes are induced, often referred to as sickness behaviour (Dantzer, 2004b; Rabin, 1999).

1.2.3.3.1 Cytokine effects on the HPA axis

Several cytokines, including IL-1, IL-2, IL-6, IFNγ and TNFα, have been demonstrated to directly influence the active state of the HPAA, which in turn influences immune responses as a result of alterations in glucocorticoid and other stress hormone levels (Wrona, 2006; Turnbull and Rivier, 1999; Besedovsky and Del Rey, 1996). Most extensively characterized, is IL-1; its capacity to stimulate the HPAA is well established. IL-1 has been implicated in the direct stimulation of ACTH release from pituitary cells, resulting in increased glucocorticoid output (Mašek *et al.*, 2003; Rabin, 1999; Turnbull and Rivier, 1999; Besedovsky and Del Rey, 1996). Furthermore, IL-1 has been shown to enhance the turnover of NE in the hypothalamus, in addition to its stimulatory action on CRH release, implying a regulatory role in hypothalamic neurosecretory activity (Wrona, 2006). Other cytokines reported to interact directly with the HPAA are IL-2, IL-6, TNFα and IFNγ; all have been implicated in pituitary-adrenal activity. The effects of IL-1, IL-2 and IL-

6 on the adrenals results in a generally stimulatory outcome, characterized by increased glucocorticoid plasma levels (Besedovsky and Del Rey, 1996).

1.2.3.3.2 Cytokine effects on the CNS

Cytokine exchanges with CNS tissues have been implicated in alterations of neurotransmitter concentrations in the brain; in addition, the, sometimes profound, behavioural changes that are associated with certain types of illnesses have been demonstrated to involve cytokine mediated neural changes (Dantzer, 2004b; Rabin, 1999). IL-1 is associated with decreased levels of NE in the CNS. This effect has been described for several key brain areas, including the hypothalamus, hippocampus, brain stem and spinal cord (Wrona, 2006; Besedovksy and Del Rey, 1996). Several behavioural alterations, including: a loss of appetite, decreased social interaction and a decreased reproductive drive are associated with cytokine actions on CNS structures. These effects are collectively referred to as sickness-associated behaviors (Dantzer, 2004b; Rabin, 1999) and are linked to various pathological (e.g. viral infections) and psychopathological (e.g. stress and depression) conditions. Cytokine induced sleepiness has also been characterized; implicated in this effect are IL-1 and IL-6, with the possible involvement of TNFα having been noted by some researchers (Rabin, 1999). Both IL-1 and IL-6 are thought to stimulate DA metabolism in the striatum, hippocampus, and prefrontal cortex (Besedovsky and Del Rey, 1996), although some authors have reported no statistically significant effects for IL-6 in any brain regions (Wang and Dunn, 1998). These two cytokines have also been shown to affect processes related to learning and memory (Dantzer, 2004b). Additionally, IL-1, TNF and IFNγ have been shown to inhibited long term potentiation. Altered serotonin metabolism, in the hippocampus,

has also been associated with these cytokines (Myint and Kim, 2003; Wang and Dunn, 1998). The regulation of Ach release from hippocampal tissues has been demonstrated for IL-2 (Hanisch *et al.*, 1993). IL-3, IFNγ and CSF have all been implicated in the augmentation of choline-Ach-transferase activity. A stimulatory effect on activity of the cortex and hippocampus was described for IL-2 and IFNγ (Besedovsky and Del Rey, 1996).

1.2.3.3.3 Cytokine interactions with the peripheral nervous system

Results from recent studies imply indirect cytokine effects on the CNS as a result of interactions with afferent peripheral neurons. The functional interactions between cytokines and peripheral nerves have not been studied as extensively as those involving the CNS. However, various cytokine-mediated effects on the activity of the sensory components of the ANS have been characterized. These interactions provide the nervous system with the functional means to monitor general immune responses and important invasive events (Wrona, 2006; Tracey, 2002). The cytokine-mediated activation of afferent pathways is thought to result in the secondary induction of important CNS events, including the activation of the HPAA, local release of cytokines and sickness behaviours (Hopkins, 2007; Rothwell and Hopkins, 1995). IL-1 has been shown to affect NE levels in the periphery. Specifically, NE turnover in sympathetic nerves innervating the spleen, lungs, diaphragm and pancreas are substantially accelerated in response to IL-1 stimulation. The in vitro inhibition of NE release, by IL-1, from certain nerves has also been demonstrated (Besedovsky and Del Rey, 1996). Functional exchanges between cytokines and the afferent neural pathways have been implicated in the fine tuning of inflammatory responses. Several cytokines, including IL-1, IL-6 and TNFα

have been shown to play an important role in the mediation of an ANS response to inflammation, including the sensitization of afferent pain fibers (Rothwell and Hopkins, 1995). In addition, IL-1 has been shown to increase production of SP, by cervical ganglia, in vitro (Besedovsky and Del Rey, 1996).

The full picture of the extent of cytokine activities and their influence on the neural and endocrine environments is still very much emerging. That cytokines can effect important changes in the hormonal cascades of the endocrine system and neural activity of both the CNS and PNS, is now generally well established. The continued elucidation of important regulatory interactions holds potentially important implications for contemporary clinical practices and treatment strategies for complex pathologies.

1.3 CLINICAL IMPLICATIONS OF AN INTEGRATED NEURO-ENDOCRINE-IMMUNE AXIS

The emerging understanding of functional neuro-endocrine-immune interactions is facilitating the re-evaluation, by clinicians, of traditional treatment regimes. With the demonstration of the profound effects of neuro-endocrine mechanisms on immune function and vice versa, the dividing line between pathological and psychopathological conditions has become, in many ways, a semantic one. Armed with a new understanding of whole-body homeostasis, holistic strategies that take cognizance of the immune systems role in the mediation of key neural functions, in addition to the nervous system's effect on immune function, are becoming more attractive. A better understanding of the nature of complex pathologies like HIV induced dementia (Dubé *et al.*, 2005) and Post-traumatic stress disorder (PTSD) (Newport and Nemeroff, 2000) becomes possible when approached with an

understanding of the physiological effects that these systems can exert on each other. Immune dysfunction, as a consequence of psychopathological conditions, such as chronic stress or depression, has been successfully treated by the pharmacological modulation of selective brain functions (Melmed, 2001). Several proposals for new and novel treatment strategies, based on a functional knowledge of the neuro-endocrine-immune axis, are emerging (Peedicayil, 2008; Brustolim *et al.*, 2006; Kulmatycki and Jamali, 2006).

The treatment of psychiatric disorders and the efficiency of clinical intervention strategies have to a large extent been limited by the effectiveness of diagnostic techniques (Avissar and Schreiber, 2002). Diagnoses of psychiatric disorders rely heavily on the subjective categorical identification of discreet symptom clusters (Abbot, 2008; Finn *et al.*, 2003). These categorical approaches have important drawbacks as psychopathological conditions are classified on the basis of observed symptoms. The fact that most psychiatric disorders are progressive, not showing classifiable symptoms during early stages of onset, make efficient diagnosis and early intervention difficult. The problem is further compounded by overlapping clinical features between many common disorders (Abbot, 2008; Hovatta and Barlow, 2008; Finn *et al.*, 2003). Although, the neurobiology of many psychiatric disorders is fairly well understood, direct, or invasive, interrogation of the neural milieu is, for obvious reasons, not practical.

Recently, the possibility of non-invasively (or indirectly) interrogating neural states has been highlighted as a promising diagnostic prospect. Segman *et al.* (2005) demonstrated the ability to accurately predict the onset and progression of PTSD, in a group of recently traumatized patients, via the microarray analysis of gene expression profiles of peripheral blood mononuclear cells (PBMC). The results

demonstrated that the appearance of stable expression profiles, within these immune tissues, could be linked to defined psychiatric categories, emphasising the clinical relevance, of a functionally intergrated NEI-axis. Since the appearance of this paper, interest in the potential of peripheral tissues as non-invasive diagnostic indicators of neural states has grown. Huntington's disease (Runne *et al.*, 2007), Schizophrenia (Shwarz and Bahn, 2008; Yao *et al.*, 2008; Bowden *et al.*, 2006) and Bipolar disorder (Le-Niculescu *et al.*, 2008), have all been the subject of studies attempting to identify peripheral biomarkers using a high-throughput molecular strategy. Although most findings are promising, having identified candidate biomarkers, some important questions regarding the diagnostically predictive power of peripheral tissues, within a psychiatric context, remain unanswered. Specifically, it still remains to be seen whether the diagnostic resolution of peripheral tissues is sufficiently specific in nature and not reflective of general dysregulations in the CNS. Extensive gene expression studies, using whole genome arrays, across both CNS and immune systems have been lacking and should provide contextual insight into the cross-systems nature of coordinated stress-associated biochemical events.

1.4 STUDY AIMS

Work by Segman *et al.* (2005) and others have provided promising demonstrations of peripheral psychiatric diagnostics. However, implied relationships between characterised psychiatric states and the peripheral transcriptome are inferential. Although an NEI-axis picture, explicated above, provides the theoretical framework for such an inference, it remains to be demonstrated that transcriptional changes in peripheral targets are paralleled by specific alterations in neural tissues (Mirnics *et*

al., 2006). Furthermore, the functional interpretation of such a demonstration could help to further substantiate the disease-specificity of peripheral markers.

As its primary goal, this study set out to test the hypothesis that gene expression profiles of circulating PBMCs can be used to predict psychological states. Specifically, the concomitant transcriptional characterisation of key neural structures served to further evaluate peripheral transcriptomics as a diagnostic avenue for use in molecular psychiatry. With access to human neural tissues being limited by various practical issues, a rodent model with relevant aetiological constructs was employed to elucidate, further, the psychopathological relationship between the PBMC and neural transcriptomes. Specifically, a mouse model of early childhood trauma, maternal separation (MS), was used. This paradigm has been shown to impact on fear- and anxiety-related behaviours in adulthood, with long-lasting neurochemical and neurophysiological effects in several species, including rodents, non-human primates and humans (Gutman and Nemeroff, 2003; Bakshi and Kalin, 2002; Heim and Nemeroff, 2001). The behavioural effects of MS were evaluated using well defined behavioural tests. Importantly, treatment-associated behavioural profiles served to establish a non-molecular context, providing an additional dimension for the interpretation of downstream results. In addition to behavioural assessments, plasma Corticosterone concentrations were measured at basal stress levels and in response to acute restraint stress. These measures provided insights into possible MS induced differences in the HPAA stress responsivity. The principle focus of this study, however, was the parallel whole-genome transcriptional characterisation of PBMC and central nervous system tissues, using microarray technology. The characterisation of three defined CNS regions (Paxinos and Franklin, 2004), Prefrontal association cortex, Hippocampus and Hypothalamus,

provided a molecular context within which to interpret treatment-associated PBMC data. Finally, the ability of PBMC transcriptional profiles, to predict the treatment class of individual samples, was evaluated using popular classification algorithms.

The aims of this study can be summarised as follows:

1) The deployment of a mouse model of MS, to establish a defined etiological context with relevant psycho- and physiological constructs.

2) The assessment of the effects of MS on adult behaviour using two well defined and etiologically relevant behavioural paradigms, the Open Field (OF) and the Elevated Plus Maze (EPM) tests.

3) The measurement of plasma Corticosterone concentrations at basal stress levels and in response to acute restraint stress.

4) The individual characterization of whole-genome transcriptional profiles, using microarrays of (a) three key, stress associated brain structures: prefrontal cortex (pFC), hippocampus (Hic) and hypothalamus (Hyp) and (b) PBMC populations.

5) An assessment of the diagnostic efficiency of PBMC transcriptional profiles at predicting the treatment class of MS individuals and controls.

- CHAPTER 2 -

The epigenetic manipulation of neural development:

Establishing differential neural profiles through a Maternal Separation paradigm

---

2.1 INTRODUCTION

Animal models have played a pivotal role in many of the most significant breakthroughs in modern medicine with non-human primates and rodents having formed the backbone of research focussed on neurobiological processes and psychiatric disorders (Rodgers *et al.*, 1997). The organization and anatomy of many important neural structures, including the: cerebral hemispheres, diencephalon, midbrain, cerebellum, pons and medulla have been well conserved across mammals. Principally, the major difference in neural organization relates to the expansion of the cerebral cortex (Fig. 2.1) (Cryan and Holmes, 2005). In humans this structure has been greatly elaborated with important implications for psychological constructs related to sophisticated cognitive evaluations and responses to external stimuli. Although subject to constraints of interpretive complexities, there are significant similarities between the neural structures and associated processes of non-cortical components, characterised by the functional conservation of important neural circuitries and behavioural responses (Cryan and Holmes, 2005). These properties provide a rational context within which inferential strategies are employed to elucidate the mechanistic, including epigenetic, genetic and biochemical, processes underlying important behavioural phenomena.

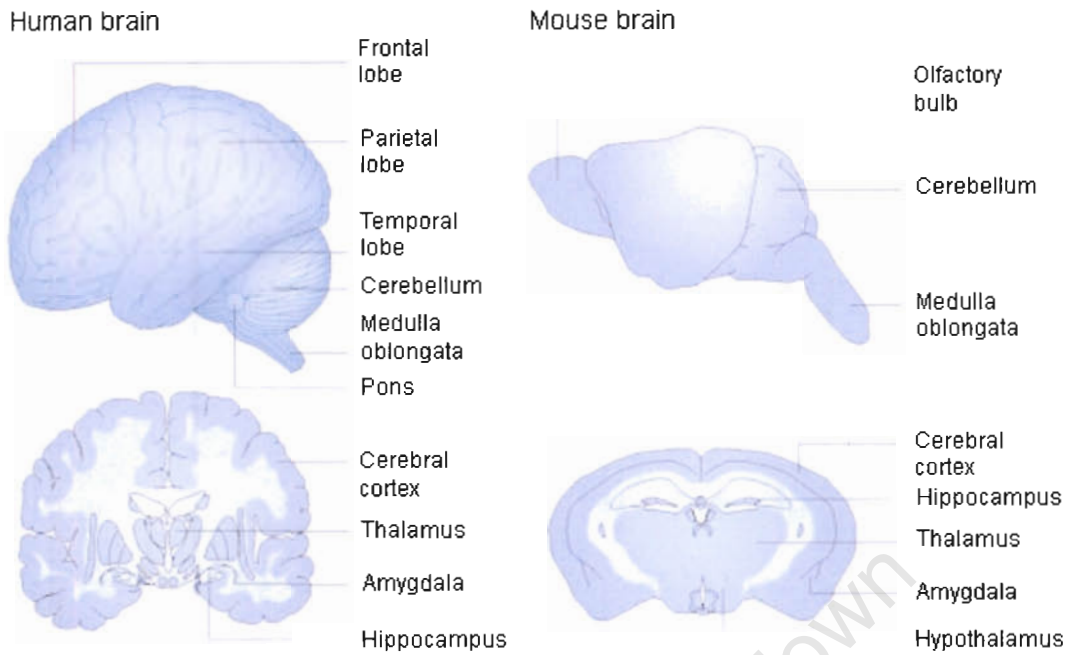

Figure 2.1: Comparison of human and mouse brains. The principle difference between the mouse and human brains pertain to the massively elaborated cerebral cortex. Other key areas have been well conserved, both structurally and functionally. Image reproduced from Cryan and Holmes (2005) with permission from A. Holmes.

2.2 MATERNAL SEPARATION – A DEVELOPMENTAL MODEL OF EARLY CHILDHOOD TRAUMA AND NEGLECT

Contemporary psychiatric paradigms are increasingly being informed by our understanding of gene x environment contributions to the susceptibility and development of many types of psychopathologies. Experimental manipulations of early developmental environments, using model organisms, have demonstrated dramatic effects on the fear- and anxiety-like behavioural phenotypes of the adult organisms (Bakshi and Kalin, 2002; Meaney, 2001). A particularly attractive developmental stress model is that of prolonged maternal separation (MS); shown to represent an ecologically valid form of early-life stress, with etiological constructs relevant to depression and anxiety (Cryan and Holmes, 2005; Bakshi and Kalin, 2002). The phenomenology of behavioural abnormalities, as a consequence of

maternal neglect or separation, has been described in rodents, non-human primates and humans (Romeo *et al.*, 2003; Bakshi and Kalin, 2002; Meany, 2001; Sanchez *et al.*, 2001; Heim and Nemerof, 2001). The early developmental environment, and specifically the quality of mother-offspring interactions, have been shown to be fundamental to the development of emotionality (Meaney and Szyf, 2005; Bakshi and Kalin, 2002). An important aspect of mother-offspring interaction is thought to involve tactile components of interaction between them (Kuhn and Schanberg, 1998). Cross-fostering studies in rats have implicated a direct relationship between the magnitude of HPA-mediated stress responses in adulthood and variations in the quality of pre-weaning maternal care received (Meaney and Szyf, 2005). Specifically, the disruption of normal mother-infant interactions represents a significant stressor that, if experienced during critical developmental phases, can have a long lasting deleterious impact on behavioural and neuroendocrine adaptations and responses to stress (Bakshi and Kalin, 2002).

A growing body of evidence, pertaining to the neurobiology of early MS, shows significant overlap in data from rodent, non-human primate and human studies (Gutman and Nemeroff, 2003; Bakshi and Kalin, 2002; Heim and Nemeroff, 2001). The exact mechanisms by which maternal deprivation affects the behavioural and neurochemical profiles of adult individuals are not fully understood, but have been shown to result in the long-term alteration of various neurotransmitter and hormone systems, with dramatic implications for emotional- and cognitive processes (Aisa *et al.*, 2007; Parfitt *et al.*, 2004; Daniels *et al.*, 2004; Bakshi and Kalin, 2002).

Recently, Meany and Szyf (2005) proposed a mechanism that involves the epigenetic modification of chromatin structures in the hippocampus, altering the transcriptional activity of genes central to the regulation of fear- and anxiety-like

the chromatin associated with the GR promoter and therefore a high degree of DNA methylation, which is maintained through to adulthood in the hippocampus.

Behavioural paradigms are classified with contextual reference to stimuli which appear to have analogous constructs in humans (Rodgers *et al.*, 1997). The elevated plus maze (EPM) and open field (OF) tests present two of the most routinely used fear- and anxiety-related paradigms. Both tests measure unconditioned (i.e. spontaneous) explorative-type behaviours and are generally considered to have a high degree of ecological validity (Rodgers *et al.*, 1997).

The EPM provides an exploratory conflict-type environment. It is shaped like a Greek cross with two "safe" runways, characterized by closed walls, and two open runways, with no walls, which present a "risk-laden" choice (Fig. 2.2A). The entire maze is raised at least 30 cm from the floor (Crawley *et al.*, 1997). Mice are allowed to spontaneously explore this environment, choosing between exposed or protected parts of the maze. The OF also assesses the conflict between innate drives to explore a novel environment and general fearfulness of exposed or unprotected spaces. The OF usually consists of a square-shaped arena with raised walls. The centre of this environment is often brightly illuminated presenting an aversive or risk-laden choice, with the thigmotactic cues of the periphery providing a safer area (Fig. 2.2B).

Conventionally, risk-based exploration is defined as a spatiotemporal component of area avoidance vs. preference (Carola *et al.*, 2002; Rodgers and Johnson, 1995). In both tests, relative levels of the riskier exploration options are quantified and taken as an index of fear- and anxiety-like behaviour. These conventional parameters have been shown to reflect important aspects of stress-associated anxiety. They have been particularly useful in highlighting important neurochemical and neurophsyiological components of anxiety-related exploration behaviours, when combined with psychopharmaceutical and/or gene-targeting

strategies (Finn *et al.*, 2003; Belzung and Griebel, 2001; Choleris *et al.*, 2001; Toye and Cox, 2001; Rodgers *et al.*, 1997; Crawley *et al.*, 1997). However, Rodgers *et al.* (1997) highlighted the fact that many of the conventional measures of animal anxiety were validated within pharmaceutical contexts. This contextually confirmed validity has certain implications for the interpretation of conventional measures of fear- and anxiety-related behaviours, when applied in divergent settings (Rodgers *et al.*, 1997).

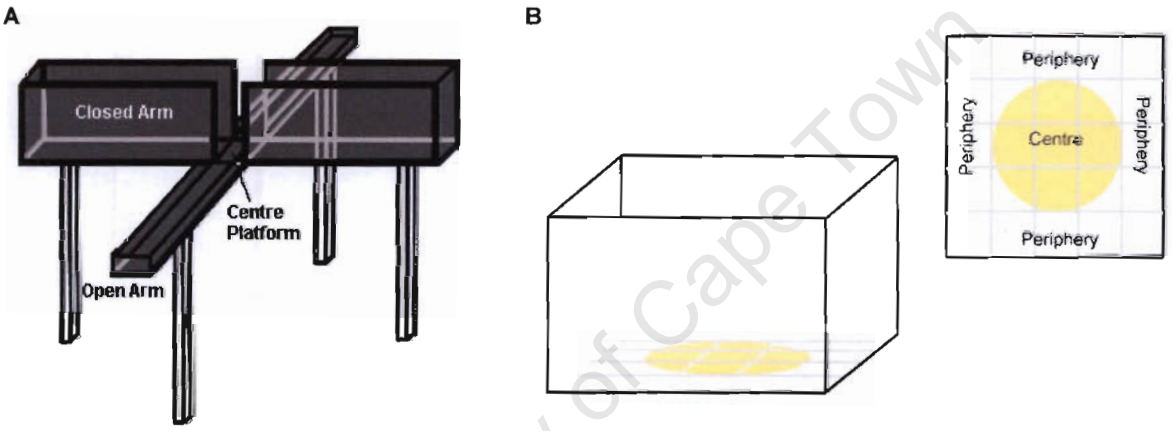

Figure 2.2: Schematic illustrations of the EPM and OF environments. (A) The EPM is typically raised of the ground with three spatially defined areas: Open arms (x2), closed arms (x2) (in this illustration with black opaque walls) and a centre platform. (B) The OF (shown here with transparent walls) has two spatially defined areas: A centre area, which is shown here as brightly illuminated (yellow), and a periphery.

Within the last decade, elaborations and critical considerations have advocated the extension of conventional spatiotemporal exploration-based measures of anxiety, to include a larger repertoire of readily observed rodent behaviours (Choleris *et al.*, 2001; Espejo, 1997; Rodgers *et al.*, 1997). These additional parameters are referred to as ethological parameters. Typically ethological strategies include aspects of risk assessment, decision making and grooming, in addition to the conventional exploration criteria. This approach has proven to be more sensitive to

subtleties in the complex behavioural responses and adaptations of rodents within these environments (Carobrez and Bertoglio, 2005; Carola *et al.*, 2002; Choleris *et al.*, 2001; Rodgers *et al.*, 1997; Rodgers and Johnson, 1995).

In addition, several authors have noted important temporal components of rodent behaviour, seen as a change in behaviour during and between test sessions (Stern *et al.*, 2008; Carobrez and Bertoglio, 2005; Bertoglio and Carobrez, 2002; Rodgers *et al.*, 1996). It is thought that these changes reflect learning processes, which underlie important aspects contextual decision making, consequently effecting the appropriate adjustment of behaviour (Bertoglio and Carobrez, 2002). These observations highlight another important dimension for the interpretation of complex behaviours. Temporal learning and adaptation components could reflect important aspects of fear- and anxiety-responses, related to information gathering and processing.

Although there are many methodological challenges, and interpretive complexities related to anthropogenic constructs of fear- and anxiety, these paradigms have proved invaluable to preclinical research on the neurobiology of psychiatric disorders, and behavioural phenotyping of rodents.

2.4 RATIONALE FOR BEHAVIOURAL TESTS IN THIS STUDY

Maternal separation was used to set up a model of anxiety in adult mice to test the hypothesis that gene expression profiles of circulating PBMCs could be used to predict psychological states. Behavioural tests were conducted to confirm that the maternally separated and control mice, used in this microarray study, showed differences in anxiety-related behaviours. The aim was to establish, using simple

empirical criteria, whether MS resulted in differential behaviours or behavioural adaptations within the context of established behavioural paradigms. Any differences were inferred to reflect divergences in the neural mechanisms underlying fear- and anxiety-related exploration behaviour.

2.5 MATERIALS AND METHODS

The experimental methods described and followed in this study have been approved by the animal ethics committee of the University of Cape Town (Ethics clearance number: 006/007) and are in accordance with National guidelines for the care and use of laboratory animals.

2.5.1 Animals and treatment

Female C57BL/6 mice were mated in a specified pathogen free (SPF) environment, and transported to the experimental facility at least three days prior to parturition. All animals were maintained under a 12 h light-dark cycle (lights on from 6h00 to 18h00). Temperature was kept at 21 ± 2 °C. All animals had ad libitum access to sterilized food and tap water. Dams[†] and litters were kept in "Shoebox" style polypropylene cages (300mm x 130mm x 120mm) with solid floors. All animals had access to bedding material, in the form of sawdust. All cages were cleaned and water bottles were changed once per week for the duration of the experiment. All animal-human interactions were limited to a single researcher. Postnatal day (PND) 0 was assigned to litters born before 15h30 each day. Litters were randomly assigned to undergo maternal separation (MS; n = 10) or to be reared under

[†] The term "Dam" is commonly used to refer to the female parent.

standard conditions, with simulated handling (control) events (SH; n = 9). The average litter size of both MS and SH groups was equal (n = 7). Overall Male-to-Female sex ratio was 1.28:1, with there being no significant sex ratio difference between MS and SH litters (recorded at weaning, PND 21).

Maternal separation was carried out as previously described (Romeo *et al.*, 2003) with some modifications. MS litters were separated from dams for 3 h a day, starting at 12h00 (6 h after lights on) and ending at 15h00, from PND 1 to 14. The MS dams were first removed from the home cage, after which the pups were moved to a clean cage, which was kept at the ambient temperature of the vivarium. The dam was placed back in the home cage and moved to a separate room for the duration of the separation, this to exclude olfactory or ultrasound vocalization exchanges between dams and their pups. After 3 h, pups and dams were reunited in their home cage. SH animals underwent daily handling. SH dams were removed from the home cage; pups were briefly moved to a clean cage and immediately returned to their home cage, followed by the dam. This procedure simulated the handling undergone by MS pups and served as a control, never lasting for more than 5 min per litter.

At PND 21, all pups were weaned and group housed by sex and treatment. All subsequent procedures were carried out using males only, as the consequences of separation are gender specific with behaviour in females being significantly influenced by the oestrus cycle (Romeo *et al.*, 2003). Post-weaning group sizes were standardized, with a minimum of three male pups per cage (to avoid isolation) and a maximum of five male pups per cage (to avoid overcrowding). The "Shoebox" style cages were used for three-pup groups and larger "Type II long" cages, made from polycarbonate (340mm x 180mm x 140mm) with solid floors, were used for groups of

four and five, thereby providing comparable mm^2 floor areas per individual. Animals were maintained in these conditions until one week before commencement of behavioural tests at which point all males were moved to individual cages. Single housing of mice has been shown to reduce HPAA activation, and therefore basal stress levels (Dalm *et al.*, 2005).

2.5.2 Behavioural assessments

Basal levels of unconditioned, exploration-based fear- and anxiety-related behaviours were assessed using both the EPM (MS n = 36, SH n = 37) and the OF (MS n = 37, SH n = 38) tests. The EPM (Med Associates Inc., St. Albans, USA) was constructed from black Plexiglas (see Fig. 2.2A for illustration) and was set atop an aluminium stand raised 72.4 cm above the ground. Both open and closed runways measured 6.1 cm wide and 34.9 cm long. The open runways had raised edges that were 0.6 cm high. The closed runways had walls that were 20.3 cm high. The OF (Med Associates Inc., St. Albans, USA) consisted of a 27.9 cm x 27.9 cm arena, with transparent walls (see Fig. 2.2B for illustration) measuring 20.3 cm high, all constructed from transparent Plexiglas. This environment was modified with a 5 x 5 grid of equal sized blocks, drawn on laminate, which was stuck on the floor of the arena. The centre of this arena was defined as the inner 3 x 3 blocks, with a perimeter consisting of 16 blocks. All behavioural tests were conducted for 5 min per individual and animals were exposed only once to each environment. A 5 min test interval was chosen, as previous studies have noted a decrease in avoidance behaviour and an increase in fatigue, after 5 to 10 min in rats (Daniels *et al.*, 2004). All experiments were recorded with a video camera, which was mounted directly above the test environment, facilitating manual scoring at a later time. The

experimenter was not present in the room during the tests. Each of the two behavioural tests was conducted over five days during a 4.5 h window, 7h30 to 12h00, with three individual sessions all lasting ± 50 min. This window was defined to control for circadian fluctuations in stress hormones and associated basal stress/anxiety levels, which are at their lowest during this part of the defined light-dark cycle (Dalm *et al.*, 2005). The environments were thoroughly wiped down with 10% ethanol between every test, ensuring that the scents from previous individuals were diminished. A low concentration of ethanol was used to avoid possible ethanol induced behavioural phenomena. Gloves were worn throughout the experiments, and changed after every session. The EPM tests were conducted before OF exposure, with 1 week between the two tests. For the EPM tests, individuals were randomly selected each day and allocated to one of the three sessions, this allocation roster was then used to assign test slots to individuals in OF tests, i.e. individual OF tests were done in the same sequence as the EPM tests ensuring a standardized time between EPM and OF exposure for each individual. Individuals were moved to the testing area ± 15 min before the commencement of tests. EPM tests were conducted using low intensity illumination, with two 15 W tungsten lights providing general illumination to the room. In contrast, a 120 W spotlight was placed vertically above the OF and focused directly at the centre of the arena; this was the only illumination in the room.

Various ethological behaviours were scored in addition to conventional spatiotemporal components. Analysed conventional EPM parameters included (see Table 2.1 for definitions): (1) Total Arm Entries (tAE); (2) % Open Arm Entries (%OAE); (3) Open Arm Time (OAT); (4) Closed Arm Time (CAT); and (5) Centre Time (CT). Due to the importance of thigmotactic cues in this environment,

ethological parameters were categorized as either protected (occurring on the closed arm or central square) or unprotected (occurring on the open arm) and were analysed as: (1) Risk assessment parameters, (a) total Stretch Attend Postures (tSAP) and (b) % protected Stretch Attend Postures (%pSAP); and (2) Exploratory scanning behaviour (a) total Head Dips (tDIP) and (b) % protected Head Dips (%pDIP). As a consequence of using C57BL/6 mice, contrast on the black Plexiglas EPM was too low to score grooming and rearing behaviours accurately.

OF parameters were similarly categorized, with the analysed conventional components comprising (see Table 2.2 for definitions): (1) Centre exploration, (a) Centre Entry Delay (CED); (b) Centre Entries (CE) and (c) Centre Time (CT); and (2) peripheral Returns (pRet). Due to the small size of the OF and unavailability of an automated tracking system, common locomotor-related variables (distance travelled and squares crossed) were not scored. The ethological parameters included: (1) Risk assessment behaviour, (a) total Stretch Attend Postures (tSAP), (b) centre Stretch Attend Postures (cSAP) and (c) peripheral Stretch Attend Postures (pSAP); (2) Grooming behaviour, (a) total Grooms (tGr); (5) centre Grooms (cGr); (6) peripheral Grooms (pGr); and (3) Rearing behaviour, (a) total Rears (tRear); (b) centre Rears (cRear) and (c) peripheral Rears (pRear).

Independent scoring was done by a researcher experienced in rodent behaviour and completely blind to treatment allocations. All parameters, except CED in the OF, were scored using min intervals, defined as a time bin (5 x 1 min bins). This allowed for the analysis of temporal adaptations in behaviour, by comparing behavioural transitions between successive time bins or differences during specific time bins.

Table 2.1 EPM parameters and definitions

	Parameters	Abbrev	Description
Conventional Spatiotemporal Parameters	Total Arm Entries	***tAE***	Total number of Open + Closed arm entries - A measure of overall locomotor activity.
	% Open Arm Entries	***%OAE***	[Open Arm Entries / Total Arm Entries] * 100 - A relative measure of Open Arm exploration, corrected for possible differences in overall locomotor activity.
	Open Arm Time	***OAT***	Total time spent on Open arm
	Closed Arm Time	***CAT***	Total time spent on Closed arm
	Centre Time	***CT***	Total time spent on Centre square
Ethological Parameters	Total Stretch Attend Postures	***tSAP***	Risk assessment behaviour, characterised by a flattened back and the forward elongation of head and shoulders, followed by retraction to original position without any locomotion.
	% Protected Stretch Attend Postures	***%pSAP***	Stretch Attend Postures from a closed arm or centre square, expressed relative to total SAP.
	Total Head Dips	***tDIP***	Exploratory scanning behaviour, characterised by scanning over the sides of the maze towards the floor.
	% Protected Head Dips	***%pDIP***	Head Dip occurring from a closed arm or centre square, expressed relative to total Head Dips.

Stringent criteria were used to quantify movement between defined spatial boundaries within the environments. For the EPM tests, the following parameters were standardised: (1) All mice were placed facing an open arm; (2) for an entry to occur an exit must have taken place; (3) both an entry and an exit occurred once all four paws crossed the entrance to an arm. For the OF tests, similar criteria were used: (1) All mice were placed in the same corner block with their backs facing the centre of the field; (2) for an entry to occur an exit must have taken place; (3) both an entry and an exit occurred once all four paws crossed the boundary defined as the

centre area. The average age of mice at the time of EPM and OF tests was 70 (range = 65 - 74) and 77 (range = 72 - 81) days, respectively.

Table 2.2 OF parameters and definitions

	Parameters	Abbrev	Description
Conventional Spatiotemporal Parameters	Centre Entry Delay	***CED***	The time taken from start of test to first centre entry.
	Centre Entries	***CE***	Total number of Centre Entries.
	Centre Time	***CT***	Total time spent in Centre area.
	Peripheral Returns	***pRet***	Entry into the Centre area with front paws only, followed by a retraction to the periphery.
Ethological Parameters	Total Stretch Attend Postures	***tSAP***	Defined as for the EPM. Here these behaviours are additionally subdivided into SAPs within centre (**c*SAP***) SAPs within Periphery (**p*SAP***).
	Total Grooms	***tGr***	Scratching or licking of fur, washing of face or licking of genitalia. Again a spatial subdivision was used to define both Centre Grooms (***cGr***) and Peripheral Grooms (***pGr***).
	Total Rears	***tRear***	Vertical exploration, characterised by an upright posture with front paws moving in the air or leaning against the wall of the environment. Centre Rears (***cRear***) and Peripheral Rears (***pRear***), were defined as above.

2.5.3 Statistical analyses

All statistical analyses were done using Statistica 8.0 (StatSoft Inc., USA.). All parameters, except CED in the OF, were analysed using a three-factor (*Treatment* [2] x *Test Day* [5] x *Time bin* [5]) repeated measures ANOVA; where *Treatment* and *Test Day* were between-subject factors and *Time bin* the within-subject factor. This strategy allowed for the assessment of time-dependent adjustments in behaviour during test sessions. In addition, main- and interaction-effects of *Test Day* were

included to highlight and account for sources of variability that arose from unanticipated or uncontrollable procedural variables. CED was analysed using a two-factor (*Treatment* [2] x *Test Day* [5]) ANOVA. Mauchly's test (Mauchly, 1940) was used to assess sphericity. Where data failed to meet this assumption, the Wilk's lambda (Mardia *et al., 1979)* multivariate approach, which does not rely on sphericity, was used to evaluate repeated measures. Fisher LSD post-hoc tests were used to confirm a main effect of *Time bin* on behaviour, in addition to evaluating any significant main effects of *Test Day* and *Treatment* x *Test Day* interactions. Planned comparisons using simple *Treatment* x *Time bin* contrasts were done to evaluate behavioural transitions, or differential changes, between successive time bins, as well as differences at each time bin. The homogeneity of variance was confirmed for all significant planned comparisons. Significance was assumed at $P < 0.05$. All graphs are shown as *Means ± 95% confidence intervals (CIs).*

2.6 RESULTS

2.6.1 Elevated Plus Maze

The results of the three-factor (*Treatment* x *Test Day* x *Time bin*) repeated measure ANOVA analyses for the EPM are shown in Table 2.3 (conventional spatiotemporal parameters) and Table 2.4 (ethological parameters). The analyses revealed no main effects of *Treatment* or *Test Day* on behaviour. In contrast, all parameters except tAE and %pDIP showed a significant effect of *Time bin* (see Fig 2.3B for a representative profile). The only parameter not significantly modified by any main- or interaction effects was tAE, confirming comparable and stable locomotor activities between MS and SH individuals. Three parameters showed significant interaction

effects between *Treatment* and *Time bin* and included: OAT, %OAE and %pSAP. In addition, OAT, %OAE, CAT as well as %pDIP behaviours all showed a significant *Treatment* x *Test Day* interaction. Significant *Time bin* x *Test Day* interactions were observed for all conventional exploration parameters except tAE. The only ethological parameter that showed a significant interaction effect for these factors was tDIP. A significant three-way *Treatment* x *Test Day* x *Time bin* effect was exclusively seen for CT.

Table 2.3 Three Factor (*Treatment* x *Test Day* x *Time bin*) repeated measure ANOVA results from a minute-to-minute analysis of the EPM showing main effects and interactions for conventional spatiotemporal parameters

Parameters / Factors	**Open Arm Time (OAT)**	**% Open Arm Entries (%OAE)**	**Closed Arm Time (CAT)**	**Centre Time (CT)**	**Total Arm Entries (tAE)**
Treatment (df = 1, 63)	F = 1.212	F = 0.381	F= 1.057	F= 0.002	F = 0.723
Day (df = 4, 63)	F = 2.217	F = 1.583	F= 1.908	F= 1.799	F = 1.005
Treatment*Day (df = 4, 63)	**F = 4.338**[b]	**F = 5.307**[c]	**F= 3.389**[a]	F= 1.788	F = 1.131
Time (df = 4, 252)	**F = 4.426**[b]	**F = 5.071**[c]	**F= 14.361**[c]	**F= 7.747**[c]	F = 1.187
Time*Treatment (df = 4, 252)	**F = 3.626**[b]	**F = 2.566**[a]	F= 1.784	F= 1.541	F = 0.516
Time*Day (df = 16, 252)	**F = 2.223**[b]	**F = 1.990**[a]	**F= 2.473**[b]	**F= 2.215**[b]	F = 0.663
Time*Day*Treatment (df = 16, 252)	F = 1.408	F = 1.434	F= 1.25	**F= 2.614**[c]	F= 0.919

df=degrees of freedom ; **a** P < 0.05 ; **b** P < 0.01 ; **c** P < 0.001

Table 2.4 Three Factor (*Treatment* x *Test Day* x *Time bin*) repeated measure ANOVA results from a minute-to-minute analysis of the EPM showing main effects and interactions for ethological parameters

Parameters / Factors	**Total Stretch Attend Postures (tSAP)**	**% Protected Stretch Attend Postures (%pSAP)**	**Total Head Dips (tDIP)**	**% Protected Head Dips (%pDIP)[§]**
Treatment (df = 1, 63)	F= 1.170	F= 0.051	F= 0.022	F= 0.666
Day (df = 4, 63)	F= 1.5791	F= 0.626	F= 2.315	F= 1.752
Treatment*Day (df = 4, 63)	F= 1.883	F= 1.031	F= 1.872	**F= 2.954[a]**
Time (df = 4, 252)	**F= 31.19[c]**	**F= 2.700[a]**	**F= 13.171[c]**	F= 2.017
Time*Treatment (df = 4, 252)	F= 0.212	**F= 3.468[b]**	F= 0.897	F= 1.418
Time*Day (df = 16, 252)	F= 0.783	F= 1.036	**F= 2.616[c]**	F= 1.267
Time*Day*Treatment (df = 16, 252)	F= 1.012	F= 0.986	F= 0.8799	F= 1.033

df=degrees of freedom ; **a** P < 0.05 ; **b** P < 0.01 ; **c** P < 0.001; **§** Failed Mauchly's test of sphericity

Interactions for *Treatment* x *Time bin* were further evaluated using planned contrasts. Contrasts were performed to evaluate transitional differences, or differential changes, between successive time bins (Tables 2.5 and 2.6), in addition to differences at specific time bins. These comparisons revealed that *Treatment* x *Time bin* interactions resulted in significant differential changes for OAT, %OAE, CT, %pSAP and %pDIP from Min 4 – 5. MS animals showed a sharp decrease in open arm exploration, measured as OAT (Fig. 2.3C) and %OAE (Fig. 2.3E), during the 5th min relative to the 4th min, in both duration and frequency. SH animals, in contrast, exhibited an opposite response. The inverse was seen for the duration of CT and frequency of %pSAP (Fig. 2.4B) and %pDIP (Fig. 2.4D) parameters, where MS animals showed a marked increase, during the 5th min relative to the 4th min. Additionally, OAT and %pSAP showed significant differences between groups at

specific time bins; at Min 4 and 5 for OAT (Fig. 2.3C) and Min 5 only for %pSAP (Fig. 2.4B).

Table 2.5 *Treatment* x *Time bin* contrasts of EPM behavioural changes between consecutive time bins for locomotor and exploration parameters

Time interval	Open Arm Time (OAT)	% Open Arm Entries (%OAE)	Closed Arm Time (CAT)	Centre Time (CT)	Total Arm Entries (tAE)
Min 1 – 2 (df = 1, 63)	F = 0.006	F = 0.041	F < 0.001	F = 0.005	F = 0.546
Min 2 – 3 (df = 1, 63)	F = 0.345	F = 0.001	F = 0.321	F = 0.034	F = 1.284
Min 3 – 4 (df = 1, 63)	F = 0.430	F = 0.024	F = 0.374	F = 3.453	F = 0.001
Min 4 – 5 (df = 1, 63)	**F = 14.011**[c]	**F = 6.160**[a]	F = 3.052	**F = 4.879**[a]	F = 0.111

df=degrees of freedom ; **a** P < 0.05 ; **b** P < 0.01 ; **c** P < 0.001

Table 2.6 *Treatment* x *Time bin* contrasts of EPM behavioural transitions between consecutive time bins for ethological parameters

Time interval	Total Stretch Attend Postures (tSAP)	% Protected Stretch Attend Postures (%pSAP)	Total Head Dips (tDIP)	% Protected Head Dips (%pDIP)[§]
Min 1 – 2 (df = 1, 63)	F = 0.144	F = 0.262	F = 0.018	F = 0.917
Min 2 – 3 (df = 1, 63)	F = 0.497	F = 0.964	F = 0.142	F = 0.008
Min 3 – 4 (df = 1, 63)	F = 0.018	F = 3.571	F = 0.037	F = 0.304
Min 4 – 5 (df = 1, 63)	F = 0.497	**F = 9.292**[b]	F = 3.149	**F = 4.098**[a]

df=degrees of freedom ; **a** P < 0.05 ; **b** P < 0.01 ; **c** P < 0.001

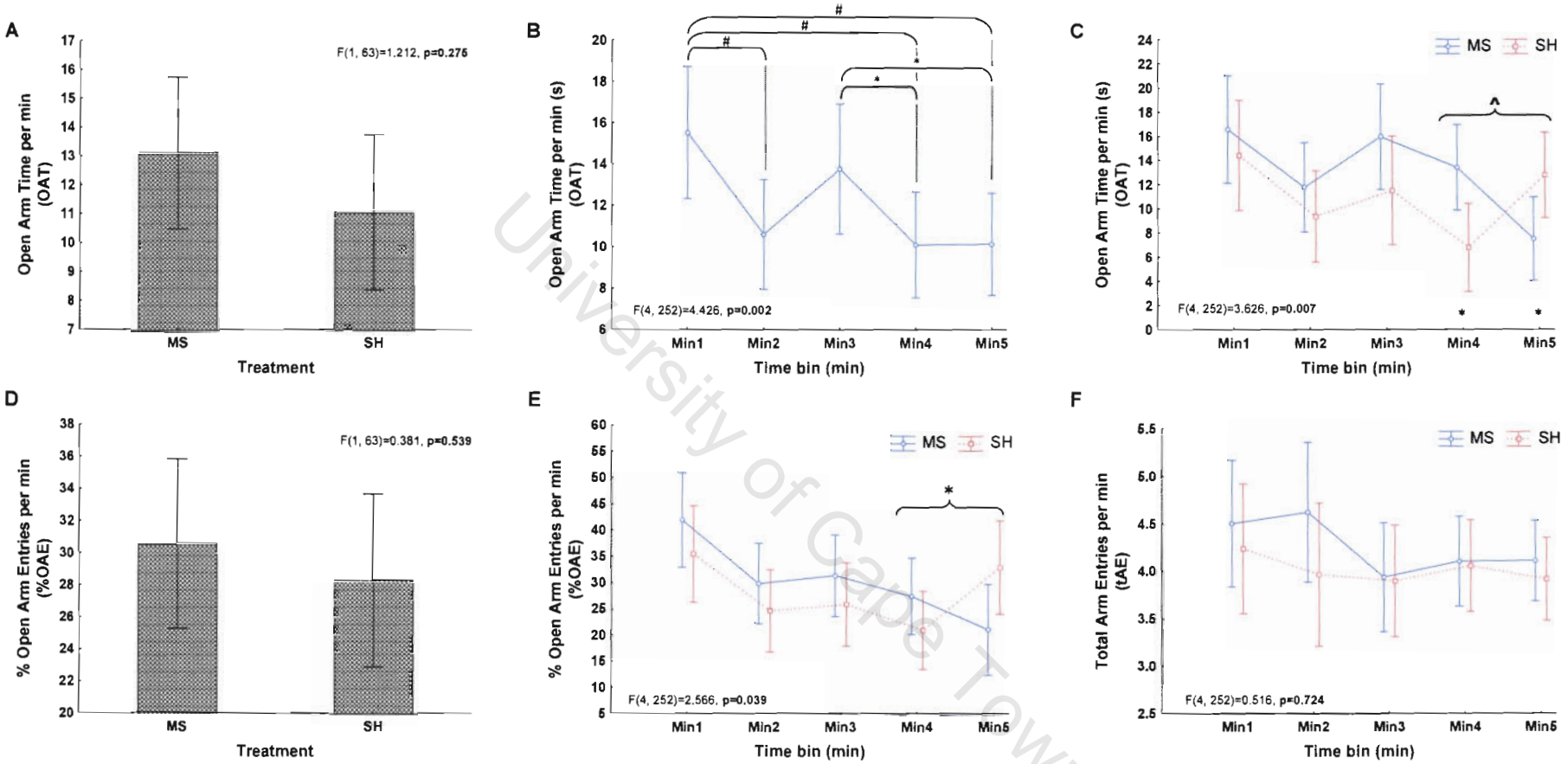

Figure 2.3: EPM open arm exploration parameters. Main effects of (A) *Treatment* on OAT and (B) *Time-bin* on OAT, (C) Interaction profiles of *Treatment* x *Time bin* for OAT, (D) Main effects of *Treatment* on %OAE and Interaction profiles of *Treatment* x *time bin* for (E) %OAE and (F) tAE. Treatment-independent significant differences between specific time bins are indicated as: ⌐¬ with *P < 0.05, #P < 0.01; Significant Treatment x Time bin interaction effects between successive time bins are indicated as: ⌐¬ with *P < 0.05 and ^P < 0.001; Significant *Treatment* x *Time bin* interaction effects at specific time bins are indicated as *P < 0.05, #P < 0.01 and ^P < 0.001. Values are Means ± 95% CIs.

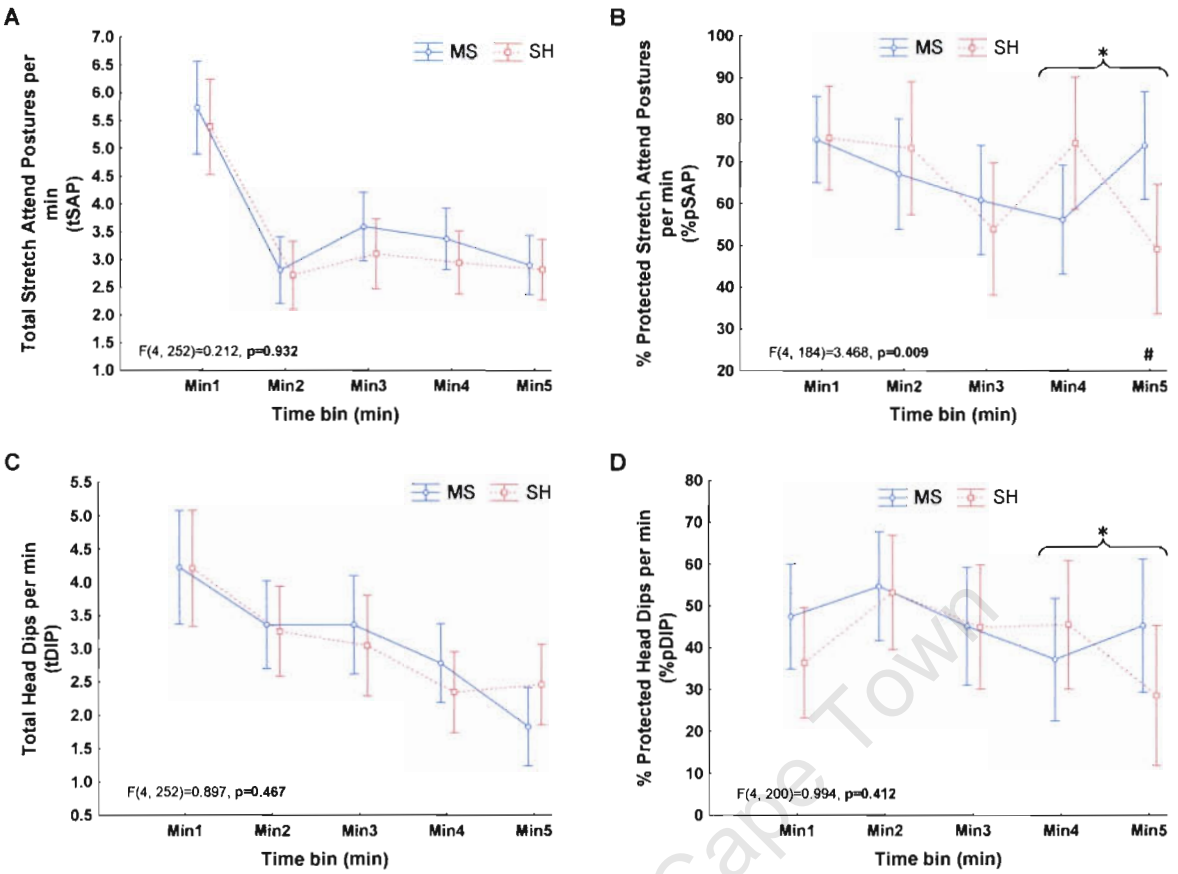

Figure 2.4: *Treatment* x *Time bin* interaction effects for EPM ethological parameters. Interaction profiles of *Treatment* x *Time bin* for (A) tSAP, (B) pSAP, (C) tDIP and (D) pDIP. Treatment-dependent significant differences between successive time bins are indicated as: ⌐¬ with *P < 0.05, #P < 0.01 and ^P < 0.001; Significant differences at specific time bins are indicated as #P < 0.01. Values are Means ± 95% CIs.

Where significant *Treatment* x *Test Day* interaction effects were observed (OAT, CAT, %OAE and %pDIP), Fisher LSD post-hoc tests were used to further evaluate this interaction at each test day only (Table 2.7). These tests showed significant between-group differences, exclusively for Test Day 1 (see Fig. 2.5 for representative profiles). MS individuals showed significantly higher open arm exploration, OAT and %OAE, behaviours relative to their SH counterparts, on Day 1. In contrast, the inverse was true for CAT duration and %pDIP.

Table 2.7 Fisher LSD post hoc results for *Treatment* x *Test Day* interaction effects at each test day

Test Day	Open Arm Time (OAT)	Closed Arm Time (CAT)	% Open Arm Entries (%OAE)	% Protected Head Dips (%pDIP)
Day 1	**P < 0.001**	**P < 0.01**	**P < 0.001**	**P < 0.05**
Day 2	NS	NS	NS	NS
Day 3	NS	NS	NS	NS
Day 4	NS	NS	NS	NS
Day 5	NS	NS	NS	NS

NS=Not significant

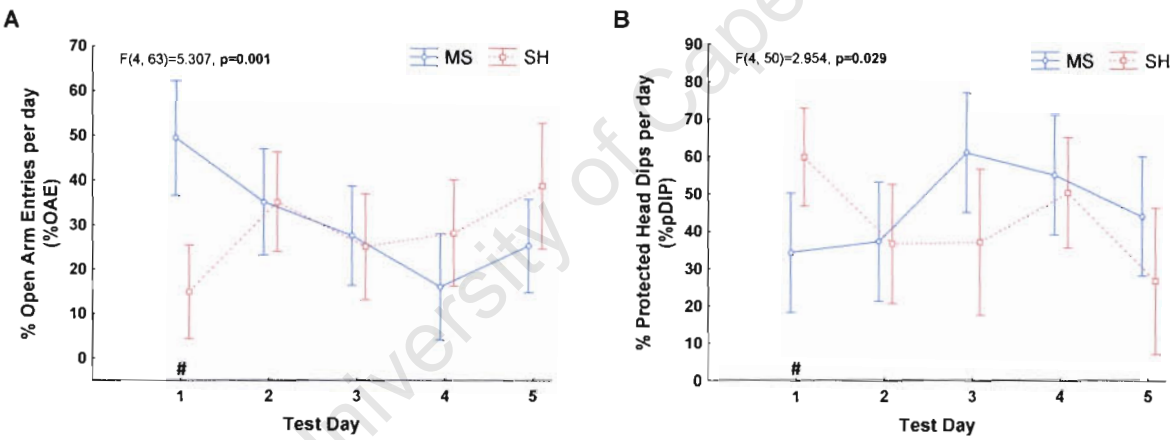

Figure 2.5: Examples of EPM *Treatment* x *Test Day* interaction effects. Interaction profiles of *Treatment* x *Test Day* for (A) %OAE and (B) %pDIP. Significant differences at specific days are indicated as *P < 0.05, #P < 0.01 and ^P < 0.001. Values are Means ± 95% CIs.

A Fisher's LSD post-hoc test dissection of significant *Test Day* x *Time bin* interactions were employed to evaluate the contribution of specific test day differences to significant *Treatment* x *Time bin* interactions. Results confirmed no significant effect of *Test Day* on any significant *Treatment* x *Time bin* interactions.

2.6.2 Open Field

The results of the three-factor (*Treatment* x *Test Day* x *Time bin*) repeated measure ANOVA analyses for the OF are shown in Table 2.8 (conventional spatiotemporal parameters) and Table 2.9 (ethological parameters). A significant main effect of *Time bin* was seen for all conventional and ethological behavioural parameters, independent of treatment (see Fig. 2.6A for a representative profile). None of the conventional exploration behaviours were modified by the *Treatment* factor. In contrast, a significant main effect of *Treatment* was seen for risk assessment behaviours, measured as tSAP, cSAP and pSAP (Table 2.9). MS individuals consistently showed significantly higher frequencies of both cSAP (Fig. 2.7C) and pSAP (Fig. 2.7D) behaviours, compared to their SH counterparts. Figure 2.7A shows that MS individuals exhibited significantly higher frequencies of tSAPs during the last 4 min of the each test session. Additionally, *Test Day* had a significant effect on rearing behaviours. Fisher LSD post-hoc tests confirmed significant differences in the frequency of tRear and pRear between test days (Results not shown).

In contrast to EPM behaviours, no OF behaviours were significantly modified as a consequence of *Treatment* x *Time bin* interactions (see Fig. 2.6B, Fig. 2.7A and Fig 2.8A and B for representative profiles). However, planned comparisons revealed a significant difference in the frequency of Total Grooms during the 1st min between MS and SH individuals (Fig. 2.8A).

Table 2.8 Three Factor (*Treatment* x *Test Day* x *Time bin*) repeated measure ANOVA results from a minute-to-minute analysis of the OF showing main effects and interactions for centre exploration and peripheral return parameters

Parameters / Factor	**Centre Entry Delay (CED)**	**Centre Entries (CE)[§]**	**Centre Time (CT)[§]**	**Peripheral Returns (pRet)**
Treatment (df = 1, 65)	F= 0.157	F= 0.241	F= 0.026	F= 0.002
Day (df = 4, 65)	F= 0.837	F= 0.871	F= 1.016	F= 1.146
Treatment*Day (df = 4, 65)	F= 0.061	F= 1.329	F= 1.984	F= 1.819
Time (df = 4, 260)	N/A	**F= 7.371[c]**	**F= 3.042[a]**	F= **5.341[c]**
Time*Treatment (df = 4, 260)	N/A	F= 0.703	F= 0.953	F= 1.229
Time*Day (df = 16, 260)	N/A	F= 0.457	F= 1.246	F= 0.696
Time*Day*Treatment (df = 16, 260)	N/A	F= 1.184	F= 1.496	F= 0.876

df=degrees of freedom ; **a** P < 0.05 ; **b** P < 0.01 ; **c** P < 0.001 ; **§** Failed Mauchly's test of sphericity ; N/A=Not Applicable

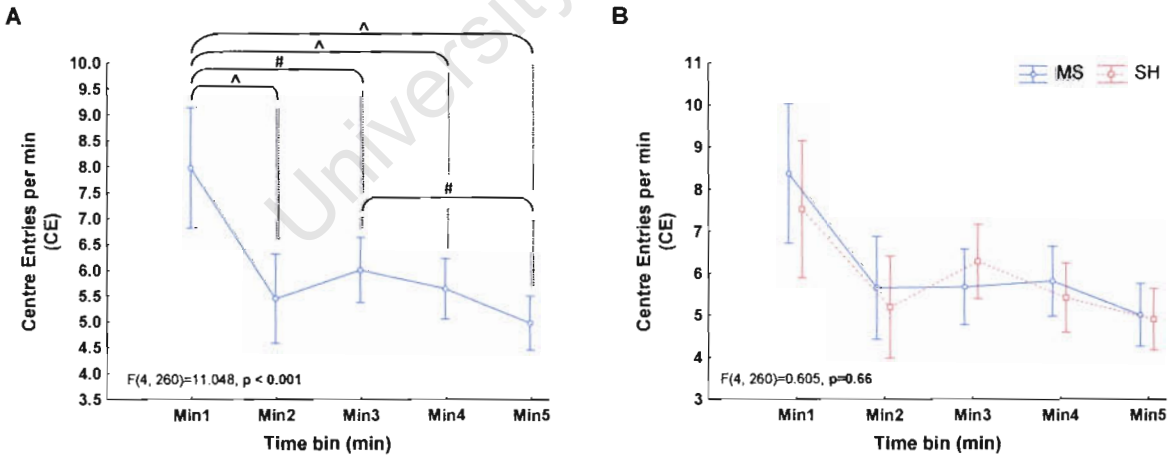

Figure 2.6: OF centre area exploration parameters. (A) Main effect of *Time bin* on CE, (B) interaction profiles of *Treatment* x *Time bin* for CE. Significant differences between specific time bins are indicated as: ⌐¬ with #P < 0.01 and ^P < 0.001. Values are Means ± 95% CIs.

Table 2.9 Three Factor (*Treatment* x *Test Day* x *Time bin*) repeated measure ANOVA results from minute-to-minute analysis of the EPM showing main effects and interactions for ethological parameters

Parameters / Factor	**Total Stretch Attend Postures (tSAP)[§]**	**Centre Stretch Attend Postures (cSAP)**	**Peripheral Stretch Attend Postures (pSAP)[§]**	**Total Grooms (tGr)[§]**	**Centre Grooms (CGr)[§]**	**Peripheral Grooms (PGr)[§]**	**Total Rears (tRear)[§]**	**Centre Rears (cRear)[§]**	**Peripheral Rears (pRear)[§]**
Treatment (df = 1, 65)	**F = 10.728[b]**	**F = 10.508[b]**	**F = 7.827[b]**	F = 2.002	F = 0.482	F = 1.530	F = 0.180	F = 0.052	F = 0.178
Day (df = 4, 65)	F = 1.186	F = 0.876	F = 1.010	F = 0.958	F = 0.545	F = 1.508	**F = 4.162[b]**	F =1.071	**F = 4.364[b]**
Treatment*Day (df = 4, 65)	F = 2.323	F = 0.635	F = 2.44	F = 0.194	F = 0.813	F = 0.08	F = 0.810	F = 1.683	F = 0.676
Time (df = 4, 260)	**F = 64.412[c]**	**F = 25.613[c]**	**F = 51.535[c]**	**F = 67.912[c]**	**F = 7.224[c]**	**F = 44.763[c]**	**F = 24.700[c]**	**F = 3.646[b]**	**F = 22.282[c]**
Time*Treatment (df = 4, 260)	F = 1.179	F = 0.668	F = 1.390	F = 1.770	F = 1.338	F = 0.978	F = 0.330	F = 0.136	F = 0.418
Time*Day (df = 16, 260)	F = 1.105	F = 0.462	F = 1.009	F = 0.823	F =1.545	F = 0.791	F = 1.271	F = 1.043	F = 1.277
Time*Day*Treatment (df = 16, 260)	F = 0.517	F = 0.625	F = 0.382	F = 1.064	F = 0.687	F = 1.247	F = 1.242	F = 1.478	F = 1.126

df=degrees of freedom ; **a** P < 0.05 ; **b** P < 0.01 ; **c** P < 0.001; **§** Failed Mauchly's test of sphericity

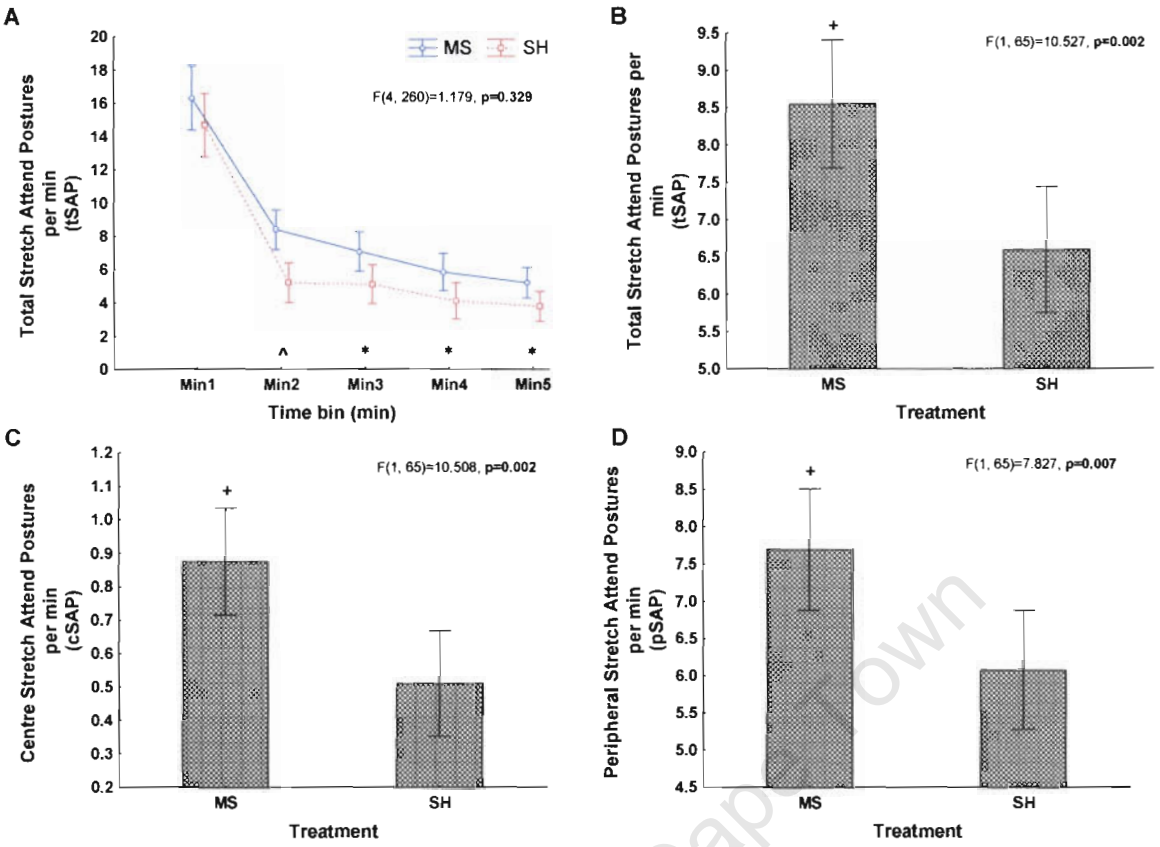

Figure 2.7: OF risk assessment behaviours. (A) Interaction profiles of *Treatment* x *Time bin* for tSAP, main effect of *Treatment* on (B) tSAP and (C) cSAP, and (D) pSAP. Significant differences at specific time bins are indicated as *P < 0.05 and ^P < 0.001. A significant main effect of *Treatment* is indicated as +P < 0.01. Values are Means ± 95% CIs.

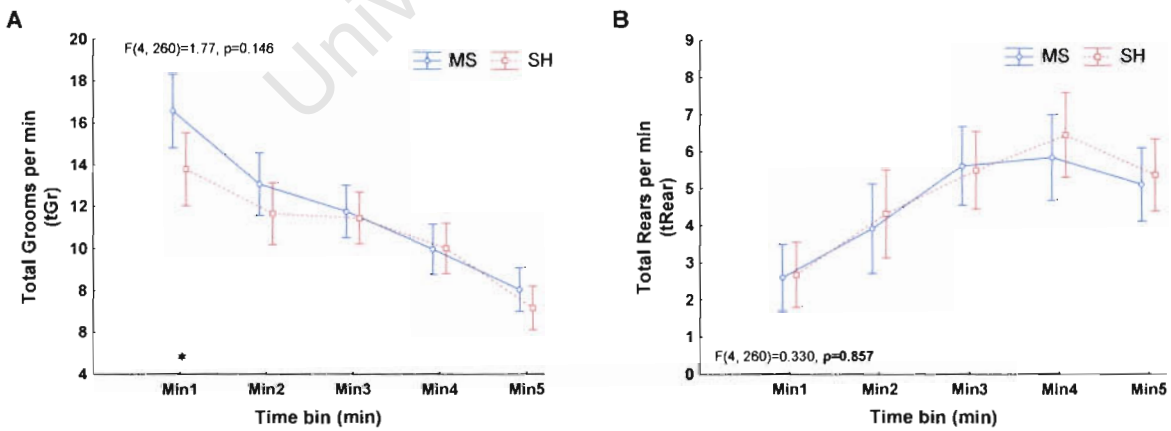

Figure 2.8: Examples of OF Treatment x Time bin interactions. Interaction profiles of Treatment x time bin for (A) tGr and (B) tRear. Significant differences at specific time bins are indicated as *P < 0.05. Values are Means ± 95% CIs.

2.7 DISCUSSION

The effects of MS on adult mouse behaviour, was evaluated using the EPM and OF tests. Several fear- and anxiety-related parameters were found to be significantly altered in response to MS. The findings presented here add to and extend current interpretations of mouse MS models, by highlighting important time-dependent aspects of mouse behaviour. The assumptions and expectations of MS outcomes in mice have mainly been informed by a large body of work done on rats, which has been extensively characterised and shown to have robust effects on common indices of stress and anxiety. Although a few studies have reported similar outcomes for MS in mice, the effects have been more difficult to consistently reproduce. Here it is shown, that the extension of conventional analysis strategies, to include time as a significant dimension of behaviour, can reveal intricate behavioural nuances, which reflect adaptive aspects of stress-driven behaviours. Context specific adjustments of behaviour are associated with neural mechanisms underlying learning and memory, which is critically affected by stress, and therefore potentially informative in behavioural interpretations of the effects of MS.

In this study, MS produced long lasting effects on adult behaviour, which were not consistently apparent when evaluated with the statistical expectations inherited from rat studies. In the EPM, conventional statistical strategies, which typically evaluate behaviour as a summed test session total, were not sufficient to reveal more complex time-dependent behavioural profiles. Specifically, the spatially more complex nature of the EPM test underscored treatment-related differences in the time-dependent adjustments of anxiogenically loaded open arm exploration and risk assessment behaviours. In contrast, the relative spatial simplicity and highly aversive

nature of the OF, elicited an immediate and consistent divergence in Risk assessment behaviours, between MS and SH individuals.

2.7.1 Elevated Plus Maze

In this study, total exploration, measured as tAE (Fig. 2.3F), was not significantly affected by any main or interaction effects, confirming that MS did not alter locomotor activity.

None of the reported parameters were significantly affected by treatment only (see Fig. 2.3A and D for representative profiles). Two recent publications also failed to describe significant MS-associated behavioural alterations of specific EPM parameters, in mice (Millstein and Holmes, 2007; Parfitt *et al.*, 2007); contrasting the robust behavioural outcomes extensively documented for rats (Parfitt *et al.*, 2004), and isolated mouse studies (Veenema *et al.*, 2007; Romeo *et al.*, 2003). The difficulty with which anticipated MS outcomes are reproduced in mouse models, bring into question the validity of current experimental implementations or interpretations. Critically, expectations of MS outcomes are largely determined by observations described in rats; possibly limiting the scope for interpretations within mouse settings. Additionally, conventional statistical strategies, typically evaluate behaviour as a summed test-session total, thereby neglecting potentially important aspects of complex behaviours.

The analytical strategy employed in this study, highlighted the importance of a temporal dimension in complex behaviour; emphasized by significant changes over time, independent of treatment, for all parameters except tAE and %pDIP. This observation is in concordance with previous studies that have highlighted significant

changes in EPM behaviour, as a consequence of learning and adaptation, during and between repeated test sessions (Stern *et al.*, 2008; Carobrez and Bertoglio, 2005; Bertoglio and Carobrez, 2002; Rodgers *et al.*, 1996). The logical extension of treatment-dependent observations, taking cognisance of time as a significant behavioural dimension, revealed treatment-specific temporal profiles for several parameters. Importantly, these interactions (*Treatment* x *Time bin*) were significant for variables that are generally taken to be important indices of fear- and anxiety-driven exploration behaviour; open arm exploration (OAT and %OAE), which correlates negatively with anxiety (i.e. increased anxiety leads to decreased open arm exploration) and risk assessment (%pSAP), which correlates positively with anxiety (Espejo, 1997; Rodgers and Johnson, 1995). The significance of these interactions could be attributed, specifically, to differential changes in these behaviours from min 4 – 5 of the test sessions (Tables 2.5 and 2.6), which highlighted a sudden increase in anxiety-related behaviour in the MS group; seen as decreased open arm exploration (Fig. 2.3C and E) and increased risk assessment (Fig. 2.4B and D). Additionally, both CT and %pDIP behaviours exhibited significant differential changes between min 4 and min 5 (Tables 2.5 and 2.6), although no significant overall interaction between treatment and time was observed for these parameters (Tables 2.3 and 2.4). The importance of behavioural shifts during the last two minutes was further emphasised by significant differences between MS and SH individuals, specifically during min 4 and min 5 for OAT (Fig. 2.3C) and min 5 for %pSAP (Fig. 2.4B). Fisher's post-hoc tests confirmed that none of the significant *Treatment* x *Time bin* interactions resulted as a consequence of any specific test days. Together these observations reveal significant time-dependent adjustments in

important fear- and anxiety-driven behaviours, highlighting differences, specifically, in stress-related adaptation strategies between the two groups.

An important aspect of behavioural adaptation involves the processing and contextual appraisal of environmental stimuli (Alexopoulos and Ric, 2007; Buckley *et al.*, 2000). The acquisition, consolidation and evaluation of environmental cues, underlie important aspects of learning and memory, which in turn is intimately link to emotion (Labar and Cabeza, 2006; Abel and Lattal, 2001). Learning processes have been shown to be a significant component of EPM fear- and anxiety-related exploration behaviours (Rodgers *et al.*, 1996), with specific time-dependent properties which can be altered through pharmaceutical manipulations of mechanisms involved in information processing and learning (Stern *et al.*, 2008, Carobrez and Bertoglio, 2005). Recently, a study by Rice and colleagues (2008) characterised specific deficits in learning and memory consolidation in adult male C57BL/6 mice, resulting from early disruptions of mother-infant interactions, in a variant early life stress model. These deficits were suggested to be as a result of impaired hippocampal function and alterations in neuroendocrine components. Similar observations were made by Fabricius *et al.* (2008), describing significant MS-associated alterations in hippocampal neuron numbers and spatial learning strategies in mice.

The significant modification of treatment by test day, observed for OAT, %OAE, CAT and %pDIP parameters, reveal yet another dimension of differential behavioural responses, in this case to unanticipated or uncontrollable environmental variables. Notably, Fisher's post-hoc LSD tests showed significant between-group differences for Day 1 of testing, only (Table 2.7). Systematic studies have shown that behavioural tests are highly sensitive to a host of environmental variables, variations

of which can have dramatic impacts on specific behavioural outcomes (Crabbe *et al.*, 1999). Although these environmental variables are difficult to identify, the most likely contribution to this observed interaction effect (*Treatment* x *Test Day*) relates to general environmental noise and activity in the two days prior to commencement of tests and those during the 5 days of testing. EPM tests were conducted over 1 week (Mon – Fri) and all animals, except for those tested on Day 1 (Monday), were exposed to a moderate amount of environmental noise and general activity in and around the animal facility. Animals tested on Day 1, would have experienced a significantly lower degree of activity and noise during the course of the weekend. Although inconsistencies between test days are generally undesirable, all animals were exposed to similar variables on each test day; therefore, this interaction effect again highlight treatment-dependent differences in behaviour. Importantly, as noted previously, test day had no statistically significant effect on any of the significant interactions between treatment and time (*Treatment* x *Time bin*).

A significant interaction between test day and time (*Test Day* x *Time bin)* for both CT and tDip highlighted a generally high degree of treatment-independent inter-day behavioural variability for specific time bins, with no implication for any of the significant treatment-dependent observations.

Lastly, a significant three-way interaction effect (*Treatment* x *Test Day* x *Time bin)* was seen for CT. However, no clear pattern was apparent, with specific significant differences between groups at min 1 on days 1 and 5, and min 2 on days 3 and 5 (Fig. 2.9). Importantly, these interactions were not associated with the significant min 4 to min 5 transition noted for CT (Table 2.5).

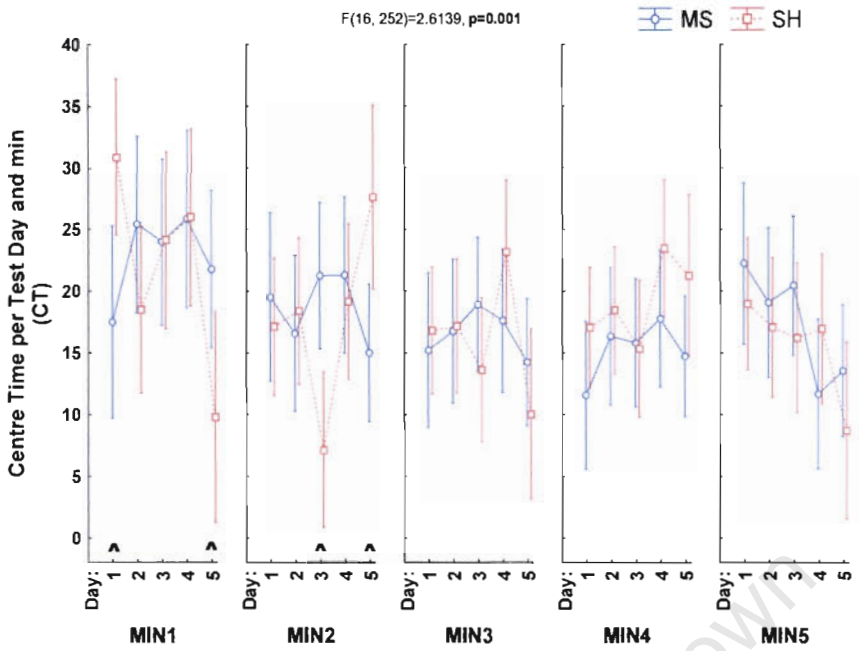

Figure 2.9: CT *Treatment* x *Test Day* x Time bin interaction effects. Significant interactions effects can be seen at min 1 on days 1 and 5, and min 2 on days 3 and 5; however, no pattern is discernable. Significant differences at specific days between groups are indicated as ^P < 0.001. Values are Means ± 95% CIs.

The time-dependent nature of open arm exploration and risk assessment parameters, in this study, reflected differences in stress-related adaptations, suggesting treatment-dependent divergences in the learning/memory mechanisms of CNS stress-associated information processing components; seen as a delayed stress response in MS animals. Several lines of evidence substantiate this interpretation. In mice, Rodgers *et al.* (1996) has described a high degree of information gathering during the first minutes of a 5 minute EPM test, which result in significant adjustments of behaviour. These within session time-dependent profiles reflect spatial learning processes, evoking increased avoidance of the potentially "dangerous" sections of the maze during the latter stages of a 5 min test (Rodgers *et al.*, 1996). Activation of the HPAA, in response to a spatial stimulus (e.g. novel environment) requires input from the hippocampus (Ziegler and Herman, 2002). Given the central role of the HPAA in driving stress- and anxiety-responses, a MS

linked deficit in hippocampus-associated information processing could result in the delayed activation of this axis; with a consequent latency in increased stress-related behaviour.

2.7.2 OF results

Consistent with EPM behaviour, all parameters changed significantly over time, independent of treatment (Tables 2.8 and 2.9). In contrast to EPM behaviour, none of the parameters, with the exception of tGrooms, showed significant interactions between treatment and time (see Fig. 2.8 for representative profiles). This interaction effect (tGrooms) was only revealed through planned comparisons and was limited exclusively to the 1st min of testing (Fig. 2.8A). Although this observation is isolated, given no other significant *Treatment* x *Time bin* interaction effects for any of the other computed parameters, grooming behaviour has been shown to reflect anxiety (Kalueff and Tuohimaa, 2005) and likely reflects an initial response to the environment (see below).

The most notable observation relates to risk assessment behaviours, where a significant treatment-dependent effect was seen for tSAP, cSAP and pSAP (Fig. 2.7B - D). Although both groups initially exhibited high levels of risk assessment, in line with previous rodent studies (Rodgers *et al.*, 1996), SAP behaviours were consistently higher (*Time bin* 1 – 5) in the MS group, with significant between-group differences from min 2 onwards (Fig. 2.7A). Importantly, this result agrees with EPM observations, which also implied a significant modification of risk assessment behaviour, albeit in a more complex manner.

Interestingly, these observations underscore key differences between the EPM and OF environments. Although systematic studies have shown overlap in stress-associated behaviours in these two environments, differences in their anxiogenic properties have been stressed (Ramos, 2008; Carola *et al.*, 2002). The EPM contains several complex spatial features (2x open arms, 2x closed arms, centre platform, elevation from floor etc.), whereas the OF present rodents with a less sophisticated choice between exposure and thigmotactic safety. In this study, the OF was significantly smaller than those used in most other studies and was modified with a brightly lit centre area (120 W spotlight illumination), a highly unpleasant stimulus for rodents; contrasted to EPM tests, which were conducted under low illumination. The small size combined with an aversion amplified centre area, presented an immediate stressor that would require limited evaluation, with a reduced dependence on spatial learning. An immediate response to environmental cues was seen in the increased risk assessment behaviours from test commencement to end (Fig. 2.7A). Significant between-group differences in grooming behaviour, during the first min of testing (Fig. 2.8A), additionally highlight an immediate response to the contextual stressors inherent in this environment; differing markedly from the more complex time-dependent responses seen in the EPM.

Additionally, test day had a significant effect on tRear and pRear behaviours. As tRear is dependent on pRear and cRear, and no effect was seen for cRear, this outcome was mainly attributable to inter-day differences in pRear behaviours. Inspection of the results show treatment-independent differences resulted mainly from an increase in pRear frequencies on Day 3 of testing (Results not shown). This observation is anomalous, as no other behavioural categories showed significant

effects of test day; combined with treatment-independence, a contextualised interpretation is not possible.

In contrast to EPM profiles, significant OF behavioural parameters were modified immediately, with risk assessment differences maintained for the duration of the test sessions. The more immediate behavioural modifications were likely due to the small size and reduced spatial complexity of this environment, combined with the highly aversive nature of the illuminated centre area.

2.8 EXTENDING CURRENT INSIGHTS – AVOIDING POTENTIAL PITFALLS

Extensive research in rats has shown that early MS can have a significant effect on stress- and anxiety-associated neuro- and endocrine mechanisms in adulthood. Behavioural studies have additionally highlighted MS-induced alterations of some of the most common indices of fear and anxiety (de Kloet *et al.*, 2005; Bakshi and Kalin, 2002; Meany, 2001; Sanchez *et al.*, 2001; Heim and Nemerof, 2001; Plotsky and Meaney, 1993). In contrast to the substantial corpus of literature describing the effects of MS on rat emotionality, there is a considerable paucity of studies describing the long-term effects on mice. Typically, in rats, a decrease in the exploration of the open arms, in the EPM, and centre area, in the OF, are the most consistently reported behavioural effects of MS. Using C57BL/6 male mice, Romeo *et al.*, (2003) reproduced these findings, showing early MS (3hrs per day, PND 1 – 14) to have long lasting and robust effects on fear- and anxiety-states of the adult males, measured as common EPM and OF parameters. The current study, however, like that of Millstein and Holmes (2007) and Parfitt *et al.* (2007), failed to replicate robust effects of MS on these conventional EPM and OF behaviours in mice.

Of contextual relevance are comments by Korte and de Boer (2002), on the robustness of EPM behavioural paradigms that assess basal levels of fear- and anxiety-related exploration behaviour. They highlight the often inconsistent and contradictory findings between laboratories. Citing as an example, an observation that differences between genetically identical inbred mice are highly correlated to testing laboratory, with many variables influencing outcome. The authors emphasized critical aspects of neural mechanisms underlying spontaneous EPM behaviour; pointing to the decisive role that cognitive appraisal processes play in determining explorative outcomes. They contrast this to fear-potentiated EPM exploration, which involve the elicitation of a fear-response prior to EPM exposure, eliminating the need for extensive contextual evaluations by animals; producing consistent and robust behavioural responses.

Recently, the disruption of mother-infant interactions in mice was shown to have profound long-term effects on spatial learning and memory (Rice *et al.*, 2008). Deficits in these processes would lead to very different behavioural profiles in environments with complex spatial features. In fact, disruptions in information processing could lead to the delayed stress responses seen in the EPM in this study. This scenario is problematic for analysis strategies that analyse behaviour as a bulk time profile, as these approaches will fail to describe time-dependent adaptations driven by the learning- and memory mechanisms associated with information processing. It is interesting to note, that although several authors have highlighted significant time-dependent changes in EPM behaviour during typical 5 min test sessions, no systematic description of the effects of MS on minute-to-minute behavioural profiles could be found. In this study, it was clearly demonstrated that EPM behaviours exhibited complex time- and treatment-interaction profiles, which

were shown to reflect significant differences in stress-related adaptive aspects of novel exploration behaviour. In contrast, OF behaviour did not reflect this treatment and time interaction. This was likely due to the reduced spatial complexity and enhanced aversiveness of this environment, eliciting a more rapid stress-response not dependent on complex spatiotemporal variables or cognitive evaluations.

The observations presented in this study raise some pertinent issues regarding the validity of current mouse MS paradigms. The difficulty with which observations from rat MS studies have been reproduced in mice (Millstein and Holmes, 2007; Parfitt *et al.*, 2007), make it clear that refinements are needed. Cryan and Holmes (2005) have noted, importantly, that "mice are not little rats", pointing to the varying degrees of success with which rat models have translated to mice. These observations have important implications for the interpretations of models such as MS, which were originally validated rats. A review of mouse MS literature shows that current implementations and behavioural assessments, using EPM and OF tests, have been adopted directly from rat studies with little or no critical revision of experimental expectations and interpretations. The behavioural parameters commonly reported are limited to a few conventional indices of stress and anxiety. Both Millstein and Holmes (2007) and Parfitt *et al.* (2007), did not report ethological parameters for the EPM and OF and EPM tests, respectively. These parameters have been shown to provide an increased resolution for the detection of alterations in important stress- and anxiety-related behaviours (Choleris *et al.*, 2001; Espejo, 1997; Rodgers *et al.*, 1997). Additionally, neither of the above two studies consider a temporal dimension of behaviour, choosing to report behaviour summed across a test session. As noted earlier, time-dependent adjustments have been shown to constitute a significant component of rodent behaviour within and between repeated

test sessions. This observation was confirmed in the current study, with most of the behavioural parameters showing significant changes across a 5 min test session, independent of treatment. Therefore it is possible that recent failures to identify any MS-associated differences in behaviour, using EPM and OF tests, is partly as a consequence of the analytical expectations inherited from rat studies. Although Millstein and Holmes (2007) have produced one of the most comprehensive assessments of MS in mice to date, it is suggested that theirs and the study by Parfitt *et al.* (2007), as well as any future mouse MS studies, might benefit from an analysis strategy that is aimed at detecting more subtle time-dependent perturbations in both conventional as well as ethological measures of behaviour.

Finally, the EPM results in the current study, suggest a delayed stress-response by MS individuals, seen as sudden changes in several anxiety-related behaviours during the last minute of these tests. Determining whether these delayed responses are as a consequence of (1) deficits in physiological components, such as HPAA mechanisms, or (2) differences in cognitive functions, is challenging when assessing behaviour at basal stress levels. Therefore, it is suggested that future studies, wishing to specifically assess physiologically driven behavioural outcomes of MS in mice might benefit from a fear-potentiation strategy, which will reduce extensive cognitive appraisals by the animals (Korte and de Boer, 2002).

In summary, MS produced differential behavioural phenotypes with complex time-dependent aspects in the EPM. Both open-arm exploration and risk assessment behaviours exhibited these profiles. Importantly, risk assessment behaviours were also significantly altered in the OF, with immediate and maintained differences between MS and SH individuals. Grooming behaviour underscored an immediate response to the OF environment. Risk assessment is a critical component of

exploration behaviour, reflecting important ethological aspects of complex responses to stress. Taken together these observations support the inference that MS modified critical stress-associated mechanisms, involving both cognitive and emotional CNS components, in the mice used in this study.

2.9 CONCLUSIONS

In this study, early MS produced long lasting effects on adult behaviour, which were not consistently apparent when time was excluded as an analytical component. In the EPM, conventional statistical strategies, which typically evaluate behaviour as a summed test session total, were not sufficient to reveal more complex time-dependent behavioural profiles. Specifically, the spatially more complex nature of the EPM test underscored treatment-related differences in the time-dependent adjustments of anxiogenically loaded open arm exploration and risk assessment behaviours. In contrast, the relative spatial simplicity and highly aversive nature of the OF, elicited an immediate and consistent divergence in risk assessment behaviours, between MS and SH individuals.

- CHAPTER 3 -

Predicting stress-related neural states using peripheral transcriptomics: Molecular descriptions of neural, endocrine and immune components

3.1 INTRODUCTION

The structural complexity of the mammalian nervous system, its dynamic nature, its integration into elaborate homeostatic networks and its in vivo inaccessibility, make the molecular dissection of neuropsychiatric disorders uniquely challenging. It is now known that psychopathological processes are multifaceted, with aetiologies characterised by cumulative contributions from large numbers of genes, involving complex interactions between several neural and non-neural components, in addition to varying degrees of dependence on non-genetic factors, such as environment. In the last decade, the simultaneous monitoring of the entire repertoire of gene products, in a specific tissue or cell population, has been made possible through (1) the completion of whole genome sequences for several organisms, including man, combined with (2) the development of high-throughput molecular technologies, such as DNA microarrays. This progress has facilitated a new generation of data-driven experimental strategies, which hold great promise for the unravelling of complex neuropsychiatric phenomena.

The application of microarray techniques in psychiatric settings is still in its infancy, facing many practical and analytical challenges. However, this strategy has begun to provide novel insights into the multi-dimensional molecular nature of complex neuropsychiatric disorders (Mirnics *et al.*, 2006). Interestingly, microarray studies have recently highlighted the possibility of non-invasively "interrogating" neural states, using peripheral tissue targets (Kawai *et al.*, 2007; Runne *et al.*, 2007;

Bowden *et al.*, 2006; Kaushik *et al.*, 2005; Segman *et al.*, 2005). This is a promising diagnostic avenue, based on the exploitation of the functional integration of neural-, endocrine- and immune-systems into a NEI-axis (see Chapter 1). Specifically, these studies have demonstrated a systematic relationship between the peripheral transcriptome of blood and defined psychiatric states. However, it remains to be established whether gene expression changes[¤] in peripheral tissue targets are paralleled by specific transcriptional alterations in neural tissues (Mirnics *et al.*, 2006). Such an undertaking is greatly complicated in human studies, where access to in vivo neural tissues is mostly impossible. In this respect, animal models provide a practical alternative which could help elaborate current insights into the diagnostic relationship between peripheral and neural transcriptomes.

In this study, it was hypothesised that gene expression profiles of circulating PBMC populations reflect specific stress-associated neural states. Using a mouse maternal separation model, microarrays were used to evaluate this relationship, by measuring, in parallel, the transcriptional profiles of PBMC and neural samples, within a stress-defined context.

3.2 THE APPEAL OF MICROARRAYS IN PSYCHIATRY

Classically, the unravelling of complex biological phenomena entails defined hypothesis-driven strategies, characterised by an iterative process of formalisation, observation and refinement. This process is generally pursued by systematically measuring and describing single components, gradually synthesising an integrated, but often limited, picture of complex processes (Wilson *et al.*, 2004). An implicit

[¤]Gene expression changes refer to changes in mRNA transcript abundance as a consequence of changes in transcription, mRNA stability or mRNA turnover.

drawback of such a strategy is that the formalisation of new hypotheses is highly dependent on existing characterisations of specific phenomena. This approach does not readily facilitate novel insights into intricate associations and interactions of components within highly dynamic environments. As a consequence, the synthesis of reasonable hypotheses about complex psychopathological mechanisms is constrained by the elaborate nature of nervous system processes and structures (Mirnics *et al.*, 2006).

In recent years, the emergence of high-throughput "–omic" technologies have stimulated a transition from "one-component-at-a-time" hypothesis-driven research strategies towards more data-driven discovery-based approaches, not dependent on *a priori* formalisations; a very appealing scenario for neuropsychiatric research. DNA microarrays have been leading the charge, by providing a highly efficient means for the simultaneous assessment of mRNA transcript abundance of all genes in specific cell or tissue populations. This approach facilitates the identification of subsets of genes whose expression can be used to predict disease states. Demonstrations of clinical applications, specifically the ability to produce disease-associated molecular fingerprints, which can be used in diagnoses, classification and prognosis (Tang *et al.*, 2005), holds considerable promise for studies of psychopathological conditions. However, the successful application of microarray technology faces several methodological challenges and limitations.

3.3 CHALLENGES, LIMITATIONS AND CRITICAL CONSIDERATIONS

Microarray technology is based on the principle of complementary hybridisation between nucleic acids. It is analogous to dot-blot systems, and can be understood as

high-throughput implementations of these (Mirnics and Pevsner, 2004). The technique allows researchers to simultaneously measure the complement of mRNA, called the transcriptome, in any tissue target. Although several specific microarray platforms exist, each with its own unique features, the general rationale is similar for all. Briefly, large numbers of DNA fragments (targets) are deposited on glass slides or membranes, at very high densities. The RNA sample (probes) to be assayed can be fluorescently labelled with various methods and are hybridised onto the microarray platform. Labelled probes hybridise to complimentary target sequences, and upon high-energy laser excitation, fluoresce with proportionality relative to abundance in the original sample. A number of excellent reviews provide detailed descriptions of the technical features of common microarray implementations (Mirnics *et al.*, 2006; Konradi, 2005; Wilson *et al.*, 2004) and will not be discussed here. Rather, focus is directed at some important practical challenges in data analysis which can have a significant impact on the interpretation of a data generated in microarray experiments.

The impact which biological and technical variation can have on the outcome of a microarray experiment has been extensively discussed. However, psychiatric settings pose several additional unique logistical and analytical challenges, which set these studies apart from many other clinical scenarios, such as studies of cancer. Gene expression studies of neuropsychiatric phenomena are greatly complicated by the nature of neural tissue targets, which are often (1) of limited supply, (2) difficult to standardise, (3) functionally complex with a high degree of cellular heterogeneity and specialisation, where (4) specific cells of interest might only constitute a small proportion of entire sample, and (5) biologically significant changes in gene expression is often modest (Barlow and Lockhart, 2002). In contrast to studies of

cancer, where tissue samples are usually (1) readily available, (2) well demarcated, with (3) a much higher degree of cellular homogeneity and (4) robust changes in gene expression (Soverchia *et al.*, 2005), the successful detection of meaningful gene expression changes in neural tissue targets pose a significant challenge.

A brief discussion of some important considerations will provide insights into the challenges and limitations posed by microarray studies of psychopathological phenomena; providing a theoretical framework within which experimental findings can be interpreted.

3.3.1 Sample collection challenges

Sample collection variables can have a decisive impact on the observed patterns of expression in a microarray experiment. Obtaining biologically meaningful results require careful experimental design, taking note of practical limitations and potential sources experimental variation.

3.3.1.1 Standardisation of sample conditions

The standardisation of sample conditions can contribute significantly to the success of a microarray experiment. When working with human samples, the ability to standardise variables, including: age, race, uniformity of disorders, medication histories and often post-mortem intervals (PMI), is only superficially possible. All of these factors can affect RNA profiles (Mirnics and Pevsner, 2004), and consequently the ability to describe transcriptome changes associated with specific neuropsychiatric disorders. These complexities are not easily circumvented as

sample supply is often limited, inevitably requiring practical compromises. Careful planning, with well defined exclusion criteria will help optimise experimental design and reduce the impact of these variables on analysis outcomes.

Gene expression studies using relevant animal models as a first step, present an alternative to human samples. A much greater degree of ease with which neural samples can be obtained, combined with better control over the ability to standardise not only genetic backgrounds, but regulate, to a large extent, specific environmental conditions, has obvious advantages. However, working with laboratory animals has its own unique challenges and as such represent a complementary alternative, not a solution, to some of the logistical challenges associated with post-mortem human tissues. The ability to standardise environmental variables, although theoretically possible, is often difficult. Minor differences in environmental cues, handling and killing procedures, and time of day when experiments are carried out, can all have a significant effect on gene expression in the brain (Soverchia *et al.*, 2005). Ideally all manipulations should be carried out by the same caretaker for the duration of an experiment. Additionally, experiments should be conducted within defined time frames, thereby minimising the effects of circadian fluctuations on gene expression and metabolism. Consideration of these factors, with appropriate procedural adjustments, will facilitate optimal experimental interpretations.

Finally, differences in tissue stabilisation variables can have a significant impact on effective transcriptional profiles. Human post-mortem samples are typically characterised by PMIs of up to 30 hours (Mirnics and Pevsner, 2004). This unavoidable feature can lead to changes in the complement of RNA transcripts in a sample. Although the post-mortem induction of transcription, as result of delayed processing, has not been systematically characterised, RNA degradation does

occur. Systematic studies have shown that RNA degradation can dramatically bias the results of downstream expression analyses, with up to three quarters of differential gene expression being due solely to differences in RNA integrity between samples (Auer *et al.*, 2003). Therefore, successful microarray analyses depend critically on the availability of high quality RNA (see below). When working with animal models, sample stabilisation can be better managed, but still require careful standardisation. In general, more rapid stabilisation of tissues will result in higher quality RNA populations and consequently, more faithful transcriptional representations of in vivo conditions.

3.3.1.2 Tissue heterogeneity

Neuropsychiatric research is usually conducted on tissue samples of neural origin. The complexities of the mammalian nervous system have already been highlighted and present a major challenge for downstream analyses and interpretations. The brain is known to exhibit a high degree of cellular and anatomical specialisation. Many important neurological events are associated with very modest transcriptional changes in specialised subpopulations of cells, with often obscured anatomical boundaries (Mirnics and Pevsner, 2004). Tissue isolation techniques are rarely able to precisely isolate such specialised populations, resulting in a high degree of functional heterogeneity in most samples. Consequently, transcriptional changes in specific subsets of cells, within larger populations, will effectively be diluted or masked by the overall profiles of these samples. In addition, specific gene expression changes in stimulatory and inhibitory neural populations might occur in opposite directions, and can contribute significantly to the "muting" of measured

transcript abundances (Soverchia *et al.*, 2005). Furthermore, imperfections in tissue dissections can be a major source of variability between samples. These factors hold important implications for the analysis of gene expression changes, where small numbers of biological replicates, as is typical in most microarray studies, make the statistical detection of modest gene expression changes, with often large amounts of biological noise, very difficult (see 3.3.3.2).

The above discussion highlighted sampling challenges and important sources of variation, which can a have an impact on microarray experimental outcomes and interpretations. Although these forms of variation cannot always be eliminated, balancing practicality with scientific rigour will help minimise their impact and improve the power to detect biologically meaningful patterns. Having considered these issues, subsequent experimental procedures and analysis strategies can be tailored to take cognisance of the challenges inherent in specific sample sets.

3.3.2 RNA processing challenges

The preparation of RNA and subsequent hybridisation, involves an elaborate series of sequential chemical manipulations. A major challenge is the faithful representation of in vivo transcript abundances, post-processing. This multi-step process is highly susceptible to the introduction of systematic inconsistencies, which can undermine the reliability of downstream analyses. Critically important is the assessment and standardisation of both quality and quantity of the RNA from which hybridisation samples will be generated.

3.3.2.1 Assessing RNA quality

As noted earlier, the delayed stabilisation of samples can have a significant impact on RNA integrity. In addition, RNA is highly susceptible to degradation during sample processing, as a consequence of nearly ubiquitous RNAases. Transcript abundances are dependent on RNA integrity, variations of which can distort measured gene expression patterns (Auer *et al.*, 2003). As such, it is important to determine RNA quality, both integrity and purity, prior to further processing, and exclude samples which do not meet study-specific quality control criteria.

Historically, the assessment of RNA integrity relied on visual interpretations of 28S:18S ribosomal RNA (rRNA) fluorescent ratios, using denaturing agarose gel electrophoresis. Intact RNA is generally expected to show double the amount of 28S compared with 18S rRNA bands. However, in recent years this method has been demonstrated to be less reliable than originally thought. Systematic studies have shown that intact samples can have a wide range of ribosomal ratios, differing between tissue types, and that the relationship between ratios and integrity is a weak one at best (Marx, 2004). Recent techniques, such as microfluidic capillary electrophoresis, allow for automated and reproducible measures of RNA integrity, with much higher molecular resolutions. Many of these methods facilitate the assessment of, not only the 28S:18S ratio, but estimates of total degradation products present across the entire range of RNA species (Schroeder *et al.*, 2006). The objective and reproducible comparison of RNA integrity between samples, has important advantages when assessing similarities in sample quality.

In addition to integrity, assessing sample purity will highlight contamination by non-nucleic acid products, such as proteins and carbohydrates, which might inhibit

the efficiency of downstream chemical reactions. Typically, A_{260}/A_{280} nm absorbance ratios are used to measure purity, which should generally be higher than 1.8 (Wilson *et al.*, 2004).

The assessment of sample integrity and purity provide important information about qualitative differences between samples, which might bias downstream interpretations. It has been noted, that lower quality samples might still be sufficient for use in a microarray experiment (Nygaard and Hovig, 2006), however, the potential impact of reduced RNA integrity and purity should be carefully considered. Importantly, the reliability of sample comparisons depends significantly on similarities in sample quality.

3.3.2.2 Amplifying limited samples

Small samples with low RNA yields usually require an amplification step to generate quantities sufficient for labelling and hybridisation. The range of sample inputs sufficient for use in a microarray experiment has been greatly extended through the introduction of RNA amplification techniques. These methods have proven very powerful, and can be used with starting amounts of RNA in the picogram (10^{-12}) range (Nygaard and Hovig, 2006). The reproducibility of these procedures has been extensively evaluated and is generally reported to be high; even improving the ability to detect various low abundance transcripts (Nygaard and Hovig, 2006). However, the potential introduction of biases during these procedures, including distortions of transcript abundance relationships, has been highlighted (Soverchia *et al.*, 2005). Systematic studies have demonstrated discrepancies between unamplified and amplified transcriptional profiles, additionally noting increased variability when input

amounts are very small. Furthermore, differences in post-amplification transcript abundances could result when RNA input amounts differ between samples (Nygaard *et al.*, 2005). This aspect is relevant to clinical settings, where the temptation to maximise sample input, rather than standardise, in an attempt to obtain optimal amplification output, could lead to biases in transcript abundances. Amplification of RNA should be done in an identical manner, including standardisation of input amounts, for samples to be reliably compared.

3.3.3 Data analysis challenges

Microarray experiments generate large amounts of data, requiring analytical strategies which are different from those employed in classic one-gene-at-a-time experiments. The microarray technique is particularly susceptible to the introduction of technical noise which can greatly obscure biological signals. The careful assessment and removal of these technical biases is an important step prior to any interpretative undertakings. Moreover, microarray data is characterised by a high-degree of multiplicity with generally thousands of observations being made for a few biological repeats, making statistical inferences challenging. Finally, a meaningful interpretation of the lists of genes typically generated, often require strategies that are able to exploit the functional annotations associated with gene products. Although these methods have advanced significantly since the early days of microarray research, there are still various inherent challenges and limitations which are worth considering.

3.3.3.1 Data normalization

The measured levels of gene expression in a microarray experiment, is significantly influenced by experimental variables, both biological and non-biological. The potentially distorting contributions by sample collection and RNA processing factors, already discussed, can be minimised through careful experimental design. However, several other non-biological variables can lead to the introduction of systematic variations in gene expression measurements. Technical variables, including: (1) labelling and hybridisation efficiencies, (2) fluorescent capacities of labels, (3) variations in array surfaces and (4) performance variables of fluorescent scanners, amongst others, can systematically bias results. Specific data manipulations, referred to as normalization can be applied to account for the effects of these potential inconsistencies. Data normalization is a fundamental step in any microarray experiment, with many different strategies available to researchers. However, deciding on a specific normalization method can be very challenging, as there is currently no consensus regarding the most appropriate strategies. The problem is compounded by the fact that different strategies can produce distinct descriptions of gene expression, resulting in divergent experimental interpretations. Critical considerations and practical guidelines are discussed in detail by the author, in Van Heerden *et al.* (2007) (Appendix A), providing a framework within which to make informed decisions regarding the suitability of different normalization methods.

3.3.3.2 Differential expression

A major goal of microarray analysis is the identification of gene expression differences associated with specific biological variables. Definitions of what

constitutes biologically meaningful differences, are currently highly subjective, and generally rely on statistical inference. Statistical tests are typically used to calculate the probability that gene expression differences reflect a significant or a true biological event. In microarray research, however, the question of statistical significance is not straightforward, as the multiplicity of experimental observations introduces a problem of multiple testing. Briefly, multiple testing issues arise when large numbers of observations are evaluated simultaneously. Here the probability of any single event (e.g. gene expression difference) occurring by chance, becomes much greater, due to the number of observations made (Griffin *et al.*, 2003). As a consequence, the lists of genes generated in these analyses are expected to contain true observations and observations resulting by chance. Enriching such a list with true discoveries (i.e. real biological events), is a major analytical challenge with little consensus.

Common strategies approach the problem, in general, by adjusting statistical thresholds, relative to the number of simultaneous observations being made (which is usually enormous), at which significance can be stated. The stringency of these methods has important implications for microarray studies of tissues, such as the brain, where small changes in the expression of various components, including neurotransmitters or receptors, can have profound biological effects. Furthermore, small numbers of biological replicates and high levels of biological variability additionally complicate matters. Gene expression changes in these tissues are rarely robust enough to meet these stringent criteria (Mirnics *et al.*, 2006).

Recently, however, systematic evaluations have brought into question the usefulness of conservative statistical philosophies, such as multiple testing corrections, for the identification of differentially expressed genes in microarray

research. The aim of significance testing is to allow researchers to say something about the probability of observing specific events again, should an experiment be repeated, by delineating boundaries of "certainty" (or uncertainty). Therefore, the issue of significance ultimately comes down to a question of reproducibility. However, one of the major challenges in microarray research, currently, is the often reported lack of concordance between replicate microarray experiments (Shi *et al.*, 2008). This situation reflects, in part, shortcomings in the criteria used to define biologically significant changes in gene expression, within microarray contexts. Many different statistical tests of significance are available to researchers, ranging from classic student's t-tests to more elaborate strategies. However, there is currently no agreement regarding the most suited algorithmic approaches for the selection of differentially expressed genes. The fact that different statistical tests of significance will tend to generate divergent lists of genes, with only partially overlapping populations (Jeffery *et al.*, 2006; Shi *et al.*, 2008), highlight a breakdown in correlations between definitions of biological and statistical relevance. Many studies have started addressing this problem, evaluating practical ways of improving the coherence between biology and statistics.

A large study called the MicroArray Quality Control (MAQC) project set out to specifically address the problem of statistical accuracy vs. biological reproducibility in microarray studies, using both statistical and non-statistical thresholds (MicroArray Quality Control Consortium, 2006). Interestingly, this study demonstrated very low levels of reproducibility when purely statistical criteria are used to identify differentially expressed genes. It was demonstrated that the additional inclusion of an empirical criterion, Fold-change, significantly improved reproducibility. Specifically, it was shown that a dual-criteria strategy, combining a relatively lenient

statistical threshold (t-test with $p < 0.05$, without multiple testing adjustments) with an empirical measure of change, such as a fold difference or fold ratio, results in optimally reproducible gene lists (Guo *et al.*, 2006). Mirnics *et al.* (2006) point out that this method eliminates genes with low variance and small magnitudes of change, a major source of false discovery.

Part of the problem of reproducibility is being able to consistently define "significant" differences; definitions of biologically relevant differences can be highly divergent. As such, combining several criteria to identify sets of genes with overlapping properties will tend to eliminate observations that are peculiar to specific statistical definitions of significance. Such a strategy of concordance will generate lists of genes with properties that are more robust and less reliant on one-dimensional interpretations of significance; an attractive alternative to the stringency of methods that correct for multiple testing issues.

3.3.3.3 Functional interpretation

A fundamental feature of any complex biological system is the fact that gene products function in large networks, with many interactions and coordinated events determining an outcome; that is, gene expression changes are not isolated events. Many analytical strategies have been developed to exploit this insight, and this is essentially where the power of microarray analysis comes into its own; allowing for more sophisticated questions and observations to be made, rather than seeking simple changes in gene expression of individual genes (Griffin *et al.*, 2003). However, extracting meaningful biological themes from long gene expression lists is

a challenging task, depending critically on the availability and contextual accuracy of functional annotations.

Continuous efforts to extend functional annotations of gene products has been greatly facilitated by the development of standardised vocabularies such as the Gene Ontology (GO) database (Gene Ontology Consortium, 2008) which provide a coherent framework within which to construct and interpret functional relationships between large numbers of gene products. This database allows for gene products to be systematically described in terms of (1) the biological processes within which they participate, e.g. signal transduction or transcription (2) the molecular functions which these products have, e.g. receptor activity or nucleic acid binding, and (3) the cellular components or anatomical structures which these products form a part of, e.g. cell membrane or ribosome. The defining feature of this vocabulary is its hierarchical structure, which allows for general gene annotations to be systematically expanded into very specific definitions, whilst maintaining relationships between different terms. As an example a gene product could be involved in metabolism, but more specifically in protein metabolism, with yet a further refinement indicating an involvement in tyrosine synthesis; the structure of this database will reflect the relationship between these increasingly detailed descriptions. It is clear that such an approach could be very useful for the identification of biological themes, where many gene products, within a list of genes, combine to participate in processes, functions or form part of some cellular structure. Typically researchers can query the GO database, using various software tools (see www.geneontology.org/GO.tools.microarray.shtml), with lists of differentially expressed genes, to determine if any GO terms are "enriched" within these lists; that is, if any of these terms are statistically over-represented in such a list relative to a

reference list. A term can be considered to be over-represented when it occurs in a list more often than would be expected by chance. However, the definition of chance depends critically on (1) the size of the sample list, (2) the size and profile of the reference list and (3) the completeness of annotations within both sample and references list. Moreover, the statistical calculations used to determine over-representation face some of the same multiple testing problems encountered in differential expression analysis. In addition to the GO database, other resources, such as the Kyoto Encyclopaedia of Genes and Genomes (KEGG) (Ogata *et al.*, 1999) allow researchers to evaluate biological themes or relationships within defined functional pathways, for example MAPK signalling or Apoptosis. Although these efforts provide indispensible tools for the functional analysis of high-throughput gene expression data, some practically relevant lirnitations require brief consideration.

Firstly, the extent to which different genes, from different organisms and different tissue targets have been annotated can vary significantly. The completeness of annotation databases, in other words, the number of gene products covered and the detail with which they are described, depend largely on the general interest of the scientific community. The ability to uncover enriched biological themes or functional pathways depend critically on the extent to which any gene list has been annotated. This is an important aspect to consider when studying phenomena that might have received limited attention.

Secondly, and related to the first issue, functional annotations are usually assigned to gene products based on specific experimental findings, or automatic annotations derived from similarities to other gene products (http://www.geneontology.org/GO.evidence.shtml). Both these methods have important implications for the interpretation of functional annotations. The fact that

many genes have divergent functions in different parts of the body, mean that the functional annotations of a gene product, within a specific context, might not apply to others. This is of particular concern for brain studies, where genes often have very different functions from those in non-neural tissues (Griffin *et al.*, 2003). In general the accuracy of any functional description is critically dependent on the extent to which a gene product has been studied.

An additional consideration pertains to the fact that functional analyses are typically conducted on lists of genes identified as differentially expressed. The successful identification of enriched functional themes in such a list, require that a significant number of genes within a particular category show expression changes robust enough to pass differential expression criteria. However, important biological phenomena can result from small but coordinated changes in the expression of several genes. Many of these genes might not pass the thresholds of differential expression, but contribute nonetheless to the outcome of some biological response. Recently, methods that recognise the more subtle and coordinated manners with which complex biological phenomena can be directed, have started to emerge (Dopazo, 2006). These methods, referred to as gene set enrichment strategies, is based on a general approach where the identification of significantly enriched biological themes does not depend on large numbers of functionally related genes with robust expression differences. Rather, an entire set of genes is evaluated in terms of the directional coordination in gene expression, within defined functional classes or terms. In other words, such approaches assess the overall behaviour of genes participating in a class, and determine whether the combined behaviours of many genes, constitute a significant response. A particular example of such an approach is that employed by the Fatiscan procedure (Al-Sharour *et al.*, 2006),

which uses a segmentation test to evaluate the coordinated distribution of functional terms within an ordered list of genes. Briefly, an ordered list of genes (e.g. highest test-statistic at the top and lowest at the bottom) is systematically partitioned (or segmented) and the asymmetric distribution and enrichment of terms are evaluated by comparing each partition to the remaining list of genes, thereby allowing for the identification of co-ordinately expressed terms. In general, these methods have an intuitive appeal and reflect a process by which many complex biological phenomena are brought about; through a series of finally tuned, but coordinated transcriptional events, often of modest magnitude.

3.3.3.4 Sample classification and prediction

One of the aims of many clinical microarray studies is the classification and prediction of disease-associated sample classes. This process is often referred to as supervised classification (Griffin *et al.*, 2003) and stands in contrast to unsupervised methods. Briefly, unsupervised methods employ a variety of clustering algorithms to discover informative expression patterns without knowledge of the specific class or label of each sample. On the other hand, supervised methods exploit *a priori* definitions of individual sample identities to find sets of genes that could be used to predict their specific sample class or label. The rationale behind these strategies essentially assumes that changes in gene expression can produce the phenotypic traits of samples (Simon *et al.*, 2003); although this relationship could be complex (see below). In other words, if specific changes in gene expression produce the phenotypic features of samples, then such expression profiles could be used to assign new arrays to a previously defined class.

A major challenge in this process is the unbiased selection of genes from a data set (Medina *et al.*, 2007). In an ideal world, it would be possible to use one data set to train and build a classifier, and then use a second independent set to test the prediction efficiency of this classifier. However, due to the small number of samples in most microarray experiments, this approach is not always possible. The conundrum which researchers face is the following: If a sample set is too small, it will contain insufficient information to build a robust predictor, on the other, if all samples in a data set are used, the genes included in the predictor will show unrealistically robust prediction efficiencies, when evaluated on the data set from which they were selected in the first place. Overcoming this problem of data maximisation, whilst minimising bias, is no trivial task. In recent years, a method called leave-one-out (LOO) cross-validation has provided an efficient workaround to this problem. Briefly, this method allows an unbiased predictor to be built through an iterative process of omitting each sample from the building stage, followed by an evaluation of prediction efficiency on the entire set. This process is completed until all such permutations have been assessed. The combined error rate of these iterations can then be used to estimate prediction efficiencies that take account of gene selection biases (Medina *et al.*, 2007).

These methods have contributed enormously to the interest in microarray technology by the clinical sector, specifically the potential of diagnostic applications. The classification and prediction of cancer classes have been demonstrated by several groups (Tibshirani *et al.*, 2003; van't Veer *et al.*, 2002). In addition, these methods have been successfully applied to various other clinical settings, including psychiatric diagnostics (Segman *et al.*, 2005); extending previous observations that showed gene expression profiles from blood samples to contain information that

could potentially be used to build predictors of neurophsyiological classes or states (Achiron *et al.*, 2004; Tang *et al.*, 2001). The classification and prediction of complex pathological conditions, using information from multiple genes, could provide novel insights into disease mechanisms, in addition to revealing new avenues for drug development, and would also stimulate the development of improved diagnostic strategies.

3.3.4 From transcriptome to proteome: an interpretive challenge

The above discussion highlighted several general methodological challenges, some specific to psychiatric settings, which are inherent in microarray techniques. Although many of these challenges are becoming less problematic, through advances in technology and analytical techniques, microarray results, by its very nature, impose an important interpretive limitation. Within clinical settings, the phenotypic characteristics of pathological processes are usually as a consequence of perturbations in protein levels or function. However, microarrays can only measure changes in mRNA transcript abundance (Konradi, 2005). The conversion of mRNA into a functional protein product is subject to an elaborate series of regulatory mechanisms, which significantly affect the relationship between gene transcript and protein levels. Consequently, correlations between mRNA and protein levels cannot be assumed. Therefore, inferences about specific phenotypic traits and their correlation to changes in gene expression should be made with caution. Nevertheless, changes in gene expression are an indication of a biological response, which might reveal important aspects of disease processes.

3.4 PROGRESS IN PSYCHIATRIC RESEARCH USING MICROARRAYS

Despite all the challenges discussed above, microarray research continuous to provide interesting insights into the complex mechanisms and functional themes underlying many neuropsychiatric disorders. Schizophrenia research has been at the forefront and has provided critical insights into transcriptional disturbances at different levels of complexity, including neural metabolism (biochemical), oligodendrocyte function and synaptic machinery (cellular), as well as components of the GABA (γ-aminobutyric acid) and glutamate systems (signalling pathways). Other studies focussing on Autism, Major Depression Disorder and Bipolar Disorder are beginning to deliver promising leads into some of the critical mechanisms associated with these disorders (Mirnics *et al.*, 2006).

Recently, the emerging use of mRNA transcript abundance data as a means of classifying complex disease profiles has gained particular appeal as a potential diagnostic tool within psychiatric settings. Specifically, the use of blood derived transcriptomics as a surrogate for the indirect interrogation of specific neurological states is an interesting prospect. The functional integration of neural, endocrine and immune systems into a NEI-axis provide a theoretical foundation for this pursuit.

Tang *et al.* (2001) provided one of the earliest demonstrations of blood derived transcriptional inferences about specific neurophsyiological states, in a proof of principle study, using rats. They showed that a series of acute neural assaults, such as intracerebral haemorrhage, kianate-induced status epilepticus and ischemic strokes, resulted in gene expression changes in peripheral blood monocytes within 24 hours. Interestingly it was found that no single gene was sufficient to predict the

specific type of brain injury, but a combination of several hundred genes made a distinction possible.

Combined with advances in classification and prediction algorithms, several recent studies have attempted to extend these observations to human neuropsychiatric contexts, which are characterised by more subtle disruptions in neurophysiology. These attempts have had some success, with a study by Tsuang *et al.* (2005) demonstrating that the microarray analysis of peripheral blood samples provide sufficient information to reliably discriminate between patients clinically diagnosed with schizophrenia or bipolar disorder and healthy controls. A combination of 8 genes (*ADSS*, *APOBEC3B*, *ATM*, *CLC*, *CXCL1*, *DATF1* and *S100A9*) was found to provide a classification accuracy of ± 95%.

In a similar study, Segman *et al.* (2005) demonstrated the ability to accurately predict the onset and, importantly, the progression of PTSD, in a group of recently traumatised patients, utilising the microarray expression profiles of PBMCs. In contrast to the small number of genes needed for diagnostic classification of samples in the Tsuang *et al.* (2005) study, more than 400 genes were needed to classify two sets of samples with ± 82% and 89% accuracy, respectively. Furthermore, it was shown that differentially expressed genes were mainly involved in (1) signal transduction, (2) transcription, (3) immune activation, (4) protein biosynthesis and degradation and (5) apoptosis. Interestingly, genes known to be active in several neural and endocrine tissues, including the amygdala, hypothalamus, pituitary and adrenal medulla, were found to be significantly enriched within the differentially expressed data set. These findings hold exciting prospects for the development of non-invasive diagnostic techniques, based on peripheral blood transcriptomics.

Although results from these studies imply a functional link between psychiatric states and gene expression changes in peripheral blood samples, an outstanding issue is the demonstration that gene expression changes in peripheral tissue targets are paralleled by transcriptional alterations in neural tissues (Mirnics *et al.*, 2006). For obvious reasons this is not possible within a clinical setting. In this respect, animal models could provide additional insights into the physiological relationship between neural responses and transcriptional alterations in peripheral tissues, within a psychopathologically defined context. In the current study, the microarray analysis of PBMC populations and selected brain regions, in a mouse MS model, was used to investigate this link.

3.5 MATERIALS AND METHODS

The experimental methods described and followed in this study have been approved by the animal ethics committee of the University of Cape Town (Ethics clearance number: 006/007) and are in accordance with National guidelines for the care and use of laboratory animals.

3.5.1 Sample collections

3.5.1.1 Acute restraint stress, sacrifice, blood collection and brain dissections

At an average age of 93 days, all mice were moved to a room adjacent to the sample processing area at least 30 min before restraint; restraint stress controls were moved at least 30 min before sacrifice. One group of mice (N_{MS} = 30, N_{SH} = 30) were subjected to 10 min of acute restraint stress by placement in a ventilated 50 ml conical

polypropylene tube, during which time individual weights were recorded. Following restraint, individuals were placed back in their home cage for a 20 min period before sacrifice. Restraint was employed as a means of acutely activating the HPAA (Daniels *et al.*, 2004), allowing for an assessment of neurophysiological stress-responsivity; the 20 min period preceding sacrifice, served additionally to allow for the HPAA initiated response to be propagated to target tissues. A second group of mice (N_{MS} = 8, N_{SH} = 8) were not subjected to restraint before sacrifice and served to provide baseline measures of HPAA activity. All mice were sacrificed, within 15 s of removal from cage, by means of cervical dislocation, immediately followed by decapitation and collection of trunk blood. Trunk blood was collected into 1.5 ml tubes pre-filled with 100 µl 3.8% (w/v) tri-Sodium-Citrate-dihydrate, which acted as an inert anticoagulant. It has been shown, that citrate-stabilised blood affords better quality RNA than other anticoagulants (Holland *et al.*, 2002). All tubes were chilled on ice prior to blood collection. Three defined brain regions (Paxinos and Franklin, 2004): the (1) prefrontal cortex, (2) hippocampus and (3) hypothalamus, were immediately dissected and stored in RNALater® (Qiagen Inc., USA), according to manufacturer's instruction, at 4°C overnight after which samples were moved to -20°C for later processing. All neural tissues were isolated and submerged in RNALater®, within 10 min of decapitation. Dissections were done under an Olympus SZ51 stereo microscope (Olympus, Japan), with LED illumination, thereby limiting heat transfer to tissues. All samples were collected within a 3.5 h window each day, starting at 7h30 and ending at 11h00. This window was defined to control for circadian fluctuations in HPAA activity and associated stress susceptibility, in addition, basal HPAA activity is at a minimum during this window (Dalm *et al*, 2005).

Mean individual weights were similar for both groups, with similar standard deviations (S.D.) (MS_{mean} = 28.00 g, S.D. = 1.93; SH_{mean} = 27.25 g, S.D. = 1.73).

3.5.1.2 PBMC separation

PBMCs were isolated using a density floatation technique. The method uses Optiprep™ (Axis-shield, Norway), a proprietary solution of Iodixanol {5,5'-[(2-hydroxy-1-3-propanediyl)-bis(acetylamino)]-bis[N,N'-bis(2,3-dihydroxypropyl-2,4,6-triiodo-1,3-benzenecarboxamide]}, with a density of 1.320 ± 0.001 g/ml at 20°C, specifically designed for the in vitro isolation of biological particles. After collecting trunk blood, 250 µl aliquots were added to 12 ml sterile test tubes (Bibby Sterilin Ltd., UK), followed by 5 ml of a prepared tricine-buffered-saline(TBS)-Iodixanol mixture (TBS: 0.85% NaCl, 10 mM Tricine, pH 7.4; TBS-iodixanol: 5 ml TBS and 1.5 ml Optiprep™). Once mixed, by gentle inversion, an additional 0.5 ml TBS was gently layered on top of the blood-TBS-Optiprep™ mixture. This layer of saline prevents cells from collecting at, and adhering to, the walls of the tube at the meniscus. Samples were centrifuged at 1000 g for 30 min at room temperature. PBMCs were collected, from the meniscus downward, in 4 ml of medium and added to a clean 12 ml tube. This suspension was diluted with two volumes of TBS. Cells were pelleted at 400 g for 10 min. The supernatant was carefully decanted, cells snap frozen in liquid N_2, and stored at -80°C until further processing.

3.5.1.3 Plasma separation

Restraint stress-subjected plasma fractions were obtained from whole blood samples after the removal of 250 µl aliquots for PBMC isolation. The entire whole blood

volume from baseline (pre-restraint) samples was processed for plasma isolation. Samples were centrifuged at 3,800 g for 7 min, and the plasma layer carefully removed. Plasma fractions were snap frozen in liquid N_2 and stored at -80°C until further processing.

3.5.2 Corticosterone assay

Circulating corticosterone concentrations were determined from 30 µl of citrate-plasma samples, using a colorimetric emzymeimmunoassay (IDS, Ltd., UK), designed for mouse and rat serum or plasma samples, with a minimum detection limit of 1.5 ng/ml. Only samples for which sufficient amounts of plasma was obtained were used for this assay. Baseline (pre-restraint), samples (N_{MS} = 8, N_{SH} = 8) and a subset of restraint subjected (N_{MS} = 17, N_{SH} = 12) samples from each treatment group were used. The assay procedure was followed as described by the manufacturer, with samples diluted 1:10 as suggested. A calibration curve was constructed as per manufacturer's instructions and concentrations were read from the curve. Values were adjusted to reflect all dilution factors. Statistical differences between samples were evaluated using the non-parametric Kruskal-Wallis test. Significance was assumed at $P < 0.05$.

3.5.3 Microarray processing

3.5.3.1 Experimental Design

Fifty five samples, 15x PFC (8x MS and 7x SH), 10x Hic and 10x Hyp (5x MS and 5x SH, each and 20x PBMC (10x MS and 10x SH) were used for microarray processing, with a two-colour common reference design (Fig. 3.1). Samples were

matched, so that 10 individuals (5x MS and 5x SH) were completely represented in all tissues (See Supplementary Material S.1 for sample details). A reference pool was constructed by combining equal amounts (0.75 µg) of pFC and Hic RNA from both groups, which was stored as single aliquots of equal concentrations.

Commercial pre-spotted, full mouse genome, microarray slides (OpArray™) were sourced from Operon (Operon Biotechnologies, Germany) which were printed with version 4.0 of the Mouse Genome Oligo Set. This set contained 35,852 longmer probes, representing ± 25,000 mouse genes and approximately 38,000 gene transcripts.

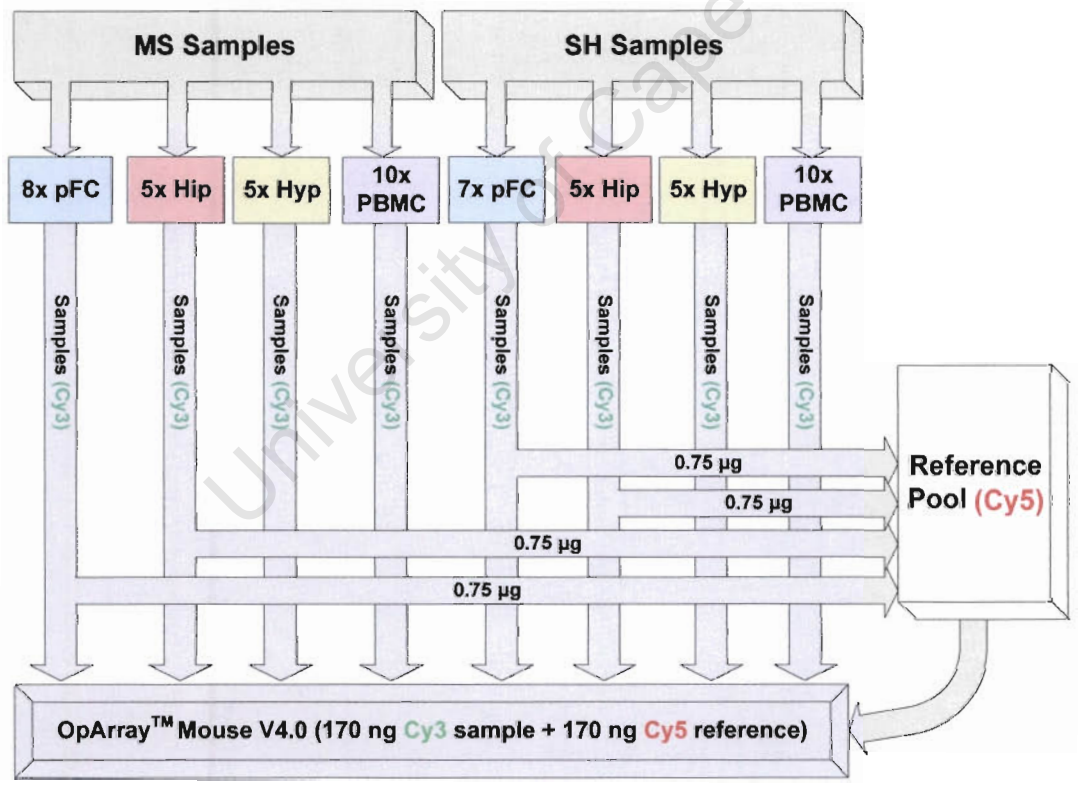

Figure 3.1: Schematic diagram of the microarray experimental design. A total of 55 two colour slides, with a common reference, were processed. Equal amounts of total RNA from pFC and Hic tissues, from both MS and SH groups, were used to construct a common reference RNA pool.

3.5.3.2 RNA purification and quantification

All RNA purifications were performed using Qiagen RNeasy® kits (Qiagen Inc., USA). Neural tissues were processed using RNeasy® Lipid Tissue Mini solution and PBMC samples using the RNeasy® Mini solution. Samples were submerged in lysis buffers and frozen for 10 min prior to homogenisation. Homogenisation was achieved using polypropylene pestles (Sigma-Aldrich Corp., USA) attached to a handheld drill. RNA extracts were quantitated using the Nanodrop ND-1000 spectrophotometer system (Nanodrop Technologies, USA).

3.5.3.3 RNA Quality Assessments

The Nanodrop ND-1000 spectrophotometer system (Thermo Scientific, USA) was used to measures sample purity as A_{260}/A_{280} nm absorbance ratios. The RNA integrity of neural samples were determined using the Agilent BioAnalyzer 2100 System (Agilent, USA), which generates a reproducible measure of integrity, called the RNA integrity number (RIN) (Schroeder *et al.*, 2006). RIN numbers for PBMC microarray samples were not generated, as total RNA obtained from these samples was critically limited. Instead, a surrogate sample set, representing total RNA from PBMC samples which were processed in parallel, with suboptimal purity measures (A_{260}/A_{280} nm ratios < 1.7) and which could not be matched to samples from other tissues, was used to determine the general integrity of RNA obtained from PBMC samples. Selection of PBMC samples used for microarray hybridisation was based on amplified RNA profiles (see RNA amplification below). Although these profiles were generated with the BioAnalyzer system they no longer have the features

necessary to calculate a RIN number; instead, samples were chosen based on visual similarities between these amplified profiles.

3.5.3.4 RNA amplification and labelling

Due to limited amounts of starting material the Amino Allyl MessageAmp™ II aRNA amplification solution (Ambion Inc., USA) was ernployed to generate sufficient RNA quantities for microarray procedures. The protocol was followed as described by the manufacturer with the only deviation being the suggested maximum length of the in vitro transcription (IVT) reaction. Briefly, the procedure consists of reverse transcription with an oligo(dT) primer, bearing a T7 promoter, to yield cDNA. A second strand is synthesised to generate a template for T7 RNA polymerase. An in vitro transcription (IVT) reaction, configured to incorporate the modified nucleotide 5-(3-aminoallyl)-UTP (aaUTP), generates large amounts of antisense RNA (aRNA). The aaUTP residues are then chemically coupled to either a Cy3™ (Green) or a Cy5 (Red) NHS ester dye during a labelling reaction. IVT incubation duration was 16 h at 37°C for all samples (maximum recommended time was 14 h). All neural tissues were amplified from 0.5 μg total RNA, whereas all PBMC samples were amplified from 0.18 μg. Reference pool samples were amplified from 0.55 μg of total RNA, generating enough aRNA for ten hybridisations. All labelling reactions were done using 6.5 μg of aRNA. Reference aRNA was labelled with Cy5 and sample aRNA with Cy3™.

3.5.3.5 Slide preparation and sample hybridisation

All OpArray™ slides were prepared and processed according to the manufacturer's instructions (Operon Biotechnologies, Germany). All hybridisations were performed using 170 ng of Cy3™ labelled sample and Cy5 labelled reference (see Fig. 3.1). This amount was found to yield dye concentrations within the range recommended by the manufacturer. All hybridisations were done at 42°C for 16 h in humidified ArrayIt® hybridisation chambers (Telechem, USA). After washing, slides were dried by centrifugation at 200 g for 5 min. Slides were kept in a light protected air tight environment and scanned on the same day.

3.5.3.6 Image acquisition

All images were acquired using an Axon 4000A dual-colour confocal laser scanner coupled to Genepix 6.0.27 Pro Software (Axon Instruments / Molecular Devices Corporation, CA, USA). Photo Multiplier Tube (PMT) settings for both 532 nm (Cy3) and 635 nm (Cy5) channels were adjusted to levels such that the ratio of signal intensities of both channels were as close to one as possible, with similar range distributions and minimal pixel saturation. Fluorescent signals were collected in Cy3 and Cy5 channels and stored as paired TIFF images.

3.5.3.7 Feature extraction

Image segmentation was done using Genepix 6.0.27 Pro Software (Axon Instruments / Molecular Devices Corporation, CA, USA). Automatic morphological feature alignment and background estimation was used and manually adjusted

where necessary. A predefined filter was used to flag features that failed to meet a set of minimum quality criteria (see Supplementary Material S2.1).

3.5.3.8 Data processing and analysis

3.5.3.8.1 Normalization

Data normalization was done in R, an environment for statistical computing and graphics (R Development Core Team, 2008; http://www.r-project.org). Specifically, the Bioconductor (Gentleman *et al.*, 2004; http://www.bioconductor.org) package, Limma (Smyth, 2005), was used to normalize all data (see Supplementary Material S.2.2.1 for generic script). All previously flagged features (see Feature extraction) were down-weighted (to 0.001) during normalization, contributing minimally to correction factor estimations. All data were assessed visually and statistically, pre- and post-normalization, to determine the most optimal normalization pipeline (see Supplementary Material S.2.2.2 for statistical functions programming code).

Although visualisations are useful for the assessment of overall behaviours, the resolution with which the outcome of normalisation can be evaluated is low. Specific statistical measures provide a much a higher resolution, with reproducibility, which can be used to compare different normalisation strategies and their efficacy at removing technical noise. For this purpose several measures were constructed and included:

(1) Pre-normalization whole array standard deviation

(2) Post-normalization whole array standard deviation

(3) Standard deviation of the pre- and post-normalization standard deviations for each array – i.e. a measure of variability between the standard deviations of

different arrays (which should be similar, as the variances of arrays are expected to be similar). This measure should be smaller after normalization. However, any improvement must not come about as a result of data compression towards a standard deviation of 0 (see next two points).

(4) Squared difference between the data set median pre-normalization standard deviation and the post-normalization standard deviation of each array – this provided a measure of standard deviation adjustments and whether the standard deviation of each array was adjusted towards the pre-normalization data set median or away (as would be the case if all distributions are compressed). The larger the squared difference the less efficiently the standard deviation was adjusted towards the pre-normalization data set median.

(5) Sum of the squared differences between the data set median pre-normalization standard deviation and the post-normalization standard deviation of each array – a measure of overall standard deviation adjustment towards a pre-normalization median. A smaller number indicates more efficient adjustments of all arrays towards the pre-normalization data set median.

(6) Average within array replicates standard deviation, pre- and post-normalization. These measures were calculated for 165 features, consisting of 35 sets of 5 replicates per array.

(7) A ratio of post-normalization:pre-normalization average standard deviation for replicate features. Replicate features are expected to have similar expression values within each array. Therefore variability is expected to be small between individual replicates. After normalization this subset of features should ideally

contain less variation than pre-normalization. A ratio of more than 1 indicated an overall removal of noise from this subset of data.

(8) The average pairwise correlation between arrays within a data set. Because most genes are expected to be expressed at similar levels, the between array correlations should generally be quite high. Reduced correlations, post-normalization, indicate the potential introduction of noise.

Combined with visualisations, methods were selected starting with improvements in inter-array correlations (point no.8), followed by the best improvement in measures of inter-array standard deviation variability (point no. 5 and then no. 3). Lastly, the efficiency of replicate feature noise removal was assessed (point no. 7).

All neural hybridisations were normalized using Global Loess adjustments combined with between array normalization, using the median absolute deviation scaling method. PBMC samples were normalized using Print-tip Loess adjustments only, with default settings. Background adjustments were not performed on any of the hybridisations, as it was not found to improve the data. Normalization yielded $\log_2$-transformed expression ratios, which were used for all subsequent procedures.

3.5.3.8.2 Duplicate merging, missing value imputation and pattern standardisation

Duplicate feature values were merged and missing values imputed using the Pre-processing module of GEPAS (Montaner *et al.*, 2006; http://www.gepas.org). A first round of duplicate merging was done based on unique oligo identifiers (i.e. all features with the same oligo sequence). Following missing value imputation, a second round of duplicate merging was performed using primary gene identifiers (ENSEMBL Mouse release 43.36d, http://www.ensembl.org; Refseq release 22,

http://www.ncbi.nlm.nih.gov; or Riken release 3.0, http://fantom.gsc.riken.jp; where available and in order of preference), thus reducing the number of duplicate gene measurements. KNN imputation (with the default of 15 nearest neighbours) was used to estimate missing values for (1) pFC patterns with a minimum of 66% unflagged features (i.e. at least 10 out of 15 slides), (2) Hic and Hyp patterns with a minimum of 70% unflagged features (i.e. at least 7 out of 10 slides) and (3) PBMC patterns with a minimum of 55% unflagged features (i.e. 11 out of 20 slides). Additionally, all gene expression patterns were standardised prior to hierarchical clustering. Standardising patterns prior to clustering is a common practice, and involves adjusting the mean of all patterns to zero and the standard deviations to one; effectively bringing all values to a similar scale. Clustering was done using the Tigr MultiExperiment Viewer V4.1 (TMEV, http://www.tm4.org) from the TM4 suite of microarray analysis tools (Saeed *et al.*, 2003).

3.5.3.8.3 Batch effect removal

Batch effects and other forms of structured noise were removed from data using ASCA-genes (Nueda *et al.*, 2007). Briefly, this method uses a principal component strategy to identify structured sources of variation attributable to experimental design, such as batch effects, which commonly occur when samples are processed as distinct subsets. Once components of variability are identified, gene expression signals can be adjusted to exclude structured sources of experimental bias. In the current study, a batch was defined as a single amplification, labelling, hybridisation and scanning run, which included 5 slides. Other forms of structured noise were identified as residual error (RE) and removed if any one factor accounted for more

than 4 times the error that would be expected by chance (an example of scripting code is shown in Supplementary material S.2.3). For example, in a set of 20 slides, if noise was random, one would expect an average of 5 % RE within each component. However, structured noise would show much higher values and indicate, therefore, some kind of systematic bias, which can be removed. In this example, any factors that accounted for more than 20 % RE were removed.

3.5.3.8.4 Differential expression

Differentially expressed genes were identified using a concordance strategy, based on overlap between three statistically divergent approaches. Genes that had a P-value < 0.05, using the Info statistic (Kaminsky and Friedman, 2002), from the ScoreGenes software package (http://www.cs.huji.ac.il/labs/compbio/scoregenes/), and a P-value < 0.05 using the Tusher *et al.* (2001) Significance Analysis of Microarrays (SAM) implementation in the T-Rex module of GEPAS (http://www.gepas.org), in addition to an absolute fold-change > 1.2 (where fold change is defined as the fold difference between MS and SH) were considered to be differentially expressed (DE).

Briefly, the Info statistic is based on misclassification rates for each gene measurement, which involve finding a threshold value of optimal separability between treatment classes. The degree of treatment-specific measurement separation by such a threshold value, whilst taking account of treatment class sizes, is used to derive a p-value. The SAM method is also referred to as a moderated t-statistic, and involves the stabilisation of the variance component of the t-statistic. This method aims to address differences in inter-gene variance, which t-distribution

derived statistics assume to be similar. Fold change is a simple empirical measure of absolute mean expression differences between treatment classes.

3.5.3.8.5 Functional enrichment

Functional enrichment of GO terms within differentially expressed gene sets was evaluated using Blast2GO (Conesa *et al.*, 2005), whilst gene set enrichment analysis was done using Fatiscan (Al-Sharour *et al.*, 2006; http://www.babelomics.org). In the current study, gene expression data was ordered according to SAM statistics. Genes over-expressed in MS samples were at the top of the list, with those at the bottom of the list being over-expressed in SH samples (or under-expressed in MS samples). The pFC and Hyp lists were evaluated using 50 partitions, the PBMC list using 55 partitions and the Hic list using 60 partitions. For each data set, the number of partitions was chosen based on optimal enrichment output.

3.5.3.8.6 Classification and prediction

The efficiency of PBMC gene expression profiles at predicting the treatment class of samples (i.e. MS or SH) was evaluated with the Prophet module in GEPAS (Medina *et al.*, 2007; http://www.gepas.org) using both the K-nearest neighbour (KNN) and Support Vector machine (SVM) algorithm options. Leave-one-out cross validation was used to counter selection bias whilst simultaneously assessing prediction efficacy.

3.6 RESULTS

3.6.1 Corticosterone assay

The HPAA is central to the physiological regulation of stress-responses, with its activation characterised by changes in several hormonal messengers and receptor targets. An important aspect of an activated HPAA is the secretion of glucocorticoids (Cortisol in human and Corticosterone in rodents) from the adrenal cortex, which serves to effect various physiological adaptations in response to a stressor (Marques-Deak *et al.*, 2005). As such, circulating levels of glucocorticoids can provide important insights into the stress-responsivity of the HPAA. Importantly, corticosterone profiles can be extended as an index of neurophsyiological responses to stress, in rodents.

Acute restraint stress was used to elicit the activation of the HPAA, and plasma corticosterone concentrations were used a means of inferring the magnitude of this response. Figure 3.2 show that there was no difference in basal (pre-restraint) plasma corticosterone concentration, between MS and SH individuals and that restraint stress produced a large increase in concentration in both groups. Twenty minutes after the cessation of restraint (post-restraint), corticosterone concentrations were significantly higher in the MS samples compared to the SH group (H (1, N= 29) =5.201771 p =0.0226).

3.6.2 Microarray analysis

All microarray data are interpreted as changes in MS sample expression, unless stated otherwise. That is, over-expression or over-representation refers to an overabundance or enrichment of gene transcripts and functional themes in the MS

group, relative to SH samples; with under-expression or under-representation designating the opposite.

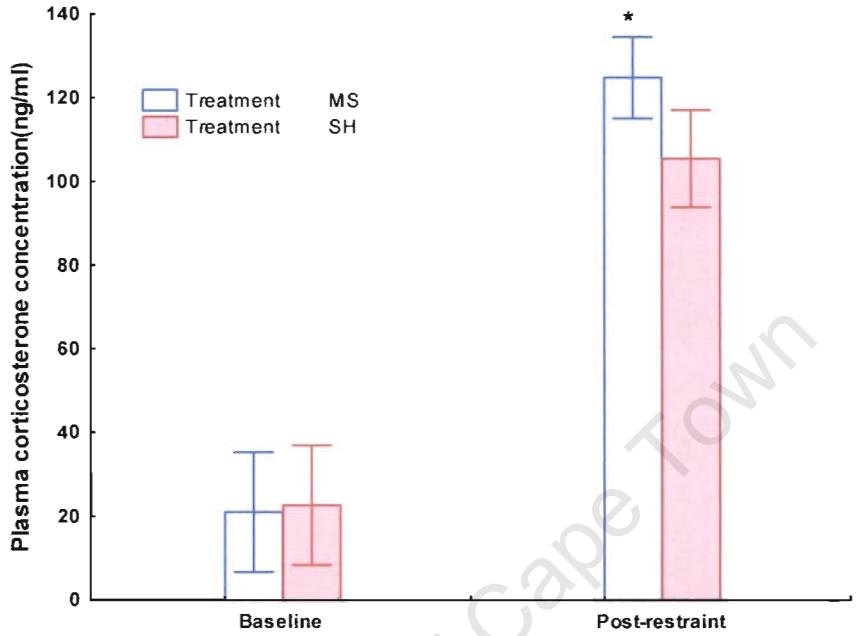

Figure 3.2: Plasma corticosterone concentrations. At baseline MS and SH individuals showed similar concentrations of corticosterone. Restraint caused an increase in concentrations which was significantly higher in the MS group, 20 min after restraint cessation. Values are means ± 95% confidence intervals. *P < 0.05.

3.6.2.1 Choice of target tissues

Neural transcriptional profiles were evaluated at three functional levels, involving defined brain structures which form part of the limbic stress pathway (Sala *et al.*, 2004). These structures participate in (1) cognitive or interpretive, (2) comparative and (3) physiological components of responses to psychological stressors. Important to stress-associated cognitive functions are cortical structures, such as the prefrontal cortex (pFC) (Sala *et al.*, 2004), whereas a "comparator system", involving the hippocampus (Hic), regulates whether a stressor requires a conditioned automatic response or higher order processing (Vinogradova, 2001). Crucial to the integration

of neuroendocrine and autonomic responses to a stressor, is the hypothalamus (Hyp), regulating HPAA axis and sympathoadrenal activation (Sanchez *et al.*, 2001). Together pFC, Hic and Hyp tissues provided a tiered description of stress-associated transcriptional profiles in the brain.

PBMC populations consist of a diverse collection of mononucleated immune cells, constituting both the innate and adaptive branches of the immune system. These cells respond to a wide range of chemical stimuli, including important stress-hormones and neurotransmitters. Specific stress-related physiological responses, such as glucocorticoid secretion, can induce important transcriptional changes within these populations (Adcock *et al.*, 2004). As such, it is expected that the transcriptome of these cells can provide insights into stress-induced physiological events, which is in part regulated by neural structures such as the hypothalamus.

3.6.2.2 RNA quality data

A Summary of sample purity (A_{260}/A_{280} nm ratios) and sample integrity measures, reported as a RIN, is provided in Table 3.1. PBMC samples used for microarray hybridisation were selected based on similar post-amplification profiles (a selection of these profiles is shown in Fig. 3.3).

Table 3.1 Summary of microarray sample purity and integrity

Tissue / Measure	**pFC**		**Hic**		**Hyp**		**PBMC**		**Surrogate* PBMC**	
	Mean	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**
A_{260}/A_{280}	2.07	0.04	2.01	0.04	2.1	0.01	1.90	0.15	1.62	0.07
RIN	7.00	1.14	5.7	0.71	6.2	0.33	N/A	N/A	7.8	0.91

*A set of PBMC samples with suboptimal A_{260}/A_{280} nm ratios (<1.7), which could not be matched with other samples from other tissues and which was used to infer PBMC sample integrity for parallel processed samples. S.D. = Standard Deviation

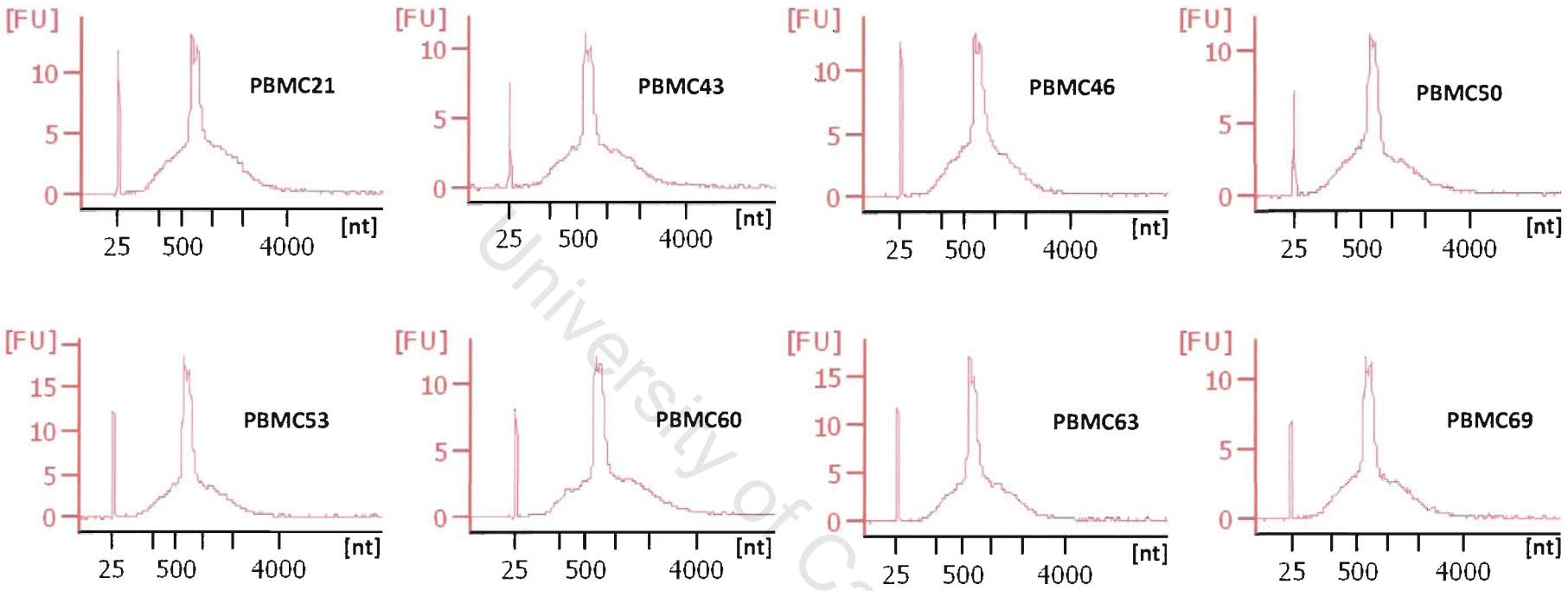

Figure 3.3: Amplified RNA profiles generated with the BioAnalyzer system. Shown here are a selection of aRNA profiles from PBMC samples (4x MS and 4x SH), which were used for microarray hybridisations. Samples were chosen based on best visual similarities in these profiles, as the features necessary to calculate RIN numbers are no longer present. Globin reduction procedures were not employed leading to the distinctive peak visible in all samples, which reflects the presence of globin mRNA (Ambion Technote 12:3, http://www.ambion.com/techlib/tn/123/10.html). The y-axis indicates relative abundance of aRNA species, measured as fluorescent units [FU], distributed across a range of nucleotide lengths [nt] (x-axis).

Although Hic and Hyp RINs were low, all samples were efficiently amplified (data not shown), which indicated intact 3`-ends as the amplification method required the presence of intact poly(A)-tails (see Materials and methods). Furthermore, the oligo's on the OpArray™ microarrays are designed to be 3`-biased (see the Mouse OpArray™ version 4.0 datasheet, http://www.operon.com). Therefore, given the efficient amplification and the microarray oligo design, the quality of Hic and Hyp RNA was deemed sufficient for microarray hybridization. In addition, all samples had very similar integrities.

3.6.2.3 Data normalization strategy

To illustrate the normalization assessment process employed for all tissues, the evaluation of a subset of normalization methods is shown for pFC samples. Visual outputs are shown as boxplot summaries and MA-plots in Figure 3.4 and Figure 3.5, respectively. Statistical assessments are summarised in Table 3.2. The results demonstrate a comparison of three general normalization methods, Median, Global Loess and Print-tip Loess (see Smyth, 2005 for technical descriptions), with or without between array scaling and standard background subtraction.

Figure 3.4A shows the boxplot summaries of pre-normalized data. Clear differences in the distributions of data between arrays are apparent from this image. Post-normalization profiles for all three methods, with between array scaling, showed a dramatic improvement in the similarities of boxplot distributions between arrays (Figure 3.4B-D). However, these figures demonstrate the difficulty with which differences in efficacies between normalization strategies are identified, as all three methods appear to have removed systematic bias from the data. Figure 3.5 highlight

the presence of a non-linear ratio-intensity bias (Quackenbush, 2002) in a representative array from the pFC data set. From these images it is clear that only Global Loess and Print-tip Loess were able to efficiently remove this bias. In addition, Print-tip Loess appeared to generally compress the data.

In contrast to the visualisations, the statistical output (Table 3.2) provides objectively comparable measures. The median normalization options were the only methods which improved inter-array correlations (measure #8); however, from the MA-plots it was clear that this approach did not deal effectively with non-linear ratio-intensity bias.

On the other hand, Print-tip Loess greatly reduced inter-array correlations (measure #8). Although it appeared as if this method very efficiently removed noise from replicate sets (measure #7), the MA-plot shown in Figure 3.5D, indicate that this was possibly as a result of data set compression.

The global loess method, combined with between array scaling, appeared to provide the best normalization outcome. Although this option slightly reduced inter-array correlations (from 0.54 pre-normalization to 0.53 post-normalization), this was not significant, given that all other measures were found to be optimally improved by this method. In addition, the MA-plot profile (Fig. 3.5C) confirmed that these improvements were not as a result of data set compression.

After normalization, replicate merging, removal of flagged features and imputation, the number of genes (unique ENSEMBL, Refseq or Riken identifiers) expressed in each tissue was: (1) PFC, 15 760; (2) Hic, 17 344; (3) Hyp, 15 794 and (4) PBMC, 13 306. All subsequent analyses were carried out on these sets.

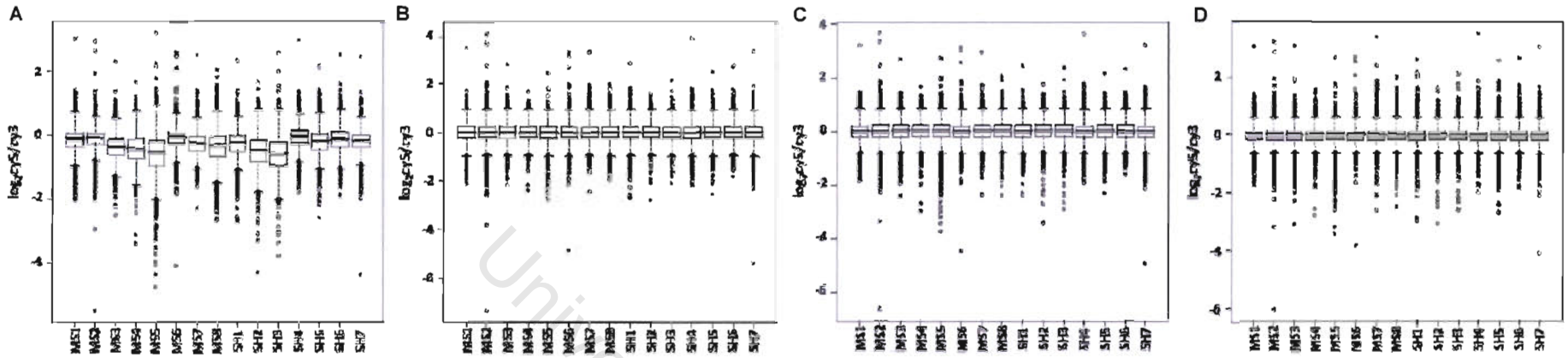

Figure 3.4: Boxplot summaries of log_2 expression ratios (Cy5/Cy3). Boxplot summaries for (A) prenormalized data, highlight differences in the data distributions between arrays. These distributions should exhibit similar profiles, regardless of treatment. Post-normalization profiles are shown for (B) Median normalization + between array scaling, (C) Global Loess + between array scaling and (D) Print-tip Loess + between array scaling. An assessment of these images suggests that all three methods have successfully normalized the data. The level of resolution provided by these profiles make a distinction between the efficacies of different methods difficult.

Table 3.2 Summary of results, for a subset of normalization methods, of statistical measures of normalization outcomes

Measure*	Pre-normalization	Median Normalization	Median Normalization + Between Arrays Scaling	Median Normalization + Between Arrays Scaling + Background Subtraction	Global Loess normalization	Global Loess Normalization + Between Arrays Scaling	Global Loess Normalization + Between Arrays Scaling + Background Subtraction	Print-tip loess normalization	Printt-ip loess Normalization + Between Arrays Scaling	Print-tip loess Normalization + Between Arrays Scaling + Background Subtraction
3	0.08	0.08	0.02	0.02	0.04	0.01	0.02	0.04	0.01	0.01
5	N/A	0.1	0.01	0.36	0.03	0.01	0.12	0.13	0.12	0.01
7	N/A	1	1.03	0.91	1.04	1.05	0.93	1.76	1.80	1.52
8	0.54	0.54	0.54	0.55	0.53	0.53	0.5	0.39	0.39	0.38

*See Materials and methods for a description of each measure; Grey shading indicates the normalization pipeline chosen for pFC tissues.

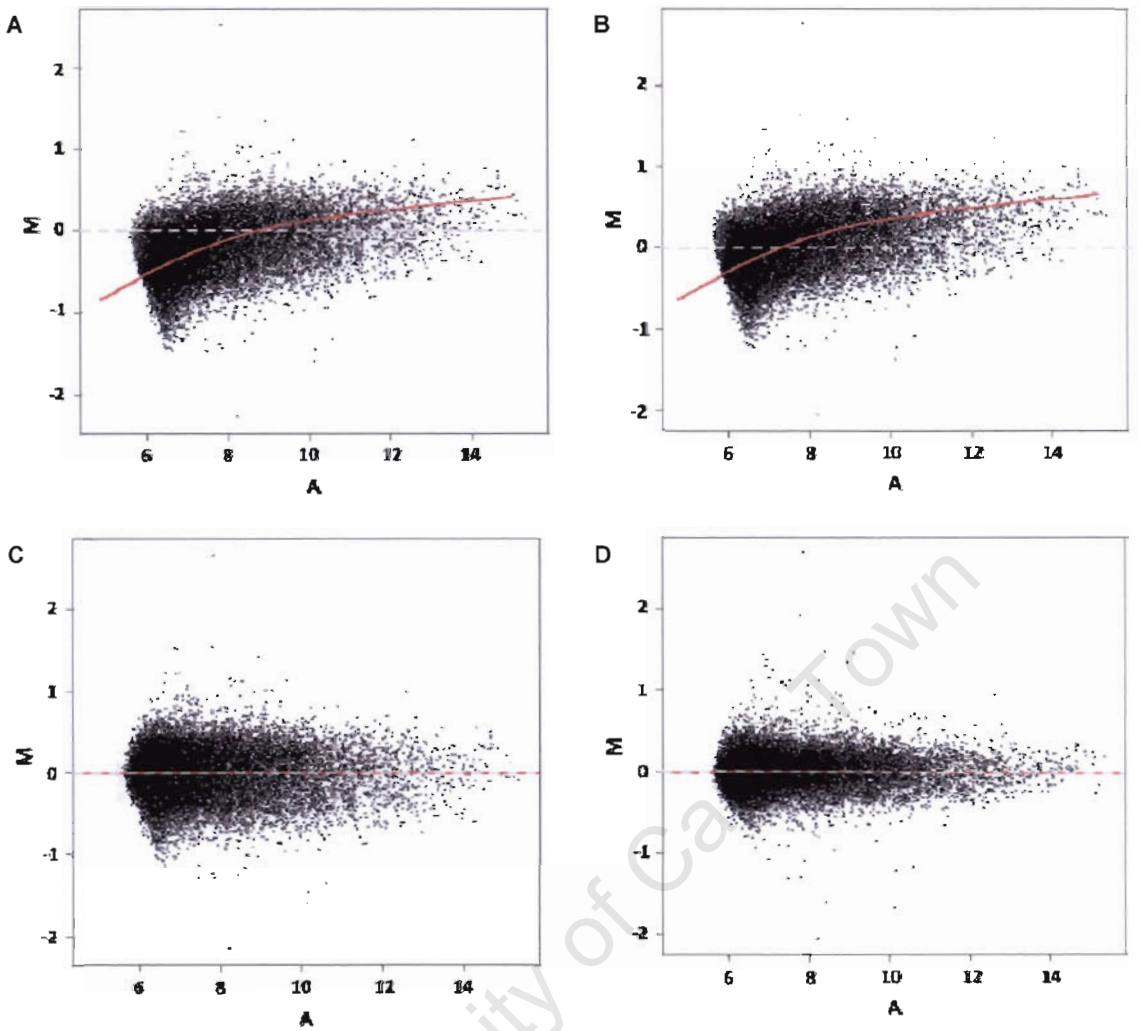

Figure 3.5: Pre- and post-normalization MA-plots of a representative pFC array. Shown here are MA-pots for (A) pre-normalized data, in addition to (B) Median, (C) Global Loess and (D) Print-tip Loess normalized data. The grey dashed-horizontal line indicates the zero ratio axis, around which the data is expected to be symmetrically distributed. The red line illustrates the general trend in the data. In (A) this trend highlights the presence of a non-linear ratio-intensity bias, which is not corrected by (B) Median normalization. In contrast, both Global (C) and Print-tip Loess (D) methods effectively removed this non-linear trend, resulting in data symmetrically distributed around the zero horizontal. Also note, that Print-tip Loess normalization (C) appears to compress the data, relative to the other methods. M = log ratio = ($\log_2$Cy5/Cy3); A = average intensity = ½Log_2[(Cy5 intensity) + (Cy3 intensity)]

3.6.2.4 Batch effect removal

Batch effects and other sources of structured noise were removed from all tissue sets, using Asca-genes (Nueda *et al.*, 2007). As an example, Figure 3.6 shows the clustering of pFC samples prior and post to the removal of batch effects. Batch

effects had a profound impact on sample profiles, as is evidenced by the fact that samples cluster clearly based on their batch allocation. After the removal of batch effects samples were more randomly distributed, as would be expected if systematic bias was removed, and most of the genes were unaffected by treatment.

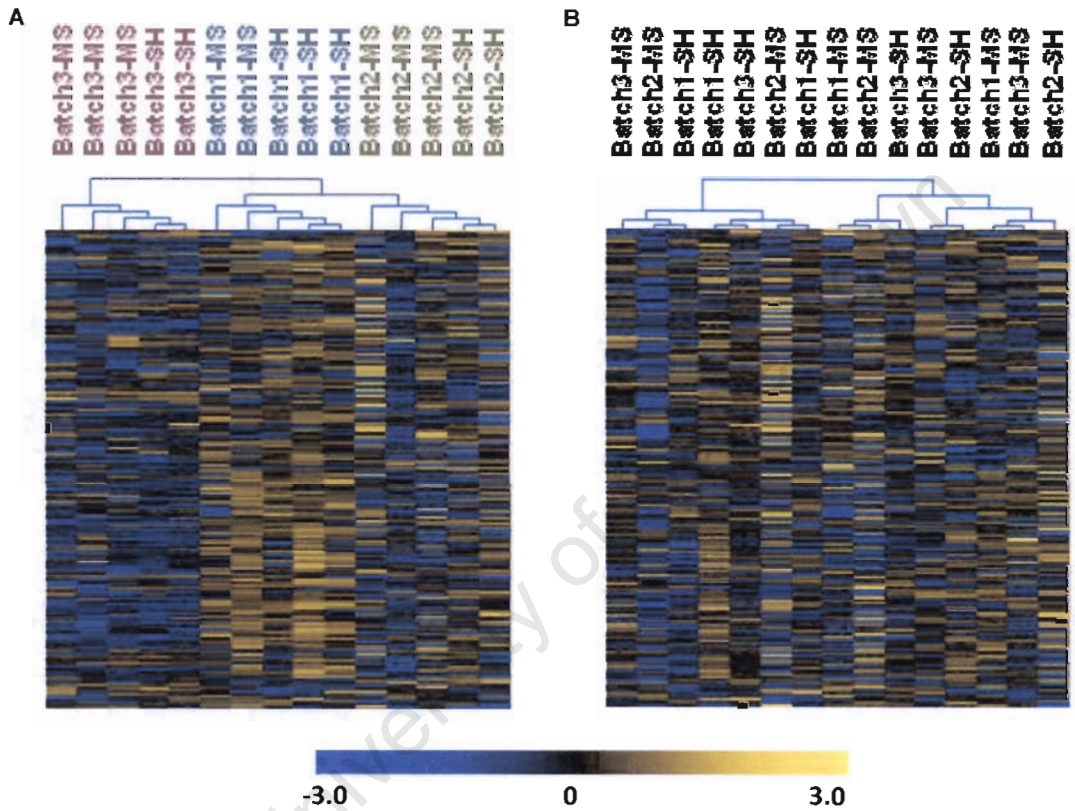

Figure 3.6: Identification and removal of batch effects from microarray data. Shown here are unsupervised hierarchically clustered pFC expression data (Pearson correlation metric with average linkage) for all genes (note: image shows only a few hundred genes), (A) prior to the removal of batch effects and (B) after the removal of batch effects. It can clearly be seen from (A) that samples cluster based on their batch allocation (i.e. processing date). After the removal of batch effects (B), samples clustered much more randomly, as would be expected if the majority of genes do not change between samples. Importantly, the removal of batch effects did not lead to the artificial creation of treatment-based clusters. Cluster images were generated using TMEV V4.1.

3.6.2.5 Differential expression

Differentially expressed (DE) genes were identified in all tissues. Importantly, the selection strategy significantly reduced the number of genes identified as DE by any

one single criterion (see Fig. 3.7). The results highlight large differences in the overlap between different criteria, emphasising selection bias by each individual method. A summary of all DE expressed genes can be found in Supplementary materials S.3.

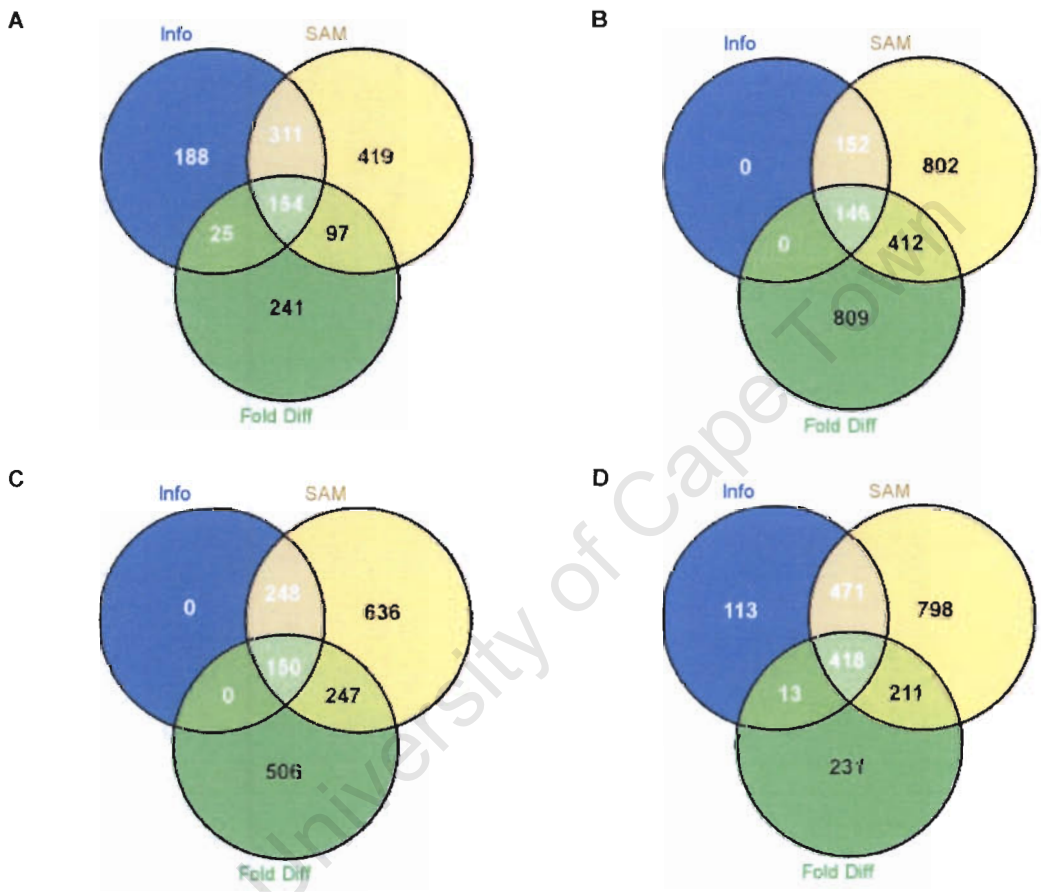

Figure 3.7: Differential expression criteria-overlap within each tissue. Venn diagrams showing the overlap between different gene selection criteria (Info and SAM $P < 0.05$ and Fold difference > 1.2) for (A) pFC, (B) Hic, (C) Hyp and (D) PBMC. Venn diagrams were generated using Venny (Oliveros, 2007).

All neural tissues exhibited approximately symmetric distributions of DE genes. In contrast, differential expression in PBMC tissues showed significant asymmetry with a majority of over-expressed genes (See Fig. 3.8 and Table 3.3). Unsupervised hierarchical sample clustering (Pearson correlation metric, with

average linkage) of differentially expressed genes, produced clear treatment (MS or SH) separations within all tissues.

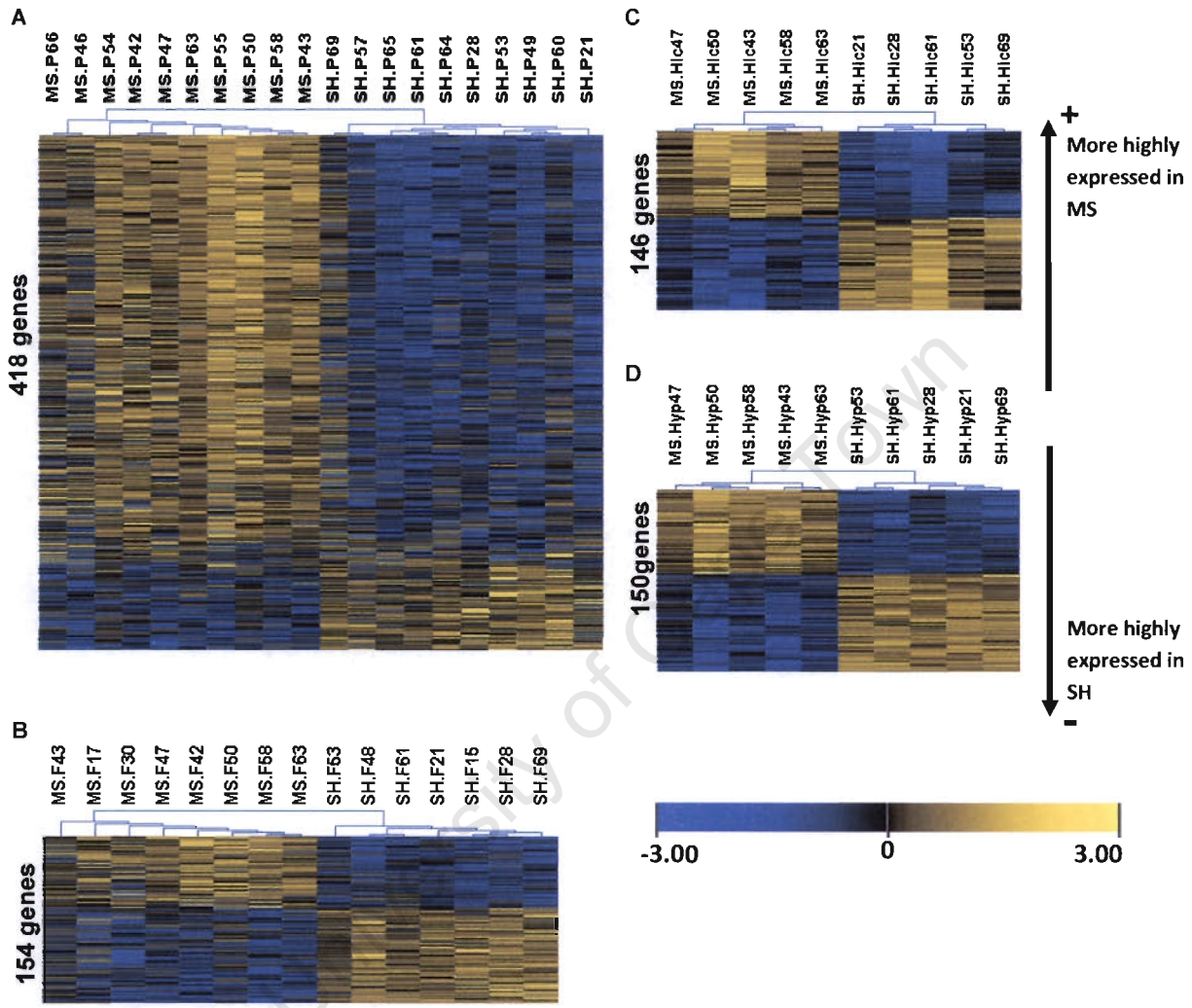

Figure 3.8: Hierarchically clustered standardised profiles of differentially expressed genes. Shown here are the false colour profiles of hierarchically clustered (Pearson correlation metric, with average linkage) differentially expressed genes for (A) PBMC samples, and neural tissues (B) pFC, (C) Hic and (D) Hyp. Genes more highly expressed in MS samples are at the top and those more highly expressed in SH samples at the bottom. The selected genes produce a clear separation between MS and SH samples. Cluster images were generated using TMEV V4.1. P = PBMC; F = pFC

Table 3.3 Summary of total differentially expressed genes within each tissue set

	Prefrontal Cortex (pFC)	**Hippocampus (Hic)**	**Hypothalamus (Hyp)**	**Peripheral Blood Mononuclear Cells (PBMC)**
Total DE / % of expressed genes	154 / ± 1%	146 / ± 1%	150 / ± 1%	418 / ± 3%
Over / under expressed	66 / 88	71 / 75	69 / 81	347 / 71

An evaluation of functional enrichment within DE neural gene sets, using Blast2GO, revealed no statistically significant terms, with genes distributed across several functional classes (results no shown). Of specific relevance to this study, however, was the identification of DE genes whose products are known to participate in stress-related signalling or behavioural regulation (see Table 3.4 for summary). In the pFC (Supplementary material S.3.1) these included important regulatory receptor components of the glutamatergic (*P2Y purinoceptor 4*, $P2ry_4$), GABAergic (*Neuropeptide VF*, *Npvf*), serotoninergic (*Serotonin receptor 3A*, $Htr3_A$) and dopaminergic (*Prostaglandin E receptor 1*, *Ptger1*) systems. The hippocampal set (Supplementary material S.3.2) included *adrenomedullin* (*Adm*) and the *nicotinic cholinergic receptor alpha 2*, *Chrna2*. DE genes with direct or inferred stress-related functions in the hypothalamus (Supplementary material S.3.3) included *cortistatin* (*Cort*), *cholecystokinins* (*Cck*), *hypocretin (orexin) neuropeptide precursor* (*Hcrt*), and *pro-melanin-concentrating hormone* (*Pmch*), all of which, except *Cort*, have been implicated directly in either the modulation of HPAA driven stress-responses or modification of stress-related behaviours.

Table 3.4 Summary of selected stress-related differentially expressed genes in CNS tissues

Symbol	Gene name	Operon Oligo ID	Tissue	Over/Under expressed in MS	Comments and References
P2ry4	*Purinergic receptor, P2Y, G-protein coupled, 4*	M200008878	pFC	Over	- Activation of P2Y4 receptors in stimulates glutamate release in PFC (Wirkner *et al.*, 2007). - Glutamate is the principle excitatory neurotransmitter and is rapidly released in the PFC in response to acute stress (Moghaddam, 2002).
Npvf	*Neuropeptide VF precursor*	M200009919	pFC	Over	- Neuropeptide VF inhibits GABAergic neurotransmission in hypothalamus (Jhamandas *et al.*, 2007). - GABAergic signalling is primarily inhibitory and serves to counter stress induced glutamatergic excitation (Morgan *et al.*, 2003).
Htr3a	*5-Hydroxytryptamine (serotonin) receptor 3A*	M400001448	pFC	Under	- Serotonin system modulates stress-related aspects of behaviour and is thought to play an important role in the suppression of behavioural and autonomic correlates of panic (Hood *et al.*, 2006). - Activation of 5-$HT3_A$ receptors enhances GABAergic mediated inhibition in the prefrontal cortex (Puig *et al.*, 2004).
Ptger1	*Prostaglandin E receptor 1 (subtype EP1)*	M300001789	pFC	Over	- Stressors such as acute restraint stress increase CNS levels of prostaglandin E2 (PGE_2) (García-Bueno *et al.*, 2008). - Ptger1 is a target of PGE_2 and is implicated in Dopamine turnover and has been shown to modulate compulsive behaviour under stress (Matsuoka *et al.*, 2005).
Adm	*Adrenomedullin*	M200000570	Hic	Over	- A regulatory peptide implicated in the regulation of anxiety-related behaviour (Fernández *et al.*, 2008). - Its actions include the activation of the HPAA and the stimulation of ACTH release (Shan *et al.*, 2003).
Chrna2	*Nicotinic cholinergic receptor alpha 2*	M200015002	Hic	Under	- Increased acetylcholine release in hippocampus associated with restraint stress (Mark *et al.*, 1996). Acetylcholine, along with nicotine, is a natural ligand for nicotinic acetylcholine receptors (nAchrs). - Activation of nAChRs is associated with the modulation of many stress associated neuronal mechanisms, including the release of GABA and glutamate (Shytle *et al.*, 2002).

Symbol	Gene name	Operon Oligo ID	Tissue	Over/Under expressed in MS	Comments and References
Cort	*Cortistatin*	M300006401	Hyp	Under	- Cortistatin (Cort) is a somatostatin (Sst) analogue preferentially expressed in the brain and binds all Sst receptors with a similar affinity. In addition it displays similar endocrine regulatory properties (Broglio *et al.*, 2008). - Although no specific stress-context studies have been undertaken, Cort has been shown to inhibit glutamate induced responses, through activation of Sst receptors in the hypothalamus (Vassilaki *et al.*, 1999); glutamate signalling in turn, is known to result in HPAA activation (Herman *et al.*, 2004).
Cck	*Cholecystokinin*	M200000995	Hyp	Under	- Cholecystokinin (Cck) has been shown to modulate behavioural and physiological components of stress-responses, with HPAA interactions having been demonstrated to result in increased levels of serum ACTH and CRH. - Additionally, maternal separation was found to produce increase the HPAA sensitivity to Cck in adulthood (Greisen *et al.*, 2005).
Hcrt	*Hypocretin (orexin) neuropeptide precursor*	M200002862	Hyp	Over	- Various stressors, including immobilisation stress, has been shown produce increased levels of orexin mRNA in the hypothalamus (Ida *et al.*, 2000). - Orexins have been found to stimulate HPAA activity and the release of corticosterone (Stricker-Krongrad and Beck, 2002).
Pmch	*Pro-melanin-concentrating hormone*	M400001802	Hyp	Over	- Melanin concentrating hormone (MCH) has been shown to have an anxiogenic effect on behaviour (Smith *et al.*, 2005). - MCH has been implicated in the stress-related activation of the HPAA and has been shown to induce the release of ACTH and corticosterone (Kennedy *et al.*, 2003).

Within the PBMC data set, DE genes (Supplementary material S.3.4) displayed a diverse range of functional classes. However, using Blast2GO, no single functional term was found to be significantly enriched (statistically) within this set. A summary of the top five most abundant Biological process (BP), Molecular function (MF) and Cellular component (CC) GO terms, highlight transcription and signalling as the major biological processes, which are differentially regulated between MS and SH PBMC samples (see Fig. 3.9). MF terms support this inference, with differences in receptor activity, ATP and ion binding; both important mechanisms in cellular signal transduction. Although cellular component terms are very general, membrane and nucleus classes, denote major sites of signalling and transcription, respectively.

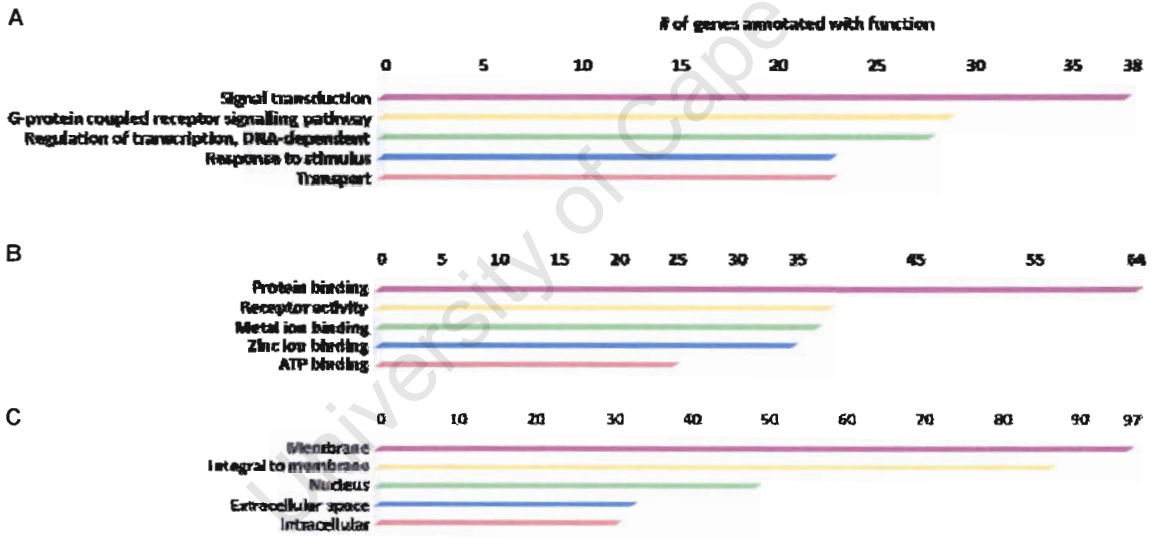

Figure 3.9: Top five BP, MF and CC GO terms within the DE PBMC gene set. (A) Biological process terms: Signal transduction (GO:0007165), 38 genes (9%); G-protein coupled receptor signalling, pathway (GO:0007186), 29 genes (7%); Regulation of transcription, DNA-dependent (GO:0006355), 28 genes (7%); Response to stimulus (GO:0050896) and Transport (GO:0006810), 23 genes each (6%). (B) Molecular function terms: Protein binding (GO:0005515), 64 genes (15%); Receptor activity (GO:0004872), 38 genes (9%); Metal ion binding (GO:0046872), 37 genes (9%); Zinc ion binding (GO:0008270), 35 genes (8%) and ATP binding (GO:0005524), 25 genes (6%). (C) Cellular component terms: Membrane (GO:0016020), 97 genes (23%); Integral to membrane (GO:0016021), 87 genes (21%); Nucleus (GO:0005634), 49 genes (12%); Extracellular space (GO:0005615), 33 genes (8%) and Intracellular (GO:0005622), 31 genes (7%). Image generate with Blast2GO (Conesa *et al.*, 2005).

Finally, no single gene was differentially expressed across all neural tissues. There was minor overlap between some tissue pairs, except for pFC and Hic samples, which shared no genes (see Fig. 3.10). None of the shared differentially expressed genes appeared to have any known neuropsychiatric or immunomodulatory functions (data not shown).

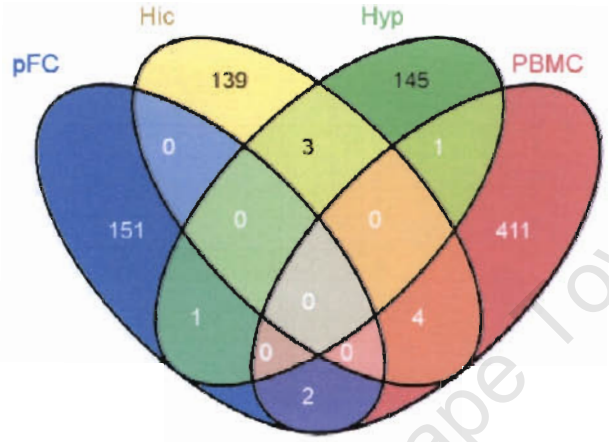

Figure 3.10: Differential expression overlap between tissues. A Venn diagram showing the overlap between differentially expressed gene lists of the different tissues. Venn diagram was generated using Venny (Oliveros, 2007)

3.6.2.6 Gene set enrichment analysis

Fatiscan was used to perform an analysis of coordinated expression changes within whole gene lists and revealed the significant enrichment of functional terms, in all tissues. In the pFC several GO terms showed significantly coordinated distributions (Fig. 3.11A) with gene products participating as synaptic membrane components, both pre- and post-synaptic (GO:0042734 and GO:0042734), ATP-binding cassette (ABC) transporter complex (GO:0043190) and protein transport and localisation (GO:0008104, GO:0006886, GO:0008565 and GO:0051234) showing general under-expression in MS samples. In contrast, ribosomal components (GO:0003735 and GO:0030529), signal transduction (GO:0004930, GO:0004872 and

GO:0007166) and cholesterol biosynthesis (GO:0006695) was found to be over-represented in MS samples.

Significant terms in hippocampal samples (Fig. 3.11B) included behavioural defense and fear responses (GO:0001662 and GO:0001662), neuropeptide signalling pathway (GO:0007218), steroid binding (GO:0005496), hormone activity (GO:0005179), ionotropic glutamate receptor activity (GO:0035255) and post-synaptic membrane (GO:0042734), which showed a general over-expression of genes participating in these classes (see Fig3.12B for example profiles). Metabolic and transcriptional terms, were also enriched, but these were ambiguously distributed between the two groups (data not shown).

Hypothalamic samples (Fig. 3.11C) were characterised by the coordinated under-expression of metabolic process terms (GO:0044237, GO:0044238 and GO:0043170) in MS samples, whilst signalling-related terms (GO:0007166, GO:0004872) were regulated in the opposite direction. In addition, both blood circulation (GO:0008015) and neurological process terms were enriched in the MS group (GO:0050877).

Lastly, in PBMC samples (Fig. 3.11D), over-expressed terms could be grouped, generally, into signalling- (GO:0004872, GO:0051606, GO:0005887, GO:0007165, GO:0007154), immune- (GO:0006955, GO:0006952, GO:0005856, GO:0007275) and, interestingly, neurologically-related (GO:0008188, GO:0050877) classes. On the other hand, under-expressed terms all displayed a metabolic theme, with terms related to RNA and protein processing (GO:0003735, GO:0016070, GO:0044267, GO:0009058, GO:0009059, GO:0015031, GO:0006412, GO:0005840, GO:0003676 and GO:0043021) and energy metabolism (GO:0005739, GO:0051187 and GO:0006099).

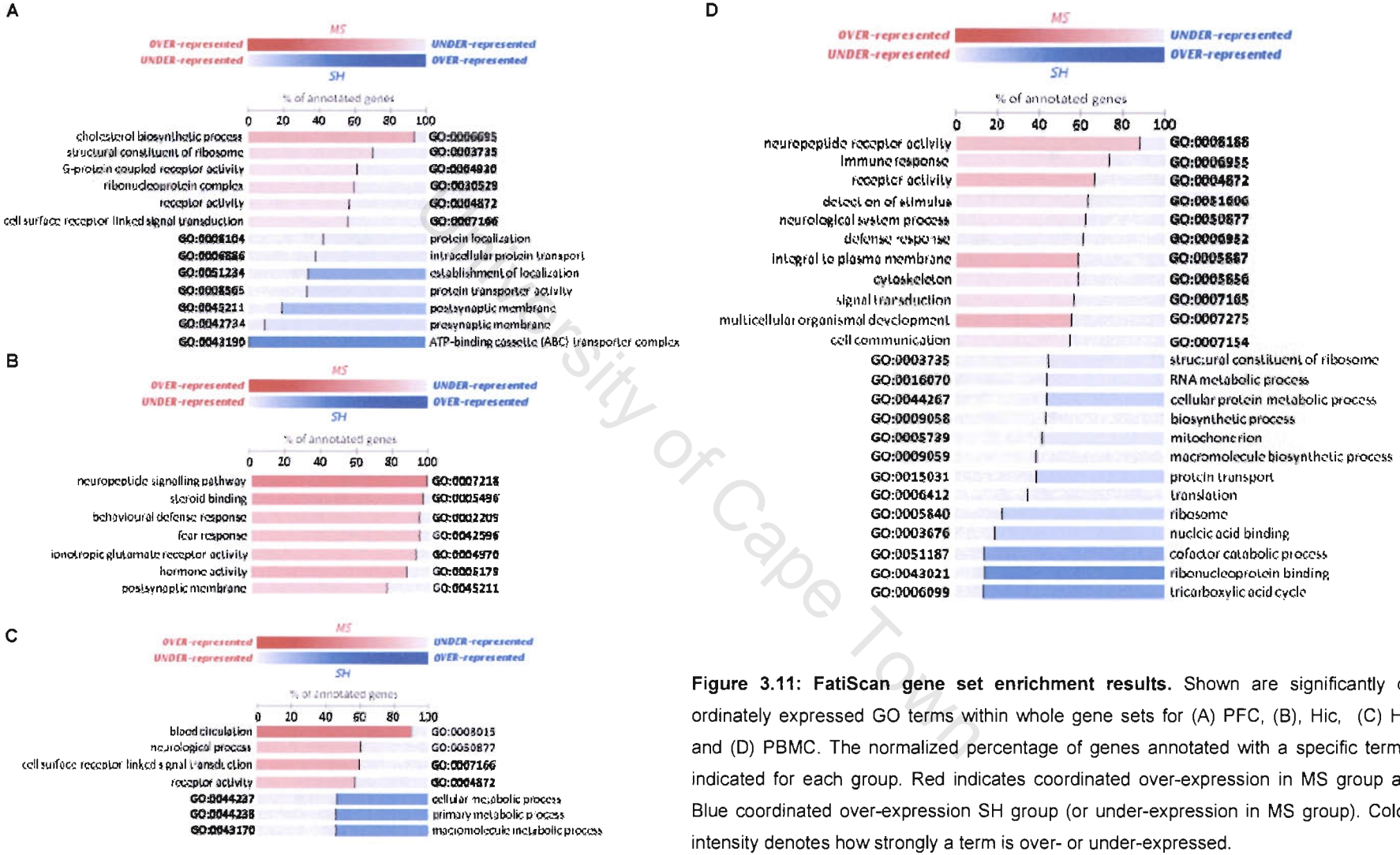

Figure 3.11: FatiScan gene set enrichment results. Shown are significantly co-ordinately expressed GO terms within whole gene sets for (A) PFC, (B), Hic, (C) Hyp and (D) PBMC. The normalized percentage of genes annotated with a specific term is indicated for each group. Red indicates coordinated over-expression in MS group and Blue coordinated over-expression SH group (or under-expression in MS group). Colour intensity denotes how strongly a term is over- or under-expressed.

Figure 3.12 shows the gene expression profiles of selected enriched terms and highlights, visually, the coordinated nature of gene expression within such functional classes. General expression themes are clearly visible, although very few of the individual genes passed the differential expression criteria used in this study.

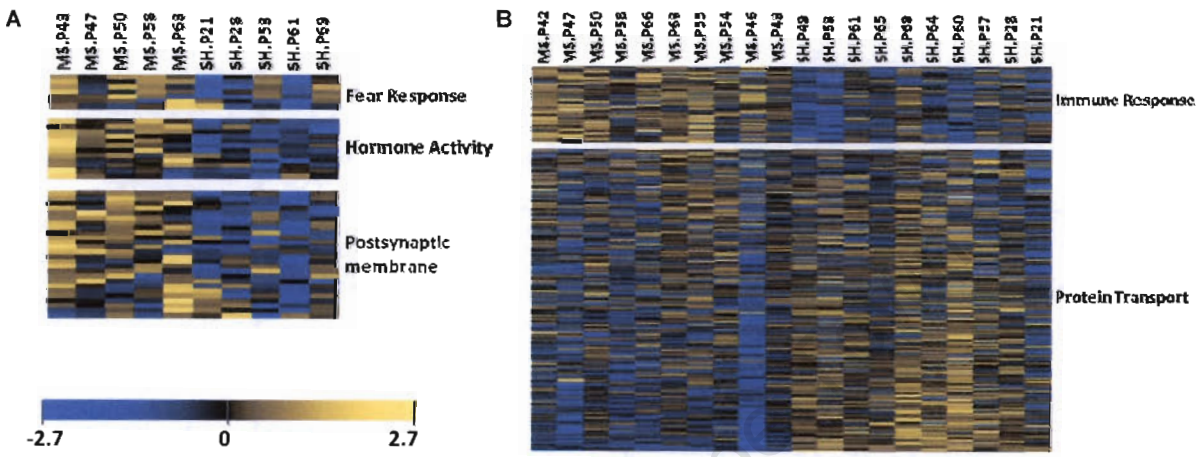

Figure 3.12: Examples of significantly coordinated functional terms. Shown here are examples from (A) Hippocampal samples and (B) PBMC tissues, where the standardized $\log_2$ expression values of significantly enriched terms within whole gene lists have been visualized. Although very few of these genes passed the differential expression criteria, general themes are clearly visible. In the hippocampal samples, coordinated differences are visible for Fear response, Hormone activity and Post-synaptic terms; genes participating in these classes are generally over-expressed (yellow) in MS samples. Similarly, immune response genes are generally over-expressed in PBMC samples, whilst genes participating in protein transport are generally under-expressed (Blue) in these tissues. Within each class, genes were sorted so that those more highly expressed (average expression) in MS samples are at the top. False colour images were generated using TMEV V4.1.

3.6.2.7 Samples classification and prediction

The classification and prediction of sample classes (MS or SH) using PBMC gene expression values, were found to be highly efficient. Both KNN and SVM algorithms produced maximum prediction efficiencies of 95%, i.e. a misclassification rate of 1/20 (see Fig. 3.13A). Using KNN (with 4 neighbours), 50 genes (see Fig. 3.13B for expression profiles and Table 3.5 for a summary of these genes) were sufficient to accurately identify sample classes 19 out of 20 times. Most of the genes included in

the predictor were over-expressed, with only 4 out of 50 showing under-expression in the MS group (Fig. 3.13B). SVM, however, only achieved this success rate using a minimum of 125 genes (with linear and radial kernels). Both algorithms converged at 125 genes, after which the prediction efficiency remained stable at 95%. Importantly, this 125 gene set consisted of the 50 genes included in Table 3.5, in addition to 75 other genes, which were the same for both algorithms (data not shown). This indicated that the KNN algorithm was better able to extract classification information from the smaller set of genes. Focus was therefore directed at this smaller set, which represent, in clinical terms, a much more manageable diagnostic collection. Notably, the same SH sample, PBMC69, was consistently misclassified as an MS sample, using both algorithms (see Table 3.3 for KNN details). Interestingly, 4 of the 50 genes included in the predictor (*Cck*, *Muted*, *St3gal6* and *Tmem63a*) were deemed not differentially expressed. Closer inspection of these genes revealed very low variance and perfect class allocation, but fold changes < 1.2.

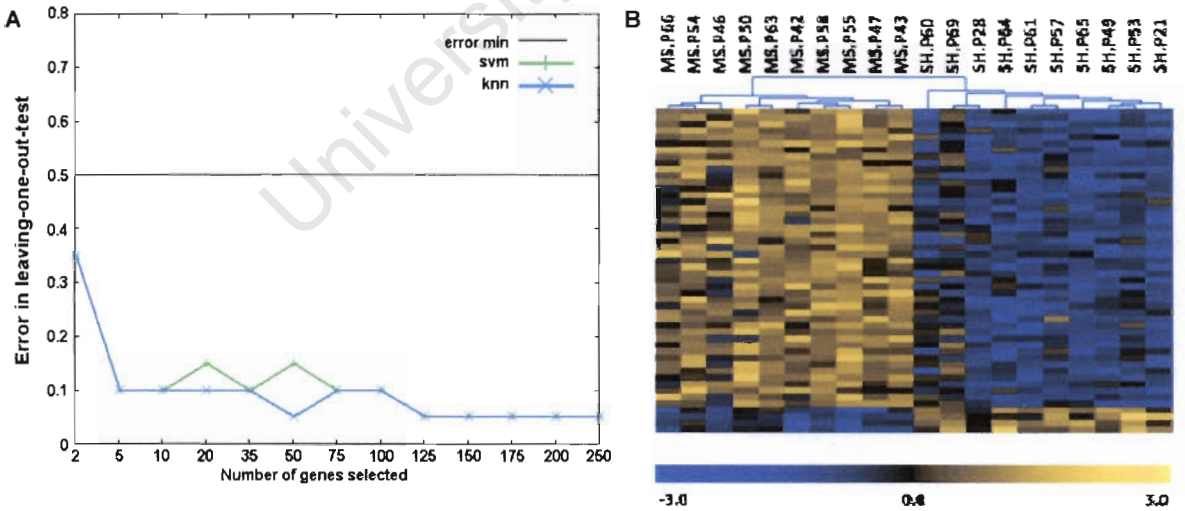

Figure 3.13. Sample classification and prediction results. (A) Leave-one-out error rates of classifiers. The KNN algorithm (blue line) reaches an optimal prediction efficiency of 95% with a minimum of 50 genes. Using 125 genes the SVM algorithm (green line) obtains this efficiency, and converges with KNN. (B) Hierarchically sample clustered (Pearson correlation metric with average linkage) profiles for the 50 gene predictor set. Notice, that although only 19 out of 20 samples were correctly classified, hierarchical clustering separates all samples into two general treatment-related clusters. Cluster images were generated using TMEV V4.1. P = PBMC

Table 3.5 Summary of 50 gene predictor set, which classified samples with 95% accuracy*

Operon Oligo ID	Description	Symbol	ENSEMBL / Refseq / Riken ID	Over / Under expressed in MS
M400008627	RIKEN cDNA 4921528I07 gene	*4921528I07Rik*	ENSMUSG00000074149	over
M200012683	Acetyl-Coenzyme A acetyltransferase 2	*Acat2*	ENSMUSG00000023832	over
M400004596	A disintegrin-like and metalloprotease with thrombospondin type 1 motif, 9	*Adamts9*	ENSMUSG00000030022	over
M200000582	Adenylate cyclase 8	*Adcy8*	ENSMUSG00000022376	over
M200005645	Actin related protein 2/3 complex, subunit 5-like	*Arpc5l*	ENSMUSG00000026755	over
M200006901	ATPase, H+ transporting, lysosomal V0 subunit E2	*Atp6v0e2*	ENSMUSG00000039347	over
M400004024	cDNA sequence BC013672	*BC013672*	ENSMUSG00000037921	over
M400008030	Bone gamma-carboxyglutamate protein, related sequence 1	*Bglap-rs1*	ENSMUSG00000074489	over
M300011602	Carbonic anhydrase 14	*Car14*	ENSMUSG00000038526	over
M200000995	Cholecystokinins precursor	*Cck*[$]	ENSMUSG00000032532	over
M200013753	Coronin 7	*Coro7*	ENSMUSG00000039637	over
M200003934	Cytochrome P450, family 2, subfamily c, polypeptide 29	*Cyp2c29*	ENSMUSG00000003053	over
M300013894	RIKEN cDNA D130054N24 gene	*D130054N24Rik*	ENSMUSG00000042790	over
M400003995	RIKEN cDNA D330050I23 gene	*D330050I23Rik*	ENSMUSG00000072569	over
M300010488	Dermokine	*Dmkn*	ENSMUSG00000060962	over
M200003607	Dedicator of cytokinesis 7	*Dock7*	ENSMUSG00000028556	over
M300014949	Endothelial differentiation, sphingolipid G-protein-coupled receptor, 5	*Edg5*	ENSMUSG00000043895	over
M400001692	Predicted gene	*EG620592*	ENSMUSG00000071719	over
M400010593	Forkhead box protein R1 (Forkhead box protein N5)	*Foxr1*	ENSMUSG00000074397	over
M300000132	Homeo box A4	*Hoxa4*	ENSMUSG00000000942	over
M400013298	LSM14 protein homolog A (Rap55)	*Lsm14a*	ENSMUSG00000066568	over
M400004821	Lysocardiolipin acyltransferase	*Lycat*	ENSMUSG00000054469	over
M400009939	Mitogen-activated protein kinase kinase kinase 9	*Map3k9*	ENSMUSG00000042724	over
M300007290	Mesoderm posterior 2	*Mesp2*	ENSMUSG00000030543	over
M200007123	Muted protein	*Muted*[$]	ENSMUSG00000038982	under
M200010626	Matrix-remodelling associated 8	*Mxra8*	ENSMUSG00000073679	over
M200007448	Nitric oxide synthase interacting protein	*Nosip*	ENSMUSG00000003421	over
M300018063	Olfactory receptor 1495	*Olfr1495*	ENSMUSG00000047207	over
M300017588	Olfactory receptor 66	*Olfr66*	ENSMUSG00000058200	over
M300015973	Olfactory receptor 669	*Olfr669*	ENSMUSG00000073916	over
M300002331	Predicted gene	*MGI:3652048*	ENSMUSG00000020682	over
M200003458	Oxytocin	*Oxt*	ENSMUSG00000027301	over
M400010890	Mus musculus polymerase (RNA) II (DNA directed) polypeptide C	*Polr2c*	ENSMUSG00000031783	over
M200000936	Peripherin 1	*Prph1*	ENSMUSG00000023484	over
M300003403	PTK2 protein tyrosine kinase 2	*Ptk2*	ENSMUSG00000022607	under
M400001722	Slingshot homolog 3 (Drosophila)	*Ssh3*	ENSMUSG00000034616	over
M300003482	Type 2 lactosamine alpha-2,3-sialyltransferase	*St3gal6*[$]	ENSMUSG00000022747	under
M200000227	Stromal interaction molecule 1	*Stim1*	ENSMUSG00000030987	over

Operon Oligo ID	Description	Symbol	ENSEMBL / Refseq / Riken ID	Over / Under expressed in MS
M300001453	Surfeit gene 5	*Surf5*	ENSMUSG00000015776	over
M400000616	Thrombopoietin precursor	*Thpo*	ENSMUSG00000022847	over
M400009774	Transmembrane BAX inhibitor motif containing 1	*Tmbim1*	ENSMUSG00000006301	over
M200013582	Transmembrane protein 25	*Tmem25*	ENSMUSG00000002032	over
M400000938	Transmembrane protein 63A	*Tmem63a*[§]	ENSMUSG00000026519	under
M400013169	Xin actin-binding repeat containing 2 isoform 2	*Xirp2*	ENSMUSG00000027022	over
M400014435	Zinc finger protein 84	*Zfp84*	ENSMUSG00000046185	over
M400018008	Novel Protein	*Not assigned*	AC160535	over
M400012711	Novel protein (I830077J02Rik)	*Not assigned*	AC121847	over
M400017112	Uncharacterised	*Not assigned*	AK054246	over
M400003712	Uncharacterised	*Not assigned*	AC122270	over
M400008575	Uncharacterised	*Not assigned*	ENSMUSG00000064159	over

*Genes are sorted by gene symbol; § Not included in differentially expressed gene list

Table 3.6 KNN sample classification summary*

Sample ID	Real class	Number of genes used / Predicted Class			
		2	50	100	125
PBMC42	MS	SH	MS	MS	MS
PBMC47	MS	MS	MS	MS	MS
PBMC50	MS	MS	MS	MS	MS
PBMC58	MS	MS	MS	MS	MS
PBMC66	MS	SH	MS	MS	MS
PBMC63	MS	MS	MS	MS	MS
PBMC55	MS	MS	MS	MS	MS
PBMC54	MS	MS	MS	MS	MS
PBMC46	MS	SH	MS	SH	MS
PBMC43	MS	MS	MS	MS	MS
PBMC49	SH	SH	SH	SH	SH
PBMC53	SH	SH	SH	SH	SH
PBMC61	SH	SH	SH	SH	SH
PBMC65	SH	SH	SH	SH	SH
PBMC69	SH	MS	MS	MS	MS
PBMC64	SH	MS	SH	SH	SH
PBMC60	SH	MS	SH	SH	SH
PBMC57	SH	MS	SH	SH	SH
PBMC28	SH	SH	SH	SH	SH
PBMC21	SH	SH	SH	SH	SH

*Classes shown in red indicates misclassification

3.7 DISCUSSION

A mouse MS model was used to set up an experimental context for the evaluation of PBMC transcriptomic-derived diagnostic classifications of stress-related neural states. The combined characterisation of corticosterone profiles and neural transcriptomes, following restraint stress, provided an independent molecular description of stress-associated differences between treatment (MS) and control groups (SH) of rnice, substantiating an interpretation of treatment-associated PBMC gene expression profiles.

Results highlighted stress-related differences in both circulating corticosterone and neural transcriptome profiles between groups. Corticosterone concentrations were significantly elevated in the MS group, 20 min after the removal of restraint stress. Neural gene expression profiles from all regions showed differences in important stress-associated genes and/or functional classes. Within this context, PBMC transcriptional profiles were found to display differences in the expression levels of many genes, in addition to the differentially coordinated regulation of several functional terms. Importantly, PBMC gene expression profiles were found to be predictive, with high efficiency, of the treatment status of individual samples.

3.7.1 Corticosterone results

Measurements of circulating corticosterone concentrations revealed significant differences in the HPAA mediated stress-responsivity between MS and SH individuals, in response to acute restraint stress (Fig. 3.2). The results obtained here are in line with previous studies which show an increased stress-induced corticosterone response in maternally separated individuals, in mice (Parfitt *et al.*,

2004) and rats (Aisa *et al.*, 2007). The significantly elevated corticosterone levels in the MS group, 20 min after stress removal, can be explained by at least two hypothetical scenarios. Assuming that after stress cessation, corticosterone concentrations start dropping back to baseline levels, either an initial hyper-activation, or the inefficient post-restraint feedback inhibition, via the HPAA, within the MS group, could produce the current outcome (Fig. 3.14). A combination of these two scenarios would produce a similar result. A clear distinction between these possibilities is not possible, as corticosterone samples were not collected at the time of stress removal, nor during the 20 min recovery period before sacrifice. However, previous studies provide some insights regarding the possible involvement of various regulatory mechanisms which might produce the observed result.

Most of the available literature, on the effects of MS on stress-induced corticosterone responses, has implicated impairments in glucocorticoid feedback mechanisms, as a major contributor to HPAA hyper-responsivity. It is thought that deficits in maternal care during early development results in reduced glucocorticoid receptor densities in the hippocampus, a major regulatory site of HPAA feedback inhibition (see Chapter 2). It is therefore likely that the observed result is due, at least in part, to developmental deficits in glucocorticoid feedback regulation of stress-induced HPAA activation. This scenario is depicted in Figure 3.14A. Furthermore, MS has also been shown to affect various neurotransmitter systems, including excitatory glutamatergic (Pickering *et al.*, 2006) and inhibitory GABAergic mechanisms (Caldji *et al.*, 2000). Deficits in the regulation of these systems, in response to stress, could lead to the hyper-activation of neural structures such as the pFC and the hippocampus (see 3.7.2.1), which are known to modulate the

stress-induced activation of the HPAA. An initially amplified response could produce the corticosterone profiles shown in Figure 3.14B.

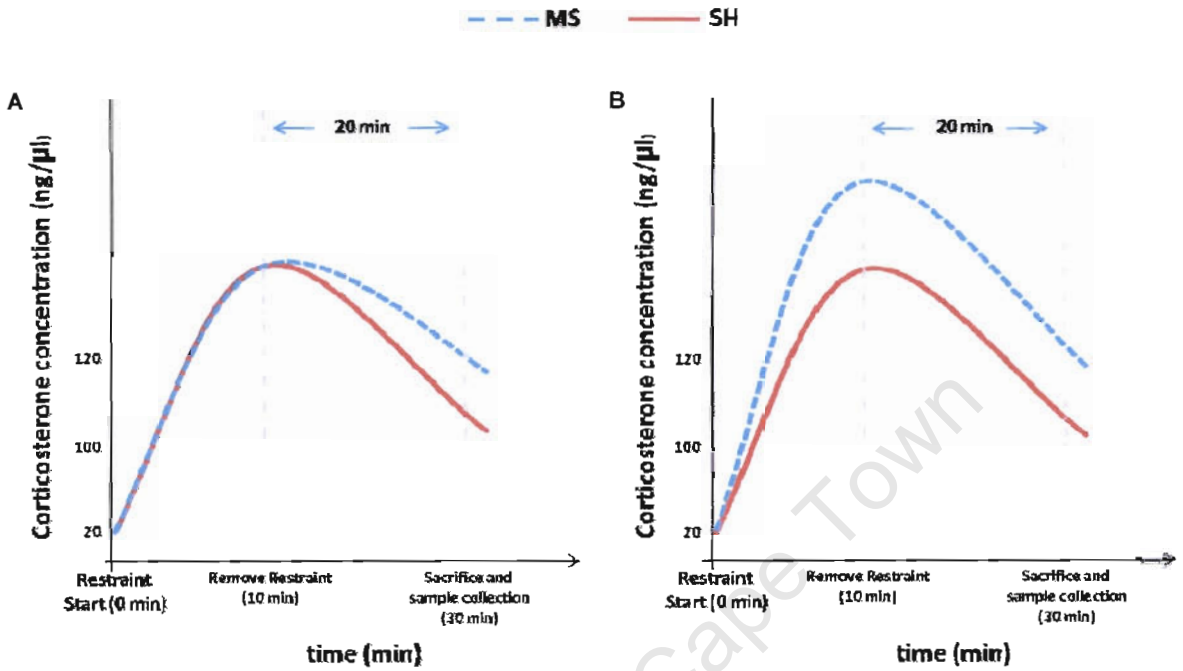

Figure 3.14: Two hypothetical stress-induced corticosterone response curves. Both profiles (A) and (B) could explain the observed differences in corticosterone concentrations 20 min after restraint cessation. Shown in (A) is a scenario where the observed differences are explained by deficits in post-restraint inhibitory feedback, rather than initial responses. The second scenario, shown in (B), describes an initial stress-induced hyper-activation of the HPAA, which results in a higher maximum response, whereas the post-restraint decline rate is similar for both groups, suggesting similar inhibitory feedback responses. Note: Axes are not drawn to scale, corticosterone concentrations at baseline (Restraint start) and sacrifice (30 min) are indicated, approximately, as experimentally obtained values, and it is assumed that corticosterone levels were higher after 10 min of restraint than at the time of sacrifice. Red line indicates SH response and blue line MS response profiles, respectively.

Although the exact nature and time course of the corticosterone stress-response is speculative, these results extend the observations made during the behavioural testing of animals, confirming increased stress-responsivity in the MS group. Additionally, these results highlight a possible mechanism through which peripheral tissues such as PBMCs, could be exposed to differential chemical milieu which could impact the biological profiles of these targets.

3.7.2 Microarray results

3.7.2.1 Neural tissues

The gene expression profiles of neural tissues revealed several differences between MS and SH individuals. The identification of genes whose products are known to be involved in the regulation of stress-related neuronal activity, in addition to the enrichment of several relevant functional terms, provided a substantial description of stress-related divergences in the mRNA profiles of MS and SH groups.

Although no individual gene was differentially expressed in all neural tissues, specific functional themes emerged. Starting with the pFC, differentially expressed genes and enriched functional terms highlighted a specific theme involving the glutamatergic and GABAergic systems. Specifically, differences between the groups pointed to a state of enhanced glutamatergic signalling, possibly as a consequence of deficiencies in GABAergic mediated inhibitory mechanisms, within the MS group. These two neurotransmitter systems constitute the major stimulatory (glutamate) and inhibitory (GABA) mechanisms of neurotransmission, and work together in a counteractive manner to ensure optimal neuronal activity in response to stress (Olson, 2002). These systems are normally in a state of equilibrium, with imbalances having been associated with several psychopathological conditions (Moghaddam, 2002; Tsapakis and Travis, 2002).

Genes whose products are involved in the modulation of these two neurotransmitter systems included *P2yr4* and *Npvf*. The activation of P2yr4 positively regulate glutamate release (Wirkner *et al.*, 2007), whereas Npvf has been shown to be an important inhibitor of GABAergic neurotransmission (Jhamandas *et al.*, 2007). The over-expression of both these genes in the MS pFC tissues, point to a

hyperactive glutamatergic system. Supporting this observation is the under-expression of *Myo6* in these MS samples. This gene is crucial for the efficient endocytosis of postsynaptic glutamate receptors, with deficits in its expression having been shown to result in increased excitatory neurotransmission (Osterweil *et al.*, 2005). Furthermore, *Htr3a* was under-expressed in MS samples. The expression of these receptors is strongly associated with GABAergic neurons and interneurons; their stimulation has been shown to play an important role in the activation of GABA mediated inhibitory neurotransmission in the prefrontal cortex (Puig *et al.*, 2004; Bloom and Morales, 1998). The identification of both pre- and post-synaptic component (GO:0042734 and GO:0042734) GO terms as co-ordinately under-expressed in MS samples, provided additional information in support of a hyperglutamatergic state. Specifically, postsynaptic terms included three $GABA_A$ receptors ($GABA_A$ alpha-1 and -3, and $GABA_A$ gamma-3). $GABA_A$ receptors are critical to the GABA mediated inhibition of neurotransmission and the regulation of anxiety-related behaviours (Wu *et al.*, 2007), with deficits in these receptors having been shown to result in enhanced anxiety (Crestani *et al.*, 1999). Presynaptic terms included two metabotropic glutamate receptors, *mGluR3* (Group II) and *mGluR7* (Group III). Importantly, both these receptors function in negative feedback mechanisms that serve to inhibit the presynaptic release of glutamate (Grueter and Winder, 2005). Taken together, these observations provide support for a glutamate-driven hyper-excitatory state in MS pFC samples.

Interestingly, hippocampal gene expression profiles extended the pFC observations, with the significantly coordinated over-expression, in MS samples, of genes whose products participate in ionotropic glutamate signalling (GO:0004970). Within the hypothalamic samples, this hyperglutamatergic theme was not readily

apparent. However, the speculative involvement of cortistatin, which was under-expressed in MS samples, would extend this theme (see below). Briefly, the product of this gene has been shown to inhibit glutamate induced responses in the hypothalamus (Vassilaki *et al.*, 1999).

These findings are highly significant given an established central role of glutamate in the regulation of several fundamental aspects of stress-responses, which include the modulation of executive and cognitive functions associated with structures such as the pFC and the hippocampus. Furthermore a central role in the regulation of HPAA activation and glucocorticoid secretion, in response to stress, has been extensively described (Moghaddam, 2002). It is thought that glutamate-driven neurotransmission within the pFC represent a common mechanism by which other neuronal stress circuits are activated. Stressors such as acute restraint have been shown to produce dramatic and rapid increases in glutamate levels primarily in the pFC, which ultimately culminates in HPAA activation and glucocorticoid secretion. In addition, the hippocampus has also been shown to be a major site of stress-associated glutamate action. The mechanisms which regulate glutamate action and release within this region appear to function downstream of prefrontal cortical processes, constituting a secondary stress-response phase, which, unlike the pFC, is sensitive to neuroendocrine modulation (Moghaddam, 2002). The pronounced glutamatergic signature specifically within the pFC, is consistent with observations which implicate prefrontal cortical glutamate signalling as a principle mechanism by which stress-responses are modulated in the brain. Furthermore, the implication of an over-expressed ionotropic glutamate signalling signature within the hippocampal tissues, add further weight to an interpretation of divergent stress-associated glutamatergic profiles. Combined, these observations suggest a

hyperglutamatergic state, within MS individuals, primarily in pFC tissues, but also in the hippocampus. Importantly, given the role of prefrontal cortical glutamate-driven activation of the HPAA (and subsequent release of glucocorticoids), a hyperglutamatergic state within MS individuals could, in part, explain the observed differences in stress-induced corticosterone profiles.

Several other observations, within pFC and Hic samples, further substantiate an inference of divergent stress-related states between MS and SH groups. In the pFC samples, *Ptger1* was differentially over-expressed in MS samples. Noteworthy is the observation that acute restraint stress leads to a significant increase in prostaglandin E2 (Garcia-Bueno *et al.*, 2008), a natural ligand of *Ptger1*, in the CNS. Activation of *Ptger1*, in turn, has been implicated in dopamine turnover in the pFC, with a specific role in stress-related behavioural inhibition having been suggested (Matsuoka *et al.*, 2005). Within hippocampal samples, the coordinated enrichment of functional terms describing stress-related behavioural processes was found. Specifically, the general over-expression of both behavioural defense response (GO:0002209) and fear response (GO:0042596) terms, within the MS group, denoted clearly the coordinated augmentation of genes whose products participate in behavioural responses to stress. Furthermore, both *Adm* and *Chrna2* have been implicated in the regulation of stress-related behaviours. Increased *Adm* levels are associated with stress exposure and its activity has been shown to result in HPAA activation (Shan *et al.*, 2003). A regulatory role in the stress-response has been proposed for *Adm*, with observations highlighting significant increases in anxiety-like behaviours in mice with *Adm* deficiencies (Fernández *et al.*, 2008). In the current study, *Adm* was over-expressed in MS individual, suggesting a possible response to enhanced stress levels within this group. *Chrna2*, a nicotinic acetylcholine receptor

(nAchr), binds acetylcholine (Ach) and nicotine, and was under-expressed in MS samples. The activation of nAchrs have been associated with the modulation of stress-related mechanisms in the CNS, including the release of several neurotransmitters, the activation of stress hormone pathways and monoaminergic transmission (Picciotto *et al.*, 2002). This modulation, however, is complex as these receptors function as heteromeric pentamers, with different subunit compositions resulting in divergent functions (Shytle, 2002). Interestingly, the unique localisation of the *Chrna2* subunit in GABAergic interneurons, within distinct hippocampal regions has been found, with a role in the regulation GABAergic neurotransmission having been suggested for these subunits (Nakauchi *et al.*, 2007). Therefore, expression differences in this gene might reflect yet another contribution to the earlier inferred deficits in GABAergic mechanisms. However, the exact role of *Chrna2* in the modulation of GABA release has not been sufficiently substantiated to include this gene as part of the previous discussion.

At this point it is pertinent to consider the extensively informed *a priori* assumption of MS-associated altered glucocorticoid receptor (GR) expression within the hippocampus (see Chapter 2). In this study, these receptors were not found to be differentially expressed between the two groups. Interestingly, a microarray study by Weaver *et al.* (2006), whose group made some of the initial contributions regarding insights into the MS-associated alteration of GR expression in the hippocampus, also failed to report significantly altered GR levels, in the hippocampus of MS rats. The lack of concordance between microarray results and results obtained by means of more focussed techniques is not uncommon. On this topic, Mirnics *et al.* (2006) emphasise the cautious interpretation of such findings as definitively negative, as

false negative observations occur frequently in microarray experiments due to the limitations of the technique (discussed earlier).

In line with pFC and Hic samples, hypothalamic tissues also displayed divergent stress-related gene expression profiles between MS and SH individuals. Importantly, several genes whose products are known to participate in the regulation of HPAA activity were found to be differentially expressed between the two experimental groups. Both *Hcrt* and *Pmch* were over-expressed in MS samples. Significantly, increased levels of *Hcrt* mRNA has been characterised in the hypothalamus in response to stress (Ida *et al.*, 2000). *Hcrt* is the neuropeptide precursor to Orexin-A and Orexin-B (collectively referred to as the hypocretins or orexins). A role for hypocretins/orexins in stress-responses has been characterised in neuroendocrine and cardiovascular studies (Ferguson and Samson, 2003). Of particular relevance to this study, is the observation that orexins modulate arousal associated with restraint stress, stimulating the HPAA, with the subsequent release of corticosterone (Stricker-Krongrad and Beck, 2002). *Pmch* is processed to produce Melanin-concentrating hormone (Mch), a hormone which has been implicated in the stress-related activation of the HPAA. Specifically, Mch has been shown to have a stress-inducing or anxiogenic effect on mouse behaviour (Smith *et al.*, 2005), with its action resulting in the release of both ACTH and corticosterone (Kennedy *et al.*, 2003). Interestingly, experimental observations have characterised functional interactions between the Mch and orexin systems within the hypothalamus, where it has been suggested that Mch activity function to regulate and fine-tune hypocretin-mediated excitation and arousal (van den Pol *et al.*, 2004).

The mRNA for two other neuroactive peptides, *Cck* and *Cort*, were found to be under-expressed in MS samples. The involvement of *Cck* in stress- and anxiety-

related responses has been extensively characterised, with studies demonstrating the Cck-induced activation of the HPAA and increased serum levels of ACTH and corticosterone. Particularly relevant to this study, is the observation that MS increased HPAA sensitivity, measured as serum corticosterone concentrations, to Cck, in rats (Greisen *et al.*, 2005). Therefore, the observed under-expression of *Cck* in MS individuals could possibly reflect and adaptation to this increased sensitivity. Lastly, *Cort*, a functional analogue of *somatostatin* (*Sst*), showed decreased levels of mRNA in MS samples. Although no specific role in the regulation of stress-responses has been described for *Cort*, several lines of evidence point to a possible role in the regulation of HPAA activity. Firstly, *Cort* has been found to inhibit glutamate induced responses, through the activation of *Sst* receptors in the hypothalamus (Vassilaki *et al.*, 1999). This observation is particularly relevant, as glutamate signalling is known to result in HPAA activation (Herman *et al.*, 2004). Therefore deficits in *Cort* could result in enhanced glutamate-mediated activation of the HPAA. This observation follows the hyperglutamatergic theme inferred from pFC and Hic samples, but is purely speculative. Additionally, Broglio *et al.* (2008) cite clinical evidence that implicate an inhibitory role for *Cort* (also called *CST*) in ACTH and corticosterone secretion in patients with Cushing's disease (CD), a disorder characterised by significantly elevated cortisol (the principle glucocorticoid in humans) levels. It was found that *Cort* and *Sst* treatment exerted similar effects, resulting in a reduction in both cortisol and ACTH, in CD patients only, and not in controls. This data suggest an inhibitory role for *Cort* in response to elevated levels of HPAA hormones. Extending these insights to the current study, a deficit in *Cort* could potentially result in the inefficient down-regulation of corticosterone levels subsequent to stress, leading to a hyperactive stress-induced HPAA profile. Notably,

this interpretation complements the implication of the inference drawn from a *Cort*-related deficit in glutamate inhibition.

A final and potentially significant observation relates to evidence that have demonstrated the direct involvement of *Hcrt*, *Pmch* and *Cck* in the regulation of feeding behaviour (Rodgers *et al.*, 2002; Forray, 2003; and Fan *et al.*, 2004, respectively). Additionally, the indirect regulatory involvement of *Cort* was inferred from observations which demonstrated an inhibitory action on ghrelin secretion (Broglio *et al.*, 2002); ghrelin is known to mediate important aspects of feeding behaviour (Nakazato *et al.*, 2001). In all vertebrates, feeding behaviour is significantly affected by stress (Carr, 2002), and as such, alterations in the mechanisms that regulate these behaviours could provide irnportant information regarding stress-responses. Particularly relevant to these observations, is the previous characterisation of MS-associated alterations in feeding behaviour, in rats (Penke *et al.*, 2001). Therefore, within the context of this study, the differential expression of these genes underscores differences in pathways that regulate an important stress-related behavioural phenomenon, which is taken as further evidence of divergences in the stress-responsive profiles of MS and SH individuals.

The above discussion highlighted important differences in the transcriptome profiles of pFC, Hic and Hyp tissues between MS and SH mice. Specifically, emphasis was directed at several important stress-related genes and regulatory themes, which provide evidence of functional divergences in mechanisms that mediate stress-responsivity. Importantly, many of these differences have specific implications for the stress-induced activation of the HPAA, providing a logical link to the significant treatment-associated restraint-induced corticosterone profiles.

3.7.2.2 PBMC tissues

PBMC gene expression was evaluated to determine whether the mRNA populations derived from these tissues exhibited treatment-specific profiles. Many genes were found to be differentially expressed between MS and SH individuals, with functional enrichment analyses revealing specific themes. Most importantly, however, the PBMC transcriptome was found to contain sufficient information to efficiently predict the treatment class of individual samples.

Prior to a discussion of PBMC gene expression profiles, a brief consideration of the potential effects of stress on the immune system and a critical limitation of the current experimental design will help delineate boundaries for the interpretation of results. Stress has been noted to have wide-ranging effects on immune-function, differentially affecting specialised cell populations. Specifically, maintained chronic stress states have been found to result in general immunosuppressive outcomes, whilst brief periods of acute stress have been demonstrated to enhance certain aspects of immune function (Dhabhar, 2002; Pruett, 2001). These effects are thought to be mediated via both neuroendocrine mechanisms and the autonomic nervous system (see Chapter 1). In the current study, the use of acute restraint prior to sample collection hold an important implication for the interpretation of immune system gene expression profiles, given the features of MS models. Specifically, MS has been shown to have long term effects on various aspects of neuroendocrine function, which in turn, can lead to the modification or adaptation of immune system parameters. These features significantly complicate an interpretation of the relative contributions of the long-term implications of MS, relative to the short-term contributions of acute restraint, to the observed PBMC gene expression profiles. As such, gene expression profiles are interpreted in general terms, with an emphasis

only on evidence suggesting functional differences in PBMC profiles; no attempt is made to dissect the specific long- (MS) and short-term (restraint) contributions to the PBMC transcriptome.

A large number of genes (418) were found to be differentially expressed between MS and SH individuals (see Supplementary material S.3.4) and included several genes whose products are known to be important modulators of immune system function. Examples include *forkhead box P3* (*Foxp3*), an essential modulator of T cell function (Kasprowicz *et al.*, 2003); *interleukin 17 receptor A* (*IL-17ra*), the receptor target for the IL-17 mediated inflammatory pathway (Gaffen, 2008); protein tyrosine phosphatase, receptor type, C (*Ptprc*) (also known as *Cd45*), which has been shown to play a critical role in lymphocyte activation and development (Hermiston *et al.*, 2003); and *chemokine (C-C motif) ligand 5* (*Ccl5*) (also known as *Rantes*), which regulates the activity of several cellular populations within the immune system (Grayson and Holtzman, 2006). A diverse range of biological functions were displayed by the set of differentially expressed genes, however, no statistically significant theme emerged. Interestingly, the majority of differentially expressed genes were over-expressed in the MS group. This was in contrast to neural tissues, where differential gene expression appeared to be approximately symmetrical. An evaluation of the functional nature of these genes highlighted a theme involving mainly signalling, but also transcriptional processes (Fig 3.5). Within these classes, genes whose products regulate both positive and negative aspects (stimulatory or inhibitory) of these processes were found. These observations suggest that the asymmetry displayed by these genes does not reflect either the general activation or inhibition of immune-system processes, but a more complex

functional state. These profiles are likely as a result of the overlapping consequences of a stress-response time course (Sapolsky *et al.*, 2000).

A surprising and unexpected observation was the identification of several differentially expressed olfactory receptor (OR) genes. Although the expression of these receptors have recently been characterised in specialised non-olfactory tissues (Feldmesser *et al.*, 2006), their function in the immune system remain unclear. Interestingly, an important role for these receptors in sperm chemotaxis was described, suggesting functionalities in addition to odorant perception (Spher *et al.*, 2003). It has been noted that the OR family contain genes whose products bind a diverse range of ligands including odorants, hormones and neurotransmitters. Furthermore, the characterisation of *odorant response abnormal 4* (*ODR-4*) expression, a gene critically required for OR trafficking and targeting, in several non-olfactory tissues, including activated T-cells, provide further insights into functional OR-mediated signalling mechanisms outside of the olfactory epithelium (Lehman *et al.*, 2005). Although the identification of these receptors could be artefactual and a consequence of the technical noise associated with microarray techniques, it is possible that they participate functionally in regulatory signalling cascades within the immune system, binding and mediating responses to various immunoactive ligands, which include hormones and neurotransmitters. Given the lack of other such experimental observations this interpretation is highly speculative, but could represent a novel insight into regulatory functions of these receptors within the immune system.

The functional enrichment analysis of all genes provided important insights into more subtle, but significantly coordinated differences in functional themes (Fig. 3.8D). A general interpretation of these results suggest a functional shift in the

immune system, characterised by the coordinated down-regulation of energy requiring processes, such as protein synthesis and transport, an interpretation substantiated further by the general down-regulation of terms associated with energy metabolism, in favour of increased signalling activities, possibly in response to the activation of signalling components by stress-induced chemical messengers. This functional shift might reflect the well characterised mobilisation of energy and inhibition of further storage in response to stress (Sapolsky *et al.*, 2000). Interestingly, the identification of neurologically-related terms (GO:0008188 and GO:0050877), suggest the involvement of neuroactive products in this observed functional shift. This is a significant observation given an established stress-related interaction between neuroactive compounds and peripheral immune tissues (see Chapter 1). Lastly, the general over-expression of immune-related terms provides an important contextualisation for these themes, specifically emphasising an immune response context.

The above observations suggest important functional differences between the PBMC-derived transcriptional profiles of the two experimental groups. However, the primary aim of this study was an evaluation of the ability to efficiently classify and predict differential stress-related neural states. The evidence obtained from the corticosterone assay and the neural transcriptomes, combined with behavioural characterisations (see Chapter 2) suggest, strongly, that pre-weaning treatment (MS or SH) resulted in differential stress-related profiles. Given this context, the sufficiency of information inherent in PBMC samples was evaluated in terms of the ability to derive accurate predictions of pre-weaning sample statuses.

Using a leave-one-out cross validation strategy, in combination with the KNN-classification algorithm, 19 out of 20 samples (95%) were accurately classified with a

minimum of 50 genes (Fig. 3.13 and Table 3.5). Notably, only one sample, belonging to the SH group, was found to be consistently misclassified (Table 3.6). Of the 50 genes included in the predictor, 46 were functionally annotated. An evaluation of functional enrichment within this set revealed no statistically significant themes. Of particular interest, however, was the identification of 3 genes, *Oxt*, *Cck* and *Adcy8* (all over-expressed), whose products are known to be important mediators of stress- and anxiety-associated behaviours (*Oxt* - Windle *et al.*, 1997 and Ring *et al.*, 2006; *Cck* - Greisen *et al.*, 2005; *Adcy8* – Schaefer *et al.*, 2000). Both Oxt and Cck are neuroactive hormones with previously described endogenous immunomodulatory properties (Csaba and Pállinger, 2007; Meng *et al.*, 2002, respectively). Notably, *Cck* was found to be differentially expressed in hypothalamic samples, although in the opposite direction. Other genes, with known immunomodulatory functions included *Thpo*, whose gene product effects changes in T-lymphocyte compositions and blood cytokine levels (Zhao *et al.*, 1998); *Stim1*, a gene critically involved in development and function of regulatory T cells (Treg) (Oh-hora *et al.*, 2008); and *Edg5*, shown to negatively regulate platelet derived growth factor induced motility and proliferation (Goparaju *et al.*, 2005), all of which were over-expressed. Furthermore, the *Nosip* gene was identified, an inhibitor of the endothelial nitric oxide synthetase (eNos) (Schleicher *et al.*, 2005; Dedio *et al.*, 2001), which is expressed in peripheral blood lymphocytes. Nitric oxide production, in turn, is associated with T-cell activation, and its inhibition significantly impairs this process (Nagy *et al.*, 2007). Lastly, although identified in the differentially expressed gene lists, the inclusion of three over-expressed olfactory receptors (*Olfr1495*, *Olfr66* and *Olfr669*) within the predictor set was surprising. A speculative role for these receptors was described earlier.

It is clear from the above results, that the transcriptional profiles of peripheral immune tissues do indeed contain sufficient information for the efficient diagnostic prediction of stress-related neural states. The significance of individual genes within this predictor, and their functional relationship to stress-induced differences in neural and endocrine systems is uncertain. However, it is possible that the products of these genes participate in complex manners in pathways that are particularly sensitive to stress-induced regulations of the immune system. Finally, the inclusion of several genes with important immunoregulatory roles, in addition to *Oxt*, *Cck* and *Adcy8*, emphasise a functional link between neural- an immune-components within stress-contexts.

- CHAPTER 4 -

Conclusions and recommendations

The complexity of psychiatric disorders combined with the inaccessibility of the brain, present unique challenges for the understanding and efficient diagnosis of these conditions. However, in recent years, evidence has begun to accumulate which suggest the possibility of indirectly or non-invasively interrogating neural states. The development of high-throughput molecular techniques, such as microarray technology, combined with the completion of whole genome sequences for several organisms, has made such novel approaches possible. Specifically, researchers have demonstrated the ability to accurately predict neuropsychiatric states from peripherally derived transcriptional profiles; a very promising prospect for psychiatric diagnostics. In the current study, the validity of this strategy was confirmed and extended. Microarray gene expression analyses demonstrated that stress-related quantitative differences in mRNA expression levels in PBMC populations were paralleled by stress-related transcriptional differences in CNS tissue targets, and that the peripheral transcriptome contain sufficient information for the accurate prediction of stress-related neural states.

As its primary goal, this study set out to evaluate the hypothesis that gene expression profiles of circulating PBMCs is predictive of stress-related neural states. A mouse model of early childhood trauma, maternal separation, with aetiological constructs relevant to stress- and anxiety disorders was used. Importantly, an animal model allowed for a multidimensional approach to the problem, including the concomitant evaluation of changes in peripheral and neural transcriptomes. A

comprehensive description of stress-related differences between maternally separated (MS) and control (SH) groups involving behavioural, endocrine and neuronal-transcriptional data provided an elaborated context for the interpretation of PBMC-derived diagnostic inferences of stress-related neural states.

Firstly, behavioural tests were conducted to characterise the effects of early MS on the stress-related behavioural profiles of adult animals. Given the literature on MS, it was anticipated that this model would affect the development of key neural structures associated with stress- and anxiety-related behaviours, producing divergent neurophysiological profiles. The rationale behind the behavioural characterisation was based on the assumption that differences in specific stress-related parameters are indicative of divergences in the neural mechanisms responsible for mediating these behaviours.

Behavioural results provided evidence of treatment-associated differences, primarily in risk assessment behaviours, a critical component of stress-related exploration behaviour. Importantly, the results emphasised a complex time-dependent dimension to these behaviours in the EPM environment, which was not apparent when evaluated with the analytical expectations inherited from rat MS studies. In contrast, divergences in risk assessment behaviours in the OF were immediate and sustained, probably as a result of the more aversive nature of this environment, in this study. A critique was given of the shortcomings of conventional analysis regimes in the identification of the more intricate behavioural patterns of mice. It was suggested that studies such as those by Millstein and Holmes (2007) and Parfitt *et al.* (2007), which failed to find any effect of MS on adult mouse behaviour, might benefit from analysis strategies which take cognisance of time as a significant behavioural dimension. Furthermore, the observed latency in behavioural

divergences in this study, along with previously characterised effects of MS on aspects of learning and memory, suggest that an extended test session (beyond 5 min) would facilitate a more detailed dissection of time-dependent aspects of behaviour within stress-related contexts.

It should be stated, that the behavioural differences described between the MS and SH groups, although found to be statistically significant, were of a modest nature. Results gained from future MS-context behavioural studies, which exmploy the analytical approach described here, should provide further insight into the robustness, or lack thereof, of these findings.

Having established differences in behavioural parameters typically associated with stress- and anxiety-behaviours; stress-related corticosterone profiles, in addition to transcriptome differences of three selected brain regions and PBMCs, were evaluated in MS and SH mice. An important feature of this component of the study was the employment of acute restraint stress, prior to sample collections. It was anticipated that the MS model would result in the altered stress-responsivity of animals, specifically in the HPAA mediated stress-response. Corticosterone profiles confirmed this expectation. While basal corticosterone levels were similar between the two groups, post-restraint concentrations were significantly elevated in MS animals. It was suggested that this difference could result, in part, from deficits in GR feedback mechanisms, a previously characterised feature of MS models. Additionally, an alternative interpretation was posited, suggesting that differences in the initial stress-response, resulting from differences in stress-related neurotransmitter systems, rather than deficits in feedback mechanisms, could produce an amplified HPAA response. This interpretation is supported by observations in the MS literature, which has described quantitative differences in

components involved in stress-associated neurotransmission, including the glutamatergic and GABAergic systems.

Combined, corticosterone and behavioural data provide evidence of divergences in the stress-related profiles of the MS and SH mice use in this study. Furthermore, the corticosterone data allow for some speculations regarding the observed behavioural patterns in the EPM and OF environments. As noted, the OF used in this study presented animals with a highly aversive environment, characterised by high intensity illumination and a very small arena space. The immediate and sustained differences in risk assessment behaviours, suggested a rapid stress-response; likely as a result of the immediate activation of stress circuitries, including the HPAA. In contrast, behaviours in the EPM were characterised by an initial latency period, during which MS and SH animals exhibited similar behaviours. Due to the spatial complexity and the relatively less aversive nature (low intensity illumination) of the EPM environment, it was suggested that this period might rely heavily on spatial appraisals and cognitive assessments by the animals. Comments by Korte and de Boer (2002) provided support for this interpretation. Such a scenario would likely result in a somewhat delayed activation of stress circuitries, pending the outcome of evaluations by animals. This initial period would therefore probably be characterised by a relatively inactive HPAA. However, once a stress-response is fully activated, differences in the regulation of stress circuitries could result in behavioural differences and explain, importantly, the observed latency in divergences, only seen during the last two minutes of the EPM tests. Although these interpretations are speculative, they provide a logical integration of the observed similarities in basal corticosterone concentrations and the

latency in increased stress-like behaviours seen for MS individuals in this environment.

Following an analysis of corticosterone profiles, microarrays were used to assess whole-genome transcriptional profiles of PBMCs and key neural structures involved in the limbic stress pathway. Importantly, the use of microarrays allowed for an unbiased discovery-driven evaluation, not dependent on *a priori* hypotheses, of potentially complex differences within prefrontal cortical, hippocampal and hypothalamic tissues - regions with critical functions in the regulation of stress-responses. Emphasis was directed at genes and/or functional themes which have previously been implicated in stress-response pathways. Combined the neural tissues provided substantial evidence of functional divergences in mechanisms that mediate stress-responsivity. The pFC and Hic tissues provided evidence of a stress-induced hyperglutamatergic state in MS samples. Moreover, coordinated gene expression changes within the hippocampal samples suggested differences in functional classes specifically associated with stress-related behavioural phenomena; a significant observation given the context of this study. Hypothalamic samples highlighted differences in genes whose products have been shown to mediate aspects of stress-induced HPAA activity. Importantly, the neural transcriptional data suggested divergences in mechanisms which could result in the stress-induced differences seen for the corticosterone response.

Within this context of established stress-related differences in neural transcriptomes, PBMC gene expression data was interpreted. Firstly, a large number of genes were found to be differentially expressed between the two groups. Although no single functional term was found to be significantly enriched, these genes were shown to participate, mainly in signalling, but also in transcriptional processes. An

analysis of coordinated gene expression suggested a functional shift in the immune system, characterised by the coordinated down-regulation of energy requiring processes, along with energy metabolism, in favour of increased signalling activities, possibly in response to the activation of signalling components by stress-induced chemical messengers. Lastly, it was shown that PBMC gene expression profiles were predictive, with high efficiency, of the treatment status of individual samples. Using a leave-one-out cross validation strategy, in combination with a KNN-classification algorithm, a set of 50 genes was found to provide sufficient information to predict sample classes with 95% accuracy. An alternative classification algorithm, SVM, also achieved this accuracy; however, a minimum of 125 genes was needed. Although both algorithms converged at this set of 125 genes, focus was directed at the 50 gene set, as it represented, in clinical terms, a logistically more manageable diagnostic collection. This set of 50 genes was shown to include not only genes with immunoregulatory roles, but also, and of particular relevance, genes whose products are known to be important mediators of stress- and anxiety-associated behaviours.

Combined, the behavioural, corticosterone and neural transcriptional data provided substantial evidence which indicate that PBMC transcriptional profiles do indeed reflect differences in the stress-related neural states, and that these differences can be exploited diagnostically.

The microarray data demonstrated the efficacy of PBMC-derived transcriptomic classifications of stress-related neural states, however, some important limitations of the current study is worth noting. Firstly, although it is assumed that the observed transcriptional profiles reflect differences in stress-responsivity as a consequence of MS, the extent to which MS and restraint contributed, respectively, to gene expression responses could, however, not be

dissected. As such, no statements regarding the effect of MS alone, on the neural transcriptome, could be made. The exclusion of an acute stressor, prior to sample collections, would facilitate a more focussed assessment of the long-term effects of MS on the neural transcriptome. Secondly, while the results highlighted some of the discovery-based benefits of microarray techniques within psychiatric settings, certain limitations were also encountered. Specifically, the extensively characterised MS-associated deficit in GR expression in the hippocampus was not confirmed in this study. It was noted that, rather than interpreting this finding as a negative result, it more likely reflects some of the limitations associated with the microarray technique, which were discussed in Chapter 3. Therefore, although many genes were found to be differentially regulated between MS and SH individuals, the results only represent genes for which differences were detectable given the limitations of microarray based assays. Thirdly, it was noted that the relationship between mRNA and protein levels is not linear, and that changes in mRNA populations cannot be assumed to result in altered protein expression. Therefore, in this study the physiological implications inferred from transcriptional profiles are speculative in nature; however it is reasonable to assert that changes in mRNA levels provide a functional basis by which alterations in protein expression might be brought about.

Lastly, future work will require the qRT-PCR validation of microarray-derived gene expression results; independently confirming the results as real and not as artefactual of the microarray technique or data analysis strategy. In addition, the relevance of the predictor set of 50 genes within clinical settings, specifically within a context of stress-disorders, needs to be evaluated. Although there are important differences between humans and mice, many of the fundamental features of stress-responses, including neural circuitries and homeostatic mechanisms, are conserved

across mammals. As such, the predictor set might represent genes which are generally sensitive and highly responsive to stress-induced alterations in mammalian homeostasis, but needs to be clinically assessed. Finally, a question not addressed in this study, but highly relevant, is whether PBMC derived diagnostic inferences provide a sufficient resolution to discern between closely related neuropsychiatric disorders. That is, to what extent are these cell populations sensitive to, and reflective of, minor differences in psychiatric dysfunction? Although Tsuang *et al.* (2005) demonstrated the ability to distinguish schizophrenia from bipolar patients; most studies have assessed peripheral transcriptional diagnostics within single disorder contexts. Therefore, future studies should consider evaluating, in parallel, the PBMC derived classification of several closely related psychiatric disorders.

In conclusion, the work presented here extends current insights into the potential of PBMCs as non-invasive diagnostic indicators of stress-related neural states. Although there are important differences in the stress-response mechanisms between humans and mice, the results strongly support the previously demonstrated viability of such an approach within psychiatric settings. Importantly, the results established that stress-related differences in the PBMC transcriptome are indeed paralleled by differences in the neural transcriptome.

- CHAPTER 5 -

References

Abbot, A. (2008). Psychiatric genetics: The brains of the family. *Nature* **454**: 154-157.

Abel, T. and Lattal, K. M. (2001). Molecular mechanisms of memory acquisition, consolidation and retrieval. *Current Opinion in Neurobiology* **11**(2): 180-187.

Achiron, A., Gurevich, M., Friedman, N., Kaminski, N. and Mandel, M. (2004). Blood transcriptional signatures of multiple sclerosis: Unique gene expression of disease activity. *Annals of Neurology* **55**: 410-417.

Adcock, I.M., and Caramori, G. (2001). Cross-talk between pro-inflammatory transcription factors and glucocorticoids. *Immunology and Cell Biology* **79**(4): 376-384.

Adcock, I.M., Ito, K. and Barnes, P.J. (2004). Glucocorticoids: Effects on Gene Transcription. *Proceedings of the American Thoracic Society* **1**: 247-254.

Aisa, B., Tordera, R., Lasheras, B., Del Río, J. and Ramírez, M.J. (2007). Cognitive impairment associated to HPA axis hyperactivity after maternal separation in rats. *Psychoneuroendocrinology* **32**(3): 256-266.

Alexopoulos, T. and Ric, F. (2007). The evaluation-behavior link: Direct and beyond valence. *Journal of Experimental Social Psychology* **43**(6): 1010-1016.

Al-Shahrour, F., Minguez, P., Tárraga, J., Montaner, D., Alloza, E., Vaquerizas, J.M., Conde, L., Blaschke, C., Vera, J. and Dopazo, J. (2006). BABELOMICS: a systems biology perspective in the functional annotation of genome-scale experiments. *Nucleic Acids Research* **34**: W472-476.

Ambion Technote 12:3. Improved Microarray Sensitivity using Whole Blood RNA Samples. URL, http://www.ambion.com/techlib/tn/123/10.html, Accessed: 03/12/08

Auer, H., Lyianarachchi, S., Newsom, D., Klisovic, M.I., Marcucci, G. and Kornacker, K. (2003). Chipping away at the chip bias: RNA degradation in microarray analysis. *Nature Genetics* **35**(4): 292-293.

Avissar, S. and Schreiber, G. (2002). Toward molecular diagnostics of mood disorders in psychiatry. *Trends in Molecular Medicine* **8**(6): 294-300.

Baigent, S. M. and Lowry, P.J. (2000). mRNA expression profiles for corticotrophin-releasing factor (CRF), urocortin, CRF receptors and CRF-binding protein in peripheral rat tissues. *Journal of Molecular Endocrinology* **25**: 43-52.

Bakshi, V.P. and Kalin, N. (2002). Animal models and endophenotypes of anxiety and stress disorders. In: Davis, K.L., Charney, D., Coyle, J.T. and C. Nemeroff, (eds.), Neuropsychopharmacology: The Fifth Generation of Progress, Lippincott, Williams and Wilkins, Philadelphia (2002), pp. 445–456.

Barkhudaryan, N. and Dunn, A. (1999). Molecular Mechanisms of Actions of Interleukin-6 on the Brain, with Special Reference to Serotonin and the hypothalamo-Pituitary-Adrenocortical Axis. *Neurochemical Research* **24**(9): 1169-1180.

Barlow, C. and Lockhart, D. J. (2002). DNA arrays and neurobiology -- what's new and what's next? *Current Opinion in Neurobiology* **12**(5): 554-561.

Bartholome, B., Spies, C. M., Gaber, T., Schuchmann, S., Berki, T., Kunkel, D., Bienert, M., Radbruch, A., Burmester, G.-R., Lauster, R., Scheffold, A. and Buttgereit, F. (2004). Membrane glucocorticoid receptors (mGCR) are expressed in normal human peripheral blood mononuclear cells and up-regulated after in vitro stimulation and in patients with rheumatoid arthritis. *The FASEB Journal* **18**: 70-80.

Belzung, C. and Griebel, G. (2001). Measuring normal and pathological anxiety-like behaviour in mice: a review. Behavioural Brain Research **125**(1-2): 141-149.

Bernik, T.R., Friedman, S.G., Ochani, M., DiRaimo, R., Ulloa, L., Yang, H., Sudan, S., Czura, C.J., Ivanova, S.M. and Tracey, K.J. (2002). Pharmacological Stimulation of the Cholinergic Antiinflammatory Pathway. *The Journal of Experimental Medicine* **195**: 781-788.

Bertoglio, L.J. and Carobrez, A.P. (2002). Behavioral profile of rats submitted to session 1-session 2 in the elevated plus-maze during diurnal/nocturnal phases and under different illumination conditions. *Behavioural Brain Research* **132**(2): 135-143.

Besedovsky, H. and Del Rey, A. (1996). Immune-Neuro-Endocrine Interactions: Facts and Hypotheses. *Endocrine Reviews* **17**(1): 64-102.

Bloom, F.E. and Morales, M. (1998). The Central 5-HT3 Receptor in CNS Disorders. *Neurochemical Research* **23**(5): 653-659.

Borovikova, L.V., Ivanova, S., Zhang, M., Yang, H., Botchkina, G.I., Watkins, L.R., Wang, H., Abumrad, N., Eaton, J.W. and Tracey, K.J. (2000). Vagus nerve stimulation attenuates the systemic inflammatory response to endotoxin. *Nature* **405**(6785): 458-462.

Bowden, N.A., Weidenhofer, J., Scott, R.J., Schal, U., Todd, J., Michie, P.T. and Tooney, P.A. (2006). Preliminary investigation of gene expression profiles in peripheral blood lymphocytes in schizophrenia. *Schizophrenia Research* **82**(2-3): 175-183.

Bradshaw, D., Groenewald, P., Laubscher, R., Nannan, N., Nojilana, B., Norman, R., Pieterse, D., Schneider, M., Bourne, D.E., Timaeus, I.M., Dorrington, R. and Johnson, L. (2003). Initial Burden of Disease Estimates for South Africa, 2000. *South African Medical Journal* **93**(9): 682-8.

Broglio, F., Koetsveld, Pv.P., Benso, A., Gottero, C., Prodam, F., Papotti, M., Muccioli, G., Gauna, C., Hofland, L., Deghenghi, R., Arvat, E., Van Der Lely, A.J. and Ghigo, E. (2002). Ghrelin Secretion Is Inhibited by Either Somatostatin or Cortistatin in Humans. *The Journal of Endocrinology and Metabolism* **87**(10):4829-32.

Broglio, F., Grottoli, S., Arvat, E. and Ghigo, E. (2008). Endocrine actions of cortistatin: In vivo studies. *Molecular and Cellular Endocrinology* **286**(1-2): 123-127.

Brustolim, D., Ribeiro-dos-Santos, R., Kast, R.E., Altschuler, E.L. and Soares, M.B.P. (2006). A new chapter opens in anti-inflammatory treatments: The antidepressant bupropion lowers production of tumor necrosis factor-alpha and interferon-gamma in mice. *International Immunopharmacology* **6**(6): 903-907.

Buckley, T. C., Blanchard, E. B. and Trammell, M.W. (2000). Information processing and ptsd: A review of the empirical literature. *Clinical Psychology Review* **20**(8): 1041-1065.

Caldji, C., Francis, D., Sharma, S., Plotsky, P.M. and Meaney, M.J. (2000). The Effects of Early Rearing Environment on the Development of GABAA and Central Benzodiazepine Receptor Levels and Novelty-Induced Fearfulness in the Rat. *Neuropsychopharmacology* **22**: 219-229.

Carobrez, A.P. and Bertoglio, L.J. (2005). Ethological and temporal analyses of anxiety-like behavior: The elevated plus-maze model 20 years on. *Neuroscience & Biobehavioral Reviews* **29**(8): 1193-1205.

Carola, V., D'Olimpio, F., Brunamonti, E., Mangia, F. And Renzi, P. (2002). Evaluation of the elevated plus-maze and open-field tests for the assessment of anxiety-related behaviour in inbred mice. *Behavioural Brain Research* **134**(1-2): 49-57.

Carr, J. A. (2002). Stress, Neuropeptides, and Feeding Behavior: A comparative perspective. *Integrative and Comparative Biology* **42**: 582-590.

Choleris, E., Thomas, A.W., Kavaliers, M. and Prato, F.S. (2001). A detailed ethological analysis of the mouse open field test: effects of diazepam, chlordiazepoxide and an extremely low frequency pulsed magnetic field. *Neuroscience & Biobehavioral Reviews* **25**(3): 235-260.

Conesa, A., Götz, S., García-Gómez, JM., Terol, J., Talón, M. and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**(18): 3674-3676.

Crabbe, J. C., Wahlsten, D. and Dudek, B.C. (1999). Genetics of Mouse Behavior: Interactions with Laboratory Environment. *Science* **284**(5420): 1670-1672.

Crawley, J.N., Belknap, J.K. Collins, A., Crabbe, J.C., Frankel, W., Henderson, N., Hitzeman, R.J., Maxson, S.C., Miner, L.L., Silva, A.J., Wehner, J.M., Wynshaw-Boris, A. and Paylor, R. (1997). Behavioral phenotypes of inbred mouse strains: implications and recommendations for molecular studies. *Psychopharmacology* **132**(2):107-124

Crestani, F., Lorez, M. Baer, K, Essrich, C. Benke, D., Laurent, J.P., Belzung, C., Fritschy, J., Lüscher, B. and Mohler, H. (1999). Decreased GABAA-receptor clustering results in enhanced anxiety and a bias for threat cues. *Nature Neuroscience* **2**(9): 833 – 839.

Cryan, J.F. and Holmes, A. (2005). The ascent of mouse: advances in modelling human depression and anxiety. *Nature Reviews: Drug Discovery* **4**(9): 775-790.

Csaba, G. and Pállinger, É. (2007). In vitro effect of hormones on the hormone content of rat peritoneal and thymic cells. Is there an endocrine network inside the immune system? *Inflammation Research* **56**(11): 447-451.

Czura, C.J. and Tracey, K.J. (2005). Autonomic neural regulation of immunity. *Journal of Internal Medicine* **257**(2): 156-166.

Daniels, W.M.U., Pietersen, C.Y., Carstens, M.E. and Stein, D.J. (2004). Maternal Separation in Rats Leads to Anxiety-Like Behavior and a Blunted ACTH Response and Altered Neurotransmitter Levels in Response to a Subsequent Stressor. *Metabolic Brain Disease* **19**(1): 3-14.

Dantzer, R., Bluthé, R.M., Kent, S. and Kelly, K.W. Cytokines and sickness behavior. In: Husband, A.J., (editor), *Psychoimmunology: CNS-Immune Interactions*. Boca Raton: CRC Press, Inc., 1993; pp. 1-16.

Dantzer, R. (2004a). Innate immunity at the forefront of psychoneuroimmunology. *Brain, Behavior, and Immunity* **18**(1): 1-6.

Dantzer, R. (2004b). Cytokine-induced sickness behaviour: a neuroimmune response to activation of innate immunity. *European Journal of Pharmacology* **500**(1-3): 399-411.

de Kloet, C.S., Vermetten, E., Bikker, A., Meulman, E., Geuze, E., Kavelaars, A., Westenberg, H.G.M. and Heijnen, C.J. (2007). Leukocyte glucocorticoid receptor expression and immunoregulation in veterans with and without post-traumatic stress disorder. *Molecular Psychiatry* **12**(5): 443-453.

de Kloet, E.R., Sibug, R.M., Helmerhorst, F.M. and Schmidt, M. (2005). Stress, genes and the mechanism of programming the brain for later life. *Neuroscience & Biobehavioral Reviews* **29**(2): 271-281.

Dedio, J., Konig, P., Wohlfart, P., Schroeder, C., Kummer, W. And Muller-Esterl, W. (2001). NOSIP, a novel modulator of endothelial nitric oxide synthase activity. *The FASEB Journal* **15**: 79-89.

Delgado, M. (2003). VIP: a very important peptide in T helper differentiation. *Trends in Immunology* **24**(5): 221–4.

Dhabhar, F.S. (2002). Stress-induced augmentation of immune function - The role of stres hormones, leukocyte trafficking and cytokines. *Brain Behaviour, and Immunity* 16(6): 785-798.

Dopazo, J. (2006). Functional interpretation of microarray experiments. *OMICS* **10**(3): 398-410.

Dubé, B., Benton, T., Cruess, D.G., Evans, D.L. (2005). Neuropsychiatric manifestations of HIV infection and AIDS. *Journal of Psychiatry and Neuroscience* 30(4): 237-246.

Elenkov, I.J., Wilder, R.L., Chrousos, G.P. and Vizi, E.S. (2000). Sympathetic Nerve - An Integrative Interface between Two Supersystems: The Brain and the Immune System. *Pharmacological Reviews* **52**(4): 595-638.

Elitsur, Y., Luk, G.D., Colberg, M., Gesell, M.S., Dosescu, J. and Moshier, J.A. (1994). Neuropeptide Y (NPY) enhances proliferation of human colonic lamina propria lymphocytes. *Neuropeptides* **26**(5): 289-295.

Espejo, E.F. (1997). Structure of the mouse behaviour on the elevated plus-maze test of anxiety. *Behavioural Brain Research* **86**(1): 105-112.

Fabricius, K., Wörtwein, G. and Pakkenberg, B. (2008). The impact of maternal separation on adult mouse behaviour and on the total neuron number in the mouse hippocampus. *Brain Structure and Function* 212(5): 403-416.

Fan, W., Ellacott, K.L.J, Halatchev, I.G., Takahashi, K., Yu, P. and Cone, R.D. (2004). Cholecystokinin-mediated suppression of feeding involves the brainstem melanocortin system. *Nature Neuroscience* **7**(4): 335-336.

Feldmesser, E., Olender, T., Khen, M., Yanai, I., Ophir, R. and Lancet, D. (2006). Widespread ectopic expression of olfactory receptor genes. *BMC Genomics*, **7**: 121.

Felten, S.Y. and Felten D.L. Innervation of lymphoid tissues. In: Ader, R., Felten, D., Cohen, N. (eds.), Psychoneuroimmunology, 2nd Ed. New York: Academic Press; 1991, pp. 429–446.

Ferguson, A.V. and Samson, W.K. (2003). The orexin/hypocretin system: a critical regulator of neuroendocrine and autonomic function. *Frontiers in Neuroendocrinology* **24**(3): 141-150.

Fernández, A. P., Serrano, J., Tessarollo, L., Cuttitta, F. and Martínez, A. (2008). Lack of adrenomedullin in the mouse brain results in behavioral changes, anxiety, and lower survival under stress conditions. *Proceedings of the National Academy of Sciences of the USA* **105**(34): 12581-12586.

Finn, D.A., Rutledge-Groman, M.T., and Crabbe, J.C. (2003). Genetic animal models of anxiety. *Neurogenetics* **4**:109-135.

Forray, C. (2003). The MCH receptor family: feeding brain disorders? *Current Opinion in Pharmacology* **3**(1): 85-89.

Francis, D.D., Diorio, J., Plotsky, P.M. and Meaney, M.J. (2002). Environmental Enrichment Reverses the Effects of Maternal Separation on Stress Reactivity. *The Journal of Neuroscience* **22**(18): 7840-7843.

Gaffen, S.L. (2008). An overview of IL-17 function and signaling. *Cytokine* **43**(3):402-407.

Garcia-Bueno, B., Madrigal, J.L.M., Perez-Nievas, B.G. and Leza, J.C. (2008). Stress Mediators Regulate Brain Prostaglandin Synthesis and Peroxisome Proliferator-Activated Receptor-{gamma} Activation after Stress in Rats. *Endocrinology* **149**(4): 1969-1978.

Gene Ontology Consortium. (2008). The Gene Ontology project in 2008. Nucleic Acids Research **36**: D440-444.

Gentleman, R., Carey, V., Bates, D., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J. and Zhang, J. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* **5**(10): R80.

Goparaju, S.K., Jolly, P.S., Watterson, K.R., Bektas, M., Alvarez, S., Sarkar, S., Mel, L., Ishii, I., Chun, J., Milstien, S. and Spiegel, S. (2005). The S1P2 Receptor Negatively Regulates Platelet-Derived Growth Factor-Induced Motility and Proliferation. *Molecular and Cellular Biology* **25**(10): 4237-4249.

Grayson, M.H. and Holtzman, M.J. (2006). Chemokine Complexity: The Case for CCL5. *American Journal of Respiratory Cell and Molecular Biology* **35**(2): 143-146.

Greisen, M.H., Bolwig, T.G. and Wörtwein, G. (2005). Cholecystokinin tetrapeptide effects on HPA axis function and elevated plus maze behaviour in maternally separated and handled rats. *Behavioural Brain Research* **161**(2): 204-212.

Griffin, R., Mills, C., Costigan, M. and Woolf, C. (2003). Exploiting microarrays to reveal differential gene expression in the nervous system. *Genome Biology* **4**(2): 105.

Grueter, B.A. and Winder, D.G. (2005). Group II and III Metabotropic Glutamate Receptors Suppress Excitatory Synaptic Transmission in the Dorsolateral Bed Nucleus of the Stria Terminalis. *Neuropsychopharmacology* **30**(7): 1302-1311.

Guo, L., Lobenhofer, E., Wang, C., Shippy, R., Harris, S.C., Zhang, L., Mei, N., Chen, T., Herman, D., Goodsaid, F.M., Hurban, P., Phillips, K.L., Xu, J., Deng, X., Sun, Y.A., Tong, W., Dragan, Y.P. and Shi, L. (2006). Rat toxicogenomic study reveals analytical consistency across microarray platforms. *Nature Biotechnology* **24**(9): 1162-1169.

Gutman, D.A. and Nemeroff, C.B. (2003). Persistent central nervous system effects of an adverse early environment: clinical and preclinical studies. *Physiology & Behavior* **79**(3): 471-478.

Hanisch, U.K., Seto, D. and Quirion, R. (1993). Modulation of hippocampal acetylcholine release: a potent central action of interleukin-2. *The Journal of Neuroscience* **13**(8): 3368-3374.

Heim, C. and Nemeroff, C.B. (2001). The role of childhood trauma in the neurobiology of mood and anxiety disorders: preclinical and clinical studies. *Biological Psychiatry* **49**(12): 1023-1039.

Herman, J.P, Mueller, N.K. and Figueiredo, H. (2004). Role of GABA and Glutamate Circuitry in Hypothalamo-Pituitary-Adrenocortical Stress Integration. *Annals of the New York Academy of Sciences* **1018**: 35-45.

Hermiston, M. L., Xu, Z. and Weiss, A. (2003). CD45: A Critical Regulator of Signaling Thresholds in Immune Cells. *Annual Review of Immunology* **21**: 107-137.

Holland, N.T, Smith, M.T, Eskenazi, B. and Bastaki, M. (2002). Biological sample collection and processing for molecular epidemiological studies. *Mutation Research* **543**: 217-234.

Holmes, A., Guisquet, A., Vogel, E., Millstein, R., Leman, S. and Belzung, C. (2005). Early life genetic, epigenetic and environmental factors shaping emotionality in rodents. *Neuroscience and Biobehavioral Reviews* **29**(8): 1335-1346.

Hood, S.D., Hince, D.A., Robinson, H., Cirillo, M., Christmas, D. and Kaye, J.M. (2006). Serotonin regulation of the human stress response. *Psychoneuroendocrinology* **31**(9): 1087-1097.

Hopkins, S.J. (2007) Central nervous system recognition of peripheral inflammation: a neural, hormonal collaboration. *Acta Biomedica* **78**(Supplement 1): 231-247.

Hovatta, I. and Barlow, C. (2008). Molecular genetics of anxiety in mice and men. *Annals of Medicine* **40**(2): 92-109.

Huang, H., Winter, E., Wang, H., Weinstock, K., Xing, H., Goodstadt, L., Stenson, P., Cooper, D., Smith, D., Alba, M.M., Ponting, C. and Fechtel, K. (2004). Evolutionary conservation and selection of human disease gene orthologs in the rat and mouse genomes. *Genome Biology* **5**: R47.

Ida, T., Nakahara, K., Murakami, T., Hanada, R., Nakazato, M. and Murakami, N. (2000). Possible Involvement of Orexin in the Stress Reaction in Rats. *Biochemical and Biophysical Research Communications* **270**(1): 318-323.

Jeffery, I., Higgins, D. and Culhane, A. (2006). Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data. *BMC Bioinformatics* **7**: 359.

Jhamandas, J.H., Simonin, F., Bourguignon, J.J. and Harris, K.H. (2007). Neuropeptide FF and neuropeptide VF inhibit GABAergic neurotransmission in parvocellular neurons of the rat hypothalamic paraventricular nucleus. *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology* **292**: R1872-R1880.

Jordaan, W. and Jordaan, J. People in context, 3rd Edition, 2nd Impression. Johannesburg : Heinemann; 2000.

Kalueff, A.V. and Tuohimaa, P. (2005). Mouse grooming microstructure is a reliable anxiety marker bidirectionally sensitive to GABAergic drugs. *European Journal of Pharmacology* **508**(1-3): 147-153.

Kaminski, N. and Friedman, N. (2002). Practical approaches to analyzing results of microarray experiments. *American Journal of Respiratory and Cell Molecular Biology* **27**: 125–132.

Karalis, K., Muglia, L.J., Bae, D., Hilderbrand, H. and Majzoub, J.A. (1997). CRH and the immune system. *Journal of Neuroimmunology* **72**: 131-136.

Kasprowicz, D.J., Smallwood, P.S., Tyznik, A.J. and Ziegler, S.F. (2003). Scurfin (FoxP3) Controls T-Dependent Immune Responses In Vivo Through Regulation of CD4+ T Cell Effector Function. *Journal of Immunology* **171**(3): 1216-1223.

Kaushik, N., Fear, D., Richards, S.C.M., McDermott, C.R., Nuwaysir, E.F., Kellam, P., Harrison, T.J., Wilkinson, R.J., Tyrrell, D.A.J., Holgate, S.T. and Kerr, J.R. (2005). Gene expression in peripheral blood mononuclear cells from patients with chronic fatigue syndrome. *Journal of Clinical Pathology* **58**(8): 826-832.

Kawai, T., Morita, K., Masuda, K., Nishida, K., Shikishima, M., Ohta, M., Saito, T. and Rokutan, K. (2007). Gene expression signature in peripheral blood cells from medical students exposed to chronic psychological stress. *Biological Psychology* **76**(3): 147-155.

Kawashima, K. and Fujii, T. (2003). The lymphocytic cholinergic system and its biological function. *Life Sciences* **72**(18-19): 2101-2109.

Kennedy, A. R., Todd, J. F., Dhillo, W. S., Seal, L. J., Ghatei, M. A., O'Toole, C. P., Jones, M., Witty, D., Winborne, K., Riley, G., Hervieu, G., Wilson, S. and Bloom, S. R. (2003). Effect of Direct Injection of Melanin-Concentrating Hormone into the Paraventricular Nucleus: Further Evidence a Stimulatory Role in the Adrenal Axis via SLC-1. *Journal of Neuoendocrinology* **15**: 268-272.

Kobayashi, H., Fukata, J., Murakami, N., Usui, T., Ebisui, O., Muro, S., Hanaoka, I., Inoue, K., Imura, H. and Nakao, K. (1997). Tumor necrosis factor receptors in the pituitary cells. *Brain Research* **758**(1-2): 45-50.

Konradi, C. (2005). Gene expression microarray studies in polygenic psychiatric disorders: Applications and data analysis. *Brain Research Reviews* **50**(1): 142-155.

Korte, S.M. and De Boer, S.F. (2003). A robust animal model of state anxiety: fear-potentiated behaviour in the elevated plus-maze. *European Journal of Pharmacology* **463**(1-3): 163-175.

Kuhn, C. and Schanberg, S. (1998). Responses to Maternal Separation: Mechanisms and Mediators. *International Journal of Developmental Neuroscience* 16(3): 10.

Kulmatycki, K.M. and Jamali, F. (2006). Drug disease interactions: role of inflammatory mediators in depression and variability in antidepressant drug response. *The Journal of Pharmacy and Pharmaceutical Sciences* **9**(3): 292-306.

LaBar, K.S. and Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience* **7**(1): 54-64.

Lee, J., Kim, H.J., Kim, J.G., Ryu, V., Kim, B., Kang, D.J. and Jeong W. (2007). Depressive behaviors and decreased expression of serotonin reuptake transporter in rats that experienced neonatal maternal separation. *Neuroscience Research* **58**(1): 32-39.

Lehman, C.W., Lee, J.D.R. and Komives, C.F. (2005). Ubiquitously expressed GPCR membrane-trafficking orthologs. *Genomics* **85**(3): 386-391.

Le-Niculescu, H., Kurian, S.M., Yehyawi, N., Dike, C., Patel, S.D., Edenberg, H.J., Tsuang, M.T., Salomon, D.R., Nurnberger, J.I., Jr. and Niculescu, A.B. (2008). Identifying blood biomarkers for mood disorders using convergent functional genomics. *Molecular Psychiatry* Feb 26;[Epub ahead of print].

Mantella, R.C., Vollmer, R.R., Li, X. and Amico, J.A. (2003). Female Oxytocin-Deficient Mice Display Enhanced Anxiety-Related Behavior. *Endocrinology* **144**(6): 2291-2296.

Mark, G.P., Rada, P.V. and Shors, T.J. (1996). Inescapable stress enhances extracellular acetylcholine in the rat hippocampus and prefrontal cortex but not the nucleus accumbens or amygdala. *Neuroscience* **74**(3): 767-774.

Marques-Deak, A. Cizza G. and Sternberg E. (2005) Brain-immune interactions and disease susceptibility. *Molecular Psychiatry* **10**: 239-250.

Marx, V. (2004). RNA Quality: Defining the Good, the Bad, the Ugly. Cover Story, Genomics and Proteomics.

Mašek, K., Slánský, J., Petrovický, P. and Hadden, J. (2003). Neuroendocrine immune interactions in health and disease. *International Immnophamacology* **3**(8): 1235-1246.

Mathers, C. D. and Loncar, D. (2006). Projections of Global Mortality and Burden of Disease from 2002 to 2030. *PLoS Medicine* **3**(11): e442.

Matsuoka, Y., Furuyashiki, T., Yamada, K., Nagai, T., Bito, H., Tanaka, Y., Kitaoka, S., Ushikubi, F., Nabeshima, T. and Narumiya, S. (2005). Prostaglandin E receptor EP1 controls impulsive behavior under stress. *Proceedings of the National Academy of Sciences of the USA* **102**(44): 16066-16071.

Meaney, M. (2001). Maternal care, gene expression, and the transmission of individual differences in stress reactivity across generations. *Annual Review of Neuroscience* **24**: 1161-1192.

Meaney, M.J. and Szyf, M. (2005). Maternal care as a model for experience-dependent chromatin plasticity? *Trends in Neurosciences* **28**(9): 456-463.

Medina, I., Montaner, D., Tárraga, J. and Dopazo, J. (2007). Prophet, a web-based tool for class prediction using microarray data. *Bioinformatics* **23**(3): 390-391.

Melmed, R.N. Mind, Body and Medicine: An integrative text. New York: Oxford University Press; 2001.

Meng, A.H., Ling, Y.L., Zhang, X.P., Zhao, X.Y. and Zhang, J.L. (2002). Anti-inflammatory effect of cholecystokinin and its signal transduction mechanism in endotoxic shock rat. *World Journal of Gastroenterology* **8**(4): 712-717.

MicroArray Quality Control Consortium (2006). The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nature Biotechnology* **24**(9): 1151-1161.

Miller, A.H., Spencer, R.L., Pearce, B.D., Pisell, T.L., Azrieli, Y., Tanapat, P., Moday, H., Rhee, R. and McEwen, B.S. (1998). Glucocorticoid Receptors Are Differentially Expressed in the Cells and Tissues of the Immune System. *Cellular Immunology* **186**(1): 45-54.

Millstein, R.A. and Holmes, A. (2007). Effects of repeated maternal separation on anxiety- and depression-related phenotypes in different mouse strains. *Neuroscience & Biobehavioral Reviews* **31**(1): 3-17.

Mirnics, K., Levitt, P. and Lewis, D.A. (2006). Critical Appraisal of DNA Microarrays in Psychiatric Genomics. *Biological Psychiatry* **60**(2): 163-176.

Mirnics, K., and Pevsner, J. (2004). Progress in the use of microarray technology to study the neurobiology of disease. *Nature Neuroscience* **7**(5): 434 – 439.

Moghaddam, B. (2002). Stress activation of glutamate neurotransmission in the prefrontal cortex: implications for dopamine-associated psychiatric disorders. *Biological Psychiatry* **51**(10): 775-787.

Montaner, D., Tárraga, J., Huerta-Cepas, J., Burguet, J., Vaquerizas, J.M., Conde, L., Minguez, P., Vera, J., Mukherjee, S., Valls, J., Pujana, M.A., Alloza, E., Herrero, J., Al-Shahrour, F. and Dopazo, J. (2006). Next station in microarray data analysis: GEPAS. *Nucleic Acids Research* **34**(Web Server issue).

Morgan, C.A., Krystal, J.H. and Southwick, S.M. (2003). Toward early pharmacological posttraumatic stress intervention. *Biological Psychiatry* **53**(9): 834-843.

Moynihan, J., Kruszewska, B., Madden, K. and Callahan, T. (2004). Sympathetic nervous system regulation of immunity. *Journal of Neuroimmunology* **147**(1-2): 87-90.

Munck, A. (2005). Glucocorticoid receptors and physiology: A personal history. *Steroids* **70**(4): 335-344.

Murray, C.J.L, and Lopez, A.D., (1996). Global Burden of Disease: A comprehensive assessment of Mortality and Morbidity from Disease, Injuries and Risk Factors in 1990 and projected to 2020, Vol I. Harvard: World Health Organization. Harvard: World Health Organization.

Myint, A.M. and Kim, Y.K. (2003). Cytokine-serotonin interaction through IDO: a neurodegeneration hypothesis of depression. *Medical Hypotheses* **61**(5-6): 519-525.

Nadeau, S. and Rivest, S. (1999). Effects of circulating tumor necrosis factor on the neuronal activity and expression of the genes encoding the tumor necrosis factor receptors (p55 and p75) in the rat brain: a view from the blood-brain barrier. *Neuroscience* **93**(4): 1449-1464.

Nagy, G., Clark, J.M., Buzás, E.I., Gorman, C.L. and Cope, A.P. (2007). Nitric oxide, chronic inflammation and autoimmunity. *Immunology Letters* **111**(1): 1-5.

Nakauchi, S., Brennan R.J., Boulter, J. and Sumikawa, K. (2007). Nicotine gates long-term potentiation in the hippocampal CA1 region via the activation of alpha2* nicotinic ACh receptors. *European Journal of Neuroscience* **25**(9): 2666-2681.

Nakazato, M., Murakami, N., Date, Y., Kojima, M., Matsuo, H., Kangawa, K. and Matsukura, S. (2001). A role for ghrelin in the central regulation of feeding. *Nature* **409**(6817): 194-198.

Nance, D.M. and Sanders, V.M. (2007). Autonomic Innervation and Regulation of the Immune System (1987-2007). *Brain, Behaviour, and Immunity* **21**(6): 736-745.

Newport, D.J. and Nemeroff, C. B. (2000). Neurobiology of Posttraumatic Stress Disorder. *Current Opinion in Neurobiology* **10**: 211–218.

Nueda, M., Conesa, A., Westerhuis, J., Hoefsloot, H., Smilde, A., Talon, M. and Ferrer, A. (2007). Discovering gene expression patterns in time course microarray experiments by ANOVA–SCA. *Bioinformatics* **23**(14): 1792-1800.

Nygaard, V., Holden, M., Loland, A., Langaas, M., Myklebost, O. and Hovig, E. (2005). Limitations of mRNA amplification from small-size cell samples. *BMC Genomics* **6**: 147.

Nygaard, V. and Hovig, E. (2006). Options available for profiling small samples: a review of sample amplification technology when combined with microarray profiling. *Nucleic Acids Research* **34**(3): 996-1014.

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H. and Kanehisa, M. (1999). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **27**(1): 29-34.

Ognibene, E., Adriani, W.,Caprioli, A., Ghirardi, O., Ali, S.F., Aloe, L. and Laviola, G. (2008). The effect of early maternal separation on brain derived neurotrophic factor and monoamine levels in adult heterozygous reeler mice. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* **32**(5): 1269-1276.

Oh-hora, M., Yamashita, M., Hogan, P.G., Sharma, S., Lamperti, E., Chung, W., Prakriya, M., Feske, S. and Rao, A. (2008). Dual functions for the endoplasmic reticulum calcium sensors STIM1 and STIM2 in T cell activation and tolerance. *Nature Immunology* **9**(4): 432 – 443.

Oliveros, J.C. (2007) VENNY. An interactive tool for comparing lists with Venn Diagrams. URL http://bioinfogp.cnb.csic.es/tools/venny/index.html.

Osterweil, E., Wells, D. G. and Mooseker, M.S. (2005). A role for myosin VI in postsynaptic structure and glutamate receptor endocytosis. *Journal of Cell Biology* **168**(2): 329-338.

Pacák, K. and Palkovits, M. (2001). Stressor Specificity of Central Neuroendocrine Responses: Implications for Stress-Related Disorders. *Endocrine Reviews* **22**(4): 502-548.

Padgett, D.A. and Glaser, R. (2003). How stress influences the immune response. *Trends in Immunology* **24**(8): 444-448.

Parfitt, D.B., Levin, J.K., Saltstein, K.P., Klayman, A.S., Greer, L.M. and Helmreich, D.L. (2004). Differential early rearing environments can accentuate or attenuate the responses to stress in male C57BL/6 mice. *Brain Research* **1016**(1): 111-118.

Parfitt, D.B., Walton, J.R., Corriveau, E.A. and Helmreich, D.L. (2007). Early life stress effects on adult stress-induced corticosterone secretion and anxiety-like behavior in the C57BL/6 mouse are not as robust as initially thought. *Hormones and Behaviour* **52**(4): 417-426.

Pavlov, V.A. and Tracey, K.J. (2004). Neural regulators of innate immune responses and inflammation. *Cellular and Molecular Life Sciences* **61**(18): 2322-2331.

Paxinos, G., and Franklin, K.B.J. The mouse brain in Stereotaxic Coordinates. California: Elsevier Academic Press; 2004.

Peedicayil, J. (2008). Epigenetic biomarkers in psychiatric disorders. *British Journal of Pharmacology* **155**(6): 795-796.

Penke, Z., Fernette, B., Nyakas, C., Max, J. P. and Burlet, A. (2002). Neonatal maternal deprivation modifies feeding in response to pharmacological and behavioural factors in adult rats. *Neuropharmacology* **42**(3): 421-427.

Picciotto, M.R., Brunzell, D.H. and Caldarone, B.J. (2002). Effect of nicotine and nicotinic receptors on anxiety and depression. *Neuroreport* **13**(9): 1097-1106.

Pickering, C., Gustafsson, L., Cebere, A., Nylander, I. and Liljequist, S. (2006). Repeated maternal separation of male Wistar rats alters glutamate receptor expression in the hippocampus but not the prefrontal cortex. *Brain Research* **1099**(1): 101-108.

Ploj, K., Roman, E. and Nylander, I. (2003). Long-term effects of maternal separation on ethanol intake and brain opioid and dopamine receptors in male wistar rats. *Neuroscience* **121**(3): 787-799.

Plotsky, P.M. and Meaney, M.J. (1993). Early, postnatal experience alters hypothalamic corticotropin-releasing factor (CRF) mRNA, median eminence CRF content and stress-induced release in adult rats. *Brain research. Molecular Brain Research* 18(3): 195-200.

Pozo, D. and Delgado, M. (2004). The many faces of VIP in neuroimmunology: a cytokine rather a neuropeptide? *The FASEB Journal* **18**(12): 1325-1334.

Pruett, S.B. (2001). Quantitative aspects of stress-induced immunomodulation. *International Immunopharmacology* **1**(3): 507-520.

Puig, M.V., Santana, N., Celada, P., Mengod, G. and Artigas, F. (2004). In Vivo Excitation of GABA Interneurons in the Medial Prefrontal Cortex through 5-HT3 Receptors. *Cerebral Cortex* **14**(12): 1365-1375.

Quackenbush, J. (2002). Microarray data normalization and transformation. *Nature Genetics* Suppl. **32**: 496-501.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rabin, S.B. Stress, Immune Function and Health: The connection. New York: John Wiley & Sons, Inc; 1999.

Ramos, A. (2008). Animal models of anxiety: do I need multiple tests? *Trends in Pharmacological Sciences* **29**(10): 493-498.

Rice, C.J., Sandman, C.A., Lenjavi, M.R. and Baram, T.Z. (2008). A novel mouse model for acute and long-lasting consequences of early life stress. *Endocrinology* **149**(10): 4892-4900.

Ring, R., Malberg, J., Potestio, L., Ping, J., Boikess, S., Luo, B., Schechter, L., Rizzo, S., Rahman, Z. and Rosenzweig-Lipson, S. (2006). Anxiolytic-like activity of oxytocin in male mice: behavioral and autonomic evidence, therapeutic implications. *Psychopharmacology* **185**(2): 218-225.

Rodgers, R.J. and Johnson, N.J.T. (1995). Factor analysis of spatiotemporal and ethological measures in the murine elevated plus-maze test of anxiety. *Pharmacology Biochemistry and Behavior* **52**(2): 297-303.

Rodgers, R.J., Johnson, N.J.T., Cole, J.C., Dewar, C.V., Kidd, G.R. and Kimpson, P.H. (1996). Plus-maze retest profile in mice: Importance of initial stages of trial 1 and response to post-trial cholinergic receptor blockade. *Pharmacology Biochemistry and Behavior* **54**(1): 41-50.

Rodgers, R.J., Cao, B.J., Dalvi, A. and Holmes, A. (1997). Animal models of anxiety: an ethological perspective. *Brazilian Journal of Medical and Biological Research* **30**(3): 289-304.

Rodgers, R.J., Ishii, Y., Halford, J.C.G. and Blundell, J.E. (2002). Orexins and appetite regulation. *Neuropeptides* **36**(5): 303-325.

Romeo, R., Mueller, A., Sisti, H., Ogawa, S., McEwen, B. and Brake, W. (2003). Anxiety and fear behaviors in adult male and female C57BL/6 mice are modulated by maternal separation. *Hormones and Behaviour* **43**(5): 561-567.

Rothwell, N.J. and Hopkins, S.J. (1995). Cytokines and the nervous system II: actions and mechanisms of action. *Trends in Neurosciences* **18**(3): 130-136.

Rothwell, N.J. (1999). Cytokines - killers in the brain? *The Journal of Physiology* **514**: 3-17.

Runne, H., Kuhn, A., Wild, E., Pratyaksha, W., Kristiansen, M., Isaacs, J., Regulier, E., Delorenzi, M., Tabrizi, S. and Luthi-Carter, R. (2007). Analysis of potential transcriptomic biomarkers for Huntington's disease in peripheral blood. *Proceedings of the National Academy of Sciences of the USA* **104**(36): 14424-14429.

Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A., Kostukovich, E., Borisovsky, I., Liu, Z., Vinsavich, A., Trush, V.and Quackenbush, J. (2003). TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* **34**: 374–378.

Sala, M., Perez, J., Soloff, P., Ucelli di Nemi, S., Caverzasi, E., Soares, J. C. and Brambilla, P. (2004). Stress and hippocampal abnormalities in psychiatric disorders. *European Neuropsychopharmacology* **14**(5): 393-405.

Salas, M.A., Brown, O.A., Perone, M.J., Castro, M.G. and Goya, R.G. (1997). Effect of the Corticotrophin Releasing Hormone Precursor on Interleukin-6 Release by Human Mononuclear Cells. *Clinical Immunology and Immunopathology* **85**(1): 35-39.

Sanchez, M.M., Ladd, C.O., and Plotsky, P.M. (2001). Early adverse experience as a developmental risk factor for later psychopathology: Evidence from rodent and primate models. *Development and Psychopathology* **13**(3): 419-449.

Sapolsky, R.M., Romero, L.M. and Munck, A.U. (2000). How Do Glucocorticoids Influence Stress Responses? Integrating Permissive, Suppressive, Stimulatory, and Preparative Actions. *Endocrine Reviews* **21**(1): 55-89.

Schaefer, M.L., Wong, S.T., Wozniak, D.F., Muglia, L.M., Liauw, J.A., Zhuo, M., Nardi, A., Hartman, R.E., Vogt, S.K., Luedke, C.E., Storm, D.R. and Muglia, L.J. (2000). Altered Stress-Induced Anxiety in Adenylyl Cyclase Type VIII-Deficient Mice. *The Journal of Neuoscience* **20**(13): 4809-4820.

Schleicher, M., Brundin, F., Gross, S., Muller-Esterl, W. And Oess, S. (2005). Cell Cycle-Regulated Inactivation of Endothelial NO Synthase through NOSIP-Dependent Targeting to the Cytoskeleton. *Molecular and Cellular Biology* **25**: 8251-8258.

Schmidt, M., Oitzl, M.S., Levine, S. and de Kloet, E.R. (2002). The HPA system during the postnatal development of CD1 mice and the effects of maternal deprivation. *Developmental Brain Research* **139**(1): 39-49.

Schroeder, A., Mueller, O., Stocker, S., Reudiger, S., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M. and Ragg, T. (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology* **7**:3.

Schwarz, E. and Bahn, S. (2008). The utility of biomarker discovery approaches for the detection of disease mechanisms in psychiatric disorders. *British Journal of Pharmacology* **153**(S1): S133-S136.

Segman, R., Shefi, N., Goltser-Dubner, T., Friedman, N., Kaminski, N. and Shalev, A. (2005). *Peripheral blood mononuclear cell gene expression profiles identify emergent post-traumatic stress disorder among trauma survivors*. *Molecular Psychiatry* **10**(5): 500-513.

Selye, H. The stress of life. New York: McGraw-Hill; 1956.

Shan, J., Stachniak, T., Jhamandas, J.H. and Krukoff, T.L. (2003). Autonomic and neuroendocrine actions of adrenomedullin in the brain: mechanisms for homeostasis. *Regulatory Peptides* **112**(1-3): 33-40.

Shepherd, A.J., Downing, J.E.G., and Miyan, J.A. (2005). Without nerves, immunology remains incomplete - in vivo veritas. *Immunology* **116**(2): 145-63.

Shi, L., Perkins, R.G., Fang, H. and Tong, W. (2008). Reproducible and reliable microarray results through quality control: good laboratory proficiency and appropriate data analysis practices are essential. *Current Opinion in Biotechnology* **19**(1): 10-18.

Shytle, R.D., Silver, A.A., Lukas, R.J., Newman, M.B., Sheehan, D.V. and Sanberg, P.R. (2002). Nicotinic acetylcholine receptors as targets for antidepressants. *Molecular Psychiatry* **7**(6): 525-535.

Simon, R., Radmacher, M., Dobbin, K. and McShane, L. (2003). Pitfalls in the Use of DNA Microarray Data for Diagnostic and Prognostic Classification. *Journal of the National Cancer Institute* **95**(1): 14-18.

Slonim, D.K. (2002). From patterns to pathways: gene expression data analysis comes of age. *Nature Genetics* Suppl. **32**: 502-508.

Smith, D.G., Davis, R.J., Rorick-Kehn, L., Morin, M., Witkin, J.M., McKinzie, D.L., Nomikos, G.G. and Gehlert, D.R. (2005). Melanin-Concentrating Hormone-1 Receptor Modulates Neuroendocrine, Behavioral, and Corticolimbic Neurochemical Stress Responses in Mice. *Neuropsychopharmacology* **31**(6): 1135-1145.

Smith, E.M., Morrill, A.C., Meyer, W.J. and Blalock, J.E. (1986). Corticotropin releasing factor induction of leukocyte-derived immunoreactive ACTH and endorphins. *Nature* **321**(6073): 881-882.

Smyth, G. K. (2005). Limma: linear models for microarray data. In: Bioinformatics and Computational Biology Solutions using R and Bioconductor, R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, W. Huber (eds.), Springer, New York, pp. 397-420.

Solomon, E. P., Berg L. R. and Martin D. W. Biology, 6th Ed. California: Thomson Learning; 2002.

Soverchia, L., Ubaldi, M., Leonardi-Essmann, F., Ciccocioppo, R. and Hardiman, G. (2005). Microarrays - The Challenge of Preparing Brain Tissue Samples. *Addiction Biology* **10**(1): 5-13.

Spangelo, B.L., MacLeod, R.M. and Isakson, P.C. (1990). Production of interleukin-6 by anterior pituitary cells in vitro. Endocrinology. 126: 582-586.

Spehr, M., Gisselmann, G., Poplawski, A., Riffell, J.A., Wetzel, C.H., Zimmer, R.K. and Hatt, H. (2003). Identification of a testicular odorant receptor mediating human sperm chemotaxis. Science, **299**: 2054-2058.

Stern, C.A.J., Carobrez, A.P. and Bertoglio, L.J. (2008). Aversive learning as a mechanism for lack of repeated anxiolytic-like effect in the elevated plus-maze. *Pharmacology Biochemistry and Behavior* **90**(4): 545-550.

Stjernquist, M., Owman, C., Sjoberg, N.O. and Sundler, F. (1987). Coexistence and cooperation between neuropeptide Y and norepinephrine in nerve fibers of guinea pig vas deferens and seminal vesicle. *Biology of Reproduction* **36**: 149-155.

Stricker-Krongrad, A. and Beck, B. (2002). Modulation of hypothalamic hypocretin/orexin mRNA expression by glucocorticoids. *Biochemical and Biophysical Research Communications* **296**(1): 129-133.

Tang, Y., Lu, A., Aronow, B.J. and Sharp, F.R. (2001). Blood genomic responses differ after stroke, seizures, hypoglycemia, and hypoxia: Blood genomic fingerprints of disease. *Annals of Neurology* **50**(6): 699-707.

Tang, Y., Gilbert, D.L., Glauser, T.A., Hershey, A.D. and Sharp, F.R. (2005). Blood Gene Expression Profiling of Neurologic Diseases: A Pilot Microarray Study. *Archives of Neurology* **62**(2): 210-215.

Tecott, L.H. (2003). The Genes and Brains of Mice and Men. *The American Journal of Psychiatry* **160**: 646-656.

Tibshirani, R., Hastie, T., Narahimhan, B. and Chu, G. (2003). Class Prediction by Nearest Shrunken Centroids, with Application to DNA Microarrays. *Statisical Science* **18**(1): 104–117.

Toye, A.A. and Cox, R (2001). Behavioral genetics: Anxiety under interrogation. *Current Biology* **11**(12): R473-476.

Tracey, K.J. (2002). The inflammatory reflex. *Nature* **420**(6917): 853-859.

Tracey, K.J. (2007). Physiology and immunology of the cholinergic antiinflammatory pathway. *The Journal of Clinical Investigation* **117**(2): 289-296.

Tsapakis, E.M. and Travis, M.J. (2002). Glutamate and psychiatric disorders. *Advances in Psychiatric Treatment* **8**(3): 189-197.

Tsuang, M.T., Nossova, N., Yager, T., Tsuang, M.M., Guo, S.C., Shyu, K.G., Glatt, S.J. and Liew, C.C. (2005). Assessing the validity of blood-based gene expression profiles for the classification of schizophrenia and bipolar disorder: a preliminary report. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics* **133**(B): 1–5.

Turnbull, A.V. and Rivier, C.L. (1999). Regulation of the Hypothalamic-Pituitary-Adrenal Axis by Cytokines: Actions and Mechanisms of Action. *Physiological Reviews* **79**: 1-71.

Tusher, V.G., Tibshirani, R. and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences of the USA* **98**(9): 5116-5121.

van den Pol, A.N., Acuna-Goycolea, C., Clark, K.R. and Ghosh, P.K. (2004). Physiological properties of hypothalamic mch neurons identified with selective expression of reporter gene after recombinant virus infection. *Neuron* **42**(4): 635-652.

van Heerden, J., Walford, S.A., Shen, A and Illing, N. (2007). A framework for the informed normalization of printed microarrays. *South African Journal of Science* **103**(9-10): 381-390.

van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A.M., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., Schreiber, G.J., Kerkhoven, R.M., Roberts, C., Linsley, P.S., Bernards, R. and Friend, S.H. (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**: 530-536.

Vassilaki, A., Lanneau, C., Dournand, P., de Lecea, L., Gardette, R. and Epelbaum, J. (1999). Cortistatin affects glutamate sensitivity in mouse hypothalamic neurons through activation of somatostatin receptor subtype. *Neuroscience* **88**(2): 359-364.

Veenema, A.H., Bredewold, R. and Neumann, I.D. (2007). Opposite effects of maternal separation on intermale and maternal aggression in C57BL/6 mice: link to hypothalamic vasopressin and oxytocin immunoreactivity. *Psychoneuroendocrinology* **32**(5): 437-450.

Vinogradova, O.S. (2001). Hippocampus as comparator: Role of the two input and two output systems of the hippocampus in selection and registration of information. *Hippocampus* **11**(5): 578-598.

Wang, J. and Dunn, A.J. (1998). Mouse interleukin-6 stimulates the HPA axis and increases brain tryptophan and serotonin metabolism. *Neurochemistry International* **33**(2): 143-154.

Ward, H., Vigues, S., Poole, S. and Bristow, A.F. (2001). The rat interleukin 10 receptor: cloning and sequencing of cDNA coding for the alpha-chain protein sequence, and demonstration by western blotting of expression in the rat brain. *Cytokine* **15**(5): 237-240.

Weaver, I.C., Cervoni, N., Champagne, F.A., D'Alessio, A.C., Sharma, S., Seckl, J.R., Dymov, S., Szyf, M. and Meaney, M.J. (2004). Epigenetic programming by maternal behavior. *Nature Neuroscience* **7**(8): 847-584.

Weaver, I.C., Meaney, M.J. and Szyf, M. (2006). Maternal care effects on the hippocampal transcriptome and anxiety-mediated behaviors in the offspring that are reversible in adulthood. *Proceedings of the National Academy of Sciences of the USA* **103**(9): 3480-3485.

Webster, E.L. and De Souza, E.B. (1988). Corticotropin-releasing factor receptors in mouse spleen: identification, autoradiographic localization, and regulation by divalent cations and guanine nucleotides. *Endocrinology* **122**(2): 609-617.

Werb, Z., Foley, R. and Munck, A. (1978). Glucocorticoid Receptors and Glucocorticoid-Sensitive Secretion of Neutral Proteinases in a Macrophage Line. *Journal of Immunology* **121**(1): 115-121.

Wilson, K.E., Ryan, M.M., Prime, J.E., Pashby, D.P., Orange, P.R., O'Beirne, G., Whateley, J.G., Bahn, S. and Morris, C.M. (2004). Functional genomics and proteomics: application in neurosciences. *Journal of Neurology, Neurosurgery, and Psychiatry* **75**(4): 529-538.

Windle, R.J., Shanks, N., Lightman, S.L. and Ingram, C.D. (1997). Central Oxytocin Administration Reduces Stress-Induced Corticosterone Release and Anxiety Behavior in Rats. *Endocrinology* **138**(7): 2829-2834.

Wirkner, K., Gunther, A., Weber, M., Guzman, S.J., Krause, T., Fuchs, J., Koles, L., Norenberg, W. and Illes, P. (2007). Modulation of NMDA Receptor Current in Layer V Pyramidal Neurons of the Rat Prefrontal Cortex by P2Y Receptor Activation. *Cerebral Cortex* **17**(3): 621-631.

World Health Organization (2001). The World Health Report 2001, Mental health: new understanding, new hope. World Health Organization, Geneva.

Wrona, D. (2006). Neural-immune interactions: An integrative view of the bidirectional relationship between the brain and immune systems. *Journal of Neuroimmunology* **172**(1-2): 38-58.

Wu, L.J., Ko, S.W., Toyoda, H., Zhao, M.G., Xu, H., Vadakkan, K.I., Ren, M., Knifed, E., Shum, F., Quan, J., Zhang, X.H. and Zhuo, M. (2007). Increased Anxiety-Like Behavior and Enhanced Synaptic Efficacy in the Amygdala of GluR5 Knockout Mice. *PLoS ONE* **2**(1): e167.

Yao, Y., Schröder, J. and Karlsson, H. (2008). Verification of proposed peripheral biomarkers in mononuclear cells of individuals with schizophrenia. *Journal of Psychiatric Research* **42**(8): 639-643.

Zhao, J.Z., Mei, Y.J., Guo, Z.K. and Chen, H.R. (1998). Thrombopoietin: a potential T-helper lymphocyte stimulator. Change in T-lymphocyte composition and blood cytokine levels in thrombopoietin cDNA transferred mice. *Haematologica* **83**(6): 572-573.

Ziegler, D.R. and Herman, J.P. (2002). Neurocircuitry of stress integration: anatomical pathways regulating the hypothalamo-pituitary-adrenocortical axis of the rat. *Integrative and Comparative Biology* **42**(3): 541-551.

Zimring, J.C., Kapp, L.M., Yamada, M., Wess, J. and Kapp, J.A. (2005). Regulation of CD8+ cytolytic T lymphocyte differentiation by a cholinergic pathway. *Journal of Neuroimmunology* **164**(1-2): 66-75.

S.1 CORTICOSTERONE ASSAY AND MICROARRAY SAMPLE SELECTIONS

Due to logistical limitations, such as sample quantity and quality, complete matching of all samples was not possible. A summary of sample id's and matching is provided in the table below.

Table S.1.1 Sample processing summary

Sample ID	Treatment	Corticosterone Assay	Microarray samples			
			pFC	Hic	Hyp	PBMC
5	SH	R				
14	SH	R				
15	SH	R	X			
20	SH	R				
21	**SH**		**X**	**X**	**X**	**X**
28	**SH**		**X**	**X**	**X**	**X**
29	SH	R				
41	SH	R				
48	SH	R	X			
49	SH					X
53	**SH**	**R**	**X**	**X**	**X**	**X**
57	SH	R				X
60	SH					X
61	**SH**		**X**	**X**	**X**	**X**
64	SH	R				X
65	SH	R				X
69	**SH**	**R**	**X**	**X**	**X**	**X**
7	MS	R				
10	MS	R				
16	MS	R				
17	MS	R	X			
19	MS	R				
22	MS	R				
23	MS	R				
26	MS	R				
27	MS	R				

Sample ID	Treatment	Corticosterone Assay	Microarray samples			
			pFC	Hic	Hyp	PBMC
30	MS	R	X			
42	MS	R	X			X
43	**MS**	**R**	**X**	**X**	**X**	**X**
46	MS	R				X
47	**MS**		**X**	**X**	**X**	**X**
50	**MS**	**R**	**X**	**X**	**X**	**X**
54	MS					X
55	MS	R				X
58	**MS**		**X**	**X**	**X**	**X**
63	**MS**	**R**	**X**	**X**	**X**	**X**
66	MS	R				
70	MS	R				
Baseline Corticosterone Samples						
33	SH	B				
35	SH	B				
36	SH	B				
37	SH	B				
72	SH	B				
73	SH	B				
78	SH	B				
80	SH	B				
31	MS	B				
32	MS	B				
38	MS	B				
40	MS	B				
74	MS	B				
75	MS	B				
77	MS	B				
79	MS	B				

M=Maternally Separated ; SH=Simulated Handling ; R=Restrained ; B=Baseline; pFC=Prefrontal Cortex ; Hic=Hippocampus ; Hyp=Hypothalamus ; PBMC=Peripheral Blood Mononuclear Cells ; Bold text indicates microarray matching across all tissues ; Bold text plus shading (grey) indicates complete matching across all molecular assays.

S.2 MICROARRAY PROCESSING

S.2.1 Genepix flagging criteria

A set of logical criteria were defined using Genepix 6.0.27 Pro Software (Axon Instruments / Molecular Devices Corporation, CA, USA), any features that met these criteria were flagged allowing for exclusion or down-weighting in subsequent procedures. The criteria were as follows:

[Name] = "Dye Marker" And
[Flags] = [Clear] Or
[Name] = "Buffer" And
[Flags] = [Clear] Or
[Name] = "null" And
[Flags] = [Clear] Or
[Name] = "EMPTY" And
[Flags] = [Clear] Or
[F Pixels] < 25 And
[Flags] = [Clear] Or
[F532 Median] > 65532 Or
[F635 Median] > 65532 Or
[F532 CV] < 100 And
[F532 Median] < 50 And
[Flags] = [Clear] Or
[F635 CV] < 100 And
[F635 Median] < 50 And
[Flags] = [Clear]

S.2.2 Data Normalization and statistical assessment

S.2.2.1 Generic normalization script

#Script written by J.H. van Heerden

#Intended for use in R with the Bioconductor package, LIMMA

#Set Variables (Note: variable values are examples)

```
RawData = "C:\\Microarray work\\"          #Raw data files location
target_def = "targets.txt"                 #Location of Target defintions files
flag_weights = 0.001                       #Weight assigned to flagged spots
Replicate_list = "controlCode.txt"         #List of features spotted in replicate at least 5 times per slide
backg_method = "none"                      #Background subtraction method to use
Within_array_method = "loess"              #Within Array Normalization method
Between_array_method = "none"              #Between Array Normalization method
Stats_script = "ArrayStats.R"              #Predefined function for calculation of specific data statistics
Hyb_no = 20                                #Number of arrays
```

#Set working directory

```
setwd(RawData)
```

#READ DATA

```
library(limma)
targets<-readTargets(target_def)
targets
RG<-read.maimages(targets$FileName,source="genepix",wt.fun=wtflags(flag_weights))
replicates<-readSpotTypes(Replicate_list)
RG$genes$Status<-controlStatus(replicates,RG)
```

#NO WITHIN OR BETWEEN ARRAY NORMALIZATION

```
normdata.backg <- backgroundCorrect(RG, method=none)
normdata.backg.none<-normalizeWithinArrays(normdata.backg,method="none")
```

```
normdata.backg.none$genes$Status<-controlStatus(replicates,RG)


#Replace flagged features with NA

normdata.backg.none$M[normdata.backg.none$weights != 1] <- NA


#Put Data into generic NoNorm object

NoNormData<-normdata.backg.none


#PRE-NORM VISUALIZATIONS

X11()

boxplot(as.data.frame (NoNormData$M),  main="No Normalization", names=c(targets$Name))

X11()

plotDensities(NoNormData)

X11()

plotMA(NoNormData)


#LOAD STATISTICAL FUNCTIONS (see S2.2.2 for function definitions)

source(Stats_script)


#CREATE MATRIX FOR STORAGE OF STATS SUMMARY OUTPUT

mcolumns = Hyb_no +1

col<-colnames(NoNormData)

FinalSD_Calc <- matrix (,10,mcolumns,dimnames=list(c("ArraySD PreNorm","ArraySD PostNorm","SQR
Difference","AvgRepSD PreNom","AvgRepSD PostNom","Rep Ratio","Average Rep Ratio", "PreNorm Between
Array Correlations","PostNorm Between Array Correlations","SSQR Difference"),c(col,"STDEV")))


#CALULATE WITHIN ARRAY REPLICATE POOL SD

# rd = Raw Data object

# s = Status component (e.g. Data object with Gene Status component - must be called $Status )

# x = number of Replicate groups

# y = number of Arrays

# z = Indicates the first column in the data matrix for which calculation will be done (default = 1)
```

```
NoNormRep_SD<-NNWithinArrayRepSD(NoNormData, NoNormData, 35, Hyb_no)

#PUT DATA INTO GENERIC OBJECT FOR FURTHUR PROCESSING
d <- NoNormRep_SD
m.NoNorm<-FinalSD_Calc

#CALCULATE AVG STDEV FOR REP SETS WITHIN EACH ARRAY
# rstdev = data object (e.g. The matrix with the Stdev values for rep sets)
# m = A matrix object within which stats summary info is stored (see above)
# y = number of arrays
# z = Indicates the first column in the data matrix for which calculation will be done (default = 1)

FinalSD_Calc<-NNAvgArrayRepSD(d,m.NoNorm, Hyb_no,)

#PUT DATA INTO GENERIC OBJECT FOR FURTHUR PROCESSING
m.NoNorm<-FinalSD_Calc

#WHOLE ARRAY STDEV
# rd = data object (e.g. A data object with $M values)
# m = A matrix object within which stats summary info is stored (see above)
# y = number of arrays
# z = Indicates the first column in the data matrix for which calculation will be done (default = 1)

FinalSD_Calc<-NNWholeArraySD(NoNormData,m.NoNorm, Hyb_no,1)

#PUT DATA INTO GENERIC OBJECT FOR FURTHUR PROCESSING
m.NoNorm<-FinalSD_Calc

#CORRELATION PRE-NORM
corrrel.none<-cor(NoNormData$M, use="pairwise.complete.obs")
correl.none
NoNorm_CorAvg<-mean(correl.none)
```

```
m.NoNorm[8,1]<-NoNorm_CorAvg


#NORMALIZATION

#BACKGROUND CORRECTION

normdata.backg <- backgroundCorrect(RG, method=backg_method)

#WITHINARRAY NORMALIZATION

normdata.backg.wnorm<-normalizeWithinArrays(normdata.backg, method=Within_array_method,
weights=RG$weights)

#REPLACE BAD FLAGS WITH NA

normdata.backg.wnorm$M[RG$weights != 1] <- NA


#BETWEEN NORMALIZATION

normdata.backg.wnorm.bnorm<-normalizeBetweenArrays(normdata.backg.wnorm,
method=Between_array_method)

normdata.backg.wnorm.bnorm$genes$Status<-controlStatus(replicates,RG)

#PUT DATA INTO GENERIC OBJECT FOR FURTHUR PROCESSING

NormData<-normdata.backg.wnorm.bnorm


#POST-NORM VISUALIZATIONS

X11()

boxplot(as.data.frame (NormData$M),  main="No Normalization",names=c(targets$Name))

X11()

plotDensities(NormData)

X11()

plotMA(NormData)


#POST-NORM ARRAY STATISTICS

#CALULATE WITHIN ARRAY REPLICATE POOL SD

# nd = Normalized Data object

# s = Status object (e.g. RG$genes$Status)

# x = number of Replicate groups

# y = number of Arrays

# z = Indicates the first column in the data matrix for which calculation will be done (default = 1)
```

 # PNWithinArrayRepSD <- function(nd, s, x, y, z=1)

NormRep_SD<-PNWithinArrayRepSD(NormData, NormData, 35, Hyb_no)

d <- NormRep_SD

m.Norm<-m.NoNorm

#CALCULATE AVG STDEV FOR REP SETS WITHIN EACH ARRAY

 # d = object with SD values

 # m = A matrix object within which stats summary info is stored (see above)

 # y = number of arrays

 # z = Indicates the first column in the data matrix for which calculation will be done (default = 1)

 # PNAvgArrayRepSD <- function (d, m, y, z=1)

FinalSD_Calc<-PNAvgArrayRepSD(d,m.Norm, Hyb_no,)

#PUT DATA INTO GENERIC OBJECT FOR FURTHUR PROCESSING

m.Norm<-FinalSD_Calc

 # WHOLE ARRAY STDEV

 # nd = Normalized Data object

 # m = A matrix object within which stats summary info is stored (see above)

 # y = number of arrays

 # z = Indicates the first column in the data matrix for which calculation will be done (default = 1)

 # PNWholeArraySD <- function(nd,m,y,z)

FinalSD_Calc<- PNWholeArraySD (NormData,m.Norm, Hyb_no,1)

#PUT DATA INTO GENERIC OBJECT FOR FURTHUR PROCESSING

m.Norm<-FinalSD_Calc

#SQR STDEV and SSQR STDEV DISTANCE

tbl= data object (e.g. The matrix with the Stdev values)

m = A matrix object within which stats summary info is stored (see above)

y = number of arrays

z = Indicates the first column in the data matrix for which calculation will be done (default = 1)

#SSqrdiff<- function(tbl,m,y,z)

FinalSD_Calc<-SSqrdiff m.Norm,m.Norm, Hyb_no,1)

m.Norm[10,1]<-sum(m.Norm[3,1:Hyb_no])

#PUT DATA INTO GENERIC OBJECT FOR FURTHUR PROCESSING

m.Norm<-FinalSD_Calc

#REPLICATE RATIOS

tbl= data object (e.g. The matrix with the Stdev values)

m = A matrix object within which stats summary info is stored (see above)

y = number of arrays

z = Indicates the first column in the data matrix for which calculation will be done (default = 1)

#WholeArrayRatio <- function(tbl,m,y,z)

FinalSD_Calc<-RepRatio(m.Norm,m.Norm, Hyb_no,1)

m.Norm[7,1]<-mean(m.Norm[6,1:Hyb_no])

#PUT DATA INTO GENERIC OBJECT FOR FURTHUR PROCESSING

m.Norm<-FinalSD_Calc

m.Norm[7,1]<-mean(m.Norm[6,1:Hyb_no])

#STDEV OF OBSERVATIONS AND CORRELATION

#STDEV

m.Norm[1, mcolumns]<-sd(m.Norm[1,1: Hyb_no])

m.Norm[2, mcolumns]<-sd(m.Norm[2,1: Hyb_no])

m.Norm[4, mcolumns]<-sd(m.Norm[4,1: Hyb_no])

m.Norm[5, mcolumns]<-sd(m.Norm[5,1: Hyb_no])

```
#CORRELATION

correl<-cor(NormData$M, use="pairwise.complete.obs")

correl

NormData_CorAvg<-mean(correl)

m.Norm[9,1]<-NormData_CorAvg


 #WRITE NORM M VALUES TO A FILE

#INVERT DATA SO THAT RATIOS ARE SAMPLE / REFERENCES (I.E. CY3 / CY5)


NormData$genes$M<-NormData$M*-1


write.table(NormData$genes,paste("NormData.",backg_method,".",Within_array_method,".",Between_array_met
hod,".txt",sep="\t")
```

################################## END ##################################

S.2.2.2 Custom statistical calculation functions

```
#Script written by J.H. van Heerden

#Intended for use with Generic Normalization script (see S2.2.1)


#PRE-NORMALIZATION CALCULATIONS (denoted NN  - NoNorm)


#CALCULATE WITHIN ARRAY REPLICATE POOL SD

# rd = Raw Data object

# s = Status object (e.g. RG$genes$Status)

# x = number of Replicate groups

# y = number of Arrays

# z = Indicates the first column in the data matrix for which calculation will be done (default = 1)


NNWithinArrayRepSD <- function(rd, s, x, y, z=1) {
```

```
	NoReplicates <- x

	ArrayRun <- z

	NoArrays <- y

	Replicates_SD <- matrix(,NoReplicates,NoArrays)

	ReplicateID <- "Replicate"


	RepRun <- z

	while (RepRun < x +1) {

		while (ArrayRun < y + 1) {

			ReplicateNo = paste(ReplicateID,RepRun)

			Replicates_SD[RepRun,ArrayRun]<-sd(rd$M[,ArrayRun][rd$genes$Status ==
ReplicateNo],na.rm=TRUE)

			ArrayRun = ArrayRun +1

		}

		RepRun = RepRun +1

		ArrayRun <- z

	}

	return(Replicates_SD)

}


#CALCULATE AVG STDEV FOR REP SETS WITHIN EACH ARRAY

# rstdev = data object (e.g. The matrix with the Stdev values for rep sets)

# m = A matrix object within which stats summary info is stored (see above)

# y = number of arrays

# z = Indicates the first column in the data matrix for which calculation will be done (default = 1)


NNAvgArrayRepSD <- function (rstdev, m, y, z=1) {

while (z < y +1) {

m[4,z]<-mean(d[,z],na.rm=TRUE)

z = z +1

}
```

```
return(m)
}

#WHOLE ARRAY STD
# rd = data object (e.g. The matrix with the Stdev values for rep sets)
# m = A matrix object within which stats summary info is stored (see above)
# y = number of arrays
# z = Indicates the first column in the data matrix for which calculation will be done (default = 1)
NNWholeArraySD <- function(rd,m,y,z) {
        while (z < y +1) {
                m[1,z]<-sd(rd$M[,z],na.rm=TRUE)
                z = z + 1
        }
        return(m)
}

#POST-NORMALIZATION CALCULATIONS (denoted PN – PostNorm)

#Calulate Within Array Replicate Pool SD
# nd = Normalized Data object
# s = Status object (e.g. RG$genes$Status)
# x = number of Replicate groups
# y = number of Arrays
# z = Indicates the first column in the data matrix for which calculation will be done (default = 1)

PNWithinArrayRepSD <- function(nd, s, x, y, z=1) {
        NoReplicates <- x
        ArrayRun <- z
        NoArrays <- y
        Replicates_SD <- matrix(,NoReplicates,NoArrays)
        ReplicateID <- "Replicate"
```

```
	RepRun <- z
	while (RepRun < x +1) {
		while (ArrayRun < y + 1) {
			ReplicateNo = paste(ReplicateID,RepRun)
			Replicates_SD[RepRun,ArrayRun]<-sd(nd$M[,ArrayRun][nd$genes$Status ==
ReplicateNo],na.rm=TRUE)
			ArrayRun = ArrayRun +1
		}
		RepRun = RepRun +1
		ArrayRun <- z
	}
	return(Replicates_SD)
}
```

#CALCULATE AVG STDEV FOR REP SETS WITHIN EACH ARRAY

 # d = data object with SD calcs

 # m = A matrix object within which stats summary info is stored (see above)

 # y = number of arrays

 # z = Indicates the first column in the data matrix for which calculation will be done (default = 1)

```
PNAvgArrayRepSD <- function (d, m, y, z=1) {
while (z < y +1) {
m[5,z]<-mean(d[,z],na.rm=TRUE)
z = z +1
}
return(m)
}
```

#WHOLE ARRAY STD

 # rd = data object (e.g. The matrix with the Stdev values for rep sets)

 # m = A matrix object within which stats summary info is stored (see above)

 # y = number of arrays

```
# z = Indicates the first column in the data matrix for which calculation will be done (default = 1)


PNWholeArraySD <- function(rd,m,y,z) {
        while (z < y +1) {
                m[2,z]<-sd(rd$M[,z],na.rm=TRUE)
                z = z + 1
        }
        return(m)
}
#SQUARED DIFFERENCE BETWEEN PRE-NORM MEDIAN ARRAYS SD AND POST-NORM SD
# tbl = data object with SD calcs
# m = A matrix object within which stats summary info is stored (see above)
# y = number of arrays
# z = Indicates the first column in the data matrix for which calculation will be done (default = 1)


SSqrdiff <- function(tbl,m,y,z) {
        while (z < y +1) {
                m[3,z]<-(median(m[1,1:y])-m[2,z])*(median(m[1,1:y])-m[2,z])
                z = z + 1
        }
        return(m)
}


#REPLICATE RATIOS
RepRatio <- function(tbl,m,y,z) {
        while (z < y +1) {
                m[6,z]<-m[4,z]/m[5,z]
                z = z + 1
        }
        return(m)
}
            ################################## END ##################################
```

S.2.3 ASCA-genes example script

```
#ASCA-genes batch effect removal from PBMC data set (2 Treatments, 4x 5 slide batches)

#Set Variables

RawData = "C:\\DataDir\\"

filename = "inputData.txt"

setwd(RawData)


# Load ASCA-genes functions

source("ASCA2f.R")

source("ASCA3f.R")

source("ASCAfun1.R")

source("ASCAfun2.R")

source("ASCAfun12.R")

source("ASCAfunres.R")

source("ASCAfun-triple.R")

source("PCA-GENES.R")

source("sceeplot.R")

source("show.var.R")

source("spe.lims.R")

source("leverage.lims.R")


#Read data

data <- read.delim (filename, row.names=1)


# Treatment Design

Designa <-cbind(MS=c(1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0), SH=c(0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1))

rownames(Designa) <- colnames(data)


#4x BATCH Design

Designb <-
cbind(B1=c(0,0,0,0,0,1,0,1,0,1,0,0,0,0,0,0,0,0,1,1),B2=c(0,0,0,0,0,0,1,0,1,0,0,0,0,0,0,1,1,1,0,0),B3=c(0,1,0,1,0,0,
0,0,0,0,1,0,1,0,1,0,0,0,0,0),B4=c(1,0,1,0,1,0,0,0,0,0,0,1,0,1,0,0,0,0,0,0))
```

```
rownames(Designb) <- colnames(data)


#PCA component plot of "raw-data" (pre-ASCA-Genes processed)

raw.pca <- princomp(data)

l_raw <- raw.pca$loadings

plot(l_raw[,1],l_raw[,2])

text(l_raw[,1],l_raw[,2], labels=colnames(data) , col=c(4,3,4,3,4,1,2,1,2,1,3,4,3,4,3,2,2,2,1,1))


# Run ASCA

#4 Batches - Remove 1st RE component (>20%)

#Calculate variability components

asca_mouse <-ASCA.2f(X = t(data),Designa = Designa, Designb = Designb,Designc = NULL, Fac =
c(2,4,0,0),type = 2, showvar=TRUE, showscree=TRUE)


# Look at RE variability components

asca_mouse$Model.res$var.exp


#Adjust Data

newdata_0 <-asca_mouse$Model.a$TP+asca_mouse$Model.res$E

otropca_0 <- princomp(t(newdata_0))

l_0 <- otropca_0$loadings


#PCA component plot of corrected data

X11()

plot(l_0[,1],l_0[,2])

text(l_0[,1],l_0[,2], labels=colnames(data) , col=c(4,3,4,3,4,1,2,1,2,1,3,4,3,4,3,2,2,2,1,1))


output0<-t(newdata_0)

write.table(output0,"output.txt",sep="\t")
```

S.3 DIFFERENTIALLY EXPRESSED GENES

Note: All genes are sorted alphabetically according to gene symbol

Table S.3.1 Frontal Association Cortex differentially expressed genes

Operon Oligo ID	Description	ENSEMBL / Refseq / Riken ID	Symbol	Info p-value	SAM p-value	Fold Difference	Over / Under expressed in MS
M400001545	RIKEN cDNA 1700047E16 gene	ENSMUSG00000033282	1700047E16Rik	0.002	0.02	1.2	under
M200004634	RIKEN cDNA 1700080E11 gene	ENSMUSG00000032566	1700080E11Rik	0.033	0.02	1.2	under
M200005944	RIKEN cDNA 2510049I19 gene	ENSMUSG00000003039	2510049I19Rik	0.013	0.04	1.2	under
M400005234	RIKEN cDNA 2610024B07 gene	ENSMUSG00000055593	2610024B07Rik	0.033	0.02	1.2	over
M200008798	RIKEN cDNA 4632404H12 gene	ENSMUSG00000042579	4632404H12Rik	0.013	0.02	1.2	over
M300006913	RIKEN cDNA 6330503C03 gene	ENSMUSG00000029861	6330503C03Rik	0.033	0.02	1.2	over
M400015122	RIKEN cDNA 9530048O09 gene	ENSMUSG00000073771	9530048O09Rik	0.005	0.02	1.2	under
M300011861	RIKEN cDNA 9630058J23 gene	ENSMUSG00000038949	9630058J23Rik	0.002	0.02	1.2	over
M400000471	RIKEN cDNA A530054K11 gene	ENSMUSG00000021510	A530054K11Rik	0.033	0.02	1.3	under
M300017486	RIKEN cDNA A930001M12 gene	ENSMUSG00000046593	A930001M12Rik	0.033	0.02	1.2	under
M300003418	ATP-binding cassette, sub-family D (ALD), member 2	ENSMUSG00000055782	Abcd2	0.023	0.04	1.2	under
M200004178	ATP-binding cassette, sub-family F (GCN20), member 2	ENSMUSG00000028953	Abcf2	0.013	0.02	1.2	under
M400008055	Abl interactor 1	ENSMUSG00000058835	Abi1	0.005	0.02	1.2	under
M200005686	1-acylglycerol-3-phosphate O-acyltransferase 4 (lysophosphatidic acid acyltransferase, delta)	ENSMUSG00000023827	Agpat4	0.023	0.02	1.2	over
M300013424	agrin	ENSMUSG00000041936	Agrin	0.023	0.02	1.2	under
M200006316	ankyrin repeat domain 47	ENSMUSG00000042099	Ankrd47	0.023	0.02	1.2	over
M400009479	Rho guanine nucleotide exchange factor (GEF) 1 (Arhgef1), mRNA.	ENSMUSG00000040940	Arhgef1	0.005	0.02	1.3	under
M200004507	ARV1 homolog (yeast)	ENSMUSG00000031982	Arv1	0.033	0.02	1.2	over
M200014150	N-acylsphingosine amidohydrolase 3-like	ENSMUSG00000038007	Asah3l	0.002	0.02	1.2	over
M300010107	cDNA sequence BC005764	ENSMUSG00000035835	BC005764	0.002	0.02	1.2	under
M300020407	cDNA sequence BC023055	ENSMUSG00000049670	BC023055	0.023	0.02	1.2	under
M400003323	cDNA sequence BC029214	ENSMUSG00000047617	BC029214	0.013	0.04	1.2	over
M400002960	cDNA sequence BC050099	ENSMUSG00000044903	BC050099	0.013	0.02	1.2	over
M200014376	bicaudal C homolog 1 (Drosophila)	ENSMUSG00000014329	Bicc1	0.013	0.02	1.2	over
M400003801	basic leucine zipper and W2 domains 1	ENSMUSG00000051223	Bzw1	0.002	0.02	1.2	under
M300001373	CAS1 domain containing 1	ENSMUSG00000015189	Casd1	0.013	0.02	1.3	under
M200013352	cask-interacting protein 2	ENSMUSG00000034471	Caskin2	0.002	0.02	1.2	over
M300006792	cyclin-dependent kinase 8	ENSMUSG00000029635	Cdk8	0.023	0.02	1.2	under
M200014768	CDKN2A interacting protein	ENSMUSG00000038069	Cdkn2aip	0.005	0.02	1.2	under
M300004820	complexin 2 (Cplx2), mRNA.	ENSMUSG00000025867	Cplx2	0.013	0.02	1.2	under

Operon Oligo ID	Description	ENSEMBL / Refseq / Riken ID	Symbol	Info p-value	SAM p-value	Fold Difference	Over / Under expressed in MS
M200002925	cAMP responsive element binding protein 3-like 1	ENSMUSG00000027230	Creb3l1	0.013	0.02	1.2	over
M200006376	cytochrome c-1	ENSMUSG00000022551	Cyc1	0.002	0.02	1.2	over
M300012286	Uncharacterized protein C9orf114 homolog.	ENSMUSG00000039660	D2Wsu81e	0.013	0.02	1.2	over
M200012931	DEAD (Asp-Glu-Ala-Asp) box polypeptide 27	ENSMUSG00000017999	Ddx27	0.013	0.02	1.3	under
M200006582	degenerative spermatocyte homolog 1 (Drosophila)	ENSMUSG00000038633	Degs1	0.002	0.02	1.2	under
M300003367	diacylglycerol O-acyltransferase 1	ENSMUSG00000022555	Dgat1	0	0.02	1.2	over
M400015808	Homeobox protein DLX-4	ENSMUSG00000020871	Dlx4	0.033	0.02	1.3	over
M300005971	DnaJ homolog subfamily B member 14	ENSMUSG00000074212	Dnajb14	0.023	0.02	1.2	under
M200006237	dermatopontin	ENSMUSG00000026574	Dpt	0.005	0.02	1.2	under
M300006038	RIKEN cDNA E130309F12 gene	ENSMUSG00000063446	E130309F12Rik	0.013	0.02	1.2	over
M400005913	similar to calmodulin 1 (LOC640703), mRNA	ENSMUSG00000072757	EG640703	0.013	0.02	1.2	under
M400002512	EH-domain containing 2	ENSMUSG00000074364	Ehd2	0.002	0.02	1.2	under
M300013452	ELMO domain containing 1	ENSMUSG00000041986	Elmod1	0.033	0.04	1.2	under
M400004407	predicted gene, ENSMUSG00000053270	ENSMUSG00000053270	ENSMUSG00000053270	0.013	0.02	1.2	over
M200003203	forkhead box C1	ENSMUSG00000050295	Foxc1	0.013	0.02	1.2	over
M400000808	FYVE and coiled-coil domain containing 1	ENSMUSG00000025241	Fyco1	0.013	0.02	1.2	under
M300002807	G protein-coupled receptor 98	ENSMUSG00000069170	Gpr98	0.013	0.02	1.2	over
M300007070	G protein-coupled receptor, family C, group 5, member D (Gprc5d), mRNA.	ENSMUSG00000030205	Gprc5d	0.013	0.02	1.2	over
M200002939	histidyl-tRNA synthetase	ENSMUSG00000001380	Hars	0.013	0.02	1.2	under
M300002684	HECT, C2 and WW domain containing E3 ubiquitin protein ligase 1	ENSMUSG00000021301	Hecw1	0.005	0.02	1.2	under
M400009328	heat shock protein 8	ENSMUSG00000062019	Hspa8	0.013	0.04	1.2	under
M400001448	5-hydroxytryptamine (serotonin) receptor 3A	ENSMUSG00000032269	Htr3a	0.033	0.02	1.2	under
M400006262	isopentenyl-diphosphate delta isomerase	ENSMUSG00000070508	Idi1	0.013	0.02	1.2	under
M200009096	interleukin 1 family, member 6	ENSMUSG00000026984	Il1f6	0.023	0.02	1.2	under
M400001046	interleukin 6 receptor, alpha (Il6ra), mRNA.	ENSMUSG00000027947	Il6ra	0.033	0.04	1.2	under
M200009706	inositol polyphosphate phosphatase-like 1	ENSMUSG00000032737	Inppl1/SHIP2	0	0.02	1.2	over
M400017224	Potassium voltage-gated channel subfamily H member 8	ENSMUSG00000035580	Kcnh8	0	0.02	1.2	under
M400018370	kelch domain containing 6	ENSMUSG00000033182	Klhdc6	0.005	0.02	1.2	under
M300007101	v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog	ENSMUSG00000030265	Kras	0.002	0.02	1.2	under
M300001470	ligand dependent nuclear receptor corepressor-like	ENSMUSG00000015882	Lcorl	0.005	0.02	1.2	under
M400002899	Predicted gene	ENSMUSG00000043462	LOC669217	0.013	0.02	1.2	under
M400010651	ENSEMBL pseudogene	ENSMUSG00000059634	LOC670761	0.023	0.02	1.2	over
M400006738	similar to 40S ribosomal protein S19 (LOC674024), mRNA	ENSMUSG00000072548	LOC676948	0.002	0.02	1.2	under
M300002004	leucine-rich repeats and immunoglobulin-like domains 3	ENSMUSG00000020105	Lrig3	0.033	0.02	1.2	over
M200005883	leucine zipper, putative tumor suppressor 2	ENSMUSG00000035342	Lzts2	0.005	0.02	1.2	over
M400000440	membrane protein, palmitoylated 5 (MAGUK p55 subfamily member 5)	ENSMUSG00000021112	Mpp5	0.005	0.02	1.2	under

Operon Oligo ID	Description	ENSEMBL / Refseq / Riken ID	Symbol	Info p-value	SAM p-value	Fold Difference	Over / Under expressed in MS
M300001339	mutS homolog 3 (E. coli)	ENSMUSG00000014850	Msh3	0.013	0.02	1.2	under
M200013308	mitochondrial translational initiation factor 2	ENSMUSG00000020459	Mtif2	0.013	0.04	1.4	over
M400005534	NADH dehydrogenase 4, mitochondrial	ENSMUSG00000065947	mt-Nd4	0	0.02	1.3	under
M300008841	myosin VI	ENSMUSG00000033577	Myo6	0.013	0.02	1.2	under
M200006647	N-acetylglucosamine kinase	ENSMUSG00000034744	Nagk	0.013	0.02	1.2	under
M300000428	nicastrin	ENSMUSG00000003458	Ncstn	0.013	0.02	1.2	over
M200000742	neurofilament, light polypeptide	ENSMUSG00000022055	Nefl	0.002	0.02	1.2	under
M200009919	neuropeptide VF precursor	ENSMUSG00000029831	Npvf	0	0.02	1.2	over
M300013607	nuclear transport factor 2-like export factor 2	ENSMUSG00000042271	Nxt2	0.013	0.02	1.2	under
M300014294	olfactory receptor 315	ENSMUSG00000056959	Olfr315	0.013	0.02	1.2	over
M200008363	oligophrenin 1	ENSMUSG00000031214	Ophn1	0.023	0.02	1.2	over
M200008878	P2Y purinoceptor 4	ENSMUSG00000044359	P2ry4	0.023	0.02	1.2	over
M200004380	phosphoglucomutase 2	ENSMUSG00000025791	Pgm2	0.023	0.02	1.2	over
M200008268	progesterone receptor membrane component 2	ENSMUSG00000049940	Pgrmc2	0.023	0.02	1.2	over
M300007869	PH domain and leucine rich repeat protein phosphatase-like	ENSMUSG00000031732	Phlppl	0.033	0.02	1.2	under
M200014786	phospholipase C, delta 3	ENSMUSG00000020937	Plcd3	0.002	0.02	1.2	over
M200013679	pleckstrin homology domain-containing, family A (phosphoinositide binding specific) member 3	ENSMUSG00000002733	Plekha3	0.013	0.04	1.2	under
M400009342	protein phosphatase 1, catalytic subunit, beta isoform	ENSMUSG00000014956	Ppp1cb	0.013	0.04	1.2	under
M300001789	prostaglandin E receptor 1 (subtype EP1)	ENSMUSG00000019464	Ptger1	0.013	0.02	1.2	over
M400004572	peptidyl-tRNA hydrolase 1 homolog (S. cerevisiae)	ENSMUSG00000053746	Ptrh1	0.023	0.04	1.3	over
M300005596	RAS guanyl releasing protein 1	ENSMUSG00000027347	Rasgrp1	0.023	0.04	1.4	under
M300008524	RNA binding motif protein 11	ENSMUSG00000032940	Rbm11	0.013	0.02	1.4	under
M200004426	ring finger protein 123	ENSMUSG00000041528	Rnf123	0.023	0.02	1.2	over
M200004286	ring finger protein 14	ENSMUSG00000060450	Rnf14	0.013	0.02	1.2	under
M200013488	ring finger protein 141	ENSMUSG00000030788	Rnf141	0.013	0.02	1.2	under
M400001747	ring finger protein 44	ENSMUSG00000034928	Rnf44	0.005	0.02	1.2	over
M400000190	ribosomal protein L15	ENSMUSG00000067184	Rpl15	0.002	0.02	1.2	under
M300003103	sciellin (Scel), mRNA.	ENSMUSG00000022123	Scel	0.013	0.02	1.2	under
M300000269	splicing factor, arginine/serine-rich 16 (suppressor-of-white-apricot homolog, Drosophila)	ENSMUSG00000061028	Sfrs16	0.013	0.02	1.4	under
M300002987	SH3-domain binding protein 5 (BTK-associated)	ENSMUSG00000021892	Sh3bp5	0.013	0.04	1.2	under
M300009754	solute carrier family 24, member 5	ENSMUSG00000035183	Slc24a5	0.033	0.02	1.2	under
M400005506	slit homolog 3 (Drosophila)	ENSMUSG00000056427	Slit3	0.013	0.02	1.2	over
M300017587	SLIT and NTRK-like family, member 4	ENSMUSG00000046699	Slitrk4	0.005	0.02	1.3	under
M300001093	spire homolog 2 (Drosophila)	ENSMUSG00000010154	Spire2	0.033	0.02	1.2	over
M400000434	signal recognition particle 54	ENSMUSG00000021020	Srp54	0.002	0.02	1.2	under
M200013889	stromal antigen 1	ENSMUSG00000037286	Stag1	0.013	0.02	1.2	under

Operon Oligo ID	Description	ENSEMBL / Refseq / Riken ID	Symbol	Info p-value	SAM p-value	Fold Difference	Over / Under expressed in MS
M200009718	serine/threonine/tyrosine interacting-like 1	ENSMUSG00000019178	Styxl1	0.033	0.02	1.2	under
M400011296	synaptotagmin XI	ENSMUSG00000068923	Syt11	0	0.02	1.2	over
M300007924	TATA box binding protein (Tbp)-associated factor, RNA polymerase I, C	ENSMUSG00000031832	Taf1c	0	0.02	1.2	over
M200000759	transgelin	ENSMUSG00000032085	Tagln	0	0.02	1.2	over
M400007874	taste receptor, type 2, member 109	ENSMUSG00000062528	Tas2r109	0.033	0.02	1.2	over
M200005526	tafazzin	ENSMUSG00000009995	Taz	0.033	0.02	1.2	over
M300010804	THAP domain containing, apoptosis associated protein 1	ENSMUSG00000037214	Thap1	0.023	0.02	1.2	over
M400011234	translocase of inner mitochondrial membrane 8 homolog a1 (yeast)	ENSMUSG00000045455	Timm8a1	0.002	0.02	1.2	under
M400003795	transmembrane protein 157	ENSMUSG00000051185	Tmem157	0.023	0.02	1.2	under
M300010080	transmembrane protein 161B	ENSMUSG00000035762	Tmem161b	0.023	0.02	1.2	under
M300006002	transmembrane protein 68	ENSMUSG00000028232	Tmem68	0.013	0.02	1.2	under
M200012053	TNF receptor-associated protein 1	ENSMUSG00000005981	Trap1	0.023	0.02	1.2	under
M200000017	tripartite motif protein 27	ENSMUSG00000021326	Trim27	0.013	0.02	1.2	over
M300011152	tetraspanin 14	ENSMUSG00000037824	Tspan14	0.013	0.02	1.2	under
M200012963	tubulin, beta 2a	ENSMUSG00000045136	Tubb2a	0.005	0.02	1.2	under
M300000203	tubulin, beta 5	ENSMUSG00000001525	Tubb5	0	0.02	1.2	over
M200004461	ubiquitin-conjugating enzyme E2 variant 1	ENSMUSG00000027694	Ube2v1	0.033	0.04	1.2	over
M200005785	UBX domain containing 1	ENSMUSG00000019578	Ubxd1	0.033	0.02	1.2	under
M200000707	ubiquitin fusion degradation 1 like	ENSMUSG00000005262	Ufd1l	0.033	0.02	1.2	under
M400010601	Serine/threonine-protein kinase ULK3	ENSMUSG00000032308	Ulk3	0.013	0.02	1.2	over
M400007489	Probable ubiquitin carboxyl-terminal hydrolase FAF-X	ENSMUSG00000031010	Usp9x	0.013	0.02	1.2	under
M300006387	vesicle-associated membrane protein 3	ENSMUSG00000075034	Vamp3	0.013	0.02	1.3	under
M200005389	visinin-like 1	ENSMUSG00000054459	Vsnl1	0.013	0.02	1.3	under
M200003872	wingless-related MMTV integration site 4	ENSMUSG00000036856	Wnt4	0.013	0.02	1.2	over
M400005315	zinc finger protein 1	ENSMUSG00000055835	Zfp1	0.013	0.02	1.2	under
M400014353	No longer in ENSEMBL	ENSMUSG00000073100	Not available	0.013	0.02	1.2	over
M400003406	No longer in ENSEMBL	ENSMUSG00000048353	Not available	0.013	0.02	1.2	over
M400003455	No longer in ENSEMBL	ENSMUSG00000048689	Not available	0.013	0.02	1.2	under
M400005962	No longer in ENSEMBL	ENSMUSG00000057538	Not available	0.023	0.04	1.3	under
M400012721	Predicted Gene	ENSMUSG00000072605	Not available	0.013	0.02	1.2	over
M400006392	Predicted gene	ENSMUSG00000058567	Not available	0.013	0.02	1.2	under
M400005868	Predicted gene	ENSMUSG00000071320	Not available	0.023	0.02	1.2	under
M400018537	Uncharacterised	AC158565	Not available	0.013	0.02	1.6	over
M400017480	Uncharacterised	AC123702	Not available	0.005	0.02	1.3	over
M400016982	Uncharacterised	AC131909	Not available	0.013	0.02	1.2	over
M400015807	Uncharacterised	AC122117	Not available	0.013	0.02	1.2	over
M400014211	Uncharacterised	AC108399	Not available	0.013	0.02	1.2	over

Operon Oligo ID	Description	ENSEMBL / Refseq / Riken ID	Symbol	Info p-value	SAM p-value	Fold Difference	Over / Under expressed in MS
M400013886	Uncharacterised	AC100491	Not available	0.023	0.02	1.2	over
M400014069	Uncharacterised	AC155164	Not available	0.033	0.02	1.2	over
M400003167	Uncharacterised	AC090007	Not available	0.023	0.02	1.2	over
M300004345	Uncharacterised	AC133523	Not available	0.013	0.02	1.2	over
M400013189	Uncharacterised	AC108401	Not available	0.013	0.02	1.2	over
M400016217	Uncharacterised	AC131592	Not available	0.005	0.02	1.2	over
M400017693	Uncharacterised	AC105163	Not available	0.013	0.02	1.2	under
M400013935	Uncharacterised	AC123614	Not available	0.033	0.02	1.2	under
M400000476	Uncharacterised	XM_145500	Not available	0.013	0.02	1.2	under
M400018682	Uncharacterised	AC107806	Not available	0.002	0.02	1.2	under
M400015408	Uncharacterised	AC126426	Not available	0.005	0.02	1.2	under
M400009919	Uncharacterised	AC130824	Not available	0.013	0.02	1.2	under
M400003628	Uncharacterised	XM_488078	Not available	0.005	0.02	1.2	under
M400001153	Uncharacterised	AC115007	Not available	0.002	0.02	1.2	under
M400015069	Uncharacterised	AC166316	Not available	0.005	0.02	1.2	under

Table S.3.2 Hippocampus differentially expressed genes

Operon Oligo ID	Description	ENSEMBL / Refseq / Riken ID	Symbol	Info p-value	SAM p-value	Fold Difference	Over / Under expressed in MS
M200009369	RIKEN cDNA 1110032O16 gene	ENSMUSG00000060538	1110032O16Rik	0.008	0.02	1.2	under
M400000838	5-hydroxyisourate hydrolase (Transthyretin-related protein)	ENSMUSG00000025481	1190003J15Rik	0.008	0.02	1.3	over
M200013718	RIKEN cDNA 1810043G02 gene	ENSMUSG00000020284	1810043G02Rik	0.008	0.02	1.2	over
M400001942	RIKEN cDNA 2610305D13 gene	ENSMUSG00000044501	2610305D13Rik	0.008	0.02	1.3	over
M300019683	RIKEN cDNA 2810453I06 gene	ENSMUSG00000048910	2810453I06Rik	0.008	0.02	1.2	over
M200008808	RIKEN cDNA 3632413A11 gene	ENSMUSG00000028373	3632413A11Rik	0.008	0.02	1.2	over
M200012628	RIKEN cDNA 4930425N13 gene	ENSMUSG00000049526	4930425N13Rik	0.008	0.02	1.3	under
M400000609	RIKEN cDNA 4930435E12 gene	ENSMUSG00000022798	4930435E12Rik	0.008	0.02	1.2	over
M300017297	RIKEN cDNA 4930474M22 gene	ENSMUSG00000046390	4930474M22Rik	0.008	0.02	1.2	under
M200011890	expressed sequence AU022252	ENSMUSG00000028642	4930538K18Rik	0.008	0.02	1.2	under
M200006247	RIKEN cDNA 4930570C03 gene	ENSMUSG00000022473	4930570C03Rik	0.008	0.02	1.3	over
M300007995	RIKEN cDNA 4932417I16 gene	ENSMUSG00000031951	4932417I16Rik	0.008	0.02	1.2	under
M400001873	RIKEN cDNA 6030443O07 gene, transcript variant 1	ENSMUSG00000036097	6030443O07Rik	0.008	0.02	1.2	under
M400004875	RIKEN cDNA 9130019P16 gene	ENSMUSG00000073067	9130019P16Rik	0.008	0.02	1.2	over
M300015653	RIKEN cDNA A330102K23 gene	ENSMUSG00000044647	A330102K23Rik	0.008	0.02	1.2	over
M200000570	adrenomedullin	ENSMUSG00000030790	Adm	0.008	0.02	1.2	over

Operon Oligo ID	Description	ENSEMBL / Refseq / Riken ID	Symbol	Info p-value	SAM p-value	Fold Difference	Over / Under expressed in MS
M400019210	Proteasome-associated protein ECM29 homolog	ENSMUSG00000050812	AI314180	0.008	0.02	1.2	over
M300012821	cDNA sequence AK129341	ENSMUSG00000040729	AK129341	0.008	0.02	1.3	over
M200015617	type 1 tumor necrosis factor receptor shedding aminopeptidase regulator	ENSMUSG00000021583	Arts1	0.008	0.02	1.2	under
M400003965	expressed sequence AW146154	ENSMUSG00000068130	AW146154	0.008	0.02	1.3	over
M400013237	Protein FAM102B	ENSMUSG00000040339	B430201A12Rik	0.008	0.02	1.2	over
M300013123	cDNA sequence BC057627	ENSMUSG00000059273	BC057627	0.008	0.02	1.3	under
M300007701	cDNA sequence BC068171	ENSMUSG00000031403	BC068171	0.008	0.02	1.2	under
M300002423	bleomycin hydrolase	ENSMUSG00000020840	Blmh	0.008	0.02	1.2	under
M300008173	bruno-like 6, RNA binding protein (Drosophila)	ENSMUSG00000032297	Brunol6	0.008	0.02	1.2	over
M300009137	RIKEN cDNA C130022K22 gene	ENSMUSG00000034083	C130022K22Rik	0.008	0.02	1.2	over
M400002913	RIKEN cDNA C530008M17 gene	ENSMUSG00000044552	C530008M17Rik	0.008	0.02	1.2	over
M200005696	calcium binding and coiled coil domain 1	ENSMUSG00000023055	Calcoco1	0.008	0.02	1.3	under
M200009111	coiled-coil domain containing 123	ENSMUSG00000023072	Ccdc123	0.008	0.02	1.2	under
M200002831	coiled-coil domain containing 91	ENSMUSG00000030301	Ccdc91	0.008	0.02	1.2	under
M400000141	CEA-related cell adhesion molecule 9	ENSMUSG00000007209	Ceacam9	0.008	0.02	1.2	over
M200000117	centromere protein K	ENSMUSG00000021714	Cenpk	0.008	0.02	1.2	under
M200015002	cholinergic receptor, nicotinic, alpha polypeptide 2 (neuronal)	ENSMUSG00000022041	Chrna2	0.008	0.02	1.2	under
M200000166	cell death-inducing DNA fragmentation factor, alpha subunit-like effector A	ENSMUSG00000024526	Cidea	0.008	0.02	1.2	over
M400000012	creatine kinase, mitochondrial 1, ubiquitous	ENSMUSG00000000308	Ckmt1	0.008	0.02	1.2	over
M200013297	chloride channel CLIC-like 1	ENSMUSG00000027884	Clcc1	0.008	0.02	1.2	under
M200004901	camello-like 2	ENSMUSG00000033634	Cml2	0.008	0.02	1.3	under
M300009490	contactin 4	ENSMUSG00000064293	Cntn4	0.008	0.02	1.3	over
M200006625	coilin	ENSMUSG00000033983	Coil	0.008	0.02	1.2	under
M200011974	procollagen, type IV, alpha 2	ENSMUSG00000031503	Col4a2	0.008	0.02	1.2	over
M300016924	Collagen alpha-4(IV) chain precursor	ENSMUSG00000067158	Col4a4	0.008	0.02	1.2	over
M300006050	coronin, actin binding protein 2A	ENSMUSG00000028337	Coro2a	0.008	0.02	1.2	under
M400019066	Carboxypeptidase D precursor	ENSMUSG00000020841	Cpd	0.008	0.02	1.2	under
M200012508	Nogo-B receptor precursor (NgBR) (Nuclear undecaprenyl pyrophosphate synthase 1 homolog).	ENSMUSG00000023068	D10Ertd438e	0.008	0.02	1.2	over
M400008947	Integrator complex subunit 9 (Int9).	ENSMUSG00000021975	D14Ertd231e	0.008	0.02	1.2	under
M300019282	Uncharacterized protein C20orf142 homolog precursor	ENSMUSG00000048486	D930001I22Rik	0.008	0.02	1.5	under
M300019075	DIP2 disco-interacting protein 2 homolog C	ENSMUSG00000048264	Dip2c	0.008	0.02	1.2	over
M200016042	ectodysplasin-A receptor	ENSMUSG00000003227	Edar	0.008	0.02	1.2	under
M200004170	ER degradation enhancer, mannosidase alpha-like 1	ENSMUSG00000030104	Edem1	0.008	0.02	1.2	under
M400003376	Ensembl pseudogene.	ENSMUSG00000048071	EG245651	0.008	0.02	1.2	over
M300006229	eukaryotic translation initiation factor 2B, subunit 3, transcript variant 1	ENSMUSG00000028683	Eif2b3	0.008	0.02	1.2	over

Operon Oligo ID	Description	ENSEMBL / Refseq / Riken ID	Symbol	Info p-value	SAM p-value	Fold Difference	Over / Under expressed in MS
M300012352	F-box and leucine-rich repeat protein 5	ENSMUSG00000039753	Fbxl5	0.008	0.02	1.2	over
M400010214	F-box and WD-40 domain protein 2	ENSMUSG00000035949	Fbxw2	0.008	0.02	1.2	under
M200001490	FK506 binding protein 10	ENSMUSG00000001555	Fkbp10	0.008	0.02	1.2	over
M200001958	glucosidase, alpha, acid	ENSMUSG00000025579	Gaa	0.008	0.02	1.3	over
M400018644	Gap junction alpha-10 protein	ENSMUSG00000051056	Gja10	0.008	0.02	1.2	under
M400000327	GLE1 RNA export mediator-like (yeast	ENSMUSG00000019715	Gle1l	0.008	0.02	1.2	under
M300000848	geminin	ENSMUSG00000006715	Gmnn	0.008	0.02	1.2	under
M200007425	glucosamine-6-phosphate deaminase 2	ENSMUSG00000029209	Gnpda2	0.008	0.02	1.2	over
M400004613	G protein-coupled receptor 112	ENSMUSG00000053852	Gpr112	0.008	0.02	1.2	under
M300007072	germ cell-specific gene 1 isoform b	ENSMUSG00000030206	Gsg1	0.008	0.02	1.2	over
M300004771	HGF-regulated tyrosine kinase substrate	ENSMUSG00000025793	Hgs	0.008	0.02	1.2	under
M400007914	histone cluster 4, H4	ENSMUSG00000069306	Hist4h4	0.008	0.02	1.2	under
M400005716	hippocalcin-like 1	ENSMUSG00000071379	Hpcal1	0.008	0.02	1.2	over
M200001986	hyaluronoglucosaminidase 2	ENSMUSG00000010047	Hyal2	0.008	0.02	1.2	under
M300021243	Hydrolethalus syndrome protein 1 homolog	ENSMUSG00000050555	Hyls1	0.008	0.02	1.2	over
M200009822	interleukin 1 receptor accessory protein-like 2	ENSMUSG00000059203	Il1rapl2	0.008	0.02	1.3	under
M300004762	inter-alpha (globulin) inhibitor H5	ENSMUSG00000025780	Itih5	0.008	0.02	1.2	under
M200004029	jerky	ENSMUSG00000046380	Jrk	0.008	0.02	1.2	under
M300008412	potassium voltage gated channel, Shab-related subfamily, member 2	ENSMUSG00000072447	Kcnb2	0.008	0.02	1.2	under
M200004603	kin of IRRE like 3 (Drosophila)	ENSMUSG00000032036	Kirrel3	0.008	0.02	1.2	over
M400008709	leukocyte immunoglobulin-like receptor, subfamily B, member 4	ENSMUSG00000043675	Lilrb4	0.008	0.02	1.2	over
M400011064	leukocyte specific transcript 1	ENSMUSG00000073412	Lst1	0.008	0.02	1.2	under
M200005128	lysophospholipase 3	ENSMUSG00000031903	Lypla3	0.008	0.02	1.2	under
M200003618	mitogen activated protein kinase kinase 3	ENSMUSG00000018932	Map2k3	0.008	0.02	1.2	over
M300002849	multiple C2 domains, transmembrane 1	ENSMUSG00000021596	Mctp1	0.008	0.02	1.2	over
M300013620	MKL/myocardin-like protein 1	ENSMUSG00000042292	Mkl1	0.008	0.02	1.2	over
M300006679	matrix metallopeptidase 17	ENSMUSG00000029436	Mmp17	0.008	0.02	1.2	over
M400014876	Metal-response element-binding transcription factor 2	ENSMUSG00000029267	Mtf2	0.008	0.02	1.2	over
M300007916	methenyltetrahydrofolate synthetase domain containing	ENSMUSG00000031816	Mthfsd	0.008	0.02	1.2	under
M200008114	Max dimerization protein 4	ENSMUSG00000037235	Mxd4	0.008	0.02	1.3	under
M400007248	Cysteine desulfurase, mitochondrial precursor	ENSMUSG00000027618	Nfs1	0.008	0.02	1.3	under
M300017902	non imprinted in Prader-Willi/Angelman syndrome 1 homolog (human)	ENSMUSG00000047037	Nipa1	0.008	0.02	1.2	under
M400012050	olfactory receptor 1316	ENSMUSG00000074944	Olfr1316	0.008	0.02	1.2	under
M400007337	olfactory receptor 151	ENSMUSG00000061165	Olfr151	0.008	0.02	1.2	under
M400003632	olfactory receptor 350	ENSMUSG00000050015	Olfr350	0.008	0.02	1.2	over
M300020522	olfactory receptor 642	ENSMUSG00000049797	Olfr642	0.008	0.02	1.2	under

Operon Oligo ID	Description	ENSEMBL / Refseq / Riken ID	Symbol	Info p-value	SAM p-value	Fold Difference	Over / Under expressed in MS
M400000091	peroxisomal biogenesis factor 11c	ENSMUSG00000069633	Pex11c	0.008	0.02	1.3	under
M400012257	putative homeodomain transcription factor 2	ENSMUSG00000039987	Phtf2	0.008	0.02	1.2	under
M200007926	phosphatidylinositol glycan anchor biosynthesis, class N	ENSMUSG00000056536	Pign	0.008	0.02	1.2	over
M200009054	phospholipase C, beta 4	ENSMUSG00000039943	Plcb4	0.008	0.02	1.2	over
M300017035	pleckstrin homology domain containing, family G (with RhoGef domain) member 3	ENSMUSG00000052609	Plekhg3	0.008	0.02	1.2	under
M400016044	phosphatidic acid phosphatase type 2 domain containing 1A	ENSMUSG00000070366	Ppapdc1a	0.008	0.02	1.2	under
M400015671	Protein phosphatase 1-like	ENSMUSG00000027784	Ppml1	0.008	0.02	1.2	under
M400010878	retinoic acid receptor, alpha	ENSMUSG00000037992	Rara	0.008	0.02	1.3	under
M200005471	RNA binding motif, single stranded interacting protein 1	ENSMUSG00000026970	Rbms1	0.008	0.02	1.2	over
M400006369	ring finger protein 207	ENSMUSG00000058498	Rnf207	0.008	0.02	1.2	over
M400003932	R-spondin 2 homolog (Xenopus laevis)	ENSMUSG00000051920	Rspo2	0.008	0.02	1.2	under
M300009182	scleraxis	ENSMUSG00000034161	Scx	0.008	0.02	1.2	over
M200002927	succinate dehydrogenase complex, subunit D, integral membrane protein	ENSMUSG00000000171	Sdhd	0.008	0.02	1.2	over
M200008074	septin 9	ENSMUSG00000059248	Sept9	0.008	0.02	1.2	over
M400010613	SET domain containing 2	ENSMUSG00000044791	Setd2	0.008	0.02	1.2	over
M300008664	Src homology 2 domain containing F	ENSMUSG00000033256	Shf	0.008	0.02	1.2	under
M300001399	signaling lymphocytic activation molecule family member 1	ENSMUSG00000015316	Slamf1	0.008	0.02	1.2	under
M200008890	solute carrier family 14 (urea transporter), member 1	ENSMUSG00000059336	Slc14a1	0.008	0.02	1.2	over
M200006592	solute carrier family 25 (mitochondrial carnitine/acylcarnitine translocase), member 20	ENSMUSG00000032602	Slc25a20	0.008	0.02	1.2	under
M300006332	solute carrier family 9 (sodium/hydrogen exchanger), member 1	ENSMUSG00000028854	Slc9a1	0.008	0.02	1.2	over
M400002608	stabilin 1	ENSMUSG00000042286	Stab1	0.008	0.02	1.3	under
M200005692	synaptonemal complex central element protein 2	ENSMUSG00000003824	Syce2	0.008	0.02	1.3	over
M300001502	TH1-like homolog (Drosophila)	ENSMUSG00000016253	Th1l	0.008	0.02	1.2	under
M200016461	transmembrane protein 55A	ENSMUSG00000028221	Tmem55a	0.008	0.02	1.2	over
M200004150	tumor necrosis factor receptor superfamily, member 19	ENSMUSG00000060548	Tnfrfs19	0.008	0.02	1.2	under
M200005594	ubiquitin-conjugating enzyme E2A, RAD6 homolog (S. cerevisiae)	ENSMUSG00000016308	Ube2a	0.008	0.02	1.2	over
M400003253	ubiquitin-conjugating enzyme E2N	ENSMUSG00000047099	Ube2n	0.008	0.02	1.2	over
M200003397	upstream transcription factor 2	ENSMUSG00000058239	Usf2	0.008	0.02	1.2	under
M200002100	Inactive ubiquitin carboxyl-terminal hydrolase 53	ENSMUSG00000039701	Usp53	0.008	0.02	1.2	under
M300014679	vacuolar protein sorting 37D (yeast)	ENSMUSG00000043614	Vps37d	0.008	0.02	1.2	under
M200006215	WD repeat domain 40A	ENSMUSG00000028436	Wdr40a	0.008	0.02	1.2	over
M300013345	WAS/WASL interacting protein family, member 1	ENSMUSG00000075284	Wipf1	0.008	0.02	1.2	under
M300001346	Yamaguchi sarcoma viral (v-yes) oncogene homolog 1	ENSMUSG00000014932	Yes1	0.008	0.02	1.2	under
M400003554	zinc finger protein 260	ENSMUSG00000049421	Zfp260	0.008	0.02	1.3	over

Operon Oligo ID	Description	ENSEMBL / Refseq / Riken ID	Symbol	Info p-value	SAM p-value	Fold Difference	Over / Under expressed in MS
M400001713	zinc finger, FYVE domain containing 9	ENSMUSG00000034557	Zfyve9	0.008	0.02	1.2	under
M400002192	No longer in ENSEMBL	ENSMUSG00000038785	Not available	0.008	0.02	1.2	over
M400002878	No longer in ENSEMBL	ENSMUSG00000072839	Not available	0.008	0.02	1.2	over
M300010161	No longer in ENSEMBL	ENSMUSG00000073011	Not available	0.008	0.02	1.2	under
M400018084	Uncharacterised	AK142255	Not available	0.008	0.02	1.3	over
M400002170	Uncharacterised	AK139544	Not available	0.008	0.02	1.3	over
M400013197	Uncharacterised	AC111124	Not available	0.008	0.02	1.2	over
M400013737	Uncharacterised	AK008796	Not available	0.008	0.02	1.2	over
M400011890	Uncharacterised	ENSMUSG00000063388	Not available	0.008	0.02	1.2	over
M400001808	Uncharacterised	AC142415	Not available	0.008	0.02	1.2	over
M400014070	Uncharacterised	AC154200	Not available	0.008	0.02	1.2	over
M400015263	Uncharacterised	AK032937	Not available	0.008	0.02	1.2	over
M400018196	Uncharacterised	AC115868	Not available	0.008	0.02	1.2	over
M400013166	Uncharacterised	AC168091	Not available	0.008	0.02	1.2	over
M400014922	Uncharacterised	AC152940	Not available	0.008	0.02	1.2	under
M400018764	Uncharacterised	AK037783	Not available	0.008	0.02	1.2	under
M400017117	Uncharacterised	AC000399	Not available	0.008	0.02	1.2	under
M400012705	Uncharacterised	AC164431	Not available	0.008	0.02	1.2	under
M400018571	Uncharacterised	AC102761	Not available	0.008	0.02	1.2	under
M400016240	Uncharacterised	AK029951	Not available	0.008	0.02	1.2	under
M400015028	Uncharacterised	AK019931	Not available	0.008	0.02	1.2	under
M400017336	Uncharacterised	AC167234	Not available	0.008	0.02	1.2	under
M400010148	Uncharacterised	CT010429	Not available	0.008	0.02	1.2	under
M400016104	Uncharacterised	AC121294	Not available	0.008	0.02	1.3	under
M400016580	Uncharacterised	AK032679	Not available	0.008	0.02	1.4	under

Table S.3.3 Hypothalamus differentially expressed genes

Operon Oligo ID	Description	ENSEMBL / Refseq / Riken ID	Symbol	Info p-value	SAM p-value	Fold Difference	Over / Under expressed in MS
M200015800	RIKEN cDNA 1700001G17 gene	ENSMUSG00000044744	1700001G17Rik	0.008	0.02	1.2	over
M300013844	RIKEN cDNA 1700012A16 gene	ENSMUSG00000042708	1700012A16Rik	0.008	0.02	1.2	over
M400003715	RIKEN cDNA 1700019N12 gene	ENSMUSG00000050623	1700019N12Rik	0.008	0.02	1.2	under
M200015125	RIKEN cDNA 1700034O15 gene	ENSMUSG00000029867	1700034O15Rik	0.008	0.02	1.2	over
M200009366	RIKEN cDNA 1700040L02 gene	ENSMUSG00000019945	1700040L02Rik	0.008	0.02	1.2	over
M400001026	RIKEN cDNA 1810062G17 gene	ENSMUSG00000027713	1810062G17Rik	0.008	0.02	1.2	under

Operon Oligo ID	Description	ENSEMBL / Refseq / Riken ID	Symbol	Info p-value	SAM p-value	Fold Difference	Over / Under expressed in MS
M400000976	RIKEN cDNA 2010317E24 gene	ENSMUSG00000026955	2010317E24Rik	0.008	0.02	1.2	over
M400013683	Predicted gene	ENSMUSG00000062826	2310038E17Rik	0.008	0.02	1.2	under
M200015932	RIKEN cDNA 2410081M15 gene	ENSMUSG00000028807	2410081M15Rik	0.008	0.02	1.2	under
M200015328	RIKEN cDNA 4833424O15 gene	ENSMUSG00000033342	4833424O15Rik	0.008	0.02	1.2	under
M400016122	RIKEN cDNA 4930423H22 gene	ENSMUSG00000073093	4930423H22Rik	0.008	0.02	1.2	over
M200007272	RIKEN cDNA 4930556P03 gene	ENSMUSG00000037525	4930556P03Rik	0.008	0.02	1.3	under
M300007995	RIKEN cDNA 4932417I16 gene	ENSMUSG00000031951	4932417I16Rik	0.008	0.02	1.3	under
M400005485	ARP3 actin-related protein 3 homolog B (yeast)	ENSMUSG00000056367	Actr3b	0.008	0.02	1.2	under
M300003877	A kinase (PRKA) anchor protein 8	ENSMUSG00000024045	Akap8	0.008	0.02	1.2	over
M200002343	ATPase, Ca++ transporting, ubiquitous	ENSMUSG00000020788	Atp2a3	0.008	0.02	1.2	over
M200016401	apoptosis, caspase activation inhibitor	ENSMUSG00000003604	Aven	0.008	0.02	1.2	under
M400013174	Homeobox protein BarH-like 2	ENSMUSG00000032033	Barx2	0.008	0.02	1.2	under
M400003304	cDNA sequence BC002199	ENSMUSG00000074238	BC002199	0.008	0.02	1.2	over
M400000099	cDNA sequence BC023814	ENSMUSG00000004896	BC023814	0.008	0.02	1.3	over
M400005994	Zinc finger 525-100	ENSMUSG00000061371	BC024063	0.008	0.02	1.2	over
M300017353	calcium/calmodulin-dependent protein kinase II inhibitor 1	ENSMUSG00000046447	Camk2n1	0.008	0.02	1.2	under
M200015884	calcium/calmodulin-dependent protein kinase kinase 2, beta	ENSMUSG00000029471	Camkk2	0.008	0.02	1.2	under
M200013299	CaM kinase-like vesicle-associated	ENSMUSG00000032936	Camkv	0.008	0.02	1.2	under
M200000995	cholecystokinin	ENSMUSG00000032532	Cck	0.008	0.02	1.2	under
M400011128	chemokine (C-C motif) ligand 9	ENSMUSG00000019122	Ccl9	0.008	0.02	1.3	under
M200001209	cyclin D2	ENSMUSG00000000184	Ccnd2	0.008	0.02	1.2	under
M400013779	T-complex protein 1 subunit delta	ENSMUSG00000007739	Cct4	0.008	0.02	1.2	over
M300015802	Cd300D antigen	ENSMUSG00000044811	Cd300d	0.008	0.02	1.2	under
M300008590	CDC14 cell division cycle 14 homolog B (S. cerevisiae)	ENSMUSG00000033102	Cdc14b	0.008	0.02	1.2	over
M400001836	chromodomain helicase DNA binding protein 8	ENSMUSG00000053754	Chd8	0.008	0.02	1.2	over
M300000926	calcium homeostasis endoplasmic reticulum protein	ENSMUSG00000052488	Cherp	0.008	0.02	1.2	over
M300000867	chordin	ENSMUSG00000006958	Chrd	0.008	0.02	1.2	over
M200006459	Carbohydrate sulfotransferase 8	ENSMUSG00000060402	Chst8	0.008	0.02	1.2	under
M400001075	chloride channel calcium activated 6	ENSMUSG00000068547	Clca6	0.008	0.02	1.2	under
M200006831	caseinolytic peptidase X (E.coli)	ENSMUSG00000015357	Clpx	0.008	0.02	1.2	over
M200004901	camello-like 2	ENSMUSG00000033634	Cml2	0.008	0.02	1.2	over
M200012882	coronin, actin binding protein 1C	ENSMUSG00000004530	Coro1c	0.008	0.02	1.2	under
M300006401	cortistatin	ENSMUSG00000028971	Cort	0.008	0.02	1.2	under
M200005258	cryptochrome 1 (photolyase-like)	ENSMUSG00000020038	Cry1	0.008	0.02	1.2	under
M400013920	CUG triplet repeat, RNA binding protein 2	ENSMUSG00000002107	Cugbp2	0.008	0.02	1.3	under
M300013668	cytochrome P450, family 2, subfamily c, polypeptide 70	ENSMUSG00000060613	Cyp2c70	0.008	0.02	1.2	under
M300004039	cytochrome P450, family 4, subfamily f, polypeptide 13	ENSMUSG00000024055	Cyp4f13	0.008	0.02	1.2	under
M400001249	RIKEN cDNA D030011O10 gene	ENSMUSG00000030313	D030011O10Rik	0.008	0.02	1.2	over

Operon Oligo ID	Description	ENSEMBL / Refseq / Riken ID	Symbol	Info p-value	SAM p-value	Fold Difference	Over / Under expressed in MS
M200003079	similar to spaghetti CG13570-PA (LOC669249), mRNA	ENSMUSG00000022466	D15Ertd682e	0.008	0.02	1.2	over
M400014840	Netrin receptor DCC precursor	ENSMUSG00000060534	Dcc	0.008	0.02	1.2	over
M300007781	dCMP deaminase	ENSMUSG00000031562	Dctd	0.008	0.02	1.2	over
M200007234	DnaJ (Hsp40) homolog, subfamily C, member 12	ENSMUSG00000036764	Dnajc12	0.008	0.02	1.2	over
M200005658	DnaJ (Hsp40) homolog, subfamily C, member 15	ENSMUSG00000022013	Dnajc15	0.008	0.02	1.2	under
M200000698	dual specificity phosphatase 6	ENSMUSG00000019960	Dusp6	0.008	0.02	1.2	under
M400005251	RIKEN cDNA E030030I06 gene	ENSMUSG00000055657	E030030I06Rik	0.008	0.02	1.2	over
M300005502	RIKEN cDNA E430002G05 gene	ENSMUSG00000027188	E430002G05Rik	0.008	0.02	1.3	under
M300008014	EGL nine homolog 1 (C. elegans)	ENSMUSG00000031987	Egln1	0.008	0.02	1.2	under
M200009058	EH-domain containing 4	ENSMUSG00000027293	Ehd4	0.008	0.02	1.2	under
M400018616	erythrocyte protein band 4.1	ENSMUSG00000028906	Epb4.1	0.008	0.02	1.2	over
M300019099	FERM domain containing 6	ENSMUSG00000048285	Frmd6	0.008	0.02	1.2	under
M200007770	glial cell line derived neurotrophic factor family receptor alpha 4	ENSMUSG00000027316	Gfra4	0.008	0.02	1.2	under
M200015701	GINS complex subunit 1 (Psf1 homolog)	ENSMUSG00000027454	Gins1	0.008	0.02	1.2	over
M400000124	Predicted gene	ENSMUSG00000006218	Gm693	0.008	0.02	1.2	under
M400009212	GNAS (guanine nucleotide binding protein, alpha stimulating) complex locus	ENSMUSG00000027523	Gnas	0.008	0.02	1.2	over
M400004037	glutamate receptor, ionotropic, NMDA2B (epsilon 2)	AC124500	Grin2b	0.008	0.02	1.2	under
M300006547	GUF1 GTPase homolog (S. cerevisiae)	ENSMUSG00000029208	Guf1	0.008	0.02	1.2	over
M200002862	hypocretin	ENSMUSG00000045471	Hcrt	0.008	0.02	1.2	over
M300002684	HECT, C2 and WW domain containing E3 ubiquitin protein ligase 1	ENSMUSG00000021301	Hecw1	0.008	0.02	1.2	over
M200008631	heparan sulfate (glucosamine) 3-O-sulfotransferase 3B1	ENSMUSG00000070407	Hs3st3b1	0.008	0.02	1.2	under
M200008813	hypoxia up-regulated 1	ENSMUSG00000032115	Hyou1	0.008	0.02	1.2	over
M200001848	intercellular adhesion molecule 5, telencephalin	ENSMUSG00000032174	Icam5	0.008	0.02	1.2	under
M200006446	isoprenylcysteine carboxyl methyltransferase	ENSMUSG00000039662	Icmt	0.008	0.02	1.2	under
M400000317	myo-inositol 1-phosphate synthase A1	ENSMUSG00000019139	Isyna1	0.008	0.02	1.2	over
M300005569	inositol 1,4,5-trisphosphate 3-kinase A	ENSMUSG00000027296	Itpka	0.008	0.02	1.3	under
M400011396	potassium voltage-gated channel member 1-like	ENSMUSG00000057263	Kcne1	0.008	0.02	1.2	over
M200015255	potassium channel, subfamily K, member 5	ENSMUSG00000023243	Kcnk5	0.008	0.02	1.2	under
M200009625	kelch-like ECH-associated protein 1	ENSMUSG00000003308	Keap1	0.008	0.02	1.2	under
M400000288	karyopherin (importin) alpha 2	ENSMUSG00000018362	Kpna2	0.008	0.02	1.2	over
M400009871	La ribonucleoprotein domain family, member 1	ENSMUSG00000037331	Larp1	0.008	0.02	1.2	under
M200009425	LIM homeobox protein 2	ENSMUSG00000000247	Lhx2	0.008	0.02	1.2	under
M400008454	40S RIBOSOMAL S27A-47	ENSMUSG00000059461	LOC670592	0.008	0.02	1.2	under
M400008543	Predicted | similar to ribosomal protein L36	ENSMUSG00000066068	LOC673026	0.008	0.02	1.2	under
M300017821	leucine rich repeat containing 16	ENSMUSG00000021338	Lrrc16	0.008	0.02	1.2	under
M400016061	mitogen activated protein kinase binding protein 1	ENSMUSG00000033902	Mapkbp1	0.008	0.02	1.2	over

Operon Oligo ID	Description	ENSEMBL / Refseq / Riken ID	Symbol	Info p-value	SAM p-value	Fold Difference	Over / Under expressed in MS
M300010898	mucin 3, intestinal	ENSMUSG00000037390	Muc3	0.008	0.02	1.2	over
M300003301	NEL-like 2 (chicken)	ENSMUSG00000022454	Nell2	0.008	0.02	1.2	under
M200001853	neurogenic differentiation 1	ENSMUSG00000034701	Neurod1	0.008	0.02	1.3	under
M400017898	neuroligin 1	ENSMUSG00000074664	Nlgn1	0.008	0.02	1.2	over
M400012137	olfactory receptor 1270	ENSMUSG00000075065	Olfr1270	0.008	0.02	1.2	over
M400008341	Olfactory receptor 486	ENSMUSG00000063610	Olfr486	0.008	0.02	1.3	under
M400002959	olfactory receptor 649	ENSMUSG00000044899	Olfr649	0.008	0.02	1.2	over
M200007302	phosphatidylinositol glycan anchor biosynthesis, class M	ENSMUSG00000050229	Pigm	0.008	0.02	1.2	under
M200001231	protein kinase inhibitor, alpha	ENSMUSG00000027499	Pkia	0.008	0.02	1.2	over
M400002331	phospholipase C, beta 2	ENSMUSG00000040061	Plcb2	0.008	0.02	1.2	under
M300002680	phospholipase D family, member 4	ENSMUSG00000052160	Pld4	0.008	0.02	1.2	under
M400001802	pro-melanin-concentrating hormone	ENSMUSG00000035383	Pmch	0.008	0.02	1.4	over
M400011108	POU domain, class 3, transcription factor 1	ENSMUSG00000043435	Pou3f1	0.008	0.02	1.2	under
M300003924	protein phosphatase 1B, magnesium dependent, beta isoform	ENSMUSG00000061130	Ppm1b	0.008	0.02	1.2	under
M200003318	protein phosphatase 1G (formerly 2C), magnesium-dependent, gamma isoform	ENSMUSG00000029147	Ppm1g	0.008	0.02	1.3	under
M400004409	Protein kinase C gamma type	ENSMUSG00000078816	Prkcc	0.008	0.02	1.2	under
M300012194	paired related homeobox 2	ENSMUSG00000039476	Prrx2	0.008	0.02	1.3	over
M200003526	proteosome (prosome, macropain) 26S subunit, non-ATPase, 10	ENSMUSG00000031429	Psmd10	0.008	0.02	1.2	under
M200000357	protein tyrosine phosphatase, non-receptor type 2	ENSMUSG00000024539	Ptpn2	0.008	0.02	1.2	over
M400003297	RNA for type IIB intracisternal A-particle (IAP) element encoding integrase, clone 111. (Fragment).	ENSMUSG00000073213	Q78E13_MOUSE	0.008	0.02	1.2	over
M200014525	glutaminyl-peptide cyclotransferase (glutaminyl cyclase)	ENSMUSG00000024084	Qpct	0.008	0.02	1.2	over
M200015891	RAB22A, member RAS oncogene family	ENSMUSG00000027519	Rab22a	0.008	0.02	1.2	over
M200002782	RAD51-like 3 (S. cerevisiae)	ENSMUSG00000018841	Rad51l3	0.008	0.02	1.3	over
M200012063	RAS-like, family 11, member B	ENSMUSG00000049907	Rasl11b	0.008	0.02	1.2	under
M400007329	resistin like alpha	ENSMUSG00000061100	Retnla	0.008	0.02	1.2	over
M300011946	relaxin 1	ENSMUSG00000039097	Rln1	0.008	0.02	1.2	over
M400000212	ring finger protein 166	ENSMUSG00000014470	Rnf166	0.008	0.02	1.2	under
M400006109	ribosomal protein S2	ENSMUSG00000044533	Rps2	0.008	0.02	1.2	under
M300013441	S100 calcium binding protein A10 (calpactin)	ENSMUSG00000041959	S100a10	0.008	0.02	1.2	under
M200008525	SERTA domain containing 2	ENSMUSG00000049800	Sertad2	0.008	0.02	1.2	under
M200009322	superkiller viralicidic activity 2-like 2 (S. cerevisiae)	ENSMUSG00000016018	Skiv2l2	0.008	0.02	1.2	under
M400001620	solute carrier family 16 (monocarboxylic acid transporters), member 2	ENSMUSG00000033965	Slc16a2	0.008	0.02	1.2	under
M200013669	solute carrier family 17 (sodium-dependent inorganic phosphate cotransporter), member 7	ENSMUSG00000070570	Slc17a7	0.008	0.02	1.2	under
M200001847	sphingomyelin phosphodiesterase 1, acid lysosomal	ENSMUSG00000037049	Smpd1	0.008	0.02	1.3	under

Operon Oligo ID	Description	ENSEMBL / Refseq / Riken ID	Symbol	Info p-value	SAM p-value	Fold Difference	Over / Under expressed in MS
M200008410	serum response factor binding protein 1	ENSMUSG00000024528	Srfbp1	0.008	0.02	1.2	over
M400013673	galactose-3-O-sulfotransferase 1	ENSMUSG00000049721	St3gal6	0.008	0.02	1.2	under
M300003441	transgelin 3	ENSMUSG00000022658	Tagln3	0.008	0.02	1.2	under
M400003765	transcription elongation factor A (SII) 1	ENSMUSG00000051015	Tcea1	0.008	0.02	1.2	over
M300004177	transcription elongation regulator 1	ENSMUSG00000024498	Terg1	0.008	0.02	1.2	over
M200006593	THAP domain containing 4	ENSMUSG00000026279	Thap4	0.008	0.02	1.2	over
M400014516	ubiquitin-conjugating enzyme E2R 2	ENSMUSG00000036241	Ube2r2	0.008	0.02	1.2	under
M300010942	UDP-glucose ceramide glucosyltransferase-like 1	ENSMUSG00000037470	Ugcgl1	0.008	0.02	1.2	under
M300004825	unc-5 homolog A (C. elegans)	ENSMUSG00000025876	Unc5a	0.008	0.02	1.2	under
M400002803	Usher syndrome 2A (autosomal recessive, mild) homolog (human)	ENSMUSG00000026609	Ush2a	0.008	0.02	1.2	over
M200005583	ubiquitin specific peptidase 21	ENSMUSG00000053483	Usp21	0.008	0.02	1.2	over
M400016001	ubiquitin specific peptidase 29	ENSMUSG00000051527	Usp29	0.008	0.02	1.2	over
M200007274	Wnt inhibitory factor 1	ENSMUSG00000020218	Wif1	0.008	0.02	1.2	under
M300007261	Zinc finger protein 235 (Zinc finger protein 93)	ENSMUSG00000055305	Zfp-93	0.008	0.02	1.2	over
M200001030	zinc finger protein of the cerebellum 1	ENSMUSG00000032368	Zic1	0.008	0.02	1.2	over
M400010646	No longer in ENSEMBL	ENSMUSG00000056304	Not available	0.008	0.02	1.2	under
M400007070	Predicted gene	ENSMUSG00000072489	Not available	0.008	0.02	1.2	under
M400012685	Predicted gene	ENSMUSG00000059645	Not available	0.008	0.02	1.3	over
M400019365	Uncharacterised	AC004093	Not available	0.008	0.02	1.3	over
M400018289	Uncharacterised	AK046257	Not available	0.008	0.02	1.3	over
M400013972	Uncharacterised	AF217545	Not available	0.008	0.02	1.2	over
M400017862	Uncharacterised	AK043124	Not available	0.008	0.02	1.2	over
M400012918	Uncharacterised	AK021030	Not available	0.008	0.02	1.2	over
M400016287	Uncharacterised	AC151982	Not available	0.008	0.02	1.2	over
M400016104	Uncharacterised	AC121294	Not available	0.008	0.02	1.2	over
M400014302	Uncharacterised	AC131734	Not available	0.008	0.02	1.2	over
M400013081	Uncharacterised	CT030641	Not available	0.008	0.02	1.2	over
M400015871	Uncharacterised	AK030486	Not available	0.008	0.02	1.2	over
M400006946	Uncharacterised	AC164642	Not available	0.008	0.02	1.2	under
M400016523	Uncharacterised	AL669857	Not available	0.008	0.02	1.2	under
M400013955	Uncharacterised	AK014152	Not available	0.008	0.02	1.2	under
M400018551	Uncharacterised	AK083146	Not available	0.008	0.02	1.2	under
M400018451	Uncharacterised	XM_135150	Not available	0.008	0.02	1.2	under
M400006139	Uncharacterised	CT027693	Not available	0.008	0.02	1.2	under
M400019234	Uncharacterised	AK033031	Not available	0.008	0.02	1.2	under
M400002772	Uncharacterised	AC093175	Not available	0.008	0.02	1.3	under

Table S.3.4 Peripheral Blood Mononuclear Cells differentially expressed genes

Operon Oligo ID	Description	ENSEMBL / Refseq / Riken ID	Symbol	Info p-value	SAM p-value	Fold Difference	Over / Under expressed in MS
M300004883	RIKEN cDNA 1110028C15 gene	ENSMUSG00000026004	1110028C15Rik	0.012	0.02	1.2	over
M200006106	RIKEN cDNA 1500010J02 gene	ENSMUSG00000020898	1500010J02Rik	0.026	0.02	1.2	over
M300004934	RIKEN cDNA 1700019A02 gene	ENSMUSG00000060715	1700019A02Rik	0.001	0.02	1.3	over
M200011015	RIKEN cDNA 1700020C07 gene	ENSMUSG00000017767	1700020C07Rik	0.002	0.02	1.3	over
M200010107	RIKEN cDNA 1700028J19 gene	ENSMUSG00000038782	1700028J19Rik	0.012	0.02	1.4	over
M400009581	RIKEN cDNA 1700055M20 gene	ENSMUSG00000031794	1700055M20Rik	0.026	0.02	1.2	over
M200011472	RIKEN cDNA 1700104B16 gene	ENSMUSG00000049476	1700104B16Rik	0.007	0.02	1.2	under
M200014164	RIKEN cDNA 1810049H13 gene	ENSMUSG00000039670	1810049H13Rik	0.026	0.02	1.2	over
M400006062	Novel KRAB box and zinc finger, C2H2 type domain containing protein	ENSMUSG00000074529	2210418O10Rik	0.026	0.02	1.2	over
M200014492	RIKEN cDNA 2310038H17 gene	ENSMUSG00000025956	2310038H17Rik	0.001	0.02	1.2	over
M300021962	RIKEN cDNA 2310047D13 gene	ENSMUSG00000069808	2310047D13Rik	0.001	0.02	1.3	over
M400009316	RIKEN cDNA 2610305D13 gene	ENSMUSG00000066000	2610305D13Rik	0.001	0.02	1.2	over
M200015309	RIKEN cDNA 2700046A07 gene	ENSMUSG00000041789	2700046A07Rik	0.012	0.02	1.3	over
M400002602	RIKEN cDNA 4732466D17 gene	ENSMUSG00000042251	4732466D17Rik	0.026	0.02	1.2	over
M200006497	RIKEN cDNA 4733401H18 gene	ENSMUSG00000070283	4733401H18Rik	0.026	0.02	1.2	under
M400008627	RIKEN cDNA 4921528I07 gene	ENSMUSG00000074149	4921528I07Rik	0.001	0.02	1.3	over
M300017458	RIKEN cDNA 4930430F08 gene	ENSMUSG00000046567	4930430F08Rik	0.012	0.02	1.2	under
M300016909	RIKEN cDNA 4930451I11 gene	ENSMUSG00000045989	4930451I11Rik	0.001	0.02	1.4	over
M300020486	RIKEN cDNA 4930479M11 gene	ENSMUSG00000049761	4930479M11Rik	0.012	0.02	1.2	under
M200013517	RIKEN cDNA 4932422M17 gene	ENSMUSG00000062946	4932422M17Rik	0.012	0.02	1.4	over
M300017039	RIKEN cDNA 5830418K08 gene	ENSMUSG00000046111	5830418K08Rik	0.026	0.02	1.2	under
M400007119	RIKEN cDNA 6330578E17 gene	ENSMUSG00000062395	6330578E17Rik	0.026	0.02	1.2	under
M300018310	RIKEN cDNA 8030462N17 gene	ENSMUSG00000047466	8030462N17Rik	0.026	0.02	1.2	over
M400006034	RIKEN cDNA 9630041A04 gene	ENSMUSG00000057710	9630041A04Rik	0.002	0.02	1.2	over
M200008402	RIKEN cDNA 9830124H08 gene	ENSMUSG00000040123	9830124H08Rik	0.026	0.02	1.2	over
M300010011	RIKEN cDNA A430041B07 gene	ENSMUSG00000035614	A430041B07Rik	0.012	0.02	1.2	under
M400002761	RIKEN cDNA A430060F13 gene	ENSMUSG00000043522	A430060F13Rik	0	0.02	1.5	over
M400005054	RIKEN cDNA A630012P03 gene	ENSMUSG00000055110	A630012P03Rik	0.007	0.02	1.3	over
M300001265	RIKEN cDNA A930006D11Rik gene	ENSMUSG00000014198	A930006D11Rik	0.001	0.02	1.3	over
M200004458	abhydrolase domain containing 8	ENSMUSG00000007950	Abhd8	0.001	0.02	1.2	over
M300001666	ABI gene family, member 3	ENSMUSG00000018381	Abi3	0.007	0.02	1.2	over
M200012683	acetyl-Coenzyme A acetyltransferase 2	ENSMUSG00000023832	Acat2	0.001	0.02	1.2	over
M200001948	acrosomal vesicle protein 1	ENSMUSG00000032110	Acrv1	0.026	0.02	1.3	over
M400000215	a disintegrin and metallopeptidase domain 28	ENSMUSG00000014725	Adam28	0.001	0.02	1.2	over
M400004596	a disintegrin-like and metalloprotease	ENSMUSG00000030022	Adamts9	0.001	0.02	1.2	over
M200000582	adenylate cyclase 8	ENSMUSG00000022376	Adcy8	0	0.02	1.2	over

Operon Oligo ID	Description	ENSEMBL / Refseq / Riken ID	Symbol	Info p-value	SAM p-value	Fold Difference	Over / Under expressed in MS
M200013059	adiponectin receptor 1	ENSMUSG00000026457	Adipor1	0.001	0.02	1.3	over
M400017721	Alkyldihydroxyacetonephosphate synthase, peroxisomal precursor	ENSMUSG00000042410	Agps	0.026	0.02	1.3	over
M400000445	aldo-keto reductase family 1, member C19	ENSMUSG00000071551	Akr1c19	0.001	0.02	1.4	over
M200006661	aldehyde dehydrogenase 9, subfamily A1	ENSMUSG00000026687	Aldh9a1	0.007	0.02	1.2	under
M300011333	amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 11 (human)	ENSMUSG00000072295	Als2cr11	0.012	0.02	1.2	under
M400010725	S-adenosylmethionine decarboxylase 1	ENSMUSG00000060096	Amd1	0.007	0.02	1.2	over
M400003439	adhesion molecule, interacts with CXADR antigen 1	ENSMUSG00000048534	Amica1	0.002	0.02	1.2	over
M200008803	ankyrin repeat domain 25	ENSMUSG00000032194	Ankrd25	0.001	0.02	1.3	over
M400001429	archain 1	ENSMUSG00000032096	Arcn1	0.026	0.02	1.2	under
M400000390	ADP-ribosylation factor 5	ENSMUSG00000020440	Arf5	0.026	0.04	1.2	over
M200002250	ADP-ribosylation factor 4-like	ENSMUSG00000034936	Arfl4	0.001	0.02	1.2	over
M300018941	AT rich interactive domain 4A (Rbp1 like)	ENSMUSG00000048118	Arid4a	0.012	0.02	1.2	under
M200005645	actin related protein 2/3 complex, subunit 5-like	ENSMUSG00000026755	Arpc5l	0	0.02	1.2	over
M200012829	Achaete-scute homolog 2 (Mash-2)	ENSMUSG00000009248	Ascl2	0.007	0.02	1.2	over
M200006901	ATPase, H+ transporting, lysosomal V0 subunit E2	ENSMUSG00000039347	Atp6v0e2	0	0.02	1.2	over
M300018100	expressed sequence AU020772	ENSMUSG00000047248	AU020772	0.002	0.02	1.3	over
M200009796	UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase, polypeptide 5	ENSMUSG00000074892	B3galt5	0.007	0.02	1.2	over
M300009507	brain-specific angiogenesis inhibitor 1	ENSMUSG00000034730	Bai1	0.001	0.02	1.3	over
M300004607	brain-specific angiogenesis inhibitor 1-associated protein 2	ENSMUSG00000025372	Baiap2	0.002	0.02	1.4	over
M200006059	cDNA sequence BC003331	ENSMUSG00000006010	BC003331	0.012	0.02	1.2	under
M400004024	cDNA sequence BC013672	ENSMUSG00000037921	BC013672	0	0.02	1.2	over
M200015742	cDNA sequence BC017612	ENSMUSG00000058173	BC017612	0.012	0.02	1.2	over
M400011905	cDNA sequence BC023179	ENSMUSG00000066838	BC023179	0.012	0.02	1.4	over
M400002266	cDNA sequence BC046331	ENSMUSG00000039523	BC046331	0.012	0.02	1.2	over
M400008030	bone gamma-carboxyglutamate protein, related sequence 1	ENSMUSG00000074489	Bglap-rs1	0.001	0.02	1.2	over
M200004758	biliverdin reductase B (flavin reductase (NADPH))	ENSMUSG00000040466	Blvrb	0.026	0.02	1.2	over
M200000210	bone morphogenetic protein 7	ENSMUSG00000008999	Bmp7	0.012	0.02	1.2	over
M300016134	bolA-like 3 (E. coli)	ENSMUSG00000045160	Bola3	0.001	0.02	1.3	over
M300018670	brain protein I3	ENSMUSG00000047843	Bri3	0.026	0.02	1.2	under
M400001353	Bruton agammaglobulinemia tyrosine kinase	ENSMUSG00000031264	Btk	0.007	0.02	1.2	over
M300004077	butyrophilin-like 1	ENSMUSG00000062638	Btnl1	0.001	0.02	1.2	over
M400009861	butyrophilin-like 9	ENSMUSG00000040283	Btnl9	0.007	0.02	1.3	over
M300007269	RIKEN cDNA C230052I12 gene	ENSMUSG00000030493	C230052I12Rik	0.012	0.02	1.2	over
M400008944	calcium binding protein 39-like	ENSMUSG00000021981	Cab39l	0.026	0.02	1.2	under
M400011436	chaperone, ABC1 activity of bc1 complex like (S. pombe)	ENSMUSG00000026489	Cabc1	0.002	0.02	1.2	over

Operon Oligo ID	Description	ENSEMBL / Refseq / Riken ID	Symbol	Info p-value	SAM p-value	Fold Difference	Over / Under expressed in MS
M300010414	calpain 11	ENSMUSG00000058626	Capn11	0	0.02	1.2	under
M300011602	carbonic anhydrase 14	ENSMUSG00000038526	Car14	0	0.02	1.4	over
M300011605	core-binding factor, runt domain, alpha subunit 2, translocated to, 2 (human)	ENSMUSG00000038533	Cbfa2t2	0.007	0.02	1.3	over
M300009668	chemokine (C-C motif) ligand 5	ENSMUSG00000035042	Ccl5 / Rantes	0.026	0.02	1.2	over
M300006017	cyclin C	ENSMUSG00000028252	Ccnc	0.007	0.02	1.3	under
M300006655	cyclin G2	ENSMUSG00000029385	Ccng2	0.002	0.04	1.2	under
M200014911	chemokine (C-C motif) receptor 1-like 1	ENSMUSG00000064039	Ccr1l1	0.007	0.02	1.3	over
M400011588	CD209b antigen	ENSMUSG00000065987	Cd209b	0.002	0.02	1.2	over
M400015482	M-phase inducer phosphatase 2	ENSMUSG00000027330	Cdc25b	0.001	0.02	1.3	over
M300001903	cell division cycle 2 homolog A (S. pombe)	ENSMUSG00000019942	Cdc2a / Cdk1	0.026	0.04	1.2	over
M300000845	cell division cycle 42 homolog (S. cerevisiae)	ENSMUSG00000006699	Cdc42	0.002	0.02	1.2	under
M200006823	CDC42 effector protein (Rho GTPase binding) 1	ENSMUSG00000049521	Cdc42ep1	0.026	0.02	1.2	over
M400001816	CUB domain containing protein 1	ENSMUSG00000035498	Cdcp1	0.001	0.02	1.3	over
M300004596	cyclin-dependent kinase 2	ENSMUSG00000025358	Cdk2	0.007	0.02	1.2	over
M200000658	cerebellar degeneration-related 2	ENSMUSG00000030878	Cdr2	0.012	0.02	1.2	over
M300017987	clathrin, heavy polypeptide (Hc)	ENSMUSG00000047126	Cltc	0.002	0.02	1.2	under
M200009103	component of oligomeric golgi complex 4	ENSMUSG00000031753	Cog4	0.007	0.02	1.3	over
M200013753	coronin 7	ENSMUSG00000039637	Coro7	0	0.02	1.2	over
M400001817	COX18 cytochrome c oxidase assembly homolog (S. cerevisiae)	ENSMUSG00000035505	Cox18	0	0.02	1.3	over
M200002222	copine VI	ENSMUSG00000022212	Cpne6	0.001	0.02	1.2	over
M200013430	crumbs homolog 1 (Drosophila)	ENSMUSG00000063681	Crb1	0.002	0.02	1.2	over
M300000288	Alpha-S2-casein-like B precursor	ENSMUSG00000061388	Csn1s2b	0.007	0.02	1.2	over
M200001596	casein beta	ENSMUSG00000063157	Csn2	0.007	0.02	1.2	over
M200003934	cytochrome P450, family 2, subfamily c, polypeptide 29	ENSMUSG00000003053	Cyp2c29	0	0.02	1.2	over
M400008475	Cytochrome P450, family 4, subfamily a, polypeptide 31	ENSMUSG00000028712	Cyp4a31	0.007	0.02	1.2	over
M300013894	RIKEN cDNA D130054N24 gene	ENSMUSG00000042790	D130054N24Rik	0.001	0.02	1.3	over
M200005579	Ser/Thr-rich protein T10 in DGCR region.	ENSMUSG00000013539	D16H22S680E	0.007	0.02	1.2	over
M400003995	RIKEN cDNA D330050I23 gene	ENSMUSG00000072569	D330050I23Rik	0.002	0.02	1.2	over
M400004508	RIKEN cDNA D430041B17 gene	ENSMUSG00000053550	D430041B17Rik	0	0.02	1.2	over
M400012469	RIKEN cDNA D830050J10 gene	ENSMUSG00000055396	D830050J10Rik	0.026	0.04	1.2	under
M200009125	deaminase domain containing 1	ENSMUSG00000019808	Deadc1	0	0.02	1.3	over
M200004176	dehydrogenase/reductase (SDR family) member 1	ENSMUSG00000002332	Dhrs1	0.007	0.02	1.2	under
M300010488	dermokine	ENSMUSG00000060962	Dmkn	0	0.02	1.2	over
M200016011	dedicator of cytokinesis 6	ENSMUSG00000032198	Dock6	0.007	0.02	1.2	over
M200003607	dedicator of cytokinesis 7	ENSMUSG00000028556	Dock7	0.001	0.02	1.3	over
M300017240	developmental pluripotency-associated 3	ENSMUSG00000046323	Dppa3	0.001	0.02	1.3	over
M300021745	Down syndrome cell adhesion molecule	ENSMUSG00000050272	Dscam	0.001	0.02	1.2	over

Operon Oligo ID	Description	ENSEMBL / Refseq / Riken ID	Symbol	Info p-value	SAM p-value	Fold Difference	Over / Under expressed in MS
M400001464	dynein cytoplasmic 1 light intermediate chain 1	ENSMUSG00000032435	Dync1li1	0.007	0.02	1.2	over
M400003278	RIKEN cDNA E330009P21 gene	ENSMUSG00000074061	E330009P21Rik	0.007	0.02	1.3	over
M300014949	endothelial differentiation, sphingolipid G-protein-coupled receptor, 5	ENSMUSG00000043895	Edg5	0.001	0.02	1.2	over
M400004753	Predicted gene	ENSMUSG00000054258	EG328231	0.012	0.02	1.2	under
M300009299	Predicted gene	ENSMUSG00000059301	EG432649	0.007	0.02	1.2	over
M400004436	Predicted gene	ENSMUSG00000053337	EG433873	0	0.02	1.4	over
M400001692	Predicted gene	ENSMUSG00000071719	EG620592	0.001	0.02	1.3	over
M400004338	Predicted gene	ENSMUSG00000053098	ENSMUSG00000053098	0.007	0.02	1.2	over
M400004594	Predicted gene	ENSMUSG00000053802	ENSMUSG00000053802	0.026	0.02	1.2	over
M400005267	Predicted gene	ENSMUSG00000055697	ENSMUSG00000055697	0	0.02	1.3	over
M200003518	ependymin related protein 2 (zebrafish)	ENSMUSG00000002808	Epdr2	0.026	0.02	1.2	under
M300006344	Eph receptor A10	ENSMUSG00000028876	Epha10	0.001	0.02	1.2	over
M200002177	epiphycan	ENSMUSG00000019936	Epyc	0.001	0.02	1.3	over
M400003954	glutamate-rich 1	ENSMUSG00000051978	Erich1	0.012	0.02	1.2	over
M300014790	RIKEN cDNA F830045P16 gene	ENSMUSG00000043727	F830045P16Rik	0	0.02	1.4	over
M300002811	FAST kinase domains 3	ENSMUSG00000021532	Fastkd3	0.001	0.02	1.2	over
M400004359	feline sarcoma oncogene	ENSMUSG00000053158	Fes	0.007	0.02	1.2	over
M300008976	fibrinogen, B beta polypeptide	ENSMUSG00000033831	Fgb	0.007	0.02	1.2	over
M200015539	Fgfr1 oncogene partner	ENSMUSG00000069135	Fgfr1op	0.026	0.02	1.2	over
M200012074	forkhead box P3	ENSMUSG00000039521	Foxp3	0.026	0.02	1.3	over
M400010593	Forkhead box protein R1	ENSMUSG00000074397	Foxr1	0	0.02	1.3	over
M300016731	FERM and PDZ domain containing 1	ENSMUSG00000035615	Frmpd1	0.001	0.02	1.2	over
M400003524	growth differentiation factor 2	ENSMUSG00000072625	Gdf2 / Bmp9	0.002	0.04	1.2	over
M200001885	glial cell line derived neurotrophic factor	ENSMUSG00000022144	Gdnf	0.001	0.02	1.2	over
M400011021	gap junction membrane channel protein alpha 9	ENSMUSG00000068615	Gja9	0	0.02	1.2	over
M400000124	Predicted gene	ENSMUSG00000006218	Gm693	0.001	0.02	1.2	over
M200005139	guanine nucleotide binding protein (G protein), gamma 11	ENSMUSG00000032766	Gng11	0.026	0.02	1.2	over
M400004249	G protein-coupled receptor 174	ENSMUSG00000073008	Gpr174	0.012	0.02	1.4	over
M200013878	G substrate	ENSMUSG00000002930	Gsbs	0.012	0.02	1.2	over
M300007072	germ cell-specific gene 1 isoform b	ENSMUSG00000030206	Gsg1	0	0.02	1.3	over
M300003021	guanylate cyclase 1, soluble, beta 2	ENSMUSG00000021933	Gucy1b2	0.001	0.02	1.3	over
M400000720	hydroxyacyl glutathione hydrolase	ENSMUSG00000024158	Hagh	0	0.02	1.2	over
M300002117	hemoglobin, theta 1	ENSMUSG00000020295	Hbq1	0.007	0.02	1.2	over
M200007414	histone deacetylase 7A	ENSMUSG00000022475	Hdac7a	0.001	0.02	1.3	over
M200006352	homocysteine-inducible, endoplasmic reticulum stress-inducible, ubiquitin-like domain member 1	ENSMUSG00000031770	Herpud1	0	0.02	1.2	over
M400003760	hematological and neurological expressed 1-like	ENSMUSG00000050961	Hn1l	0	0.02	1.2	over
M300000132	homeo box A4	ENSMUSG00000000942	Hoxa4	0.001	0.02	1.3	over

Operon Oligo ID	Description	ENSEMBL / Refseq / Riken ID	Symbol	Info p-value	SAM p-value	Fold Difference	Over / Under expressed in MS
M200000051	homeo box A7	ENSMUSG00000038236	Hoxa7	0.002	0.02	1.2	over
M200002749	hydroxysteroid (17-beta) dehydrogenase 2	ENSMUSG00000031844	Hsd17b2	0.002	0.02	1.3	over
M300010568	heat shock protein, alpha-crystallin-related, B6	ENSMUSG00000036854	Hspb6	0.002	0.02	1.2	over
M200000033	inhibitor of DNA binding 3	ENSMUSG00000007872	Id3	0.007	0.02	1.2	over
M200005156	immediate early response 3	ENSMUSG00000003541	Ier3	0.026	0.04	1.2	under
M400008098	Immunoglobulin heavy chain V gene segment	ENSMUSG00000076636	IGHV2-9-1	0.007	0.02	1.2	over
M400005731	Immunoglobulin heavy chain V gene segment	ENSMUSG00000076711	IGHV8-5	0.026	0.04	1.2	over
M400008278	Immunoglobulin Kappa light chain V gene segment	ENSMUSG00000076599	IGKV3-3	0	0.02	1.2	over
M400007654	Immunoglobulin Kappa light chain V gene segment	ENSMUSG00000076547	IGKV4-70	0.026	0.02	1.2	over
M400010434	Immunoglobulin Kappa light chain V gene segment	ENSMUSG00000076587	IGKV6-20	0.001	0.02	1.2	over
M400005854	Immunoglobulin Kappa light chain V gene segment	ENSMUSG00000076576	IGKV6-32	0.026	0.02	1.2	over
M400005852	Immunoglobulin Kappa light chain V gene segment	AJ235968	IGKV6-32	0.007	0.02	1.2	over
M200001766	interleukin 17 receptor A	ENSMUSG00000002897	Il17ra	0.026	0.02	1.2	over
M200007083	integrin alpha FG-GAP repeat containing 2	ENSMUSG00000001518	Itfg2	0.007	0.02	1.4	over
M200000411	integrin beta 2	ENSMUSG00000000290	Itgb2/cd18	0.001	0.02	1.3	under
M300003142	junctophilin 4	ENSMUSG00000022208	Jph4	0.026	0.02	1.2	over
M400010825	potassium inwardly-rectifying channel, subfamily J, member 8	ENSMUSG00000030247	Kcnj8	0.026	0.02	1.2	over
M300017316	potassium inwardly-rectifying channel, subfamily K, member 6	ENSMUSG00000046410	Kcnk6	0.012	0.02	1.2	over
M200000318	kinesin family member C2	ENSMUSG00000004187	Kifc2	0.012	0.02	1.2	over
M200012730	Kruppel-like for 6	ENSMUSG00000000078	Klf6	0.026	0.04	1.2	under
M200004399	keratin 18	ENSMUSG00000023043	Krt18	0.007	0.02	1.2	over
M300014589	keratin 80	ENSMUSG00000037185	Krt80	0.026	0.02	1.2	over
M400008980	LAG1 longevity assurance homolog 5	ENSMUSG00000023021	Lass5	0.001	0.02	1.2	over
M300001812	leukocyte receptor cluster (LRC) member 1	ENSMUSG00000019734	Leng1	0	0.02	1.3	over
M300013305	lipoma HMGIC fusion partner-like 1	ENSMUSG00000041700	Lhfpl1	0.001	0.02	1.2	over
M400004334	lines homolog 2 (Drosophila)	ENSMUSG00000053091	Lins2	0.026	0.02	1.2	over
M400003122	Predicted gene	ENSMUSG00000046157	LOC669382	0.007	0.02	1.2	over
M400002939	Predicted gene	ENSMUSG00000070609	LOC671056	0.001	0.02	1.3	over
M300008195	lysyl oxidase-like 1	ENSMUSG00000032334	Loxl1	0.007	0.02	1.2	over
M300016061	leucine rich repeat neuronal 6C	ENSMUSG00000045083	Lrrn6c	0.007	0.02	1.2	under
M400013298	LSM14 protein homolog A	ENSMUSG00000066568	Lsm14a	0	0.02	1.3	over
M400004821	lysocardiolipin acyltransferase	ENSMUSG00000054469	Lycat	0	0.02	1.2	over
M400001813	mitogen-activated protein kinase kinase kinase 7 interacting protein 3	ENSMUSG00000035476	Map3k7ip3	0.012	0.02	1.2	over
M400009939	mitogen-activated protein kinase kinase kinase 9	ENSMUSG00000042724	Map3k9	0.001	0.02	1.4	over
M300003717	membrane-associated ring finger (C3HC4) 5	ENSMUSG00000023307	March5	0.026	0.02	1.2	over
M400010843	mannan-binding lectin serine peptidase 1	ENSMUSG00000022887	Masp1	0.002	0.02	1.2	over

Operon Oligo ID	Description	ENSEMBL / Refseq / Riken ID	Symbol	Info p-value	SAM p-value	Fold Difference	Over / Under expressed in MS
M200002705	mannose binding lectin (A)	ENSMUSG00000037780	Mbl1	0.012	0.02	1.2	over
M300020874	melanin-concentrating hormone receptor 1	ENSMUSG00000050164	Mchr1	0.012	0.02	1.3	over
M300002849	multiple C2 domains, transmembrane 1	ENSMUSG00000021596	Mctp1	0.007	0.02	1.2	over
M300005462	mediator of RNA polymerase II transcription, subunit 19 homolog (yeast)	ENSMUSG00000027080	Med19	0.001	0.02	1.2	over
M300001453	Mediator of RNA polymerase II transcription subunit 22	ENSMUSG00000015776	Med22	0.001	0.02	1.2	over
M300007290	mesoderm posterior 2	ENSMUSG00000030543	Mesp2	0	0.02	1.2	over
M200004869	meteorin, glial cell differentiation regulator-like	ENSMUSG00000039208	Metrnl	0.007	0.02	1.2	under
M200007783	Glycogen [starch] synthase, brain	ENSMUSG00000003865	MGI:107378	0.007	0.02	1.2	over
M400004452	Taste receptor type 2 member 107 (T2R107) (mT2R43) (T2R4) (STC5-1).	ENSMUSG00000053389	MGI:2681207	0.001	0.02	1.3	over
M200002498	makorin, ring finger protein, 1	ENSMUSG00000029922	Mkrn1	0.012	0.02	1.2	over
M200008501	mutL homolog 3 (E coli)	ENSMUSG00000021245	Mlh3	0.007	0.02	1.3	over
M300016225	MORN repeat containing 2.	ENSMUSG00000045257	Morn2	0.026	0.02	1.2	over
M200000644	Moloney leukemia virus 10	ENSMUSG00000002227	Mov10	0.001	0.02	1.3	over
M200001877	myeloperoxidase	ENSMUSG00000009350	Mpo	0.026	0.04	1.3	under
M400000871	musculin	ENSMUSG00000025930	Msc	0.001	0.02	1.2	over
M400011075	homeo box, msh-like 3	ENSMUSG00000025469	Msx3	0.026	0.02	1.2	over
M400015537	Translation initiation factor IF-2, mitochondrial precursor	ENSMUSG00000020459	Mtif2	0.002	0.02	1.2	over
M400011262	metaxin 2	ENSMUSG00000027099	Mtx2	0.026	0.02	1.2	under
M200010626	matrix-remodelling associated 8	ENSMUSG00000073679	Mxra8	0	0.02	1.2	over
M400005730	myosin, heavy polypeptide 3, skeletal muscle, embryonic	ENSMUSG00000020908	Myh3	0.026	0.02	1.2	over
M400013879	Myelin transcription factor 1-like protein (Zinc finger protein Png-1)	ENSMUSG00000061911	Myt1l	0.007	0.02	1.2	over
M300001550	neighbor of Brca1 gene 1	ENSMUSG00000017119	Nbr1	0.001	0.02	1.3	over
M300000428	nicastrin	ENSMUSG00000003458	Ncstn	0.002	0.02	1.2	under
M200013973	nasal embryonic LHRH factor	ENSMUSG00000006476	Nelf	0.012	0.02	1.2	over
M200003759	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor-like 2	ENSMUSG00000059323	Nfkbil2	0.007	0.02	1.2	over
M400009579	Nucleotide-binding oligomerization domain-containing protein 2 (Caspase recruitment domain-containing protein 15)	ENSMUSG00000055994	Nod2	0.012	0.02	1.2	over
M200007448	nitric oxide synthase interacting protein	ENSMUSG00000003421	Nosip	0	0.02	1.4	over
M400003480	nuclear receptor subfamily 1, group H, member 5	ENSMUSG00000048938	Nr1h5	0.007	0.02	1.2	over
M300007293	nuclear receptor subfamily 2, group F, member 2	ENSMUSG00000030551	Nr2f2	0.012	0.02	1.2	over
M300006242	NOL1/NOP2/Sun domain family, member 4	ENSMUSG00000028706	Nsun4	0.001	0.02	1.3	over
M300000164	oculocerebrorenal syndrome of Lowe	ENSMUSG00000001173	Ocrl	0.026	0.02	1.2	over
M300008556	2-oxoglutarate and iron-dependent oxygenase domain containing 1	ENSMUSG00000033009	Ogfod1	0.026	0.02	1.2	under
M400011972	olfactory receptor 1080	ENSMUSG00000075177	Olfr1080	0.026	0.02	1.2	under

Operon Oligo ID	Description	ENSEMBL / Refseq / Riken ID	Symbol	Info p-value	SAM p-value	Fold Difference	Over / Under expressed in MS
M400006264	olfactory receptor 1137	ENSMUSG00000075150	Olfr1137	0.001	0.02	1.6	over
M300022061	olfactory receptor 1184	ENSMUSG00000051424	Olfr1184	0.001	0.02	1.3	over
M400012127	olfactory receptor 1226	ENSMUSG00000075097	Olfr1226	0.001	0.02	1.3	over
M400012067	olfactory receptor 1245	ENSMUSG00000075083	Olfr1245	0.001	0.02	1.2	under
M400002669	olfactory receptor 1353	ENSMUSG00000042774	Olfr1353	0.026	0.02	1.2	over
M400005952	olfactory receptor 1459	ENSMUSG00000057503	Olfr1459	0.026	0.02	1.2	over
M300018063	olfactory receptor 1495	ENSMUSG00000047207	Olfr1495	0	0.02	1.3	over
M200016327	olfactory receptor 159	ENSMUSG00000044801	Olfr159	0.007	0.02	1.3	over
M300014112	olfactory receptor 275	ENSMUSG00000070984	Olfr275	0.007	0.02	1.2	over
M400008417	olfactory receptor 282	ENSMUSG00000063780	Olfr282	0	0.02	1.2	over
M400006280	olfactory receptor 290	ENSMUSG00000070460	Olfr290	0.012	0.02	1.2	over
M400006011	olfactory receptor 328	ENSMUSG00000057654	Olfr328	0.007	0.02	1.2	over
M400003632	olfactory receptor 350	ENSMUSG00000050015	Olfr350	0.001	0.02	1.3	over
M400003260	olfactory receptor 64	ENSMUSG00000047163	Olfr64	0.001	0.02	1.2	over
M300017588	olfactory receptor 66	ENSMUSG00000058200	Olfr66	0	0.02	1.3	over
M300015973	olfactory receptor 669	ENSMUSG00000073916	Olfr669	0	0.02	1.3	over
M300018080	olfactory receptor 684	ENSMUSG00000047225	Olfr684	0	0.02	1.4	over
M400002773	olfactory receptor 706	ENSMUSG00000036744	Olfr706	0.001	0.02	1.4	over
M400005909	olfactory receptor 804	ENSMUSG00000044537	Olfr804	0.007	0.02	1.2	over
M400007302	olfactory receptor 944	ENSMUSG00000060905	Olfr944	0	0.02	1.2	over
M200009524	opticin	ENSMUSG00000010311	Optc	0.012	0.02	1.2	over
M300002331	Predicted gene	ENSMUSG00000020682	OTTMUSG00000000934	0	0.02	1.2	over
M200003458	oxytocin	ENSMUSG00000027301	Oxt	0	0.02	1.4	over
M300004783	par-3 (partitioning defective 3) homolog (C. elegans)	ENSMUSG00000025812	Pard3	0.001	0.02	1.2	over
M400004003	platelet-derived growth factor receptor-like	ENSMUSG00000052122	Pdgfrl	0.007	0.02	1.2	over
M200001101	preproenkephalin 1	ENSMUSG00000045573	Penk1	0.026	0.02	1.2	under
M200001837	peroxisomal biogenesis factor 14	ENSMUSG00000028975	Pex14	0.001	0.02	1.2	over
M200009514	peroxisome biogenesis factor 19	ENSMUSG00000003464	Pex19	0.001	0.02	1.2	over
M400010863	Phosphoglycerate kinase 1	ENSMUSG00000062070	Pgk1	0.026	0.04	1.2	under
M400001534	pleckstrin homology-like domain, family B, member 2	ENSMUSG00000033149	Phldb2	0.012	0.02	1.2	over
M400000850	phosphatidylinositol glycan anchor biosynthesis, class Q	ENSMUSG00000025728	Pigq	0.007	0.02	1.2	over
M200001202	phosphatidylinositol transfer protein, alpha	ENSMUSG00000017781	Pitpna	0.012	0.02	1.2	over
M400003793	1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase beta-1 (Phospholipase C-beta-1)	ENSMUSG00000051177	Plcb1	0.012	0.02	1.2	under
M200013786	phospholipid scramblase 3	ENSMUSG00000019461	Plscr3	0.001	0.02	1.2	over
M300010461	plexin B2	ENSMUSG00000036606	Plxnb2	0.001	0.02	1.2	under
M200004817	pyridoxine 5'-phosphate oxidase	ENSMUSG00000018659	Pnpo	0.012	0.02	1.2	over
M400010890	polymerase (RNA) II (DNA directed) polypeptide C	ENSMUSG00000031783	Polr2c	0	0.02	1.2	over
M300019159	POU domain, class 4, transcription factor 1	ENSMUSG00000048349	Pou4f1	0.007	0.02	1.3	over

Operon Oligo ID	Description	ENSEMBL / Refseq / Riken ID	Symbol	Info p-value	SAM p-value	Fold Difference	Over / Under expressed in MS
M200008140	peptidylprolyl isomerase (cyclophilin)-like 4	ENSMUSG00000015757	Ppil4	0.012	0.02	1.2	under
M200006133	preimplantation protein 4	ENSMUSG00000027346	Prei4	0.007	0.02	1.2	over
M200007913	Prolactin-3B1 precursor (Chorionic somatomammotropin hormone 2) (Placental lactogen II)	ENSMUSG00000038891	Prl3b1 / Pl-II	0.012	0.02	1.3	over
M200008092	PRP18 pre-mRNA processing factor 18 homolog (yeast)	ENSMUSG00000039449	Prpf18	0.012	0.02	1.2	over
M200000936	peripherin 1	ENSMUSG00000023484	Prph1	0	0.02	1.3	over
M300010472	protease, serine, 34	ENSMUSG00000056399	Prss34	0.001	0.02	1.2	over
M400001933	pre T-cell antigen receptor alpha	ENSMUSG00000036858	Ptcra	0.026	0.02	1.2	over
M300003403	protein tyrosine kinase 2	ENSMUSG00000022607	Ptk2 / FAK1	0	0.02	1.2	under
M200009499	protein tyrosine phosphatase, receptor type, C	ENSMUSG00000026395	Ptprc / CD45	0.012	0.02	1.2	over
M400005872	Ultra-high sulphur keratin homolog	ENSMUSG00000059845	Q9D141_MOUSE	0	0.02	1.2	over
M300001595	RAB11 family interacting protein 4 (class II)	ENSMUSG00000017639	Rab11fip4	0.007	0.02	1.2	over
M200012096	RAB1B, member RAS oncogene family	ENSMUSG00000024870	Rab1b	0.026	0.02	1.2	over
M300012918	RAB3D, member RAS oncogene family	ENSMUSG00000040883	Rab3d	0.001	0.02	1.2	over
M300007367	RAB6, member RAS oncogene family	ENSMUSG00000030704	Rab6	0.026	0.02	1.2	under
M300003258	RAN GTPase activating protein 1	ENSMUSG00000022391	Rangap1	0.001	0.02	1.2	over
M300000176	retinoic acid receptor, gamma	ENSMUSG00000001288	Rarg	0.001	0.02	1.3	over
M200005588	arginyl-tRNA synthetase	ENSMUSG00000018848	Rars	0.012	0.02	1.2	under
M400000339	receptor accessory protein 3	ENSMUSG00000019873	Reep3	0.026	0.02	1.2	over
M400011721	renin 2 tandem duplication of Ren1	ENSMUSG00000070645	Ren2	0	0.02	1.2	over
M200014010	ral guanine nucleotide dissociation stimulator-like 2	ENSMUSG00000041354	Rgl2	0.001	0.02	1.2	over
M300002981	ribonuclease, RNase A family 4	ENSMUSG00000021876	Rnase4	0.007	0.02	1.3	under
M200013340	arginyl aminopeptidase (aminopeptidase B)	ENSMUSG00000041926	Rnpep	0.026	0.02	1.2	under
M400002098	sno, strawberry notch homolog 1 (Drosophila)	ENSMUSG00000038095	Sbno1	0.007	0.02	1.2	under
M300003928	spermine binding protein	ENSMUSG00000024128	Sbp	0.026	0.02	1.3	over
M300008276	SREBF chaperone	ENSMUSG00000032485	Scap	0	0.02	1.3	over
M300000957	secernin 3	ENSMUSG00000008226	Scrn3	0.026	0.04	1.2	under
M200003654	SEC31 homolog A (S. cerevisiae)	ENSMUSG00000035325	Sec31a	0.012	0.02	1.2	over
M400010022	SEC61, gamma subunit	ENSMUSG00000066757	Sec61g	0	0.02	1.3	over
M400009613	SEC61, gamma subunit	ENSMUSG00000071731	Sec61g	0.002	0.02	1.2	over
M400005768	SEC61, gamma subunit	ENSMUSG00000071731	Sec61g	0.026	0.02	1.3	over
M200004341	SUMO/sentrin specific peptidase 3	ENSMUSG00000005204	Senp3	0.012	0.02	1.3	over
M400010913	serine (or cysteine) peptidase inhibitor, clade A, member 1a	ENSMUSG00000071178	Serpina1c	0.007	0.02	1.3	over
M400002918	serine (or cysteine) peptidase inhibitor, clade B, member 3C	ENSMUSG00000073601	Serpinb3c	0.026	0.02	1.2	under
M300002741	serine (or cysteine) peptidase inhibitor, clade B, member 9b	ENSMUSG00000021403	Serpinb9b	0.007	0.02	1.2	over
M400018648	SERTA domain-containing protein 4.	ENSMUSG00000016262	Sertad4	0.002	0.02	1.3	over
M300004390	Splicing factor 1 (Zinc finger protein 162)	ENSMUSG00000024949	Sf1	0.007	0.02	1.2	over

Operon Oligo ID	Description	ENSEMBL / Refseq / Riken ID	Symbol	Info p-value	SAM p-value	Fold Difference	Over / Under expressed in MS
M400000050	splicing factor 3a, subunit 1	ENSMUSG00000002129	Sf3a1	0.007	0.02	1.3	over
M300008504	sphingosine-1-phosphate phosphotase 2	ENSMUSG00000032908	Sgpp2	0.002	0.02	1.2	over
M400000393	serine hydroxymethyl transferase 1 (soluble)	ENSMUSG00000020534	Shmt1	0.012	0.02	1.2	over
M200004083	SNF2 histone linker PHD RING helicase	ENSMUSG00000039282	Shprh	0.007	0.02	1.2	under
M300002683	SIVA1, apoptosis-inducing factor	ENSMUSG00000061516	Siva1	0.026	0.02	1.2	under
M300004285	solute carrier family 15, member 3	ENSMUSG00000024737	Slc15a3	0.007	0.02	1.2	over
M200014065	solute carrier family 25, member 39	ENSMUSG00000018677	Slc25a39	0	0.02	1.3	over
M200002499	solute carrier family 27 (fatty acid transporter), member 1	ENSMUSG00000031808	Slc27a1	0.026	0.02	1.2	over
M400006285	solute carrier family 27 (fatty acid transporter), member 4	ENSMUSG00000059316	Slc27a4	0.007	0.02	1.2	over
M400004909	solute carrier family 2 (facilitated glucose transporter), member 9	ENSMUSG00000005107	Slc2a9	0.012	0.02	1.3	over
M200008829	solute carrier family 43, member 3	ENSMUSG00000027074	Slc43a3	0.012	0.02	1.3	over
M300001120	solute carrier family 5 (sodium/glucose cotransporter), member 1	ENSMUSG00000011034	Slc5a1	0.001	0.02	1.3	over
M200004429	solute carrier family 6 (neurotransmitter transporter, betaine/GABA), member 12	ENSMUSG00000030109	Slc6a12	0.002	0.02	1.2	over
M300002270	structural maintenance of chromosomes 6	ENSMUSG00000020608	Smc6	0.026	0.02	1.2	under
M200008355	Sperm-specific antigen 2 homolog	ENSMUSG00000027007	Ssfa2	0.007	0.02	1.2	under
M400001722	slingshot homolog 3 (Drosophila)	ENSMUSG00000034616	Ssh3	0	0.02	1.3	over
M400003740	somatostatin receptor 5	ENSMUSG00000050824	Sstr5	0.012	0.02	1.2	over
M200001529	signal transducing adaptor molecule (SH3 domain and ITAM motif) 1	ENSMUSG00000026718	Stam	0.002	0.02	1.2	under
M300002124	stanniocalcin 2	ENSMUSG00000020303	Stc2	0.012	0.04	1.2	over
M200000227	stromal interaction molecule 1	ENSMUSG00000030987	Stim1	0	0.02	1.3	over
M300008876	synaptic nuclear envelope 2	ENSMUSG00000054397	Syne2	0.007	0.02	1.3	over
M300012846	tetratricopeptide repeat, ankyrin repeat and coiled-coil containing 2	ENSMUSG00000053580	Tanc2	0.012	0.02	1.2	over
M200004934	TAO kinase 1	ENSMUSG00000017291	Taok1	0.001	0.02	1.2	over
M300003655	TAR (HIV) RNA binding protein 2	ENSMUSG00000023051	Tarbp2	0.002	0.02	1.2	over
M200000720	tubulin cofactor a	ENSMUSG00000042043	Tbca	0.007	0.02	1.2	under
M300013716	transcription factor AP-2, epsilon	ENSMUSG00000042477	Tcfap2e	0.026	0.02	1.2	over
M200003582	T-cell lymphoma breakpoint 1	ENSMUSG00000041359	Tcl1	0.007	0.02	1.2	over
M200011387	Tektin-3	ENSMUSG00000042189	Tekt3	0.012	0.02	1.2	over
M400000616	Thrombopoietin precursor (Megakaryocyte colony-stimulating factor)	ENSMUSG00000022847	Thpo	0.001	0.02	1.2	over
M200007179	thrombospondin, type I, domain 1	ENSMUSG00000031480	Thsd1	0.007	0.02	1.2	over
M400009774	transmembrane BAX inhibitor motif containing 1	ENSMUSG00000006301	Tmbim1	0	0.02	1.2	over
M400004303	transmembrane protein 1	ENSMUSG00000000374	Tmem1	0.002	0.02	1.3	over
M300014461	transmembrane protein 130	ENSMUSG00000043388	Tmem130	0.007	0.02	1.2	over
M200013582	transmembrane protein 25	ENSMUSG00000002032	Tmem25	0.001	0.02	1.2	over

Operon Oligo ID	Description	ENSEMBL / Refseq / Riken ID	Symbol	Info p-value	SAM p-value	Fold Difference	Over / Under expressed in MS
M200015689	transmembrane protein 34	ENSMUSG00000031617	Tmem34	0.007	0.02	1.2	under
M400005236	tensin 3	ENSMUSG00000020422	Tns3	0.026	0.02	1.4	over
M200000672	topoisomerase (DNA) II beta binding protein	ENSMUSG00000032555	Topbp1	0.026	0.02	1.2	under
M300001830	triadin	ENSMUSG00000019787	Trdn	0.026	0.02	1.2	over
M400004294	tripartite motif protein 10	ENSMUSG00000073400	Trim10	0.012	0.04	1.2	over
M400002465	tripartite motif-containing 62	ENSMUSG00000041000	Trim62	0.007	0.02	1.2	over
M300000316	tuberous sclerosis 2	ENSMUSG00000002496	Tsc2	0.007	0.02	1.2	under
M400007283	testis specific 10	ENSMUSG00000060771	Tsga10	0.007	0.02	1.2	over
M200008468	tetratricopeptide repeat domain 28	ENSMUSG00000044237	Ttc28	0.007	0.02	1.2	over
M200009973	tubulin, delta 1	ENSMUSG00000020513	Tubd1	0.026	0.02	1.2	over
M200015843	thioredoxin domain containing 8	ENSMUSG00000038709	Txndc8	0.002	0.02	1.3	over
M300011994	ubiquitin-conjugating enzyme E2H	ENSMUSG00000039159	Ube2h	0.001	0.02	1.2	over
M400011348	ubiquitin-conjugating enzyme E2L 6	ENSMUSG00000027078	Ube2l6	0.026	0.02	1.2	over
M400013127	E3 ubiquitin-protein ligase UBR1	ENSMUSG00000027272	Ubr1	0.026	0.02	1.2	over
M400008661	UDP glucuronosyltransferase 1 family, polypeptide A7C	ENSMUSG00000054545	Ugt1a6a	0.002	0.02	1.2	over
M300000168	ubiquitin-like, containing PHD and RING finger domains, 1	ENSMUSG00000001228	Uhrf1	0.026	0.02	1.3	under
M400006134	unc-13 homolog D	ENSMUSG00000057948	Unc13d	0.002	0.02	1.3	over
M200007213	vesicle-associated membrane protein 1	ENSMUSG00000030337	Vamp1	0.001	0.02	1.2	over
M400010938	vomeronasal 2, receptor, 3	ENSMUSG00000000606	Vmn2r88	0.002	0.02	1.2	over
M300008881	vacuolar protein sorting 8 homolog (S. cerevisiae)	ENSMUSG00000033653	Vps8	0	0.02	1.2	under
M400001818	WD repeat domain 20b	ENSMUSG00000035560	Wdr20b	0.001	0.02	1.2	over
M400013169	xin actin-binding repeat containing 2 isoform 2	ENSMUSG00000027022	Xirp2	0	0.02	1.4	over
M300012466	zinc finger CCCH type containing 12D	ENSMUSG00000039981	Zc3h12d	0.012	0.02	1.2	over
M400013605	zinc finger protein 207	ENSMUSG00000017421	Zfp207	0.026	0.02	1.2	over
M200003451	zinc finger protein 62	ENSMUSG00000046311	Zfp62	0.026	0.02	1.2	under
M400014435	zinc finger protein 84	ENSMUSG00000046185	Zfp84	0	0.02	1.2	over
M200007919	zinc fingers and homeoboxes protein 1	ENSMUSG00000022361	Zhx1	0.026	0.02	1.2	under
M400002768	zinc finger and SCAN domain containing 4D	ENSMUSG00000054272	Zscan4c	0.007	0.02	1.2	under
M400008087	No longer in ENSEMBL	ENSMUSG00000059582	Not available	0.007	0.02	1.3	over
M400002333	No longer in ENSEMBL	ENSMUSG00000040068	Not available	0.012	0.02	1.2	over
M400002629	No longer in ENSEMBL	ENSMUSG00000042454	Not available	0	0.02	1.3	over
M400000686	No longer in ENSEMBL	ENSMUSG00000044936	Not available	0.012	0.02	1.2	over
M400003573	No longer in ENSEMBL	ENSMUSG00000049558	Not available	0.007	0.02	1.2	over
M400005737	No longer in ENSEMBL	ENSMUSG00000074450	Not available	0.026	0.02	1.3	over
M400002071	Predicted gene	XM_146372	Not available	0.007	0.02	1.3	over
M400000200	Predicted gene	ENSMUSG00000013419	Not available	0	0.02	1.2	under
M400006204	Predicted gene	ENSMUSG00000057157	Not available	0.007	0.02	1.2	over
M400008575	Predicted gene	ENSMUSG00000064159	Not available	0.001	0.02	1.3	over
M400011890	Predicted gene	ENSMUSG00000063388	Not available	0.007	0.02	1.2	under

Operon Oligo ID	Description	ENSEMBL / Refseq / Riken ID	Symbol	Info p-value	SAM p-value	Fold Difference	Over / Under expressed in MS
M200010436	Predicted gene	ENSMUSG00000078618	Not available	0.026	0.02	1.3	over
M400016356	Uncharacterised	AC109203	Not available	0.012	0.02	1.2	over
M400016708	Uncharacterised	AC114558	Not available	0.012	0.02	1.2	over
M400014814	Uncharacterised	AC114984	Not available	0.001	0.02	1.2	over
M400015514	Uncharacterised	AC119269	Not available	0.026	0.02	1.2	under
M400012711	Uncharacterised	AC121847	Not available	0.001	0.02	1.2	over
M400017277	Uncharacterised	AC122243	Not available	0.012	0.02	1.2	over
M400003712	Uncharacterised	AC122270	Not available	0	0.02	1.3	over
M400010301	Uncharacterised	AC122439	Not available	0.012	0.02	1.2	over
M400013324	Uncharacterised	AC122458	Not available	0.007	0.02	1.2	under
M400014189	Uncharacterised	AC122733	Not available	0.002	0.02	1.2	over
M400008184	Uncharacterised	AC124739	Not available	0.026	0.02	1.2	under
M300020131	Uncharacterised	AC127254	Not available	0.007	0.02	1.2	over
M400008816	Uncharacterised	AC129022	Not available	0.012	0.02	1.2	over
M400006014	Uncharacterised	AC134443	Not available	0.002	0.02	1.2	over
M400014137	Uncharacterised	AC139350	Not available	0.002	0.02	1.2	over
M400014702	Uncharacterised	AC139942	Not available	0.007	0.02	1.3	over
M300018330	Uncharacterised	AC140210	Not available	0.007	0.02	1.2	over
M400004733	Uncharacterised	AC140381	Not available	0.001	0.02	1.3	over
M400004993	Uncharacterised	AC142267	Not available	0.012	0.02	1.2	over
M400017578	Uncharacterised	AC148007	Not available	0.007	0.02	1.2	over
M400016369	Uncharacterised	AC154767	Not available	0.026	0.02	1.2	over
M400015588	Uncharacterised	AC156606	Not available	0.026	0.02	1.2	under
M400000841	Uncharacterised	AC158224	Not available	0.012	0.02	1.2	over
M400018008	Uncharacterised	AC160535	Not available	0	0.02	1.3	over
M400013221	Uncharacterised	AC163628	Not available	0.007	0.02	1.2	over
M200014069	Uncharacterised	AC164883	Not available	0.007	0.02	1.2	over
M400004880	Uncharacterised	AK128967	Not available	0.012	0.02	1.2	over
M400010648	Uncharacterised	AL670953	Not available	0.012	0.02	1.2	over
M400016344	Uncharacterised	AL807794	Not available	0.001	0.02	1.4	over
M400013686	Uncharacterised	CT030173	Not available	0.001	0.02	1.2	under
M400018037	Uncharacterised	CT030709	Not available	0.026	0.02	1.2	over
M300016779	Uncharacterised	NM_175192	Not available	0.001	0.02	1.2	over
M400014623	Uncharacterised	AK007003	Not available	0.026	0.02	1.3	over
M400013091	Uncharacterised	AK008615	Not available	0.002	0.02	1.3	over
M400013384	Uncharacterised	AK011969	Not available	0.007	0.02	1.3	over
M400014478	Uncharacterised	AK016281	Not available	0.012	0.02	1.3	over
M400015687	Uncharacterised	AK031519	Not available	0.001	0.02	1.2	over
M400015468	Uncharacterised	AK032614	Not available	0.001	0.02	1.2	over

Operon Oligo ID	Description	ENSEMBL / Refseq / Riken ID	Symbol	Info p-value	SAM p-value	Fold Difference	Over / Under expressed in MS
M400016084	Uncharacterised	AK033233	Not available	0.001	0.02	1.2	under
M400018479	Uncharacterised	AK034373	Not available	0.001	0.02	1.2	over
M400018477	Uncharacterised	AK034409	Not available	0	0.02	1.4	over
M400016126	Uncharacterised	AK034614	Not available	0.007	0.02	1.2	under
M400018109	Uncharacterised	AK037255	Not available	0	0.02	1.4	over
M400016981	Uncharacterised	AK037322	Not available	0	0.02	1.2	over
M400017345	Uncharacterised	AK037581	Not available	0.026	0.02	1.3	over
M400017209	Uncharacterised	AK041437	Not available	0.007	0.02	1.2	under
M400018234	Uncharacterised	AK043175	Not available	0.001	0.02	1.4	over
M400017902	Uncharacterised	AK044082	Not available	0.012	0.02	1.2	over
M400004425	Uncharacterised	AK044672	Not available	0.012	0.02	1.3	over
M400014846	Uncharacterised	AK021191	Not available	0.012	0.02	1.2	over
M400019395	Uncharacterised	AK051215	Not available	0.012	0.02	1.3	over
M400019419	Uncharacterised	AK051341	Not available	0.007	0.02	1.3	over
M400017112	Uncharacterised	AK054246	Not available	0	0.02	1.3	over
M400009426	Uncharacterised	X67211	Not available	0.026	0.02	1.4	over
M400017440	Uncharacterised	XM_488999	Not available	0.012	0.02	1.2	over
M400013234	Uncharacterised	XM_489246	Not available	0.007	0.02	1.4	over
M400015117	Uncharacterised	AC102258	Not available	0.012	0.02	1.2	over

- APPENDIX A -

van Heerden *et al.*, 2007

A framework for the informed normalization of printed microarrays

Johan van Heerden*, Sally-Ann Walford*, Arthur Shen* and Nicola Illing*‡

Microarray technology has become an essential part of contemporary molecular biological research. An aspect central to any microarray experiment is that of normalization, a form of data processing directed at removing technical noise while preserving biological meaning, thereby allowing for more accurate interpretations of data. The statistics underlying many normalization methods can appear overwhelming to microarray newcomers, a situation which is further compounded by a lack of accessible, non-statistical descriptions of common approaches to normalization. Normalization strategies significantly affect the analytical outcome of a microarray experiment, and consequently it is important that the statistical assumptions underlying normalization algorithms are understood and met before researchers embark upon the processing of raw microarray data. Many of these assumptions pertain only to whole-genome arrays, and are not valid for custom or directed microarrays. A thorough diagnostic evaluation of the nature and extent to which technical noise affects individual arrays is paramount to the success of any chosen normalization strategy. Here we suggest an approach to normalization based on extensive stepwise exploration and diagnostic assessment of data prior to, and after, normalization. Common data visualization and diagnostic approaches are highlighted, followed by descriptions of popular normalization methods, and the underlying assumptions they are based on, within the context of removing general technical artefacts associated with microarray data.

Introduction

DNA microarrays allow for the large-scale quantitation of gene expression by inference of mRNA transcript abundance.[1] Since its inception, the technology has developed to become an essential item in the biologist's arsenal of tools. Microarray-based techniques rely heavily on various statistical methods for the preparation and analysis of the high-throughput data generated in these experiments. The large numbers, and nature, of variables associated with microarray experiments require novel statistical procedures. These methods present a new challenge to the molecular biologist, requiring a paradigm shift from classic one-gene-at-a-time approaches, to techniques that evaluate thousands of genes simultaneously. An aspect central to any microarray experiment is that of normalization, a form of data processing directed at removing technical noise or systematic variation while preserving biological meaning, thereby allowing for more accurate interpretations of data.[2]

In recent years, biologists have been confronted with a multitude of publications detailing purportedly new and advanced algorithms for the normalization of microarray data. The effectiveness of many algorithms, at reducing error, has been evaluated by using data sets of which sample ratios are known *a priori*.[3,4] Prompted by these studies, we provide an introductory review on the performance and robustness of several commonly used algorithms, highlighting the assumptions that these methods are based on and suggesting an approach to normalization that could be useful when encountering microarray-based techniques for the first time. An *ad hoc* approach is encouraged, recognizing that each microarray experiment will have unique requirements, which have to be identified before deciding on a normalization strategy.

*Department of Molecular and Cell Biology, University of Cape Town, Private Bag, Rondebosch 7701, South Africa.
‡Author for correspondence. E-mail: nicola.illing@uct.ac.za

The need for normalization

Microarray experiments allow biologists to investigate gene expression patterns of thousands of genes in a single assay. The observed patterns of expression in any microarray experiment are affected by several sources of variation, which can obscure true biological values and impede meaningful interpretations.[5] Variation can be divided into two broad categories: (1) the interesting kind, which has biological meaning and is of value to the researcher; (2) the obscuring kind, also referred to as noise or systematic variation, which has no meaning and is a result of the technical error rather than the experimental design.[6] The aim of normalization is to account for these artefactual contributions while preserving the true biological meaning of the observed expression values.

Sources of experimental noise[7] have been well documented,[5,7] and considering their effect on microarray data is an important aspect of any normalization strategy. Systematic errors can be introduced at various points during a microarray experiment, from sample preparation to hybridization and scanning. These errors appear as inconsistencies in the generated data, which can be identified by various diagnostic and visualization tools.[5] Failure to correctly identify and correct for systematic error can lead to results becoming obscured to the point of not containing any biological meaning.[8] There are many alternative methods of normalization available to a researcher. Deciding which normalization algorithms to apply to their data, and being able to substantiate a chosen strategy, are among the challenges faced by researchers.

Experimental design, normalization and the role of controls

It is pertinent to note that printed microarrays commonly come in one of two flavours, referred to as single- or dual-channel arrays. With the former, a single biological sample is labelled and hybridized to an array surface; the latter involves two independent biological samples, each labelled with different fluorophores. This distinction is important, as certain types of technical artefacts occur only in dual-channel arrays. Where necessary, these differences will be highlighted in the text.

When approaching normalization strategies, it is important to realize that underlying many of the algorithms are certain assumptions about the nature of data being normalized. Commonly used normalization algorithms often assume two things: (1) that the majority of genes in a microarray experiment are not differentially regulated, i.e. remain unchanged, and (2) that the number of up-regulated genes are more or less equal to the number of down-regulated genes.[2,8,9] While these assumptions might be accurate for arrays that include all or most of the genes in a genome, it cannot be assumed to be valid for arrays that

include only a subset of genes, often referred to as custom or directed arrays.[5] When the above assumptions are invalid, control spots will be essential to any chosen normalization strategy, where they can be used as stable references for the validation and normalization of microarray data.[8]

Careful experimental design, at the outset of an experiment, cannot be overemphasized. Researchers should, *a priori*, consider possible normalization strategies based on the content of their array slides. It is important that there is no bias in the types of gene targets printed in different sections of the slide. No normalization strategy will be reliable if the appropriate controls are not included as part of the design. The choice of control spots will differ, depending on the type of array platform and the facility at which the array is produced; common approaches include the use of synthetic spots, housekeeping genes or the identification of a set of genes that are known to be invariant or unresponsive across conditions assayed. All these approaches aim to provide some kind of calibration reference, i.e. a set of spots for which expression values can be predicted beforehand, which can be used to validate and normalize microarray data. Any deviations from expected or predicted values can be considered a result of systematic bias. A bias factor can then be calculated using the control spots and its effect extrapolated to the rest of the spots on the array.[5,8]

For control spots to be effective, they should (1) span the entire intensity range, (2) be distributed randomly across the surface of an array, and (3) be numerous enough to provide a statistically robust reference.[8] It is desirable to include, as part of the design of an experiment, flexibility with regard to several possible normalization options.

Data diagnostics and visualization

Identifying systematic bias via data exploration

A first step, in deciding upon the most appropriate method of normalization, is to visualize the patterns of variation in the raw data.[5] Identifying the nature of technical interference allows the researcher to decide on a directed normalization strategy, one that preserves biological data while reducing the noise specific to the array data set. Approaching normalization blindly, without assessing raw data, introduces a real danger of silencing or removing biological information of interest.[8,10] This is as detrimental to the outcome of an experiment as the non-removal of technical noise.

Systematic errors, which have discrete local effects on subsets of data on an array (e.g. all the spots printed by a specific pin), are called local biases. In contrast to this, systematic errors which have a general or global effect across an entire data set are referred to as global biases. Accordingly, normalization algorithms address the contribution of systematic errors either globally or locally.[2] The implication of this is that global normalization methods assume a general smooth error trend across a data set, while local methods assume that the source of bias affects discrete subsets of data independently from other such discrete units. The scope of normalization chosen—global or local—should be dictated by, and complement, the nature of a specific bias.[8] Using diagnostic and visualization methods to explore data, allows the researcher to determine whether technical noise or systematic errors produce within- and/or between-slide variations, and whether these biases exhibit global or local behaviours.

Several diagnostic and visualization tools exist and are available in most microarray analysis software packages. Most common and useful among these are: 1) box plots, 2) histograms, 3) scatter and MA-plots, and 4) false-colour plots. Each method allows for the identification of specific traits of the data and facilitates an evaluation of the contribution of unwanted noise. Interpretation of visualizations depends on experimental design and requires careful consideration. For directed or custom arrays, it is imperative that all diagnostic interpretations are substantiated by control spots[c] or known invariant genes, as any predictions or assumptions regarding the distribution of feature values can be problematic.

Box plots[5]

Box plots are commonly used to assess the relative spread of data, usually log ratios[d] or feature intensities, and are therefore a convenient way of identifying scale differences within or between arrays (Fig. 1). These plots provide a graphical overview of the so-called five-number summary of a data set, which includes information about the three quartiles [i.e. 25th percentile, 50th percentile (also called the mean), and the 75th percentile] and the minimum and maximum values. This tool can be used to compare the spread of data points from different print-tips, different microtitre plates or the overall spread of data from different arrays.

Overall scale differences between blocks can be the result of inconsistencies between pins or the non-random distribution of genes during printing. The latter can be eliminated by good experimental design. Assuming that experimental design has allowed for an unbiased distribution of gene targets, scale differences can result from variations in the amount of target deposited by different print-tips, which results in differences in the relative brightness of blocks of spots.[10] A second source of variation that could contribute to scale differences within cDNA arrays is that of microtitre plates. Different cDNA amplification batches are usually associated with different microtitre plates; inconsistent conditions show up as variations in spot intensities between replicate spots picked from different batches.[11]

Any differences in the spread of data between arrays could be a result of differences in scanner settings used to scan each array,[12] i.e. photomultiplier tube (PMT) and laser voltage settings, differences in mRNA concentrations isolated from samples, or differences in the labelling conditions of samples. Care should be taken when interpreting scale differences between arrays, as these differences could also reflect real experimental conditions which, if corrected for, will introduce rather than remove noise.[11]

Figure 1 shows an example of treatment-induced scale differences, where these differences reflect biological responses, rather than technical inconsistencies. Control spots can be a useful calibration guide for making sure that scanner settings are set correctly at the point of data capture of fluorescent signals from custom slides, to reduce between-slide scale differences.

Identifying scale differences and their probable sources will allow the researcher to adjust correctly for these. Emphasis is again directed at informed interpretation of the observed behaviour. Correcting for scale differences between print-tips[e] or microtitre plates, when these differences are a result of the non-random distribution of genes, will do nothing but introduce noise and silence biological meaning. Similarly, between-array scale differences could be condition- or treatment-specific and should be judiciously evaluated.

Histograms[5]

It is often useful to visualize information regarding the shape of the distribution of generated data. Histograms are plots of the frequency of feature intensity values or log ratios. Information regarding distributional density, i.e. number of values and their relative occurrence, across a data set, can be gleaned from these plots. Such information is useful when comparing the equivalence of distributions between two data sets. Some between-array normalization algorithms assume the data between arrays to

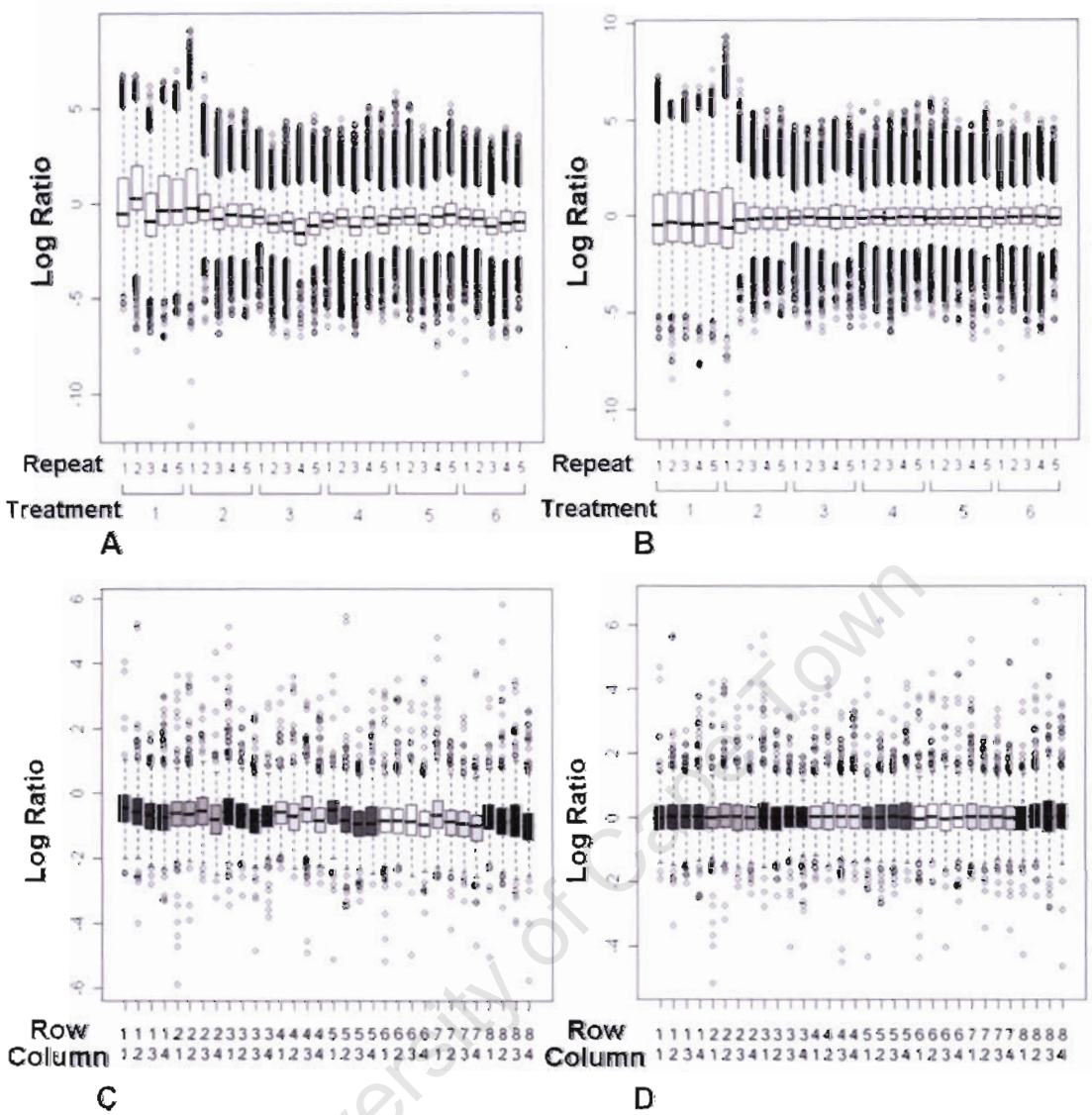

Fig. 1. Box plots illustrating $\log_2$ scale comparisons between arrays, (**A**) raw and (**B**) normalized, and blocks within a single array, (**C**) raw and (**D**) normalized, of microarray data from dual-channel (i.e. $\log_2$ Red/Green) custom arrays for the resurrection plant, *Xerophyta humilis*. **A** and **B** illustrate microarray data from six treatments, with 5 biological repeats per treatment. Each experimental sample was labelled with Cy3 and hybridized with a reference sample, labelled with Cy5, against a *X. humilis* cDNA slide representing 3400 cDNAs. Note that there is a consistent difference in the spread of the data for the biological repeats (1–5) for treatment 1, compared with treatments 2–6. This biological variation has been maintained during normalization (**B**). **C** and **D** illustrate that, within a slide of 32 blocks in a matrix of 8 rows and 4 columns, there is no difference between the overall spread of data but, rather, differences in the overall intensities of blocks. The trend appears to be spatial in nature, with a slight decrease in overall log ratios from blocks in column 1 through to those in column 4, for each row. This spatial trend is corrected following normalization (**D**).

be equally distributed.[8] Histograms also provide information regarding the central tendencies and absolute values of data sets, similar to box plots. In addition, these plots are useful when trying to ascertain whether the given data are normally distributed, as this is a requirement for many parametric statistical analysis techniques to be valid. The visualization of intensity distributions from custom arrays is particularly important, where assumptions regarding the distributional nature of data can be problematic.

Scatter plots[5,11]

Scatter plots provide a useful means for comparing the behaviour of different dyes in dual-channel experiments or, alternatively, comparing the relative overall behaviour between arrays (single- or dual-channel arrays). When comparing two arrays, the ratio values of features from each array are plotted on the *x*- and *y*-axis, respectively. If two arrays behave similarly, that is, the overall log ratios or feature intensities of individual features are comparable, then the points within such a plot will approximate a straight line with a slope of one and an intercept of zero. When comparing replicate arrays (or control spots from non-replicate arrays), any deviation from the expected straight line is indicative of systematic error.

More commonly, the same logic is used to compare the behaviour of two different dyes (Fig. 2A). Any deviations from a straight line with slope one and intercept zero is indicative of systematic differences between the two dyes. A useful type of scatter plot is the so-called 'MA-plot' (also referred to as the ratio-intensity, or RI, plot), which is used to identify inconsistencies or biases in the behaviour of two different dyes, across the entire feature intensity range, in dual-channel experiments. The MA-plot is essentially a normal scatter plot, of which the axis has been shifted by 45° and then scaled (Fig. 2B). The average log intensity $\{A = \frac{1}{2}[\log(\text{Ch1i}) + \log(\text{Ch2i})]\}$ of features is plotted against the $\log_2$ ratio $[M = \log(\text{Ch1i}) - \log(\text{Ch2i})]$ of these features, yielding a horizontal axis around which points are distributed. The *x*-axis

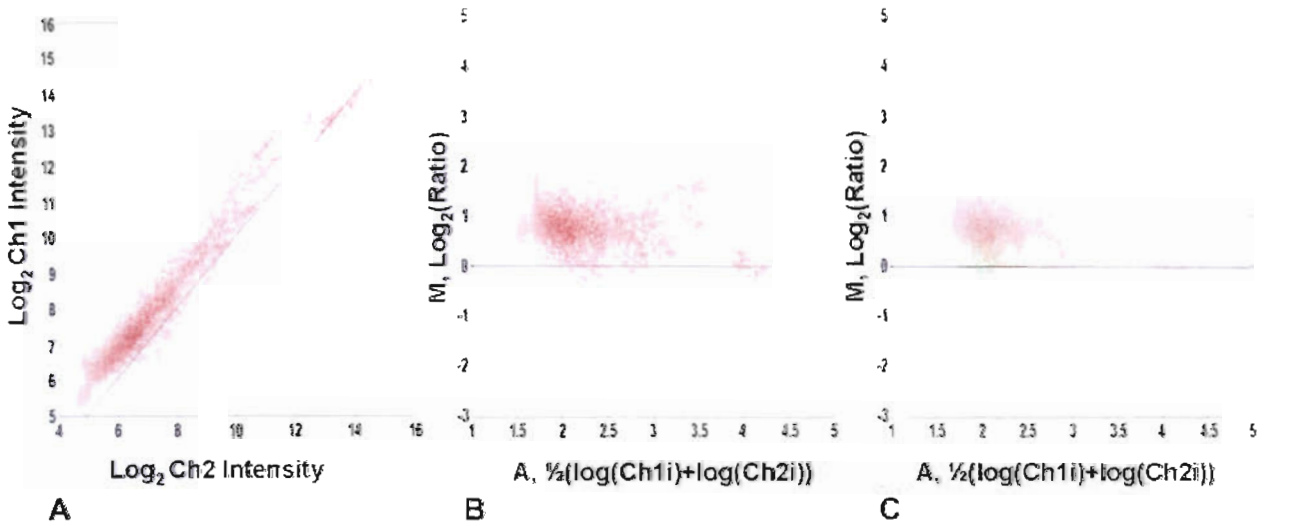

Fig. 2. (**A**) Scatter and (**B**) MA-plots of raw data for channel 1 (Ch1), labelled with Cy3, and channel 2 (Ch2), labelled with Cy5, hybridized to a *X. humilis* custom array. The diagonal line shown in **A** and the horizontal lines in **B** and **C** indicate the axis around which the data points should be, more or less, equally distributed if general whole-genome assumptions regarding differential expression in microarray data are true. Shown in (**C**) are normalized (green) and raw (red) data MA-distributions [where $M = \log_2(Ch1/Ch2)$]. Illustrated is the more even distribution of the normalized data points (green) around the $y = 0$ horizontal line. Interpretation of these graphs should lead to the conclusion that the Cy3 channel (Ch1) is over-represented across most of the intensity range. The validity of this interpretation should, however, be additionally confirmed by plotting control spots separately from experimental data, as the data come from a small custom *Xerophyta* array, and general assumptions regarding the symmetric nature of data can be problematic.

values (average intensity, or A) can be calculated as $\log_2$ or $\log_{10}$. For large random data sets, the assumption is that points should be distributed more or less symmetrically around a log ratio of zero. This assumption should carefully be considered when working with small or custom arrays. It is not uncommon to see a tailing of values at extreme intensity ranges, often referred to as the 'banana-effect'. This type of artefact can be ascribed to differences in the fluorescent capacities, or quantum yields, at different intensities, and differential incorporation of the dyes, due to differences in the size of Cy3 and Cy5 molecules.[5,8,12]

Identifying this type of bias is clearly important if some kind of reliable comparison, between samples labelled with different dyes, is to be made. When comparing the behaviour of dyes, the MA-plot has an important advantage over a normal scatter plot: points are plotted along a horizontal axis rather than a diagonal one – the human eye and brain are more efficient at interpreting horizontal distributions than diagonal ones[11] (Fig. 2B).

False-colour plots[8,11]

These kinds of plots are commonly used to identify spatial bias, which has been found to affect many arrays. Spatial bias refers to the effect that a specific feature's two-dimensional position has on its intensity value. False-colour plots can be generated by plotting the log transformed ratio or intensity value of a feature, as a function of its xy-coordinates in an array, or alternatively as a rank value, again as a function of its xy-coordinates. Spatial trends can easily be identified in this kind of plot and can be seen as a non-random distribution of log transformed ratios or intensity values. This type of bias can be introduced as a result of differences between microtitre plates or print-tips, hybridization artefacts, inserting slide into scanner at an angle, imperfections on the glass slide or any other effects related to the optical properties of microarray technology. Differences between print-tips lead to a specific type of spatial bias, where discrete blocks of features appear to be distinctly different from other blocks. Differences between microtitre plates lead to a similar discrete pattern of spatial bias. Imperfections on the glass slide and scanner-based variables can cause either discrete or smooth spatial effects (Fig. 3). It is important to identify and correctly interpret the behaviour of spatial bias. Choosing a normalization algorithm that corrects for discrete spatial effects, when these effects are smooth and global, will introduce noise. Similarly, a global spatial correction method should be chosen if the bias does not exhibit discrete behaviour and feature distribution is known to be random. Special care should be taken when interpreting false-colour plot data from small custom arrays, where clusters of differentially regulated genes might appear as spatial artefacts. A global normalization method will silence the

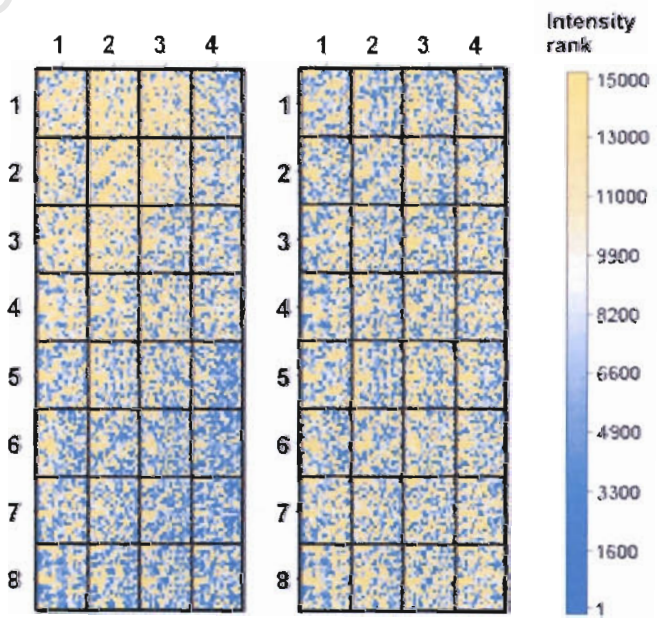

Fig. 3. A false-colour plot of microarray data generated from a custom *X. humilis* cDNA set, representing the $\log_2$ ratio (calculated as Cy3/Cy5), i.e. M-value, ranks of ~15 000 spots (i.e. rank 1 is represented by the gene with the highest $\log_2$(Cy3/Cy5) ratio and rank 15 000 the gene with the lowest ratio). (**A**) Pre-normalization image of raw data, highlighting the presence, assuming features were randomly arranged on the array, of a graded spatial bias of high $\log_2$ ratio values in the upper-middle-to-left portion of the slide, indicative of hybridization or scanning noise. (**B**) The same slide after spatial normalization shows a more even distribution of $\log_2$ ratio values across the slide. The chosen normalization method eliminated most of the spatial bias observed in the raw data.

biologically meaningful information contained within these small, differentially regulated clusters.

Measures of replicate variability

The visualization tools outlined above form a central part of pre-analysis data exploration, assisting the researcher in identifying the nature and extent to which data are affected by systematic error or technical noise. As illustrated above, these visualization tools are equally useful for the post-normalization state evaluation of data, providing some insight regarding the efficacy of a chosen method at removing noise. The visualization and subsequent interpretation of normalized data, however, should be approached cautiously. Applying inappropriate normalizations to data can potentially introduce noise which might not necessarily show up as an increase in noise when using the visualization tools. Logic, however, dictates that replicate features within arrays, or between biological repeats, should exhibit equivalent behaviour, with any deviations being indicative of the effects of unwanted systematic error or technical noise. The extent to which normalization minimizes or reduces variation across replicate features can therefore be used as a reliable means of assessing the efficacy of any normalization strategy in addition to the visualization methods already noted. Estimations of variability for replicate features are commonly obtained using pooled variances or ANOVA models, amongst others. It is assumed that the better a normalization method addresses the specific biases present in a data set, the smaller the variation among replicated observations will become.[12]

The issue of background[9]

More than any other bias factor, background contributions and their bias effect have been hotly debated in the literature. Although background subtraction is not the focus of this review, a brief consideration of its effect on data and normalization is warranted. Background refers to the contribution to overall spot intensity by targets binding non-specifically to the support matrix as well as fluorescence by the glass slide itself;[11] it is therefore reasoned that this leads to an over-estimation of target abundance for specific features. It has been shown that methods aimed at removing this bias often introduce, rather than remove, noise. In addition, the choice of array platform can greatly affect the performance of background subtraction and its effect on overall noise reduction.[4] Khojasteh *et al.*[4] noted significant differences in the efficacy of background subtraction, when applied to copy number (CGH) data generated by SMRT (Sub Mega base Resolution Tiling) array and cDNA array platforms, respectively. Data from the SMRT arrays showed a higher degree of reliability when background values were subtracted, whereas the cDNA array data showed the opposite behaviour. Khojasteh *et al.*[4] ascribed the apparent need to subtract background values from SMRT array-derived data to platform specific behaviour in addition to the specific image analysis methods used for these types of arrays.

Background bias can be measured globally or locally. Global measures assume a general linear trend or contribution across the array, while local approaches assume a more discrete contribution. Commonly used methods for the estimation of a background bias factor involve: (1) the inclusion of unrelated gene sequences or the inclusion of 'blank' spots on the slides, which are used to estimate global background fluorescence, or (2) the estimation of local background fluorescence in the area immediately surrounding a spot.[5] These approaches all aim to arrive at a reliable estimation of a non-specific contribution to foreground fluorescence. The intention of background subtraction is a valid one, but its effectiveness is debatable. First, it is questionable whether contributions made by non-specific hybridization are significant enough to warrant correction. Second, assuming that background contributions are non-trivial, local estimations of background by image analysis software are based on the incorrect assumption that the relative contribution of background to overall fluorescence is linear. This assumption is not valid, as the relative contribution of background to the overall intensity of spots within the high intensity range will be much less than the contribution experienced by spots in the lower intensity range. Third, methods attempting to determine the non-specific interaction effect between targets and unrelated gene sequences are erroneous in that background contribution is based on chemical interactions between the glass matrix and nucleic acids, the chemistry of which is profoundly different from that observed in interactions between unrelated nucleic acid species.[5] Lastly, methods based on estimations of local background are notoriously unreliable, with the choice of image analysis software having a major effect on the outcome of estimation.[4,13] Many background subtraction algorithms that aim to account for the non-linearity of local contributions are available. These algorithms modify calculated background values before subtracting them from calculated feature values. These modifications typically attempt to avoid common logical paradoxes, such as negative feature values, that arise from linear estimations and simple subtractions of background.

The effect background subtraction has on the efficiency of subsequent normalization depends very much on the accurate identification and removal of a true background bias factor. Special care should be taken not to confuse background with spatial effect, which is better corrected for by spatial normalization methods.[8] False-colour plots, discussed above, can also be used to explore the nature of background bias by plotting estimated background values against *xy*-coordinates.[10] Incorrectly estimating background contribution will do nothing but increase noise, which is clearly undesirable.[3,14] Situations resulting in negative values for features are not uncommon[8] and are indicative of the shortcomings of the specific methods. Wit and Mclure[8] propose some approaches to background subtracting that aim to avoid this logical dilemma. It is advisable, in addition to visualizing the background contribution, to assess diagnostically the effect of background subtraction on data, once applied.

The literature does not provide a clear-cut answer to the issue of background subtraction, and although its application might be theoretically warranted, it is important to consider whether attempts at estimating background bias are robust enough to be reliable. A fairly simple global strategy is highlighted by Wit and McClure.[8] The method uses empty spots as measures of the lowest achievable value on the array and then estimating a background bias factor based on the average intensity of these control spots. This central tendency, preferably median, can be subtracted from overall intensity values of features, setting any resulting negative values to zero. This method, however, requires that empty spots be sufficiently numerous and that their distribution cover a reasonable representation of the array surface, for a reliable background bias factor to be estimated. In addition, it is based on the assumption that the background bias is constant across the array. A consensus regarding the need for, and reliability of, background subtraction is yet to be reached; recent publications are, however, progressively leaning towards a 'no background subtraction' approach.[4,8]

Normalization methods

Numerous normalization algorithms exist, each one designed to correct for specific systematic errors introduced during a microarray experiment. This section will consider normalization methods in the context of the type of technical variation they

address. Many of these methods are equally applicable to one- and two-colour arrays; where this is not the case, it will be highlighted. This discussion does not aim to provide a definitive list of normalization strategies, but rather attempts to highlight the bias factors being addressed by each method and some assumptions upon which these algorithms are based.

A recommendation regarding a generally reliable and proven strategy is made, following the discussion of normalization methods. This recommendation is based on studies involving the empirical validation of various normalization methods. Empirical studies conducted by Qin *et al.*[3] and Khojasteh *et al.*[4] have provided a wealth of useful information regarding the performance of the algorithms on data sets with known expression values.

A case-by-case interactive approach is encouraged, which should involve the empirical exploration of data, via the diagnostic and visualization tools highlighted above, before and after normalization. Once the specific biases, affecting an experiment, have been identified, a stepwise normalization framework can be compiled. Emphasis is again directed at the two assumptions underlying many normalization methods: (1) that the majority of genes in a microarray experiment are not differentially regulated (they remain unchanged), and (2) that the number of up-regulated and down-regulated genes is more or less equal.[2,9] In addition, some between-array normalization algorithms also assume a similar distribution of expression values, that is, the frequency of specific intensity values (or log ratios) is approximately equivalent between arrays.[8] These methods are referred to as parametric, which means that some kind of explicit assumption regarding the distribution of the data is made, this is in contrast to non-parametric methods.[5] Whenever these assumptions are not satisfied, a set of invariant genes, or control spots, for which true values can be predicted should be used to determine bias factors to be extrapolated to the rest of the data set. This effectively allows the researcher to use empirically validated parametric methods to predict a bias effect on non-parametric data. Most normalization algorithms, where applicable, allow the researcher explicitly to define control spots or a set of invariant genes to be used.

Colour correction/Dye bias

Intensity bias, also referred to as intensity-dependent dye bias, is a result of slight differences in the properties of the commonly used Cy3 and Cy5 dyes. Using scatter plots, mentioned in the *Data diagnostics and visualization* section above, highlights these kinds of biases clearly. The effect is often seen as a 'tail' of spots in the lower and/or higher intensity ranges, indicating inconsistent behaviour of dyes.[8] This type of bias affects only dual-channel arrays, so the methods considered here will be discussed in this context. There are several different algorithms which address and correct for this type of bias.

1. *Linear regression*

Simple linear regression was one of the first methods used in early microarray experiments but is generally no longer used. These techniques are included to provide some historical perspective.

a. *Dye vs dye*

This method involves the adjustment of one channel (e.g. Cy3 vs Cy5) relative to the other (Fig. 2A), based on the distributional assumptions regarding microarray data symmetry. A best-fit line is generated through the distribution of spots on a scatter plot of Cy3 vs Cy5 (or Cy5 vs Cy3) values. The gradient and intercept of this best-fit line is then calculated, with deviations from a slope of one, and a y-intercept of zero, being taken as a reflection of inherent noise. The linear equation therefore provides a normalization function that can be used to adjust values such that the slope is equal to one with a y-axis intercept equal to zero. This is simply done by adjusting all x-axis values with the 'deviating' slope and intercept values (i.e. $x_{norm} = mx + c$). Note, however, that this type of linear regression treats Cy3 and Cy5 channels differently, which is not desirable as the assignment of dyes to different axes produces different results.[11]

b. *MA*

Linear regression-based normalization can alternatively be done on data distributed within an MA-plot (Fig. 2B). A best-fit line is again calculated through the points in the plot. Normalized M-values are calculated by subtracting the fitted value, for each feature, from the raw log ratio (i.e. M-value). Using MA-plot-based linear regression has the advantage that the two channels are treated equally, which makes these regressions more robust and reproducible.[11]

Neither of these methods is recommended for correcting the intensity-dependent bias often observed in dual-channel arrays. It has been noted that the intensity-dependent effect is non-linear, with the implication that a linear model will not be able to account effectively for this kind of bias.[11]

2. *Lo(w)ess based methods*

Lo(w)ess (LOcally WEighted Scatterplot Smoothing, or LOcally WEighted polynomial regreSSion) is a type of non-linear regression commonly used to adjust the distribution of points in an MA-plot. All Lo(w)ess algorithms employ the same strategy for the modelling of bias and subsequent adjustment of values. Adjustments are made, as with linear regression, by subtracting fitted values from raw log ratios. This method performs a series of local regressions across the MA-plot, using sliding windows of predefined size.[16] These local regressions are then combined into a single smooth curve.[5,11] The non-linear fits generated by Lo(w)ess algorithms are able to account more reliably for intensity-dependent dye bias than linear methods. The distributional assumptions highlighted earlier pertain to standard Lo(w)ess algorithms, as these employ a weighting function for determining the centre of a collection of data points. Lo(w)ess algorithms try to fit data around an M-value ($\log_2$ ratio) of zero. Points further away from this assumed mean are deemed more unreliable than those closer to it and therefore contribute less to the position of a centroid within a collection of points.[2]

Various parameters control the behaviour of a Lo(w)ess adjustment. First, a Lo(w)ess regression can be performed using, in theory, polynomials of any degree, which affects the nature of the generated best-fit line. It is common practice to fit data around polynomials of degree one, i.e. a straight line, as it has been observed that higher-degree polynomials (e.g. binomial, trinomial, etc.) tend to over-fit data, which does not capture the general trend in a population of data points.[5] The assumption is that within a certain range, the intensity-dependent bias is linear, and using a series of sliding windows will ensure the fitting of spots within a linear range. The size of the sliding window, the second parameter, influences the reliability and sensitivity of the Lo(w)ess algorithm and typically determines the proportion of points to include in each regression calculation.[5] Setting this window size too small again results in over-fitting, while setting this window too big results in a Lo(w)ess curve that does not model, effectively, the non-linear nature of the bias. The Lo(w)ess regression windows overlap, hence the designation sliding window, with the result being a very large number of overlapping regression calculations. Lo(w)ess

algorithms can be computationally demanding for this reason, but this is becoming less of a problem with powerful desktop computers. It should be noted that Lo(w)ess algorithms can be sensitive to outliers, despite the weighting approach mentioned earlier. Where data points are limiting, the statistical robustness, i.e. reliability, of each regression can become unreliable in the presence of large numbers of outliers due to an increase in the relative contribution made by each data point.

a. *Global Lo(w)ess*

This approach uses all data points within a given array to generate a non-linear curve, which is used to adjust for intensity-dependent dye bias.[4,8] Global Lo(w)ess generally performs quite well at estimating a dye bias, provided the observed bias is not a result of other systematic errors (e.g. underlying spatial bias).[8]

b. *Composite Lo(w)ess*[10]

This type of Lo(w)ess provides a slightly more advanced global method for dye bias adjustment. Concerns regarding the reliability of Global Lo(w)ess adjustments within the extreme intensity ranges have been raised, as these ranges typically contain fewer data points than intermediate intensity ranges. Composite Lo(w)ess is based on a model where control spots as well as assayed features are used to generate a non-linear best-fit line. The idea is that as the sliding window moves into extreme intensity ranges, the Lo(w)ess curve will be increasingly based on the control spots rather than the assayed features, which will contribute increasingly fewer data points to the window. Because this type of Lo(w)ess relies on both assayed features and control spots, it is not a viable normalization method for data sets that rely on control spots only for the calculation of a bias factor.

c. *Print-tip Lo(w)ess*

In contrast to the above global methods, Print-tip, also known as block-by-block, Lo(w)ess employs a discrete local strategy for the modelling and correction of dye bias. Print-tip Lo(w)ess is used to adjust feature intensity values printed by each pin separately. The principle remains the same, but the assumptions are slightly different. Print-tip Lo(w)ess assumes that each discrete block of features will behave slightly different from other blocks due to minor physical differences between pins. Print-tip Lo(w)ess can simultaneously correct for intensity-dependent and spatial bias (discussed below).[10] Concerns regarding the discrete nature of this type of Lo(w)ess approach should be noted. There is a danger of introducing bias in cases were there is no discernible difference in the overall intensity-dependent behaviour of features from different blocks.[8] It has also been noted that Print-tip Lo(w)ess is unreliable in cases where there are fewer than 150 data points per print-tip group. In these cases, a global method provides a statistically more robust approach as the number of data points used to model the bias is substantially more.[10]

3. *Dye-swap normalization*[8]

Another method commonly used to account for differences in the fluorescent capacity of the different dyes is that of dye-swap normalization. This method relies on the inclusion of technical replicates as part of an experiment's design. Typically, an array will be replicated with the samples in each replicate, labelled with opposite dyes. The intensities of the replicate features are then calculated as average intensities across both dyes.

This method can be extended to include Lo(w)ess fitting, once average intensities are calculated for replicate features, thereby providing a very robust model for intensity-dependent correction. Including dye-swap replicates, as part of a dual-channel experiment, is theoretically highly desirable, but often practically unfeasible, due to the high cost of microarray experiments.

4. *Splines*[17,h]

Spline-based algorithms present a non-parametric alternative to the Lo(w)ess-based regression fitting of data, and are commonly used to account for intensity-dependent dye bias. The main benefit of spline-based normalization methods is their independence from the assumptions underlying other parametric approaches. Spline algorithms do not make any assumptions regarding the distributional nature of data, but rather treat values simply as a collection of points and are therefore useful when normalizing directed or custom arrays.

This method is related to Lo(w)ess in that it is based on the calculation of several local regressions, across a data set, which are joined to form a smooth curve. Spline-based methods, however, are based on a discrete window approach, as opposed to the overlapping sliding window method employed by Lo(w)ess algorithms. Spline algorithms perform a fixed number of linear regressions, within a predefined number of windows, across a data set. Spline-based dye bias normalizations are usually implemented in a manner equivalent to the Lo(w)ess methods discussed above, using MA-distributions. The behaviour of a spline curve can be modified via parameters similar to those used to modify Lo(w)ess regressions. Typically, the polynomial degree of the curve is specified; as with Lo(w)ess, it has been found that higher-degree polynomials often over-fit data, with linear equations generally performing best. In addition, the number of windows, i.e. number of regression calculations, across a data set is defined. This is distinctly different from Lo(w)ess methods, where the number of regression calculations for any predefined window size can differ, depending on the number of data points within a data set. Spline-based algorithms are computationally much less intensive than Lo(w)ess algorithms, due to the usually significantly smaller number of regression calculations performed.

The various methods discussed above can be extended to include several robust parameters, which add dimensions to data sets. One such extension, suggested by Smyth and Speed,[10] involves ranking the quality of spots and assigning a reliability weight to features when applying Lo(w)ess regressions to the data. Rank Weighted Lo(w)ess involves calculating a centroid of a collection of data points based on the statistical reliability of features. Reliability is measured as a percentage of the complement of pixels that make up each spot. Spots with more pixels are deemed to be more reliable than those with fewer and consequently carry more weight in determining the centroid of a specific collection of data points.

Spatial bias

This type of bias, if present, can clearly be seen in the false-colour plots discussed earlier.[11] Unlike intensity-dependent bias, this type of systematic variation affects both single- and dual-channel arrays.

1. *Lo(w)ess based methods*

Again, Lo(w)ess based algorithms can be used to model the spatial effect observed on many arrays. The assumptions and parameters are the same as previously stated.

a. *Print-tip Lo(w)ess*

As noted above, Print-tip Lo(w)ess can be used to correct simultaneously for intensity-dependent bias as well as spatial bias. The assumption is that spatial trends are localized to discrete areas of the array and can therefore be accounted for by adjusting values within discrete units, in this case print-tip groups or

blocks.[10] The same curve used to correct for intensity-dependent bias is used to adjust spatial trends within a print-tip group, with adjustments based on MA-distribution regressions. Because of the discrete nature of this approach, there is a danger of introducing noise at the edges of a print-tip group when the underlying spatial effect is continuous across the surface of an array.[8] It is therefore important to determine whether the spatial trends observed in false-colour plots are discrete or continuous before applying Print-tip Lo(w)ess. Another important consideration is the number of features associated with each print-tip. As mentioned previously, Print-tip Lo(w)ess is unreliable in cases where there are fewer than 150 data points per print-tip group.[10]

b. *2D Lo(w)ess*

This type of spatial correction is effective for the removal of continuous spatial trends. Regression fitting of values is based on trends seen within two-dimensional false-colour plots. As with other Lo(w)ess methods, polynomial curves are used to model non-linear trends within data. Wit and McClure[8] recommend using Lo(w)ess polynomials of degree one, i.e. linear functions, when correcting for spatial trends, as it has been observed that higher-degree polynomials tend to be unstable near the edges of microarrays. 2D Lo(w)ess assumes a global spatial trend which, as mentioned above, might or might not be the case. Again, an assessment of spatial trends is necessary before making any adjustments. 2D Lo(w)ess might not be the best option in cases where imperfections on the array present sudden rather than smooth changes, or in cases where clusters of differentially expressed genes are found. 2D Lo(w)ess will confuse such clusters with spatial bias that has to be adjusted for.[11] These aspects require consideration by the researcher and care should be taken to avoid any unbiased distribution of gene targets during printing at the outset of the experiment.

2. *Median based methods*[4]

An alternative to the Lo(w)ess methods discussed above is a spatial correction method based on the central tendency of neighbourhoods of spots. For each spot, the median of $\log_2$ intensity values of spots within a spatial neighbourhood of predefined size (number of rows × number of columns), centred on that spot, is calculated. The difference between the neighbourhood median and the intensity value of the spot is considered to be a bias factor. The value of each spot is adjusted accordingly. The neighbourhood size used to adjust the value of spots is an important parameter to consider. A small neighbourhood is sensitive and corrects discrete and local artefacts, but might be problematic when adjusting for more general or global trends (compare Global vs Print-tip Lo(w)ess). A large neighbourhood size will clearly have the opposite effect. The choice of neighbourhood size depends on diagnostic interpretations of the spatial bias effect.

Scale differences

Scale biases are common to both single- and dual-channel arrays and methods correcting for this kind of systematic error are generally applicable to both platforms. All scaling methods have two things in common: (1) they adjust the means of compared data sets to be more or less equal (also known as centring) and/or (2) adjust the spread or variation of data to be more similar (also known as scaling). Data sets can consist of any collection of measurements, for example, values associated with specific microtitre plates or print-tips, values associated with control spots, and values associated with one or multiple arrays.

1. *Subtract $\log_2$ central tendency*

This method adjusts the means of all distributions to zero. It is one of the simplest forms of scale adjustment and involves subtracting either the $\log_2$ mean or median of a distribution from each feature's $\log_2$ ratio. This results in the mean of all distributions, adjusted in this way, being equal to zero. This method can also be applied to the raw ratios; in this case, all ratios are divided by the measure of central tendency (i.e. mean or median).[11] This technique works well, but ignores possible array-wide changes which might be a reflection of sample conditions or treatment. It is therefore advisable to approach global scaling methods with caution. As with all normalization methods, controls can be used to calculate an adjustment factor. Another objection, concerning this method, involves assumptions regarding the linearity of variation. Bright arrays exhibit compression of values at high intensities, whereas darker arrays show compression of values near low intensities. This behaviour is a consequence of the limit imposed on possible intensities values for any feature (0 to $2^{16}-1$), which lead to a breakdown in linearity at the extremes of the intensity range.[8]

2. *Subtract $\log_2$ central tendency and divide by standard deviation*

This form of scaling adds another dimension to that of the method discussed above. In addition to adjusting the means of all distributions to zero, the standard deviations of these distributions are brought to one. The same comments and considerations discussed above apply to this method.

3. *Quantile normalization*

Quantile normalization was proposed as a method for the scaling of replicate arrays, where assumptions based on whole-genome expression distributions are problematic, but works equally well on whole-genome arrays. This method forces the distribution of values in each array in a set of arrays to be the same. All features are ranked according to their intensity value—that is, the lowest intensity value is assigned rank 1; the second-lowest intensity value is assigned rank 2, etc.—until all features within each slide have been ranked. The ranked distributions are then compared and the mean of each rank across the arrays is calculated. This calculated mean replaces the original value and the normalized data are rearranged to have the original ordering.[8] This type of approach ensures an equivalent distribution of intensity values between distributions. Wit and McClure[8] point out that this method is able to deal with the non-linear compressions that might affect the two scaling approaches mentioned above, as the ranking approach ensures a linear distribution of features.

When adjusting the scale of replicate slides with this method, it is reasonable to assume that the distribution of feature intensities should be comparable. This assumption, however, is often problematic, as different slides usually do not have the same number of captured features. This is commonly a result of technical thresholds or noise (e.g. detection of low signals or washing artefacts) and less often biological differences.[8] A potential problem is that slides with different numbers of features have a different number of ranks. Features with missing values across arrays should therefore be excluded or imputed, prior to attempting quantile normalization. Several methods for the imputation of missing values exist,[18,19] a detailed discussion of which falls outside the scope of this review. Some implementations of the quantile algorithm[20] overcome the problem of missing values by assuming that missing values are random and not a result of low signal or technical noise. By implication, this means that the number of missing values should be proportional to the number of features on an array. Missing values that result from non-random effects (commonly due to a low signal) are therefore still problematic as these invalidate the assumption of the quantile algorithm (G.K. Smyth, pers. comm.). As previously

emphasized, stated assumptions are often erroneous and should be validated before proceeding with any type of normalization. Histograms provide a useful way for visualizing and comparing the density distributions of different data sets.

One approach to between-condition or -treatment scaling is based on the ranking of a set of controls or invariant genes, i.e. genes that are known not to change between conditions. A general ranked scale is then generated and the remaining genes are linearly distributed, known as interpolation, between the ranks of the invariant genes. This approach can also be used when there are large differences in the number of data points within distributions. For this method to be reliable, the smallest and largest values on each array have to be part of the set of invariant genes.[8]

Dual-channel arrays, which are based on a common reference design, offer an intuitively interesting solution to between-slide scale adjustments. Sample- or treatment-channels may not show a comparable distribution of intensity values across the different experimental conditions. If use is made of a common reference, however, this by definition should have the same distribution across all slides, and channel-specific implementations of quantile adjustments[20] can be considered. These methods essentially force the reference channel across all slides to be exactly the same and extrapolate an adjustment factor for the sample- or treatment-channel within each array.

4. *Cyclic Lo(w)ess*

This is an inter-array variant of the previously mentioned Lo(w)ess-based methods originally developed for cDNA microarrays. This algorithm can be applied to both single- and dual-colour arrays to adjust for scale differences between them. An MA-plot is generated, where M is defined as the $\log_2$ ratio of replicate feature values and A as the average of replicate feature log values.[6,13] Generating an MA-plot in this way allows for a comparison of features across replicate arrays. As with other Lo(w)ess methods, a non-linear regression curve is calculated and the data are fitted accordingly. This process is carried out in a pair-wise manner and iterated until differences between arrays have been removed. This procedure can be applied to sets of invariant genes or control spots in cases where the features on arrays are not expected to be directly comparable.[13] The same Lo(w)ess parameters, previously discussed, are applicable to this specific implementation of the Lo(w)ess algorithm.

5. *Qspline*

This is referred to as a baseline method and, similar to Cyclic Lo(w)ess, is an inter-array variant of its intra-array counterpart. Estimations and adjustments are dependent on the definition of a baseline array, also called a reference array. The baseline array is used as a ranking reference for subsequent adjustments. Target array features are ranked and compared to the ranked features of the baseline array. A spline-based smoothing curve (discussed above) is then calculated to relate the ranks of features from the target array to those on the baseline array.[6] The choice of baseline array is important and can have a profound effect on results. The baseline array should ideally be representative of an average behaviour of replicate features across the different arrays. This essentially means that the feature values on the baseline array should preferably be more or less equal to the mean value of those features across all arrays. It is not necessary that the baseline array contain all features; it can be constructed using a set of known invariant genes which occur across all arrays, or other forms of control spots for which expression values can be expected to be similar. Quantile ranks are then constructed using these genes, with the resultant smoothing function being extrapolated to other features of an array.[13]

ANOVA-based methods

ANOVA-based methods of normalization have been shown to be effective in modelling systematic error. These methods do technically more than just normalize data; they also provide estimations regarding the significance of condition-specific gene expression, a feature which falls outside the scope of this discussion. This method is based on a composite linear model, which contains terms for each aspect of the array and all possible sources of bias. Interactions between the different aspects of an array (e.g. genes, dyes, print-tips, spatial position, and time point) and the specific sources of bias or error are then mathematically defined. Variations between features are assessed in the context of various null hypotheses and their significance statistically determined. The null hypotheses usually state that the observed variation is not significant but a product of systematic bias. If these hypotheses are rejected, the observed variation is presumed to be biologically meaningful. In this way, ANOVA methods distinguish between interesting and obscuring variations.[5] The problem with such an approach is the non-linear nature of many artefacts associated with systematic error. An ANOVA-based model is not able to estimate these non-linear artefacts, a good example of which is the non-linear intensity-dependent dye effect described earlier.[5,8]

Evaluation of normalization methods

A comprehensive study, conducted by Khojasteh *et al.*,[9] showed that a composite stepwise approach to normalization provides the most reliable means of identifying and removing the systematic errors. They compared the efficacy of different combinations of normalization models at detecting single-copy gene changes between samples, for which the gene copy ratios were known. In total, 19 different normalization strategies were assessed across five different data sets, all with different numbers of arrays. Performance was based on a specific strategy's ability to reduce variability (standard deviation) and enhance the accuracy of predicting single-copy gene changes between samples. Normalization strategies consisted of single-step (addressing only one type of bias), two-step (addressing two sources of bias), and three-step (addressing three sources of bias) approaches. The types of biases addressed were: (1) spatial bias, (2) intensity bias, and (3) global scale bias. All normalization strategies were performed including and excluding background subtraction. The results were unanimous, indicating that a three-step strategy, one that systematically addresses the three most common sources of variation, without background subtraction, outperformed both the two- and one-step approaches. Two-step strategies, in turn, outperformed the one-step approaches, highlighting the importance of identifying and correcting for all sources of systematic error.

Conclusion

The sources of systematic variation that affect a microarray experiment are many and accounting accurately for each of them is not trivial. Although a variety of normalization strategies have been developed to identify and correct for these systematic biases, these strategies are based on stringent assumptions which require careful consideration. The conscientious design of a microarray experiment, the inclusion of appropriate controls and the unbiased printing of gene targets are imperative to the successful normalization of microarray data.

We recommend a stepwise strategy that systemically addresses the various types of biases identified with the diagnostic and visualization tools discussed earlier. It should be emphasized that a thorough diagnostic interpretation of data, prior to normalization, facilitates the compilation of a normalization strategy aimed at addressing directly the types of systematic

error present within a specific experiment. The effectiveness of each normalization strategy should be diagnostically monitored before proceeding with the next step. Wit and McClure[8] suggest that local artefacts should be corrected before progressing to normalizations that involve several or all arrays. These recommendations are empirically supported by the results of Khojasteh *et al.*[4] In particular, Wit and McClure suggest the following systematic strategy:

1. Spatial correction
2. Background correction
3. Intensity-dependent dye bias
4. Within-replicate scaling
5. Between-condition/-treatment scaling.

Wit and McClure[8] point out that, although background subtraction is included in the above list, they advise strongly against its use. This advice is supported by the study of Khojasteh *et al.*[4]

Many other more complex algorithms exist, a list too long to include here, but their reliability and performance remain questionable due to insufficient empirical validation. Many of these purportedly novel algorithms are derivations of commonly used ones and essentially address the same types of biases. A new trend towards non-parametric algorithms—algorithms that make no explicit assumption regarding the distributional nature of the data—can be seen in the literature. Whether these methods provide a real advantage over current parametric models depends on results obtained from empirical validations, where *a priori* statements regarding expression values can be made.

Recommended resources

Many commercial and open source solutions are available to microarray researchers. Choosing among the various options is often a matter of personal preference, as most of the available software packages, aimed at microarray data analysis, contain a large selection of visualization and diagnostic tools. A good starting point is one of several microarray portals (some listed below), which include descriptions of software packages as well as links to other useful resources.

1. http://www.biodirectory.com/biowiki/Microarray_portals_and_resource_pages
   A site that is used by life-science researchers to find tools and databases related to all sorts of molecular biological and bioinformatics-related activity, including experimental troubleshooting, tutorials, as well as applications and methods-related information.
2. http://www.microarrays.in/links.html
   A great one-stop resource for microarray software, protocols, links, publications and other microarray-related information including discussion forums.
3. http://cbio.uct.ac.za/arrayportal
   A portal containing introductory information on important microarray-related topics, including links to useful references.

We thank Wiesner Vos, from the UCT National Bioinformatics Node, for many useful discussions and advice on normalizing data from custom microarrays. This work was supported financially by the University of Cape Town and the National Research Foundation. J. v. H. thanks Dan Stein, Department of Psychiatry, University of Cape Town, for financial support during his Honours year in the Department of Molecular and Cell Biology.

Notes (refer to superscripts in text)

a) Experimental noise comes in many forms and is often referred to as technical, or systematic, error, variation or noise. In this paper these terms are all used to refer to the unwanted, biologically meaningless, variability observed within, or between, arrays.

b) Technical variation can exist within an array, that is, between features printed on the same slide, or between arrays, i.e. between the features from different slides.

c) Spots and features are both terms used to refer to the genes present on an array. In addition, each spot or feature can be considered a discrete collection of data, or data point.

d) It is common for intensity values to be log-scaled. Most often this is done as $\log_2$. Advantages of the log-scaling of data include the linearization and symmetrization of the distribution of feature intensities. In addition, $\log_2$-scaling makes data more amenable to subsequent interpretation.[5]

e) Print-tips refer to the pins used by the spotting robot to spot DNA sample onto the array. Commonly, each pin is associated with a discrete group, also called a block of spots.

f) It is important to consider the way in which expression ratios are represented, when interpreting MA-plots. In MA-plots ratios are commonly calculated as Cy5/Cy3, but can equally be done as Cy3/Cy5.

g) Background estimation and subtraction refer to two different procedures. Estimation refers to the process of calculating raw background values whereas subtraction refers to the method by which estimated values are processed.

h) Traditionally, a spline is defined as a long, narrow, flexible strip of timber. The definition of a mathematical spline is analogous to its timber counterpart and refers to a flexible mathematical function able to adapt to data.[17]

i) Imputation refers to the process by which values for features with missing values are derived from patterns inherent in existing data.

Received 16 April. Accepted 1 September 2007.

1. Schena M., Shalon D., Davis R. and Brown P.O. (1995). Quantitative monitoring of gene expression patterns with a cDNA microarray. *Science* 270, 467–470.
2. Quackenbush J. (2002). Microarray data normalization and transformation. *Nature Genetics* Suppl. 32, 496–501.
3. Qin L. and Kerr F. (2004). Empirical evaluation of data transformations and ranking statistics for microarray analysis. *Nucleic Acids Res.* 32, 5471–5479.
4. Khojasteh M., Lam W.L., Ward R.K. and MacAulay C. (2005). A stepwise framework for the normalization of array CGH data. *BMC Bioinformatics* 6, 274.
5. Drăghici S. (2003). *Data Analysis Tools for Microarrays*. Chapman & Hall/CRC Press, London.
6. Bolstad B.M., Irizarry R.A., Astrand M. and Speed T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19, 185–193.
7. Hartemink A., Gifford D., Jaakkola T. and Young R. (2001). Maximum likelihood estimation of optimal scaling factors for expression array normalization. SPIE International Symposium on Biomedical Optics 2001 (BiOS01). In *Microarrays: Optical Technologies and Informatics*, eds M. Bittner, Y. Chen, A. Dorsel and E. Dougherty, *Proc. SPIE*, 4266, 132–140.
8. Wit E. and McClure J. (2004). *Statistics for Microarrays: Design, Analysis and Inference*. John Wiley, Chichester, Hants.
9. Wang D., Huang J., Xie H., Manzella L. and Soares M.B. (2005). A robust two-way semi-linear model for normalization of cDNA microarray data. *BMC Bioinformatics* 6, 14.
10. Smyth G.K. and Speed T.P. (2003). Normalization of cDNA microarray data. *Methods* 31(4), 265–273.
11. Stekel D. (2003). *Microarray Bioinformatics*. Cambridge University Press, Cambridge.
12. Park T., Yi S., Lee S., Lee Y. and Simon R. (2003). Evaluation of normalization methods for microarray data. *BMC Bioinformatics* 4, 33.
13. Wu W., Dave N., Tseng G.C., Richards T., Xing E.P. and Kaminski N. (2005). Comparison of normalization methods for Codelink Bioarray data. *BMC Bioinformatics* 6, 309.
14. Wu W., Xing E.P., Myers C., Mian I.S. and Bissell J.M. (2005). Evaluation of normalization methods for cDNA microarray data by k-NN classification. *BMC Bioinformatics* 6, 191.
15. Sen A. and Srivastava M. (1990). *Regression Analysis: Theory, Methods, and Applications*. Springer-Verlag, New York.
16. Cleveland W.S. and Devlin S.J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *J. Am. Stat. Assoc.* 83, 596–610.
17. Rupert D., Wand M.P. and Carrol R.J. (2003). *Semiparametric Regression*. Cambridge University Press, Cambridge.
18. Sehgal M.S.B., Gondal I. and Dooley L.S. (2005). Collateral missing value imputation: a new robust missing value estimation algorithm for microarray data. *Bioinformatics* 21, 2417–2423.
19. Acuna E. and Rodriguez C. (2004). The treatment of missing values and its effect in the classifier accuracy. In *Classification, Clustering and Data Mining Applications*, eds D. Banks *et al.*, pp.639–648. Springer-Verlag, Berlin, Heidelberg.
20. Smyth G.K. (2005). Limma: linear models for microarray data. In *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, eds R. Gentleman, V. Carey, S. Dudoit, R. Irizarry and W. Huber, pp. 397–420. Springer, New York.