STUDY OF PROTEINS AFTER RECOVERY FROM BACTERIAL INFECTION

BY

CYNTHIA ZAVALA

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Animal Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2014

Urbana, Illinois

Advisor:

Professor Sandra Luisa Rodriguez-Zas

Abstract

The effect of Bacillus Calmette–Guérin (BCG) on the microglia of mice after recovery from infection is incompletely understood. Microglia cells of mice were collected from two groups, mice infected with BCG vs. control mice (BCG vs. control). The objective of this study was to compare the proteins and peptides present in the microglia of BCG-challenged mice one-week after treatment relative to control mice and gain insights through differential detection and enrichment analysis. Detection relied on tandem mass spectrometry proteomics. The database search software PEAKS was used to detect peptides and proteins in the treated and control samples. Differential detection was performed using SAS procedures and enrichment classification was completed using PANTHER. The consistently higher number of proteins and peptides (except for one sample) detected in the control samples suggested that BCG impacts the production of proteins. A number of proteins including F-actin-capping protein subunit alpha-2 (P47754), Alpha-enolase (P17182), and myelin basic protein (F7A0B0) were identified in greater abundance in control mice; while chitinase-like protein 3 (O35744), vesicle-associated membrane protein-associated protein A (Q9WV55), and Protein SET (Q9EQU5) were identified only in BCG-treated mice. The differential detection was consistent with similar studies on the neurological activity and inflammation response to BCG-challenged mice. Functional enrichment analysis identified enriched pathways among differentially detected proteins with a differential detection of two (BCG and control mice) associated with inflammation-mediated and microglia activation with Huntington disease (P00029) and Glycolysis (P00024).

Table of Contents

Chapter 1: Literature Review

1. Immune response

Exposure to pathogens like bacteria can lead to immediate activation of the immune system. The immune system is a system of biological processes within an organism that protects the body from infection (Figure 1.1). Although innate immune cells including macrophages and dendritic cells play important roles, nonprofessional cells such as epithelial cells, endothelial cells, and fibroblasts also contribute to innate immunity (Takeuchi and Akira, 2010). Macrophages are a type of white blood cell that are formed in response to an infection or accumulating damaged or dead cells. Macrophages are the innate immune cells that primarily respond to pathogens. Dendritic cells (DCs) are antigen-presenting cells of the immune system. An antigen is a molecule that causes the immune system to produce antibodies against it and induces an immune response. Innate immune recognition, also known as pattern recognition, is unusual in that each host pattern recognition receptor (PRR) has a broad specificity and can potentially bind to a large number of molecules that have a common structural motif or pattern (Medzhitov, 2007; Figure 1.2). The targets of pattern recognition receptors are referred to as pathogen-associated molecular patterns (PAMPs; Medzhitov, 2007). Macrophages that encounter pathogens ingest, process and display antigen fragments on their cell surfaces. Macrophages with antigen fragments displayed on their surfaces are called antigen-presenting cells. An antigen-presenting macrophage interacts with a T-helper cell that recognizes the same antigen. T-helper cells coordinate immune responses by communicating with other cells. During interaction, the macrophage releases a chemical signal called interleukin-1, which stimulates the T-helper cell to secrete interleukin-2. Inerleukin-2 causes the proliferation of certain cytotoxic T cells and B cells. From this point, the immune response follows two paths, one using cytotoxic T cells and one using B cells. Cytotoxic T cells directly attack cells carrying certain foreign or abnormal molecules on their surfaces and work to kill the infected cells. Death of the infected cells results in destruction of the pathogen. B cells work by secreting antibodies into the body's fluids and ambushing foreign antigens circulating in the bloodstream. Antibodies bind to the antigens on the surfaces of the pathogens, marking them for destruction by macrophages. Immune responses are principally mediated by T and B cells, which possess enormous diversity in antigen recognition, high antigen specificity and long-lasting immunologic memory (Sakaguchi, 2008).

The innate immune system is a major contributor to inflammation induced by microbial infection or tissue damage (Takeuchi and Akira, 2010). Inflammation is a protective response by the body to ensure removal of detrimental stimuli, as well as a healing process for repairing damaged tissues (Takeuchi and Akira, 2010). Inflammation is characterized by five symptoms: redness, swelling, heat, pain, and loss of tissue function (Takeuchi and Akira, 2010). Following inflammation, damaged cells release chemicals including

histamine and prostaglandins. The release of these chemicals causes blood vessels to leak fluid into the tissues, causing swelling, and helps isolate the foreign substance from further contact with body tissues. In terms of inflammatory responses, Toll-like receptors (TLRs) activate tissue-resident macrophages to produce pro-inflammatory cytokines, including tumor-necrosis factor (TNF), interleukin-1β (IL-1β) and IL-6, which coordinate local and systemic inflammatory responses (Medzhitov, 2007). Cytokines are small proteins released by cells that have a specific effect on the interactions between cells, on communications between cells or on the behavior of cells. Cytokine is a general name; other names include lymphokine (cytokines made by lymphocytes), monokine (cytokines made by monocytes), chemokine (cytokines with chemotactic activities), and interleukin (cytokines made by one leukocyte and acting on other leukocytes; Zhang and An, 2007). Cytokines may act on the cells that secrete them (autocrine action), on nearby cells (paracrine action), or in some instances on distant cells (endocrine action; Zhang and An, 2007).

Studies have explored innate inflammatory responses (Figure 1.3) in the murine central nervous system (CNS) using different pro-inflammatory agents such as bacteria endotoxin (lipopolysaccharide, LPS) or pro-inflammatory cytokines IL-1β and tumor necrosis factor-α (TNF-α; Matyszak, 1998) and more recently with Bacillus Calmette–Guérin (BCG). The CNS is a complex of nerve tissues that compromises the brain and spinal cord. Sickness behaviors are not mediated by the infectious pathogens themselves, but rather they are a critical component of the immune response orchestrated by the immune system via the release of cytokines (Bilbo and Schwarz, 2012). Previous studies have shown that behaviors, rather than being pathological consequences of infection, are strategies that are critical to host survival. Specifically, preventing the synthesis of cytokines, or the binding of cytokines to their receptors within the brain prevents the expression of sickness behavior even in the presence of a peripheral immune challenge; while administering individual cytokines directly into the brain, in the absence of a peripheral infection, will induce sickness behavior (Bilbo and Schwarz, 2012). Peripherally-derived immune factors are a part of the peripheral nervous system (PNS) which consists of the nerves and ganglia outside of the brain and spinal cord.

Pro-inflammatory cytokines act in the brain to induce non-specific symptoms of infection, including fever and profound psychological and behavioral changes (Kelley et al., 2003). Sick individuals experience weakness, malaise, an inability to concentrate and listlessness. These behaviors are usually considered to be results of the process that occurs during infection, but they have also proven to play a role in maintaining a natural homeostatic reaction the body uses to fight infection. These changes in behavior can be considered to be the expression of a motivational state that resets the organism's priorities to promote resistance to pathogens and recovery from infection (Kelley et al., 2003). By preventing the occurrence of

those activities that are metabolically expensive and favoring expression of those that decrease heat loss (e.g., rest) and increase heat production (e.g., shivering), sickness behavior positively contributes to recovery following infection (Kelley et al., 2003).

Peripheral administration of a cytokine inducer, such as LPS, or of recombinant cytokines, such as IL-1β or TNF-α, mimics all nonspecific symptoms of sickness, including fever, activation of the hypothalamic-pituitary-adrenal (HPA) axis, reduction of food intake and other behavioral activities, and withdrawal from the physical and social environment (Dantzer, 2009). Sickness behavior is assessed by reduction in food intake and decreased socialization. Experimental findings indicate that pro-inflammatory cytokines mediate the clinical signs of the host response to infection. The physiologic and behavioral changes that are characteristic of sickness are mediated in the CNS (Dantzer, 2009). For example, fever represents a regulated rise in body temperature resulting from increased production of heat (thermogenesis) and decreased thermal loss (thermolysis) in response to an elevated set point for the regulation of body temperature (Dantzer, 2009). Given that the body temperature set point is controlled by temperature-sensitive neurons in the preoptic hypothalamus, pyrogenic cytokines, such as IL-1β and IL-6, need to act in the CNS to induce fever (Dantzer, 2009).

There are multiple pathways by which peripherally-derived immune factors can affect the brain, and in turn by which the brain can impact peripheral immune responses which include the autonomic nervous system, activation of the "stress axis" (the hypothalamic-pituitary-adrenal axis), and cytokines, chemokines, and leukocytes that travel or signal across the blood-brain barrier (Bilbo and Schwarz, 2012). Studies of the effect of pathogen challenge on mouse macrophage and macrophage-like cells are offering insights into the corresponding immunological, physiological, and molecular impact.

1.1 Microglia

Microglia cells are macrophage-like cells located in the CNS. Microglia cells are the primary mediators of the brain's immune defense system and are integral to the subsequent inflammatory response (Loane and Byrnes, 2010). Microglia colonization of the brain follows a similar pattern in mice and human and thus mice studies of microglia are popular (Harry and Kraft, 2012). Microglia cells express a wide number of surface and nuclear receptors, including those for complement proteins, cytokines and chemokines, as well as major histocompatibility (MHC) molecules, immunoglobulins, TLRs, and cell adhesion molecules (Bilbo and Schwarz, 2012). The complement system is made up of a large number of distinct plasma proteins that react with one another to make bacteria or other cells more susceptible to the action of phagocytes. Furthermore, the complement system induces a series of inflammatory responses that help to fight infection. Microglia cells are distinct from other tissue macrophages due to their relatively quiescent

phenotype and tight regulation by the CNS microenvironment (Harry and Kraft, 2012). Microglia can develop a range of functional phenotypes, although most classification is derived mainly from animal studies based on the expression of either pro- or anti- inflammatory cytokines in association with pro- or anti- inflammatory receptors (Boche et al., 2013). In a healthy brain, microglia cells are in a state of rest. Microglia activation is essential to the response of CNS's tissue to injury. Changes in microglia phenotype relate to cellular processes including specific neurotransmitter, pattern recognition, or immune-related receptor activation (Harry, 2013). Microglia cells have a distinct morphology when compared with macrophages in other tissues, making identification possible. When the brain is injured or affected by diseases such as Alzheimer's and Huntington's disease**,** ramified microglia morphologically transform into "activated microglia", which show retracted processes and enlarged cell bodies and become proliferative at the affected site (Nakajima and Kohsaka, 2001). Microglia cells are capable not only of actively monitoring but also controlling the extracellular environment, walling off areas of the CNS from non-CNS tissue and removing dead or damaged cells (Harry and Kraft, 2012). Microglia processes extend towards those of their neighbors with the consequence that together they form a matrix extending throughout the CNS (Boche et al., 2013). The discovery that the brain of the developing mouse embryo already contains microglia cells suggested that microglia precursors may originate from the yolk sac, a structure that is present from an earlier stage of embryogenesis (Prinz and Priller, 2014). The early presence of microglia cells supports the need to understand their immunological role and response.

1.1.1 Microglia response

In response to injury or immune stimulation, microglia up-regulate a number of these surface receptors, including those for complement proteins, MHC II (important for antigen presentation), and cytokines, which in turn initiate both repair and cytotoxic processes via interactions with numerous other CNS cell types (e.g., neurons; Bilbo and Schwarz, 2012). Changes in the function of microglia, rather than simply changes in number and morphology, underlie their increased reactivity in neonatally-infected rats (Bilbo and Schwarz, 2012). Microglia is both larger and exhibited increased CD11b+ expression on a per cell basis in neonatally-infected rats (Bilbo and Schwarz, 2012). Cells identified as CD11b+ are dendritic cells that are constitutive markers of microglia. The modulation of inflammation by microglia cells is caused by interferon-gamma (IFN-γ; Boche et al., 2013). Interferons (IFN's) are a group of proteins known for their role in in stimulating the immune system to fight disease in response to pathogens. Interferon-gamma is a cytokine involved in the regulation of nearly all phases of immune and inflammatory responses, including the activation, growth and differentiation of T cells and B cells. Collective data has led to three conclusions: (1) Microglia have a critical role in learning and memory via the production of IL-1β, (2) Microglia dysfunction (exaggerated IL-1β) results in cognitive dysfunction, and (3) Early-life

events can significantly impact cognitive function later in life via long-term programming of microglia function (Bilbo and Schwarz, 2012).

Alzheimer's disease (AD) is a neurological disorder in which the death of brain cells causes memory loss and cognitive decline. Alzheimer's disease remains as one of the most common neurodegenerative disorders. Brain inflammation is a pathological trademark of AD. However, the characteristic inflammatory features such as swelling, heat, and pain are not present in the brain, and therefore inflammation is described as chronic (Rubio-Perez and Morillas-Ruiz, 2011). A key feature of chronic inflamed tissues is the presence of increased microglia cells in the CNS. Inflammatory components related to AD neuroinflammation include brain cells such as microglia and astrocytes, the complement system, as well as cytokines and chemokines (Rubio-Perez and Morillas-Ruiz, 2011). Significant activation of microglia appears to occur in a very early stage of the disease before severe cognitive decline occurs (E. McGeer and P. McGeer, 2003). The role of microglia has been found to be beneficial since activated microglia are neuropathologic features of AD. Overall, the cellular evidence of inflammation in AD consists of the appearance of activated microglia and astrocytes coupled with small accumulations of T cells marginating along postcapillary venules, with a few entering the tissue (E. McGeer and P. McGeer, 2003).

Huntington's disease (HD) is a monogenic neurodegenerative disease characterized by abnormal motor movements, personality changes, and early death (Moller, 2010). In line with the neurodegenerative process, astrocytes and microglia appear in the affected regions. Microglia cells increase considerably in individuals diagnosed with HD and expressed complement system factors as well. Inflammatory mediators such as IL-1β and TNF-α are increased in the striatum, IL-6 and IL-8 are upregulated in cortex and, surprisingly, the cerebellum, a CNS region commonly thought to be spared in HD (Moller, 2010). IL-1β is not only secreted by microglia but is also typically expressed in dendritic cells and is verifiably increased in several mouse models of HD and in the serum of HD patients (Ellrichmann et al., 2013). Previous research supports interactions between immune cells and neurodegeneration in HD. Of note, migration deficits seen in immune cells in both, HD patients and mouse models, may be an initial step in early immunological changes and may lead to elevation of proinflammatory cytokines and chemokines in HD (Ellrichmann et al., 2013). The relationships associated with neurodegenerative diseases, such as Alzheimer's and Huntington's disease, can be unraveled by understanding the role of microglia and inflammation.

1.2 Bacillus Calmette–Guérin (BCG)

Bacillus Calmette–Guérin (BCG) is an attenuated form of *Mycobacterium bovis*. Bacillus Calmette–Guérin inoculation of mice is a reliable rodent model to study the impact of the inflammatory response in the brain. Bacillus Calmette–Guérin can induce regulatory T cells (Tregs), which are neuroprotective in models of neurodegenerative diseases (Lacan et al., 2013). Particularly, Tregs limit inflammatory responses and tissue damage during persistent infection with Bacillus Calmette–Guérin (Lacan et al., 2013). The ability of BCG to modulate responses on an immune system can be attributable to the initial pro-inflammatory-anti-inflammatory cytokine balance, leading to tissue damage (Méndez-Samperio et al., 2008).

1.2.1 Sickness, behavior changes, and depression in response to BCG

Challenge with BCG induces changes in sickness, behavior, and depression in response to BCG. Sickness is a condition of being ill due to the presence of pathogens, with sickness behaviors described as the way an organism responds to pathogen stimulus. Depression is a mental health condition that is characterized by a mood disorder in which feelings of sadness, loss, anger, or frustration are present. Depressive phenotypes such as loss of experience in pleasurable experiences (anhedonia) and learned helplessness are observed in response to immunological challenges, including BCG (Rodriguez-Zas et al., 2014). The BCG challenge model is an effective model of acute sickness followed by depressive-like behaviors (Moreau et al., 2008; O'Connor et al., 2009).

Sickness behavior was prolonged in aged mice, as assessed by both locomotor and rearing activity when compared to adult-age mice (Kelley et al., 2013). Adult mice displayed depression-like behaviors at one day and seven days after exposure to BCG; however, aged mice continued to express both of these depression-like behaviors at three weeks following infection (Kelley et al., 2013). Depressive-like behaviors included a reduction in sucrose consumption (anhedonia) as well as increased immobility time during tail suspension as learned helplessness. Aged mice eventually recovered from the behavioral changes caused by BCG at both 14 and 21 days following infection (Kelley et al., 2013). The peripheral cytokine response to an inflammatory signal such as LPS is elevated in aged mice. The greater change in peripheral cytokines results in an exaggerated inflammatory response in aged animals, as illustrated by heightened cytokine expression in the brains of aged mice (Kelley et al., 2013). Furthermore, when microglia cells from mice injected peripherally with LPS were isolated, cells showed a large reduction in expression of fractalkine, a chemokine expressed by neurons and binds to its receptor, CX_3CR1 on microglia. This reduction was sustained longer if the microglia were derived from aged mice (Kelley et al., 2013). This data is consistent with the idea that microglia cells in the CNS become deregulated during

the aging process and support the emerging concept that hyperactivity of microglia mediates inflammation-induced behavioral changes (Kelley et al., 2013). Aged mice demonstrated prolonged sickness and depressive-like behaviors, an illustration that aging effects the duration of post-challenge behaviors (Kelley et al., 2013). A comprehensive analysis of the changes in behavior associated with BCG challenge was undertaken using univariate and multivariate approaches (Rodriguez-Zas et al., 2014). These techniques were applied to accomplish two goals: the identification of groups of mice and the identification of groups of sickness and depression-like indicators (Rodriguez-Zas et al., 2014). Mice treated with BCG exhibited sickness as indicated by changes in body weight during the first days after the challenge (Rodriguez-Zas et al., 2014). Sickness caused by BGC was supported by mice exhibiting sickness behaviors encompassing reduction in body weight and locomotor activity (Rodriguez-Zas et al., 2014). Although the difference in sickness indicators between BCG-treatment groups subsided by Day 5, differences in depression-like indicators were detected in subsequent days (Rodriguez-Zas et al., 2014).

Behavioral consequences of chronic infection of mice with BCG has been shown to chronically activate brain indoleamine 2,3-dioxygenase (IDO), a tryptophan-catabolizing enzyme that mediates the occurrence of depressive-like behavior following acute innate immune system activation (Moreau et al., 2008). IDO is the first rate-limiting enzyme in the synthesis of kynurenine and is required to mediate depressive-like behavior in response to infection with BCG (Wang et al., 2010) Activation of this enzyme by pro-inflammatory cytokines leads to the generation of several neuroactive metabolites including quinolic acid, an N-methyl-D-aspartate (NMDA) receptor agonist that has been implicated in neurodegenerative disorders (O'Connor et al., 2009). The nature of the pro-inflammatory cytokines responsible for up regulating IDO is still elusive within the context of inflammation-associated depression (O'Connor et al., 2009). Inoculation with BCG induced an acute episode of sickness (up to 5 days) that was followed by development of delayed depressive-like behaviors lasting over several weeks (Moreau et al., 2008). Intraperitoneal administration of BCG to mice was rapidly followed by long-lasting mycobacterial dissemination across organs (particularly the lungs) except the brain and by a drastic increase in circulating IFN-γ, a key cytokine for the activation of IDO (Moreau et al., 2008). Furthermore, BCG induced a sustained increased immobility up to three weeks after treatment as well as BCG-induced mice had a progressive elevation of TNF-α concentrations and an IL-6 increase that was significantly higher at 1 week after treatment (Moreau et al., 2008). Overall, the study revealed sustained depressive-like behavior for up to 3 weeks post-inoculation, i.e. well beyond the period during which signs of sickness behavior were present (Moreau et al., 2008).

With the identification of peptides and proteins susceptible to immune challenge, researchers can understand the pathways affected and develop markers and treatments to help with dysfunction. Proteins

play a role in BCG-injected mice in relation to a microglia cell response (Yong et al., 2011). This pathogen slowly replicates in the host causing a long-term increase in the levels of circulating immune factors (e.g., cytokines) which can enter the CNS (Yong et al., 2011). These immune factors have limited microglia activation and proliferation, the influx of peripheral macrophages or microglia precursor cells, or had a supportive effect on neurons in the area of injury (Yong et al., 2011). Twenty-one days after treatment, mice were still displaying effects of BCG-induced neuroprotective immune responses (Yong et al., 2011). Factors such as neurodegeneration and secondary damage to neurons were understood throughout the study given the amount of time passed after treatment. Further limitations when analyzing the effects of BCG include choosing proper analyses. Specifically, completing only univariate analysis can fail to capture sickness indicators, thus, demonstrating the importance of multivariate analysis to detect subtle differences critical for developing effective individualized therapies (Rodriguez-Zas et al., 2014). Such limitations support the need for further analyses of peptides and proteins expressed in response to BCG challenges.

2. Proteomics, proteins and peptides

Proteomics is a large-scale study of proteins in all aspects, such as their abundances and post-translational modifications (PTMs). Proteins are large molecules composed of one or more long chains of amino acids and are an essential part of all living organisms. Different combinations of amino acids allow for a unique sequence that determines the protein's shape and function. Protein functions include providing immune protection, transporting molecules such as oxygen and controlling growth and differentiation. All amino acids contain an amino (or imino) group and a carboxyl group. A peptide bond occurs when two amino acids interact leading to a series of amino acids that constitute a peptide. Proteins and peptides play a key role in the response of immune cells to challenge (Wang et al., 2010). Furthermore, proteins and peptides can undergo a variety of post-translational modifications that influence its shape, function, and expression. Proteins play a crucial role in the immune system and have become the dominant subject in medicinal research.

Post-translational modifications are covalent processing events that change the properties of a protein by proteolytic cleavage or addition of a modifying group to one or more amino acids (Mann and Jensen, 2003). Common post-translational modifications are carbamidomethylation, carboxymethyl, and oxidation. Post-translational modifications regulate activity, localization and interaction with other cellular molecules such as proteins. Post-translational modifications increase the complexity of the proteome and play a key role in functional proteomics because they have the potential to target proteins that directly cause or contribute to diseases. Proteomics approaches are well-suited to help us understand

the role proteins and peptides associated with the immune response orchestrated by the microglia in response to BCG. The most prevalent platform for proteomics is the mass spectrometer (MS).

2.1 Bottom-up proteomics

Bottom-up protein analysis refers to the characterization of proteins by analysis of peptides released from the protein through proteolysis (Zhang et al, 2013). Proteolysis is the breakdown of proteins into smaller polypeptides or amino acids. Shotgun proteomics is a special case of bottom-up proteomics where a complex mixture of proteins is digested into peptides, typically by trypsin, followed by multidimensional high performance liquid chromatography (LC) online coupled to the mass spectrometer (Matthesion and Bunkenborg, 2013). Shotgun proteomics also provides an indirect measurement of proteins through peptides derived from proteolytic digestion of intact proteins (Zhang et al, 2013). The first choice of enzyme is usually trypsin, which cleaves C-terminal to arginine and lysine, resulting in peptides with basic amino acids at the C-terminus, which are most easily sequenced by MS (Mann and Jensen, 2003). The difference between bottom-up and shotgun proteomics is that bottom-up strategies do not necessarily have LC separation of peptides prior to MS, whereas in shotgun strategy LC and typically multidimensional LC is always used to separate a complex mixture of peptides originating from many different proteins (Matthesion and Bunkenborg, 2013). The combination of bottom-up and shot gun approaches is efficient for identification of proteins due to the complex mixture of peptides fractionated by chromatography on the peptide level.

Liquid chromatography affects the quality and accuracy of all subsequent analyses, thus, making the technique crucial. Liquid chromatography produces information such as peptide masses, peptide fragment masses and peptide retention time(s) on the liquid chromatography column. LC-MS/MS has helped in developing routine methods of high sensitivity, high specificity, high throughput, and high-cost effectiveness. The mass spectrometer used for LC-MS in proteomics is most often a tandem mass spectrometer which produces MS or both MS and MS/MS data (Matthesion and Bunkenborg, 2013).

2.2 Tandem Mass Spectrometry

A tandem mass spectrometer is a device for measuring the mass-to-charge ratio (m/z) of ionized molecules. Tandem mass spectrometers consist of three parts: an ion source, two mass analyzers, and a detector (Figure 1.4). Sample proteins are first proteolytically cleaved into smaller peptides by the enzyme trypsin. The resulting peptide mixture is then subjected to reversed-phase chromatography directly coupled with a mass spectrometer (Nesvizhskii et al., 2003). Peptides are then ionized and selected ions are subjected to fragmentation in the collision cell to produce tandem mass spectra

(Nesvizhskii et al., 2003; Figure 1.4). Collision induced dissociation (CID) entails acceleration of the kinetic energy of ions to promote energetic collisions with a target gas, thus causing conversion of the ion's kinetic energy to internal energy and ultimately resulting in ion fragmentation (Zhang et al., 2013). The mass analyzer separates the ions according to their mass-to-charge ratio. At this stage, computational methods must be used to infer the peptides and proteins that gave rise to the observed spectra (Nesvizhskii et al., 2003). A simple peptide sample for MS analysis gives less risk of peak overlap and more sensitivity due to less ion suppression (Matthesion and Bunkenborg, 2013). Furthermore, label-free quantitation of complex proteomic samples using the high mass resolution and high mass accuracy is completed with a Linear Ion Trap Mass spectrometer (LTQ) mass spectrometer (Andreev et al., 2007). The LTQ-FT consists of an Ion Trap is used for measuring m/z or as fragmentation chamber and optimum pressure for CID fragmentation for scanning ions out of chamber (Andreev et al., 2007). Both protein and peptide separation techniques at different stages can be applied to obtain simple peptide mixtures (Matthesion and Bunkenborg, 2013). In a typical MS setup for protein identifications the survey scan is analyzed on the fly by the instrument software to select ions that are isolated, fragmented and analyzed by a mass analyzer to generate an MS/MS spectrum (Matthesion and Bunkenborg, 2013). To analyze these MS/MS spectra the parent ion intensity in the MS spectra over the elution profile of the peptide is correlated with the ion intensity of the fragment ions and in this way a link between parent ions and fragment ions is established (Matthesion and Bunkenborg, 2013). The detector records a signal that can be electronically amplified and stored (Figure 1.4).

For many instruments, the major peaks in the MS/MS spectrum are b-ions, where the charge is retained on the N-terminus, or y-ions, where the charge is retained on the C-terminus (Cottrell, 2011; Figure 1.5). By separating the proteins and using mass spectrometers, amino acid identification is amplified. Each amino acid has a unique mass, therefore, identifying and sequencing these smaller pieces makes protein identification possible. Database matching of MS/MS spectra identifies peptides, not proteins. Using the peptide sequences to deduce which proteins were present in the original sample is difficult because many of the peptide sequences in a typical search result can be assigned to more than one protein (Cottrell, 2011). Typically MS/MS aims to identify a minimum number of proteins that can account for the observed peptides, although some ambiguity concerning which proteins are present in data should be taken into consideration. One of the main aims of MS-based proteomics is to identify and quantify proteins and their post-translational modifications in either a purified, enriched or complex protein mixture (Matthesion and Bunkenborg, 2013). Tandem mass spectrometry can be used for absolute quantitation of peptide and proteins together with their associated post-translational modifications (Matthesion and Bunkenborg, 2013).

2.3 Improving MS-based proteomic profiling

Fractionation is the separation of a chemical compound into smaller components. Fractionation or separation of protein mixtures prior to trypsin digestion and mass spectrometry analysis reduces the complexity of samples and increases the sensitivity and likelihood of detecting less abundant proteins (Jafari et al., 2012). Fractionation techniques at the peptide and protein levels enhance the analytical power of bottom-up MS-based proteomics. Protein fractionation techniques are commonly seen as gel-based. The principle of protein fractionation in gels is based on a protein's molecular weight (MW), isoelectric point (pI) or both (Jafari et al., 2012). Gel-based fractionation techniques are used to isolate simplified protein mixtures of proteins with similar physical structures and properties. Designated fractions 1 and 2 (Fr1/Fr2) represent two groups of proteins that have, on average, low and high sizes, respectively. The size is determined by how far each protein traveled across the gel. In Fr1, proteins traveled further in the gel; thus, they had a smaller average size than Fr2 proteins.

Liquid chromatography MS/MS data often contains noise, redundancy, as well as errors due to sample preparation and instrument approximation; therefore, further data refinement is necessary. Particularly, when peptide signals are spread over a large retention time, multiple signals for the same peptide due to technical or biological variation in retention time, mass spectrometer speed and sensitivity, and background noise due to chemical interference are factors taken into consideration (Zhang et el., 2013). In LC-MS/MS analyses of complex peptide mixtures, dynamic exclusion is used to minimize repeat selections of identical peptide precursors for collision-induced dissociation. The dynamic exclusion approach is evaluated to assess the trade-off between the higher number of proteins identified and corresponding lower total spectral count per protein when dynamic exclusion was enabled or ON relative to OFF. Techniques that affect MS-based proteomic profiling improve the data generated for further analysis.

3. Detection of peptides and proteins using protein databases

The data at the end of a typical LC-MS/MS experiment consists of a large collection of precursor ion intensities and associated tandem mass spectra from which the most important information needs to be extracted (Matthesion and Bunkenborg, 2013). MS/MS spectra are processed to a peak list. The generated MS/MS are compared to the theoretical spectra of known or predicted peptides such as those compiled in public databases of proteins. A number of protein and peptide databases can be used to identify peptides using database search programs including Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB), the Worldwide Protein Data Bank (wwPDB), and Universal Protein Resource (UniProt).

The RCSB PDB archive is the single worldwide repository of information about the 3D structures of large biological molecules such as proteins that was established in 1971 (Berman et al., 2000). The RCSB PDB is updated weekly by an advisory committee that is made up of an international team of experts in X-ray crystallography, cryo-electron microscopy, nuclear magnetic resonance, bioinformatics and education. The RCSB PDB allows users to perform simple and complex queries on the data, analyze, and visualize the results (Berman et al., 2000). In 2003, worldwide wwPDB was formed to maintain a single PDB archive of macromolecular structural data that is freely and publicly available to the global community (Berman et al., 2003). Three organizations have formed a collaboration to oversee the newly formed wwPDB includes: The Research Collaboratory for Structural Bioinformatics, the Macromolecular Structure Database at the European Bioinformatics Institute and the Protein Data Bank Japan at the Institute for Protein Research in Osaka University (Berman et al., 2003). The creation of the wwPDB ensures that the PDB archive remains uniform while providing a mechanism to ensure consistent data for software developers and users worldwide.

With the increasing volume and variety of protein sequences and functional information, UniProt serves as a central resource of protein sequence and function, providing a cornerstone for scientists active in modern biological research, especially in the field of proteomics (Apweiler et al., 2004). In 2002 three institutes decided to form the UniProt consortium which includes: the European Bioinformatics Institute, the Swiss Institute of Bioinformatics and the Protein Information Resource. For convenient sequence searches, UniProt also provides several non-redundant sequence databases (Apweiler et al., 2004) that are updated daily. UniProt is the most widely used database because it provides more annotations than any other sequence database with a minimal level of redundancy through human input or integration with other databases (Xu, 2012).

4. Database search programs and PEAKS

Database search programs including X! Tandem, OMSSA, Crux, Sequest, and PEAKS (Akhtar et al., 2012; 2014a,b) are commonly used to identify the peptides by comparing the observed spectra to the theoretical spectra from the peptides in a target database. X!Tandem (http://www.thegpm.org/tandem; Version 2010.12.01.1 released on December 01, 2010) was developed to optimize speed and to minimize the computational requirements. The algorithm preprocesses the observed spectra to remove noise, processes database peptide sequences with cleavage reagents, PTMs and scores the peptide matches between the observed and predicted spectra (Akhtar et al., 2012). The scores are converted to hyperscores and the distribution of hyperscores of all matches is used to translate the hyperscore of each match into an E-value (Akhtar et al., 2012).

The Open Mass Spectrometry Search Algorithm (OMSSA; Version 2.1.7 released on June 15, 2010; http://pubchem.ncbi.nlm.nih.gov/omssa) centers on optimizing the speed of database searching approach. The scoring of each match assumes that the number of matches between observed and predicted peaks for a peptide sequence follows a Poisson distribution (Akhtar et al., 2012). The lambda (or average) parameter of the Poisson distribution is calculated as a function of the fragment ion tolerance, the number of predicted and observed peaks and the neutral mass of the precursor ion. OMSSA provides E- and P-values based on the dimensions of the target database (Akhtar et al., 2012).

Among these, PEAKS (http://www.bioinfor.com; Version 7.0.2 released on September 18, 2014) was used and will be discussed in detail. PEAKS was developed to effectively identify peptides and proteins, PTMs, as well as result validation, visualization and reporting (Ma et al., 2003). PEAKS performs *de novo* sequencing directly from the MS/MS data and therefore does not rely on a protein database. PEAKS computes the best possible sequences among all possible amino acid combinations. PEAKS scoring is calculated with a programming algorithm to perform the computational methods that compute peptides whose ions correspond to as many high abundance peaks in the spectrum as possible (Me at al., 2003).

4.1 *De novo* sequencing and scoring

De novo sequencing develops the peptide sequence from the tandem mass spectrum without the need of a sequence database. The independence to a sequence database makes *de novo* sequencing a method for identifying novel peptides. Even when the peptide sequence is in the database, *de novo* sequencing improves the database-search based peptide identification since the match between the *de novo* sequence and the database sequence confirms the correctness of the identification. PEAKS assigns a local confidence score for each amino acid in *de novo* sequences. PEAKS is the only software that reports a local confidence score for each amino acid of the *de novo* sequence which filtration at the amino acid level. To filter at peptide level, PEAKS additionally calculates the average local confidence (ALC) and total local confidence (TLC) scores for each peptide by averaging and summing up the local confidences.

The confidence score computed is a score for each of the top-scoring peptide sequences. The refined scores can be seen as non-normalized measures of the likelihood of correctness for each peptide, and the distribution of scores gives a measure of the overall probability of successful sequencing (Ma et al., 2003). PEAKS first converts the refined score x of each peptide sequence to a raw confidence X by the formula $X=\exp(cx)$, where c is a parameter that is estimated from the spectrum by PEAKS (Ma et al., 2003). The raw confidence scores for all the top-scoring peptide sequences are then normalized to be the final confidence scores so that they sum up to 1. Finally, the local confidences for each residue are derived from consensus among the globally top-scoring sequences (Ma et al., 2003). The local confidence

score ranges from 0% to 99%, indicating how confident the algorithm considers a particular amino acid is correctly sequenced (Ma et al., 2003). In relation to amino acid confidence scores, PEAKS assigns high confidence score as ≥ 80%. Furthermore, peptide sequences are evaluated by an ALC score. The ALC score is calculated as the total of the residue local confidence scores in the peptide divided by the peptide length (Ma et al., 2003).

To report a *de novo* only peptide, the ALC scores must be better than or equal to a specified threshold. Meanwhile, the spectrum's best database search result's score should be no greater (Ma et al., 2003). Furthermore, the significance value is expressed as $-10*\log_{10}$(P-value). For a given score, the corresponding P-value is defined as the probability that a false identification in the current search achieves the same or better matching scores (Zhang et al., 2011). The P-value attempts to predict the false positive rate (FPR), i.e., the ratio between the number of false identifications above the given score *T* and the total number of false identifications (Zhang et al., 2011). If the P-value is *P*, the final peptide score (called the significance score) is a $-10\log_{10}$P-value. Overall, the more significant match will have a higher -10*log10(P-value). Typically the $-10\log_{10}$P-value threshold used for *de novo* peptides is the same as the $-10\log_{10}$P-value threshold used for filtering peptides. The *de novo* only peptides are represented by *de novo* sequence tags and remain unidentified by the database search algorithms.

PEAKS scoring also includes coverage, mass error, and linear discriminative function (LDF) score. PEAKS defines coverage as the percentage of the protein sequence covered by supporting peptides. The protein coverage visualizes the mapping of supporting peptides and *de novo* tags to the protein selected. The coverage view also displays all identified sites with modifications to assist with protein characterization at the amino acid level. Furthermore, proximity between observed and expected mass is necessary for accurate peptide identification. PEAKS provides a visual inspection of trends in differences between observed and expected masses. Precursor mass error, calculated as 10^6 × (precursor mass - peptide mass) / peptide mass. (Ma et al., 2003). Precursor mass errors are expected to center at 0 ppm across the range of m/z to illustrate a well-calibrated instrument. Finally, PEAKS uses a LDF score to measure the quality of the peptide-spectrum match. The LDF score considers the adequacy of the match between the fragment ions and the peaks in the spectrum and other factors such as the similarity between the *de novo* sequencing peptide and the database peptide (Zhang et al., 2012). For easier interpretation, the LDF is converted to a P-value ($-10*\log_{10}$(P-value)). P-value is defined as for a given score *x*, its equivalent P-value is the probability that a false identification has score ≥ *x*. The false identification is the result of many random peptides in the database, thus, the definition in PEAKS DB is more useful for the controlling of result quality with P-value. The smaller the P-value, the less likely the peptide-spectrum

match is random. Adequacy is illustrated by many target matches and very few decoy matches in a graph showing the x-axis as P-values and the y-axis as the number of peptide-spectrum matches with that score.

5. Differential Detection Analysis

Differential detection analyses identified the differences in peptides and proteins between BCG vs. control group. Following the identification of peptides and proteins in the groups observed, the analysis of interest is to uncover differences in the presence of these molecules between groups. Spreadsheet and statistical software can be used to count the detection of peptides and proteins across samples. Statistical Analysis System (SAS) 9.3 (Cary, 2011) is a statistics program used to analyze data. SAS 9.3 program routines can include data steps and statements to execute user-specified procedures. Each data step includes an informat, infile and input statement. For example, routines can be written to count the presence of peptides and proteins with the use frequency, sort, and transpose procedures and the where function. The procedure steps process the SAS data sets and allows for the generation of final output.

Individual samples relative to pooling of samples allow for pros and cons when compared to one another. Although the cost of analyzing individual samples is greater than the cost of pooling samples, consideration should be given for both. Individual samples generally produce lower error rates, while pooling strategies generally lead to higher error rates (Kusonmano et al., 2012). Specifically, larger pool sizes cause higher error rates for classifiers that produce misclassification of data. Furthermore, storing specimens individually allows for specimen variance to be analyzed regularly (Bignert et al., 2014) which allows for a basis for studies made later in time. On the other hand, pooling samples has become widely accepted among the scientific community. Pooled samples are deemed suitable in many situations, especially when chemical analytical costs are the largest proportion of the budget, or if variation due to specimen variance needs to be reduced (Bignert et al., 2014). Scientists are encouraged to make informed decisions to leverage pooling designs as a valid strategy to compensate for limited amounts of samples or high biological variation, or as a remedy to improve time efficiency (Kusonmano et al., 2012). Identification of peptides and proteins through differential detection allows for further investigation of the relationships among proteins using enrichment analysis.

6. Functional category enrichment and Protein ANalysis THrough Evolutionary Relationships (PANTHER)

Common functional categories can be identified among the peptides and proteins detected using functional enrichment analysis. A number of programs are available to identify enriched categories among lists of proteins or genes. The programs differ on the annotations accepted as input, categories

considered, and approaches to identify enrichment of categories. The Database for Annotation, Visualization and Integrated Discovery (DAVID) is a bioinformatic resource that consists of an integrated biological knowledgebase and analytic tools aimed at systematically extracting biological meaning from large protein lists (Huang et al., 2009). DAVID Version 6.7 (release date Jan 27, 2010) includes DAVID Functional Annotation Clustering, a feature to the DAVID Functional Annotation Tool. The Functional Classification Tool generates a gene-to-gene similarity matrix based shared functional annotation using over 75,000 terms from 14 functional annotation sources. This function uses a novel algorithm to measure relationships among the annotation terms based on the degrees of their co-association genes to group the similar, redundant, and heterogeneous annotation contents from the same or different resources into annotation groups (Huang et al., 2009). The threshold of EASE Score, a modified Fisher Exact P-Value, for gene-enrichment analysis is calculated and ranges from 0 to 1. Fisher Exact P-Value equalling 0 represents perfect enrichment. Usually P-Value is equal or smaller than 0.05 to be considered strongly enriched. This reduces the burden of associating similar redundant terms and makes the biological interpretation more focused in a group level (Huang et al., 2009).

The Protein ANalysis THrough Evolutionary Relationships (PANTHER) Classification System (Mi et al., 2013) web tool frequently used to identify enriched categories. PANTHER Version 9.0 (release date Jan 20, 2014) contains 7180 protein families, divided into 52,768 functionally distinct protein subfamilies. The enrichment test takes a list of genes, each with a numerical value, optimally this list is genome wide (i.e. there is a value for as many genes in a genome as possible); it then finds functional classes for which the genes of that class have values that are non-randomly selected from the genome-wide distribution of values (Mi et al., 2013). The output of the tool is a list of P-values for under- or overrepresentation of each functional category in each of the input lists. From this output page, the user can export the statistics, or follow links to graphically view (as pie charts or bar graphs) the data used to compute the P-values, or to look at the list of genes/proteins in any functional group (Mi et al., 2013).

PANTHER and DAVID accept protein identifiers from many repositories including the Mouse Genome Informatics (MGI) database and evaluate enrichment of protein family and subfamilies, Gene Ontology molecular functions, biological processes and pathways among a list of genes or proteins provided.

Mouse Genome Informatics is an international database resource for the laboratory mouse, providing integrated genetic, genomic, and biological data (Blake et al., 2014). Projects that contribute to the Mouse Genome Informatics system are: Mouse Genome Database Project, Gene Expression Database Project, Mouse Tumor Biology Database Project, Gene Ontology Project at MGI, and MouseMine Project. The

MGI system provides the highest quality and most comprehensive consensus and experimental views of [the] laboratory mouse as an experimental organism (Blake et al., 2014).

Gene Ontology (GO) database is a collaborative effort to address the need for consistent descriptions of gene products across databases. An ontology comprises a set of well-defined terms with well-defined relationships (Ashburner et al., 2000). The Gene Ontology project provides a controlled vocabulary of terms for describing gene product characteristics and gene product annotation data from GO Consortium members, as well as tools to access and process these data. The project began in 1998 as a collaboration between three model organism databases: Flybase (*Drosophila*), the *Saccharomyces* Genome Database, and the Mouse Genome Informatics projects, but has since grown to include many databases (Gene Ontology Consortium, 2004). The GO database is constantly being expanded as collaborating databases increase their knowledge in accumulating and changing protein roles in cells. Data can be annotated to varying levels depending on the amount and completeness of available information (Ashburner et al., 2000). Gene Ontology requires that all gene associations to the ontologies must be attributed to the literature; for each citation the type of evidence will be encoded (Ashburner et al., 2000).

The GO molecular function (MF) categories refer to the function of a protein by itself or with directly interacting proteins at a biochemical level. The molecular function of individual proteins is complex to their physical structures and unique to their amino acid sequences. Gene Ontology molecular function terms represent activities rather than the entities (molecules or complexes) that perform the actions, and do not specify where, when or in what context the action takes place (Gene Ontology Consortium, 2004). This definition also applies to the capability that a gene product (or gene product complex) carries as a potential. Examples of individual molecular function terms are the broad concept ‘kinase activity’ and the more specific ‘6-phosphofructokinase activity’, which represents a subtype of kinase activity (Gene Ontology Consortium, 2004).

The GO biological process (BP) categories refer the function of the protein in the context of a larger network of proteins that interact to accomplish a process at the level of the cell or organism. Biological Process describes biological goals accomplished by one or more ordered assemblies of molecular functions (Gene Ontology Consortium, 2004). Processes often involve a chemical or physical transformation, in the sense that something goes into a process and something different comes out of it (Ashburner et al., 2000). High-level processes such as ‘cell death’ can have both subtypes, such as ‘apoptosis’, and subprocesses, such as ‘apoptotic chromosome condensation’ (Gene Ontology Consortium, 2004).

The PANTHER Pathway consists of over 176, primarily signaling, pathways, each with subfamilies and protein sequences mapped to individual pathway components (Mi et al., 2013). A component is usually a single protein in a given organism, but multiple proteins can sometimes play the same role. The PANTHER Pathway system uses information from databases that consist of Reactome and Kyoto Encyclopedia of Genes and Genomes (KEGG). Reactome is a collaboration among groups at the Ontario Institute for Cancer Research, Cold Spring Harbor Laboratory, New York University School of Medicine and The European Bioinformatics Institute, to develop an open source curated bioinformatics database of human pathways and reactions (Croft et al., 2011). The goal of Reactome is to provide intuitive bioinformatics tools for the visualization, interpretation and analysis of pathway knowledge to support basic research, genome analysis, modeling, systems biology and education (Croft et al., 2011). The Scientific Advisory Board provides the Reactome group with independent, expert advice on scientific research, emerging issues and trends, and on scientific partnerships and linkages (Croft et al., 2011). The Scientific Advisory Board meets annually to discuss Reactome's scientific agenda, explore ways to expand its research efforts and the critical review of our database, and software development.

Kyoto Encyclopedia of Genes and Genomes (KEGG) is a database resource that integrates genomic, chemical and systemic functional information (Kanehisa et al., 2012). The KEGG database was developed in collaboration with the following two centers: Bioinformatics Center, Institute for Chemical Research, Kyoto University and the Human Genome Center, Institute of Medical Science, University of Tokyo (Kanehisa et al., 2012). The KEGG database is the most suitable of databases and associated software because it integrates current knowledge on molecular interaction networks in biological processes (PATHWAY database) in PANTHER. The Kyoto Encyclopedia of Genes and Genomes consists of three databases: PATHWAY for representation of higher order functions in terms of the network of interacting molecules, GENES for the collection of gene catalogs for all the completely sequenced genomes and some partial genomes, and LIGAND for the collection of chemical compounds in the cell, enzyme molecules and enzymatic reactions (Kanehisa and Goto, 2000). The KEGG pathway representation focuses on the network of gene products, mostly proteins (Kanehisa and Goto, 2000). A metabolic pathway is a network of indirect protein-protein interactions, which is a network of enzyme-enzyme relations (Kanehisa and Goto, 2000). In contrast, the regulatory pathway often consists of direct protein-protein interactions, such as binding and phosphorylation, and another class of indirect protein-protein interactions, which are relations of transcription factors and transcribed gene products via gene expressions (Kanehisa and Goto, 2000). Kyoto Encyclopedia of Genes and Genomes database also provides information about the universe of proteins (GENES/SSDB/KO databases), and the information

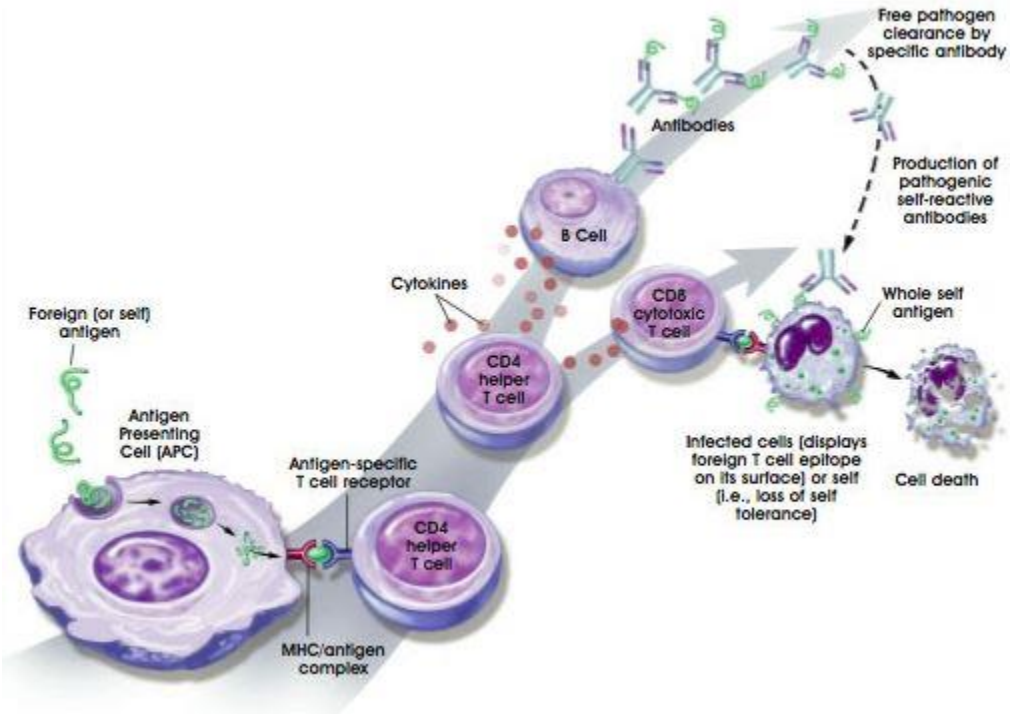
about the universe of chemical compounds and reactions (COMPOUND/GLYCAN/REACTION databases).

7. Research Objective

The objectives of my study were to identify the proteins and peptides in the microglia cells of BCG-treated and control mice, compare the detected proteins and understand the differences through enrichment analysis of differentially detected proteins. The study of proteins and functional categories affected by the BCG-treatment allows for relationships among proteins related to the neurological immune response in mice to be understood. With studies such as this, researchers can aim to develop approaches to minimize the impact of pathogen challenge in neurological functions.

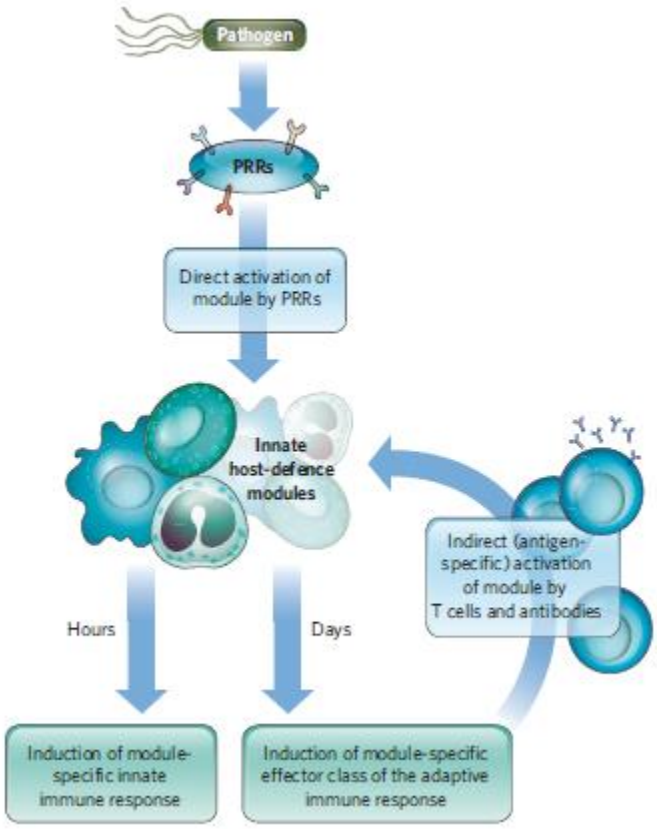
8. Figures

Figure 1.1 Immune Response



Source: Autoimmune Diseases and the Promise of Stem Cell-Based Therapies. In Stem Cell Information. Bethesda, MD: National Institutes of Health, U.S. Department of Health and Human Services, 2009 http://stemcells.nih.gov/info/scireport/pages/chapter6.aspx

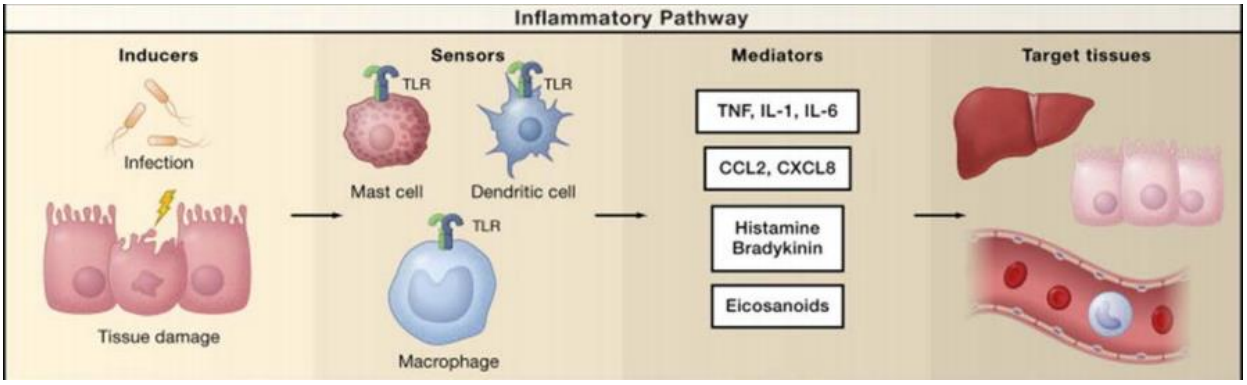
Figure 1.2 Host Defense initiated by Patter-Recognition Receptor



Host-defense mechanisms can be induced directly, by engagement of PRRs, or indirectly, by T cells and/or antibodies. Each module is characterized by distinct antimicrobial defense mechanisms and can instruct the adaptive immune system to mount a response involving a module-specific effector class. After an adaptive immune response has been initiated, it results in antigen-specific activation of the same innate immune module that instructed the adaptive immune response.

Source: Ruslan Medzhitov. October 2007. Nature 449, 819-826.

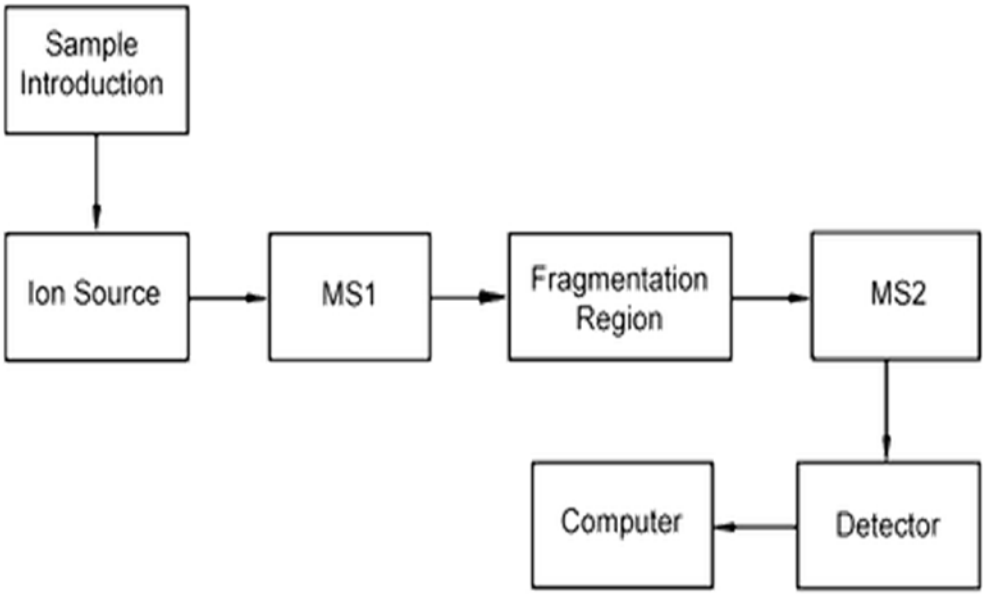
Figure 1.3 Inflammatory Pathway



Inflammation is an essential immune response that enables survival during infection or injury and maintains tissue homeostasis under a variety of noxious conditions. Inflammation comes at the cost of a decline in tissue function, which can in turn contribute to the pathogenesis of diseases of altered homeostasis.

Source: Ruslan Medzhitov. Inflammation 2010: New Adventures of an Old FlameCell, Volume 140 , Issue 6 , 771 – 776.

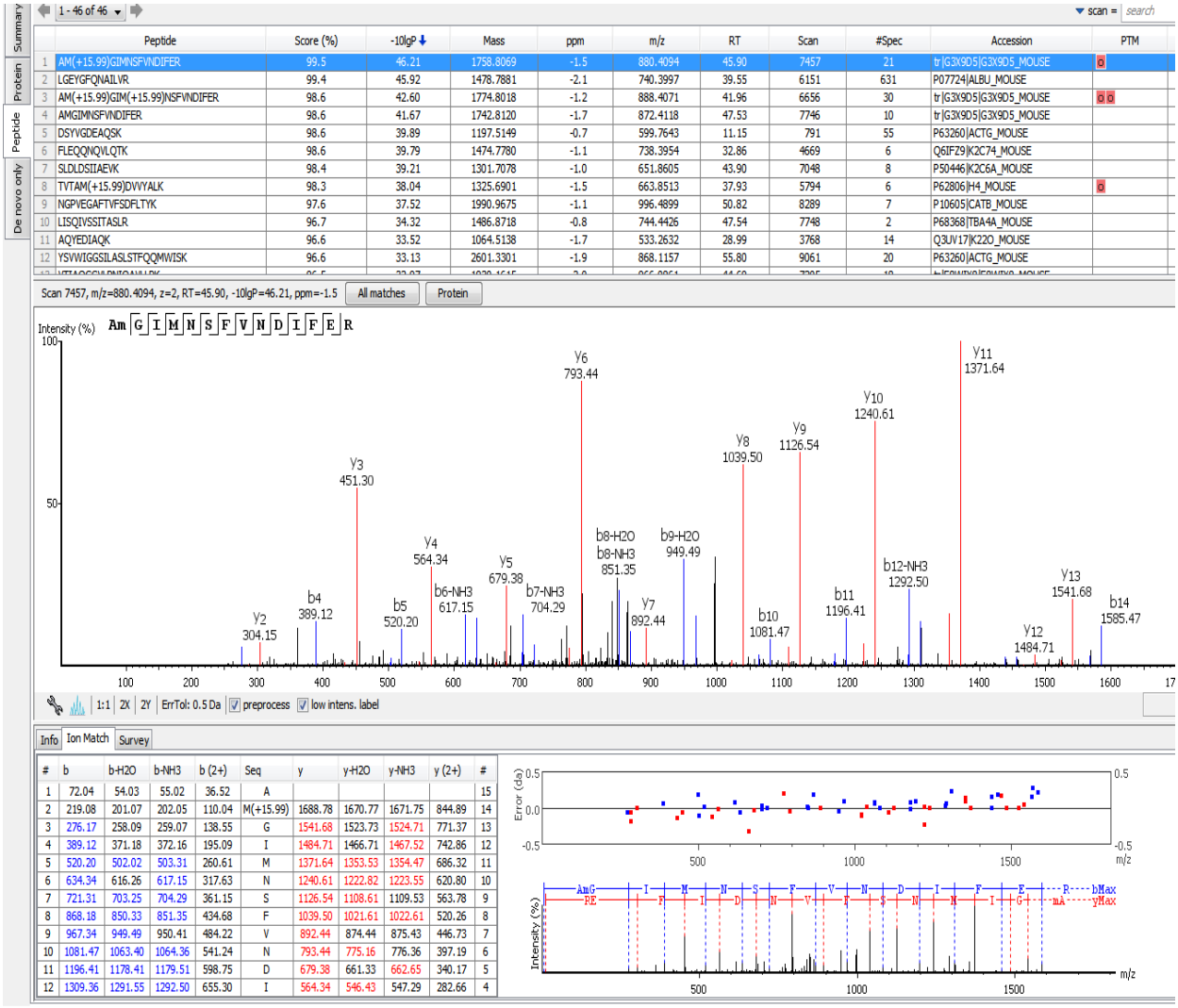
Figure 1.4 Tandem Mass Spectrometry



Sample proteins are first introduced. Peptides are then ionized and selected ions are subjected to fragmentation in the collision cell to produce Tandem mass spectrometer consists of two mass spectrometers linked in such a way that ions preselected by the first mass analyzer (MS1) are chemically or energetically modified and the results analyzed by the second mass analyzer (MS2). To analyze the spectra the parent ion intensity in the MS spectra over the elution profile of the peptide is correlated with the ion intensity of the fragment ions. The detector records a signal that can be electronically amplified and stored in a computer.

Source: Kahkashan Zaidi. U.S. Pharmacopeia. General Chapter 5 USP29-NF24 Page 2705.

Figure 1.5 MS/MS Spectrum b- and y-ions



Representation of major peaks (b-and y-ions) in the MS/MS spectrum from PEAKS software.

Chapter 2: Research Chapter

Study of proteins after recovery from bacterial infection

Abstract

The effect of Bacillus Calmette–Guérin (BCG) on the microglia of mice after recovery from infection is incompletely understood. Microglia cells of mice were collected from two groups, mice infected with BCG vs. control mice (BCG vs. control). The objective of this study was to compare the proteins and peptides present in the microglia of BCG-challenged mice one-week after treatment relative to control mice and gain insights through differential detection and enrichment analysis. Detection relied on tandem mass spectrometry proteomics. The database search software PEAKS was used to detect peptides and proteins in the treated and control samples. Differential detection was performed using SAS procedures and enrichment classification was completed using PANTHER. The consistently higher number of proteins and peptides (except for one sample) detected in the control samples suggested that BCG impacts the production of proteins. A number of proteins including F-actin-capping protein subunit alpha-2 (P47754), Alpha-enolase (P17182), and myelin basic protein (F7A0B0) were identified in greater abundance in control mice; while chitinase-like protein 3 (O35744), vesicle-associated membrane protein-associated protein A (Q9WV55), and Protein SET (Q9EQU5) were identified only in BCG-treated mice. The differential detection was consistent with similar studies on the neurological activity and inflammation response to BCG-challenged mice. Functional enrichment analysis identified enriched pathways among differentially detected proteins with a differential detection of two (BCG and control mice) associated with inflammation-mediated and microglia activation with Huntington disease (P00029) and Glycolysis (P00024).

Keywords: Proteomics – MS/MS spectrometry – Bacillus Calmette–Guérin (BCG) – Functional enrichment

1. Introduction

An immune response occurs when a stressor, such as bacterial infection, is detected by the body. The inflammatory response following this detection is mediated by activated microglia. Microglia are the primary mediators of the central nervous system's (CNS) immune defense system and are integral to the subsequent inflammatory response (Loane and Byrnes, 2010). Microglia are resident cells of the brain involved in regulatory processes critical for development, maintenance of the neural environment, response to injury, and subsequent repair (Harry, 2013). Microglia cells sense pathological events,

causing innate immune activation with the appropriate physiological and behavioral response. Bacillus Calmette-Guérin (BCG) induces an acute episode of sickness (five days after injection) that is followed by development of delayed depressive-like behaviors lasting over several weeks (Moreau et al., 2008). The BCG challenge model is an effective model of acute sickness (Moreau et al., 2005; O'Connor et al., 2009). Proteins also play a role in BCG-injected mice in relation to a microglia cell response (Yong et al., 2011). Twenty-one days after treatment, mice were still displaying effects of BCG-induced neuroprotective immune responses (Yong et al., 2011).

No study of the proteins in microglia cells has been completed a week after bacterial challenge. The objective of this study was to understand the impact of BCG challenge in the microglia cells during recovery from bacterial infection. A proteomic study was undertaken to detect proteins associated with recovery. Tandem mass spectrometry (MS/MS) analysis was used to detect peptides and proteins by comparing observed and theoretical spectra using the database software PEAKS. Proteins were compared between treatment groups and functional enrichment analysis using PANTHER. Results can help identify specific peptides and proteins expressed during a neurological, immune response. Protein information can be used for medicinal purposes aiming at development of approaches to minimize the impact of pathogen challenge in neurological functions.

2. Materials and Methods

2.1 Experiment and spectra collection

An experiment was conducted to identify differences in the brain microglia between BCG-inoculated and control mice. Male mice (C57 BL/6J strain) mice were injected into the peritoneum with TICE (Organon USA Inc., USA) Bacille Calmette-Guérin (10 mg; n= 12/group) at 21 weeks of age. The control group (n= 12/group) were injected with phosphate-buffered saline solution. Mice housing consisted of shoebox cages with corn-cobb bedding and wire tops. Mouse diets consisted of Tek Chow food and water ad libitum (Teklad 8640 chow, Harlan Laboratories, Indianapolis, IN, USA; Rodriguez-Zas et al., 2014). A reversed 12 hour light/dark cycle was used (lights turned OFF at 10am, ON at 10pm) because mice activity increases in darkness (O'Connor et al., 2009). Brain microglia cells were isolated from both groups one-week post injection. Mice were euthanized using CO_2 asphyxiation following protocols institutionally approved. Brain extraction consisted of perfusing a saline solution through the right ventricle to clear out blood, followed by removal of the brain. A combination of enzymatic dissociation using trypsin, Percoll float (removed myelin), and anti-CD-11b magnetic beads was used to isolate the microglia (Nikodemova and Watters, 2012). CD-11b is a marker on microglia for identification and separation of the homogenous mixture (Nikodemova and Watters, 2012). Primary mixed glial cultures

were established from brains of mice using the following techniques (Wang et al., 2010). After removal of the meninges, brains were mechanically minced and dissociated with 0.25% trypsin/0.5 mM EDTA. After inactivation of trypsin, the tissue suspension was passed through a 70 µm nylon cell strainer to remove debris. This cell suspension was centrifuged at 100g for 15 min. Cells were resuspended in 30% Percoll (GE Healthcare, Princeton, NJ) which allowed myelin separation by floating the myelin above pelleting cells during centrifugation. Purity of microglia was confirmed as >95% CD11b+ cells as verified by flow cytometry (Wang et al., 2010). The microglia from the mice were pooled into three groups (n= 3pools/group) such that each pool included 4 mice because of the limited sample that was obtained from individual mice. Microglia cells were suspended in TRIzol® Reagent for protein separation. (Hummon et al., 2007).

A bottom-up proteomics approach was used starting with the digestion of each microglia cell pool using trypsin. This enzymatic cleavage provides a complex peptide mixture (Matthesion and Bunkenborg, 2013). The bottom-up proteomic strategy for identification of proteins and their associated modifications generates thousands of MS/MS spectra. Following identification, MS/MS spectra are matched against a protein sequence database. Liquid chromatography allowed for simpler work flows and significantly faster analytical turnaround times (Grebe and Singh, 2011) relative to peptide identification. The liquid chromatography stage produces fractionated peptides and regulates the flow of peptides. Tandem MS spectrometry consisted of three parts: ion source, a mass analyzer, and detector. The bottom-up LTQ-FT approach allowed for fragmentation with detection involving b-ions that retain the charge on the N-terminus and y-ions that retained the charge on the C-terminus (Matthesion and Bunkenborg, 2013; Cottrell, 2011). Consideration of the ion fragments and match of the masses to a database of possible peptides enabled the identification of peptides and associated proteins. The mass to charge (m/z) value of the precursor ion and the set of masses corresponding to the resulting fragments enabled the uniquely identification of the sequence of the starting peptide, using the database software PEAKS.

2.2 Detection of peptides and proteins

A preliminary study evaluated the impact of fractionation and dynamic exclusion on the capability of database search programs to identify protein in the data set available. Fractionation is a strategy used to improve the sensitivity in mass spectrometry-based proteomic profiling (Jafari et al., 2012). The designated fractions 1 and 2 (Fr1/Fr2) represent two groups of proteins that have, on average, low and high sizes, respectively. The size was determined by how far each protein traveled across the gel. In Fr1, proteins traveled further in the gel; thus, they had a smaller average size than Fr2 proteins. Dynamic exclusion picks the most intense full (or the first stage) mass scan peak(s) for the succeeding (second

stage) MS/MS scans. In LC tandem mass spectrometry analysis, dynamic exclusion is used to minimize repeat selections of identical peptide precursors for collision-induced dissociation (CID). The dynamic exclusion approach was evaluated to assess the trade-off between the higher number of proteins identified and corresponding lower total spectral count per protein when dynamic exclusion was enabled or disabled.

The PEAKS v7 database search software package was used to detect peptides in the samples (Ma et al., 2003). Features of PEAKS7 that were critical for precise peptide and protein identification included: automatic faster *de novo* sequencing and validation as well as *de novo* residual local confidence scores (Zhang et al., 2011). *De novo* sequencing is the identification of a peptide that is carried out without the use of either a sequence database or spectral library (Cottrell, 2011). *De novo* sequencing derives the peptide sequence directly from the MS/MS spectrum, whereas a database search queries a sequence database for the best peptide to explain the peaks in the MS/MS spectrum (Zhang et al., 2012). Unlike the traditional database search approach, PEAKS incorporated *de novo* sequencing results to improve both the speed and accuracy of the database search. When the peptide sequence is in a database, *de novo* sequencing can greatly help improve the database-search based peptide identification since the match between the *de novo* sequence and the database sequence confirms the correctness of the identification (Zhang et al., 2012). The target database chosen for protein detection was UniProt (Magrane, 2011). Experimental MS/MS spectra input are compared against the target database of theoretical fragmentation patterns to find a statistically significant match and thus identify the peptide and series of peptides from a protein.

The parameter specification for *de novo* sequencing were: parent mass error tolerance (15.0 Da), fragment mass error tolerance (0.1 Da), and the enzyme specified was trypsin. The parent mass error tolerance is defined as the ion mass errors that PEAKS will allow for during the analysis, while the fragment mass error tolerance is defined as the error tolerance for the peaks in the tandem mass spectrum (Ma et al., 2003). The adequacy of the parameter specification was supported by the resulting mass to charge ratio (m/z) and mass error figures which were both concentrated around 0 ppm. The precursor mass error is calculated in ppm as 10^6 × (precursor mass - peptide mass) /peptide mass (Ma et al., 2003). Post-translational modifications (PTMs) considered in the search encompassed: carbamidomethylation, carboxymethyl, and oxidation due to their frequent appearance in peptides. The use of variable modifications increases the size of the computational search space for the *de novo* sequencing algorithm (Ma et al., 2003). In order to retain the authenticity of *de novo* sequencing, the maximum amount of variable PTMs of three was chosen.

PEAKS assigned a local confidence score (0% to 99%) to each amino acid in *de novo* sequences indicating the confidence level of each amino acid that was correctly sequenced (Zhang et al., 2011). Quality control filtering was performed prior to protein identification. To filter at peptide level, PEAKS additionally calculates the average local confidence (ALC) and total local confidence (TLC) scores for each peptide by averaging and summing up the local confidences (Zhang et al., 2011). Filtering of detected peptides that had low confidence included the *de novo* total local confidence (TLC) ≥ 3 and the *de novo* average local confidence (TLC divided by the peptide length, ALC)) $\geq 30\%$. The ALC value adjusted for the trend that larger proteins tend to have higher TLC values. Local confidence scores that were high (greater than 80%) and low (less than 60%) indicated the sequence accuracy.

PEAKS also assigned a linear discriminative function (LDF) that measures the quality of each peptide-spectrum match compiled. The LDF score uses the matching between the fragment ions and the peaks in the spectrum and factors such as the similarity between the *de novo* sequencing peptide and the database peptide (Zhang et al., 2012). The LDF score is converted to a P-value and the smaller the P-value, the less likely the peptide-spectrum match is due to a random match (Zhang et al., 2012). The statistical significance value of the peptide-spectrum match is a $-10*\log_{10}$(P-value). Peptides with a $-10*\log_{10}$(P-value) ≥ 15 and proteins with a $-10*\log_{10}$(P-value) ≥ 20 were considered for further study.

The PEAKS output consists of a table that lists protein identifications, coverage, peptides and *de novo* only peptides. With a well-calibrated instrument, precursor mass errors should center at 0 ppm across the range of m/z and was achieved in this experiment through the chosen parameters. The coverage output helps visualization of the mapping of supporting peptides and *de novo* tags representing the *de novo* only peptides with sequence tags matched to the protein (Ma et al., 2003). The coverage view also displays all identified sites with PTMs to assist with protein characterization at the amino acid level.

2.3 Differential detection analysis

Following the identification using PEAKS, a comparison of the detected peptides or proteins between treatment groups was undertaken with the use of a Statistical Analysis System (SAS, Cary, 2011). The SAS software was used to count the number of times a protein was detected in each treatment group. The count of peptides or proteins in the treatment group ranged from 0 (undetected in the three samples from a group) to 3 (detected in all three samples from a group). For each peptide (or protein) present in at least one sample, interest was in peptides or proteins that could be differentially detected between treatment groups. This study focused on peptides (and proteins) that differ between groups by a minimum count of two. This condition encompassed two scenarios: 1) peptides identified in at least two samples in one

group and no sample in the other group, and 2) peptides identified in three samples in one group and one sample in the other group.

2.4 Functional enrichment analysis

Annotation of proteins relied on two databases: Uniprot (Magrane, 2011) and Mouse Genome Informatics (MGI; Blake et al., 2014) databases. Uniprot provides a resource of protein sequence and functional information for multiple species. Mouse Genome Informatics is an integrated mouse genetic, genomic, and other biological databases. The Mouse Genome Informatics was used because it provided Gene Ontology (GO) terms that were further needed for enrichment classification.

Subsequent to the identification of proteins that were differentially detected between treatment groups, the enrichment of functional categories among those molecules was undertaken. Enrichment analysis is a means to characterize biological attributes in a given a data set. The Protein ANalysis THrough Evolutionary Relationships (PANTHER) Classification System (Mi et al., 2013) web tool was used to identify protein class and other categories including Gene Ontology (GO) molecular function (MF), biological process BP) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway (Nikolsky and Bryant, 2009). For protein class, terms are taken from the PANTHER Protein Class ontology. Protein class derives from the PANTHER/X ontology and represents common grouping terms for families and subfamilies (Mi et al., 2013). The GO MF is the function of the protein by itself or with directly interacting proteins at a biochemical level. The GO BP is the function of the protein in the context of a larger network of proteins that interact to accomplish a process at the level of the cell or organism. The information in KEGG pathway is similar to the biological process, but a pathway also explicitly specifies the relationships between the interacting molecules.

The three enrichment categories (GO BP, GO MF, and KEGG pathway) were studied in PANTHER. Enriched pathways among differentially detected proteins with a differential detection of two, both BCG and control mice, were uploaded into PANTHER. The differentially detected proteins were uploaded using the UniProt Accession annotation. PANTHER contains 7180 protein families, each with a phylogenetic tree used to relate organisms, such as *Mus musculus*. The core of the system is a large collection of phylogenetically defined protein families and subfamilies generated by computational algorithms and curated by expert biologists using an extensive software system for associating ontology terms (Nikolsky and Bryant, 2009). Each family and subfamily are represented as a hidden Markov model, which can be used to classify new sequences to an existing subfamily. The output of PANTHER enrichment classification includes statistics and graphic views (such as bar graphs).

3. Results and Discussion

The results from the preliminary study that focused on the impact of fractionation and dynamic exclusion demonstrated that both factors impacted analysis. Differences were observed in the number of proteins detected between Fr1 and Fr2. In Fr1, the number of proteins totaled to 306, while in Fr2 the number of proteins totaled to 124. When dynamic exclusion was enabled, fewer peptides were identified. This was expected because the repeated m/z- value was excluded from the CID analysis, decreasing the amount of peptides identified. The parameters set were as follows: parent mass error tolerance (15.0 Da), fragment mass error tolerance (0.1 Da), and the enzyme specified was trypsin. The accuracy of these parameters was supported by the resulting mass to charge ration (m/z) and mass error figures which were both concentrated around 0 ppm. This study demonstrated that factors such as fractionation and dynamic exclusion when combined with user-specific parameters impact protein identification.

Peptides and proteins were detected in the microglia of BCG-treated and control mice. The detected peptides and proteins were compared between treatment groups. Functional enrichment analysis of the differentially detected proteins with a differential detection of two (BCG and control mice) offered insights into the processes associated with microglia response to BCG treatment after seven days. The following discussion focuses on the reported roles of detected proteins in neuroinflammatory and related processes.

3.1 Sample characterization

The number of tandem mass spectrometry scans ranged from 9934 to 10780 for the BCG-treated group and from 8905 to 10809 for the control group. Table 2.1 summarizes the number of protein and peptides detected in the three BCG and control sample pools. Furthermore, Table 2.1 illustrates how the control mice, overall, have higher total and unique total proteins and peptide-spectrum matches compared to the BCG-treated mice. After comparison between the two treatment groups, analysis showed a total of one protein and 11 peptides common within the BCG group and 66 proteins and 88 peptides common within the control group. The higher number of proteins and peptides detected in the control suggests that BCG impacts the production of proteins at the transcription or translation stages.

3.2 Differential detection

Table 2.2 summarizes the number of proteins and peptides that were differentially detected (two or more samples difference) between the BCG and control groups and the number of differing samples (Control – BCG). No proteins or peptides were found for the comparison between control vs. BCG group of 0 vs. 3 or 1 vs. 3. Proteins that differed in the detection in at least two samples between treatment groups were

considered differentially detected. The use of detection or non-detection within each sample is a qualitative characterization of the protein detection. Like with any arbitrary threshold, subsequent functional and pathway analyses may be altered due to more or less stringent differential thresholds which could alter the identification of enriched functional categories. Table 2.3 lists the identification of the proteins that exhibited differential detection between the BCG and the control groups. The differential detected proteins were as expected between the control and the BCG groups due to their relationships with similar studies on neurological activity and inflammation response to BCG-challenged mice (Rodriguez-Zas et al., 2014; O'Connor et al., 2009).

Proteins that were observed in all control samples and were not detected in the BCG-treated samples included: F-actin-capping protein subunit alpha-2 (P47754) and Histone H2A type 2-C (Q64523; Table 2.3). F-actin-capping protein subunit alpha-2 acts as a structural component of the cytoskeleton and Histone H2A type 2-C is a DNA binding protein that is a core component of the nucleosome (Magrane, 2011). Previous studies demonstrated that cytoskeletal networks made up of F-actin microfilaments and tubulin microtubules are indispensable for normal neurite development and regenerative responses to injury and neurodegenerative stimuli, while cytoskeletal abnormalities characterize neurodegenerative diseases (Suszynska-Zajczyk et al., 2014). Significant activation of microglia cells occur in the early stages of the neurodegenerative disease referred to as Alzheimer's disease (E. McGeer and P. McGeer, 2003). Specifically, symptoms of cognitive impairment and memory loss that are associated with the Alzheimer's disease have increasing evidence showing that the stability of dendritic spines and actin remodeling participating in the pathology of the disease (Suszynska-Zajczyk et al., 2014). Actin plays a physiological role in controlling neurons and a pathophysiological role in neuroinflammation and stress responses (Correa and Eales, 2012). The presence of F-actin-capping protein subunit alpha-2 in control mice is due to the primary role in up keeping in stability for cytoskeleton formation. Cytoskeleton formation is necessary for cell shape maintenance, motility (movement) of the cell and motility of organelles within a cell. Extracellular core histones, such as Histone H2A type 2-C, are not neurotoxic. Histones are highly basic proteins that are present in the nucleus where they form major components of eukaryotic chromatin and function to regulate transcription (Gilthorpe et al., 2013). Studies have shown that Histone H1, but not core histones, selectively kills embryonic cortical neurons and causes microglia to become reactive (Gilthorpe et al., 2013). Therefore, the core histone, H2A type 2-C protein, is present in control mice, but does not cause microglia cell reaction.

Proteins present in two more samples in the control treatment group relative to the BCG group included: Alpha-enolase (P17182) and Destrin (Q9R0P5; Table 2.3). Alpha-enolase interacts in the neuronal plasma membrane and promotes activation, and also stimulates immunoglobulin production; Destrin acts

independently as an actin-depolymerizing protein (Magrane, 2011). Alpha-enolases binds cytoskeletal and chromatin structures indicate that enolase may play a crucial role in transcription and a variety of pathophysiological processes (Pancholi, 2001). Enolase is expressed on the surface of brain cells and is reasonable to expect anti-enolase circulating autoantibodies may play an important role in the development of neurological disorders (Pancholi, 2001). Alpha-enolase has been detected on the surface of hematopoietic cells such as monocytes, T cells and B cells, and neuronal cells (Diaz-Ramos et al., 2012). Alpha-enolase also binds with high affinity to other glycolytic enzymes: pyruvate kinase, phosphoglycerate mutase, which are adjacent to enolase in the glycolytic pathway, and to aldolase, which is known to associate with cytoskeletal proteins (Diaz-Ramos et al., 2012). Alpha-Enolase has been reported as a strong plasminogen receptor within the brain and is upregulated in the Alzheimer's disease brain and has been proposed as a promising therapeutic target for this disease (Diaz-Ramos et al., 2012). Previous studies such as these demonstrate that Alpha-enolase play a crucial role in neurological development and activity. Destrin regulates the balance between two pools of cytoplasmic actin, F-actin and G-actin (Verdoni et al., 2008). When Destrin proteins were disrupted, inflammation associated terms became over-represented (Verdoni et al., 2008). Overall, expression of genes related to the structure of the cytoskeleton is strongly altered. Destrin mutant mice serve as models to further investigate cell signaling mechanisms and transcriptional regulation affected by actin dynamics. When unaffected, the presence of destrin plays a primary role as a cytoskeletal component; thus, abundance in the control group is expected.

The proteins that were present in two BCG-treated samples and absent from all three control samples were Chitinase-like protein 3 (O35744) and Vesicle-associated membrane protein-associated protein A (Q9WV55; Table 2.3). Chitinase-like protein 3 has chemotactic activity for T-lymphocytes and plays a role in inflammation. Chitinases/chitinase-likeproteins (C/CLPs) play a pivotal role in both innate and adaptive immune responses (Lee et al., 2011). Chitin is a stimulator of innate immune responses and subsequent tissue injury. Chitin stimulates inflammatory cytokines, including IL-1β and interferon-γ (IFN-γ; Lee et al., 2011). Chitin-free antigens are also able to induce the innate reactions that, in turn, lead to adaptive immune responses. Overall, chitinase-like proteins play roles that (1) trigger a response to infections and/or antigen challenge; (2) function directly as chemotactic agents or indirectly by inducing other chemokines that attract T cells to sites of infection; and/or (3) modulate tissue inflammation, immunity, and/or remodeling (Elias et al., 2005). Differential transcription profiles have been linked to disease-specific genes, notably including those encoding proteins involved in vesicle trafficking (Smith et al., 2009). Vesicle-associated membrane protein-associated protein A plays a role in vesicle trafficking with the possibility of a cytoskeleton association (Magrane, 2011) Identification of Vesicle-associated

membrane protein-associated protein A illustrates involvement in vesicle trafficking and cytoskeletal organization which suggests that the secretion of inflammation-related cytokines may lie in posttranscriptional processing (Smith et al., 2009). Together, these two proteins share a relationship that is also connected by the similarity of inflammation that was expected due to BCG inoculation and that may still be present one week after inoculation.

Differential abundance was considered for peptides within proteins as well. A protein that was detected in all three in the control group and was not detected in the BCG group included Histone H2B type 1-M (P10854). Histone H2B type 1-M protein is a core component of nucleosome and plays a central role in transcription regulation and DNA repair (Magrane, 2011). Previous studies have shown that AMP-activated protein kinase (AMPK) activates transcription through direct association with chromatin and phosphorylation of Histone H2B (Bungard et al., 2010). AMPK is a highly conserved protein kinase that plays a role in inflammation and immunity and provides to be an enticing prospect of new therapeutic approaches for inflammatory diseases O'Neill and Hardie, 2013). AMPK is directly associated with chromatin to regulate transcriptional programs required to survive a wide variety of metabolic and environmental stresses (Bungard et al., 2010). Furthermore, studies have shown that activation of AMPK in cells of the immune system promote the switch from a pro-inflammatory to an anti-inflammatory phenotype partly by causing a switch away from rapid glucose uptake and glycolysis towards mitochondrial oxidative metabolism, including fatty-acid oxidation (O'Neill and Hardie, 2013). This protein in control mice plays a role in maintaining cellular homeostasis and is readily available to respond to various metabolic stresses including involvement with the immune system; therefore, detection in control mice is expected.

Among the proteins present in two more samples in the control group relative to the BCG group is myelin basic protein (F7A0B0). Myelin basic protein is a structural constituent on the myelin sheath (Magrane, 2011). Myelin has a direct association with inflammation, specifically, inflammation associated with CNS demyelination (Setzu et al., 2006). Multiple sclerosis, the most widely occurring demyelinating disease in humans, is an immune-mediated disease in which various immune cells target CNS specific antigens leading to a destructive inflammatory environment that damages both myelinating cells and axons (Setzu et al., 2006). Formation of compact myelin is required for maturation of the axonal cytoskeleton (Brady et al., 1999). Myelin deficiencies in the CNS lead both to local consequences on the axon cytoskeleton (i.e., reduced neurofilament phosphorylation) and to altered gene expression in the cell body (Brady et al., 1999). Due to playing a crucial role in the cytoskeletal composition and the CNS, myelin was detected in control mice.

Finally, SET (Q9EQU5) was present in two BCG-treated samples and absent from all three control samples. SET is a multitasking protein, involved in transcription, nucleosome assembly and histone chaperoning that binds to $p35^{nck5a}$ and $p35^{nck5ai}$; both proteins are activators of the neuronal Cdk5 (Magrane, 2011; Canela et al., 2003). Cdk5 promotes neuronal death under conditions of stress (Huang et al., 2010). Deregulated Cdk5 is also implicated in the pathogenesis of neurodegenerative disorders such as Alzheimer's disease and Parkinson's disease (Huang et al., 2010). The detection of SET in the BCG-treated mice suggests evidence of current or past neuronal stress in this group of mice.

The proteins differentially detected illustrate the relationship between neuronal activity and inflammation one week after challenge. The time from challenge to sample may be a reason for the absence of differentially detected cytokines in this study.

3.3 Functional enrichment analysis

Enrichment of functional categories among the differentially detected proteins using Table 2.3 was undertaken with PANTHER. Figure 2.1 and Table 2.4 depict proteins classes as PANTHER/X ontology terms. Figure 2.2 and Table 2.5 depict enriched GO MF categories among differentially detected proteins. Figure 2.3 and Table 2.6 depict enriched GO BP among differentially detected proteins. Finally, Figure 2.4 and Table 2.7 depict KEGG pathway among differentially detected proteins. The focus of the interpretation of the enrichment analysis will be on the pathway processes because the detected MF and BP are commonly seen in living organisms.

Among the differentially detected proteins, enriched pathways (Figure 2.4 and Table 2.7) included Huntington disease (P00029) and Glycolysis (P00024). In relation to Huntington disease (P00029), the disease is a fatal, neurodegenerative disorder. The causative mutation leads to widespread brain neurodegeneration, with cell loss mainly in the striatum and cerebral cortex, although neuronal abnormalities are also found in many other brain regions (Zuccato and Cattaneo, 2007). Cognitive dysfunction in manifest Huntington disease is characterized by impairment in attention, planning and memory, visuospatial abilities, and language (Reading et al., 2004). Huntington disease affects neurotrophic factors, which are essential contributors to the survival of peripheral and CNS neurons. The reduced presence in diseased brains indicates that neurotrophic factors play a role in various neurological disorders (Zuccato and Cattaneo, 2007). Microglia cell activation may play a role in the pathogenesis of Huntington's disease as demonstrated in presymptomatic gene carriers (Tai et al., 2007). Widespread microglia activation in presymptomatic Huntington disease gene carriers is associated with subclinical striatal neuronal loss of dopamine D2 receptor binding (Tai et al., 2007). Findings such as these indicate that microglia cell activation is an early event in the pathogenic processes of Huntington disease and is

associated with the subclinical progression of the disease (Tai et al., 2007). The identification of Huntington pathway enrichment in this study is consistent with the altered processes previously associated with this disorder.

The glycolysis pathway has been linked to neurodegenerative disorders associated with inflammation and microglia as well (Volobоueva et al., 2013). Glycolysis is the breakdown of glucose enzymes, which results in the release of energy (i.e. ATP) and pyruvic acid. Microglia cell activation triggers a metabolic reprogramming based on an increased glucose uptake and a strengthening of anaerobic glycolysis (Gimeno-Bayon et al., 2014). Overall, microglia adaptation enhanced glycolytic demand with a high lactate production and high metabolic rate (Gimeno-Bayon et al., 2014). Furthermore, lipopolisaccharide-induced inflammatory response of microglia cells BV-2 demonstrate that LPS-induced activation promotes significant metabolic changes increasing glycolysis (Volobоueva et al., 2013). Overall, increased glycolytic activity contribute to pro-inflammatory responses and microglia cell activation. The identification of Glycolysis enrichment in this study is consistent with the altered processes previously associated with this metabolic pathway.

4. Conclusion

Tandem mass spectra analysis enabled the detection of peptides and proteins in the microglia of mice infected with BCG versus control. Analysis of mass spectra was followed by protein and peptide detection and enrichment analysis. A number of proteins associated with inflammation response in the brain were differentially detected between treatment groups. These findings offer insights into long lasting response to inflammation during recovery. The timing of the sampling could explain the absence of differential detection of cytokines and other proteins associated with immediate response to bacterial challenge. Alpha-enolase and myelin basic protein were detected only in control mice. These detections are consistent with the role of the former protein in neurological development and activity and the role of the latter in cytoskeletal composition in the CNS. Chitinase-like protein 3 and SET were more abundant in BCG-treated than control mice. These results are consistent with the role of the former protein in stimulating the innate immune responses subsequent tissue injury and of the latter protein promoting neuronal death under conditions of stress. Enrichment analysis identified pathways related to neurodegenerative disorders and metabolism that have been associated with inflammation and microglia, consistent with the challenge studied. Discussion focused on relevant roles of detected proteins in neuroinflammatory and related processes; however the proteins also participate on other processes. In summary, this study added proteomic insights into understanding the impact of proteins related to recovery in BCG-treated mice using tandem mass spectrometry.

5. Tables and Figures

Table 2.1 Number of total and unique protein and peptides detected in BCG and control mice

Type	Protein	*Peptide-Spectrum Match (PSM)
Total:		
BCG	463	1,645
Control	733	2,344
Unique total:		
BCG	194	431
Control	239	438

* multiple PSMs result from a peptide matched to many spectra

Table 2.2 Number of proteins and peptides that were differentially detected between control and BCG mice

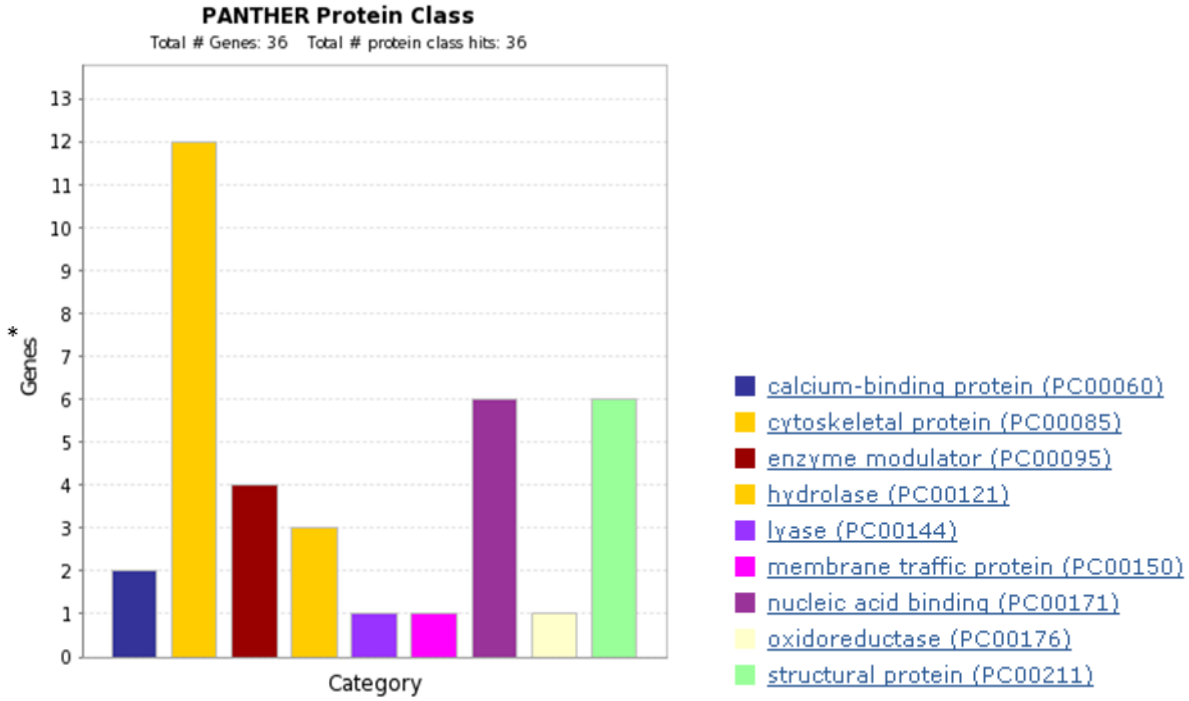
Type	Control[1] - BCG	Differential abundance	Number
Protein	3 - 0	3	3
Protein	3 - 1	2	16
Protein	2 - 0	2	30
Protein	0 - 2	-2	2
Peptide	3 - 0	3	21
Peptide	3 - 1	2	63
Peptide	2 - 0	2	102
Peptide	0 - 2	-2	10

[1] Number of control samples – number of BCG-treated samples

Table 2.3 Proteins and their differential detection (control vs. BCG)

Uniprot Accession	Differential Detection	Uniprot Accession	Differential Detection
P47754	3	E9Q2G1	2
Q64523	3	P68033	2
Q6IFZ6	3	E9PYP8	2
D3YYE1	2	Q9WVJ5	2
O35381	2	Q7TPM0	2
D3Z7M9	2	Q9CYJ8	2
P17182	2	Q9R0P5	2
P62204	2	P05064	2
Q3UKW2	2	A6ZI44	2
P16858	2	P62874	2
F8WIX8	2	P62880	2
Q6GSS7	2	E9QKR0	2
Q6ZWY9	2	D3YZX3	2
P10854	2	Q8CGP7	2
P02535	2	P54071	2
A2A513	2	P97822	2
S4R1W1	2	Q60605	2
P97822	2	Q61781	2
E9Q0X5	2	Q9QWL7	2
P04104	2	Q6NXH9	2
Q8BGJ5	2	Q922I7	2
Q8CB58	2	Q9CWF2	2
Q60605	2	O35744	-2
A2AFI4	2	Q9WV55	-2
Q7TMM9	2		

Figure 2.1 PANTHER/X ontology



* genes stand for protein

Table 2.4 PANTHER/X ontology

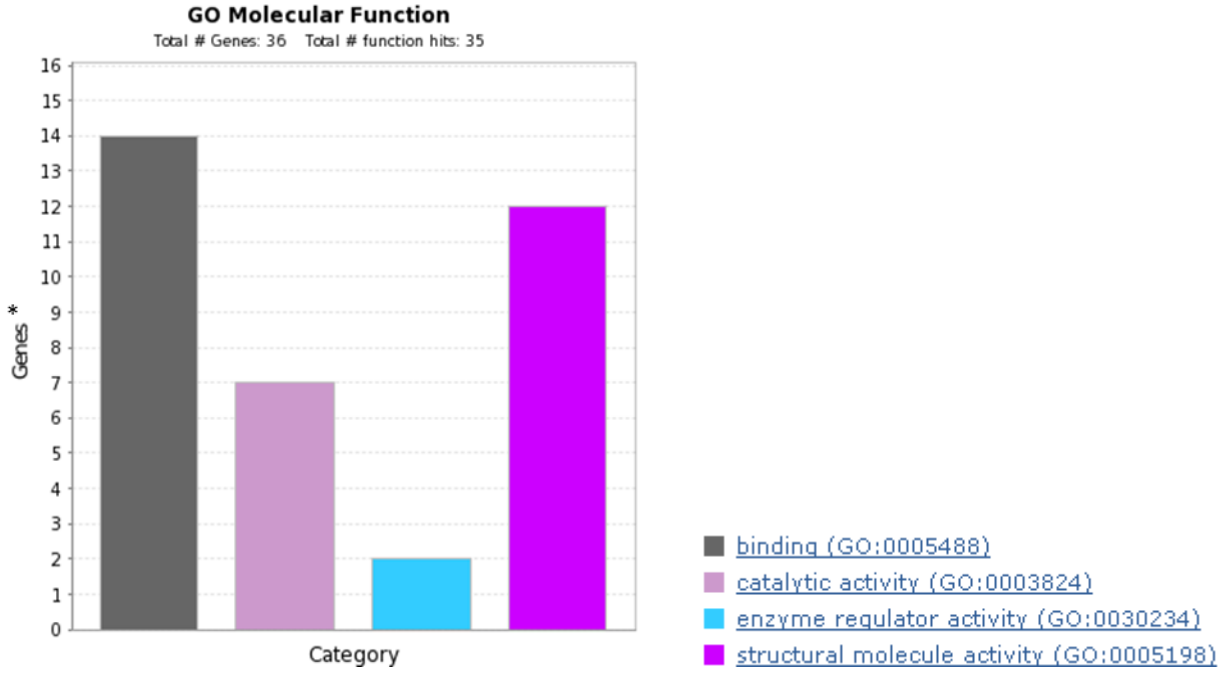
Category name (Accession)[a]	# genes*	Percent of gene hit against total # genes	Percent of gene hit against total # Protein Class hits
membrane traffic protein (PC00150)	1	2.80%	2.80%
hydrolase (PC00121)	3	8.30%	8.30%
oxidoreductase (PC00176)	1	2.80%	2.80%
enzyme modulator (PC00095)	4	11.10%	11.10%
lyase (PC00144)	1	2.80%	2.80%
nucleic acid binding (PC00171)	6	16.70%	16.70%
calcium-binding protein (PC00060)	2	5.60%	5.60%
cytoskeletal protein (PC00085)	12	33.30%	33.30%
structural protein (PC00211)	6	16.70%	16.70%

[a] catergory defined as PANTHER/x ontology

symbol denotes number of

* genes stand for protein

Figure 2.2 Gene Ontology Molecular Function among differentially detected proteins



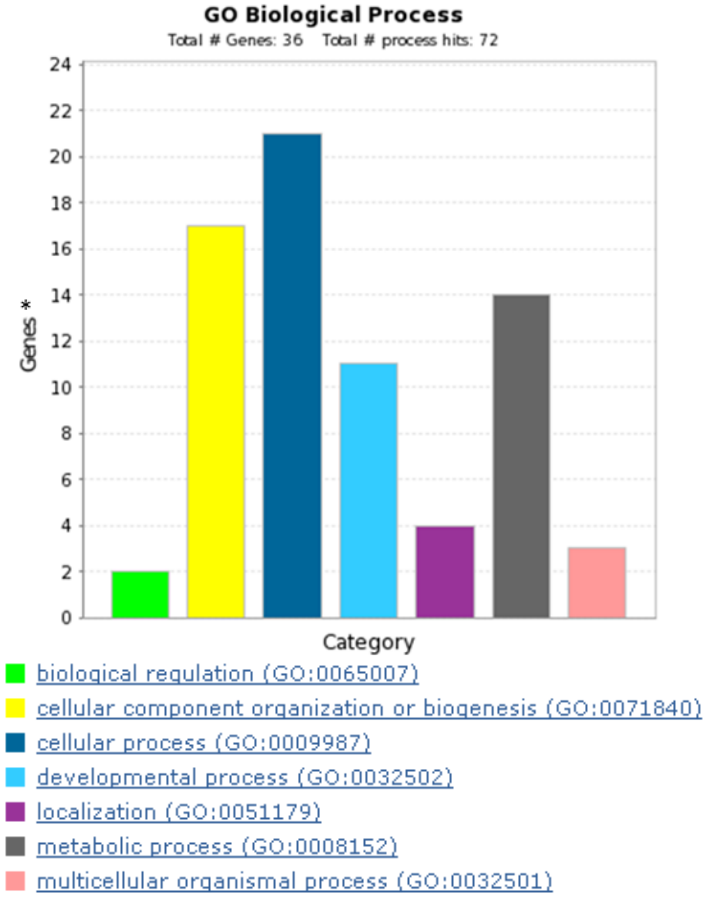
* genes stand for protein

Table 2.5 Gene Ontology Molecular Function among differentially detected proteins

Category name (Accession)[a]	# genes*	Percent of gene hit against total # genes	Percent of gene hit against total # Process hits
binding (GO:0005488)	14	38.90%	40.00%
enzyme regulator activity (GO:0030234)	2	5.60%	5.70%
structural molecule activity (GO:0005198)	12	33.30%	34.30%
catalytic activity (GO:0003824)	7	19.40%	20.00%

[a] category defined as GO MF
GO denotes Gene Ontology
symbol denotes number of
* genes stand for protein

Figure 2.3 Gene Ontology Biological Process among differentially detected proteins



* genes stand for protein

Table 2.6 Gene Ontology Biological Process among differentially detected proteins

Category name (Accession)[a]	# genes*	Percent of gene hit against total # genes	Percent of gene hit against total # Process hits
developmental process (GO:0032502)	11	30.60%	15.30%
cellular process (GO:0009987)	21	58.30%	29.20%
multicellular organismal process (GO:0032501)	3	8.30%	4.20%
metabolic process (GO:0008152)	14	38.90%	19.40%
biological regulation (GO:0065007)	2	5.60%	2.80%
cellular component organization or biogenesis (GO:0071840)	17	47.20%	23.60%
localization (GO:0051179)	4	11.10%	5.60%

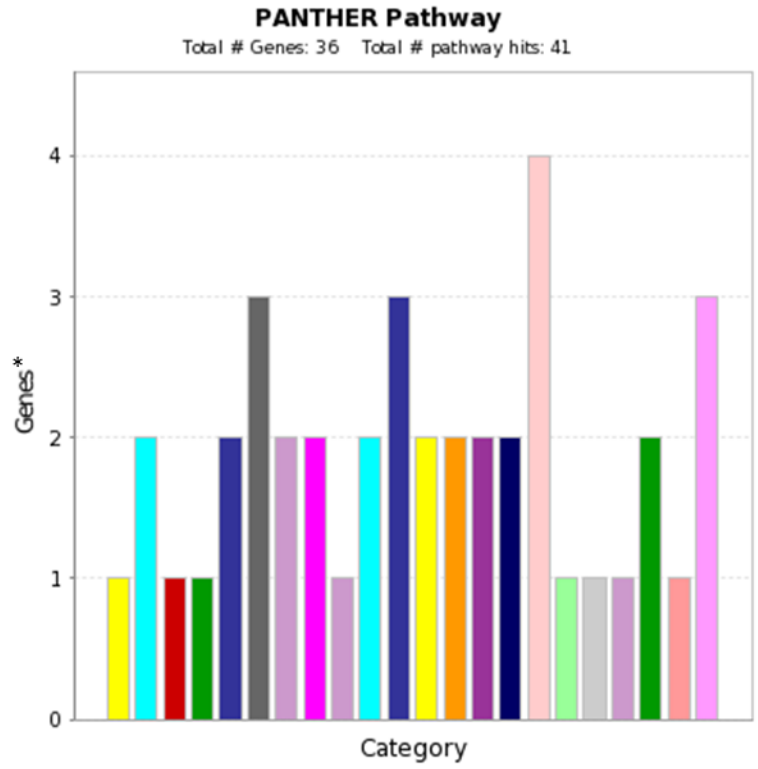
[a] category defined as GO BP

GO denotes Gene Ontology

symbol denotes number of

* genes stand for protein

Figure 2.4 KEGG Pathway processes among differentially detected proteins



- Alzheimer disease-presenilin pathway (P00004)
- Angiotensin II-stimulated signaling through G proteins and beta-arrestin (P05911)
- B cell activation (P00010)
- Cadherin signaling pathway (P00012)
- Cortocotropin releasing factor receptor signaling pathway (P04380)
- Cytoskeletal regulation by Rho GTPase (P00016)
- Dopamine receptor mediated signaling pathway (P05912)
- Endogenous cannabinoid signaling (P05730)
- Fructose galactose metabolism (P02744)
- GABA-B receptor II signaling (P05731)
- Glycolysis (P00024)
- Gonadotropin releasing hormone receptor pathway (P06664)
- Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway (P00026)
- Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway (P00027)
- Heterotrimeric G-protein signaling pathway-rod outer segment phototransduction (P00028)
- Huntington disease (P00029)
- Inflammation mediated by chemokine and cytokine signaling pathway (P00031)
- Nicotine pharmacodynamics pathway (P06587)
- Nicotinic acetylcholine receptor signaling pathway (P00044)
- PI3 kinase pathway (P00048)
- T cell activation (P00053)
- Wnt signaling pathway (P00057)

* genes stand for protein

Table 2.7 KEGG Pathway processes among differentially detected proteins

Category name (Accession)[a]	# genes[*]	Percent of gene hit against total # genes	Percent of gene hit against total # Process hits
Cytoskeletal regulation by Rho GTPase (P00016)	3	8.30%	7.30%
Huntington disease (P00029)	4	11.10%	9.80%
Wnt signaling pathway (P00057)	3	8.30%	7.30%
Glycolysis (P00024)	3	8.30%	7.30%

[a] category defined as KEGG Pathway

GO denotes Gene Ontology

symbol denotes number of

* genes stand for protein

Chapter 3: Future Studies and Conclusions

The completed study focused on comparing the proteins and peptides detected in the microglia of BCG-treated and control mice, through database search and enrichment analysis. The results were supported by previous studies. The success of this study suggests the opportunity to use the same approach with a different focus. Specifically, challenging mice with a virus such as using an influenza vaccine could be studied.

Due to its popularity during flu season, which peaks between winter months (December and February), the vaccine has become prevalent. Seasonal flu is a contagious respiratory illness caused by an influenza virus. The flu ranges in its ability to be spread due to its constantly changing virus forms and ability to adapt. The flu can be caught by coughing and sneezing or as easily as touching a surface that has the flu virus on it and then touching your mouth, eyes, or nose. The flu is known to spread quickly between people and causes mild to severe illness, and in some cases can lead to death. Flu symptoms include a fever, cough/sore throat, runny nose, headaches, chills, fatigue, nausea and/or diarrhea. Although the flu can affect all ages, it is most dangerous to children and elderly. The influenza vaccine is increasingly distributed without charge throughout the United States, but it is not received by all individuals.

No study has focused on the effects of the influenza vaccine on proteins and peptides in the mice through detection and enrichment analysis. The study can focus on the effect on microglia as well with a comparison of influenza vaccine-treated vs. control mice. Furthermore, there are multiple types of influenza vaccines available. Common vaccines include the flu shot and flu nasal spray. The flu shot contains killed (inactive) viruses while the nasal spray flu vaccine uses live, weakened flu viruses. With the option of different vaccines, it is possible to compare the vaccines against one another, as well as the vaccines against a control group. With the increase of choices, more results would be available for analysis.

Furthermore, the effect of the vaccine can be tested among mice of different ages. The vaccine is commonly given to children (older than six months) and is effective. By comparing vaccine distribution between young mice vs. older mice, protein and peptide detection may vary. By incorporating the influenza vaccine with an array of variables, protein and peptide that are present can be understood based on molecular function, biological process, and pathway relationships as well. Once the influenza vaccine is tested, the expected proteins detected would be made up mostly of antibodies that provide protection against infection. Also, proteins that elicit direct immune responses against pathogens would be detected as well. Previous studies have shown that the immune response to infection by influenza virus in mice

includes alterations in the kinetics of the antibody response and the inflammatory response (Kiecolt-Glaser et al., 1996).

In conclusion, protein and peptide detection in microglia of mice (or other cells) can help understanding the relationship between a vaccinated vs. control mice. With studies such as this, understanding the relationship between the protein and effects of the vaccine can be understood for medicinal purposes.

References

Abolins, SR, MJ Pocock, JC Hafalla, EM Riley, ME Viney. Measures of immune function of wild mice, Mus musculus. Mol Ecol. 2011 Mar;20(5):881-92.

Andreev, VP, L Li, L Cao, Y Gu, T Rejtar, SL Wu, BL Karger. A New Algorithm Using Cross-Assignment for Label-Free Quantitation with LC/LTQ-FT MS. J Proteome Res. 2007 Jun;6(6):2186-94.

Apweiler, R, A Bairoch, CH Wu, WC Barker, B Boeckmann, S Ferro, E Gasteiger, H Huang, R Lopez, M Magrane, MJ Martin, DA Natale, C O'Donovan, N Redaschi, LS Yeh. UniProt: the Universal Protein knowledgebase. Nucleic Acids Res. 2004 Jan 1;32(Database issue):D115-9.

Akhtar, MN, BR Southey, PE Andrén, JV Sweedler, SL Rodriguez-Zas. Evaluation of database search programs for accurate detection of neuropeptides in tandem mass spectrometry experiments. J Proteome Res. 2012 Dec 7;11(12):6044-55.

Akhtar, MN, BR Southey, PE Andrén, JV Sweedler, SL Rodriguez-Zas. Identification of best indicators of peptide-spectrum match using a permutation resampling approach. J Bioinform Comput Biol. 2014 Oct;12(5):1440001.

Ashburner, M, CA Ball, JA Blake, D Botstein, H Butler, JM Cherry, AP Davis, K Dolinski, SS Dwight, JT Eppig, MA Harris, DP Hill, L Issel-Tarver, A Kasarskis, S Lewis, JC Matese, JE Richardson, M Ringwald, GM Rubin, G Sherlock. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000 May;25(1):25-9.

Berman, HM, J Westbrook, Z Feng, G Gilliland, TN Bhat, H Weissig, IN Shindyalov, PE Bourne. The Protein Data Bank. Nucleic Acids Res. 2000 Jan 1;28(1):235-42.

Berman, H, K Henrick, H Nakamura. Announcing the worldwide Protein Data Bank. Nat Struct Biol. 2003 Dec;10(12):980.

Bilbo, SD and JM Schwarz. The immune system and developmental programming of brain and behavior. Front Neuroendocrinol. 2012 Aug;33(3):267-86.

Blake, JA, CJ Bult, JT Eppig, JA Kadin, JE Richardson; The Mouse Genome Database Group. The Mouse Genome Database: integration of and access to knowledge about the laboratory mouse. Nucleic Acids Res. 2014 Jan;42(Database issue):D810-7.

Boche, D, VH Perry, JAR Nicoll. Review: activation patterns of microglia and their identification in the human brain. Neuropathol Appl Neurobiol. 2013 Feb;39(1):3-18.

Brady, ST, AS Witt, LL Kirkpatrick, SM de Waegh, C Readhead, PH Tu, VM Lee. Formation of compact myelin is required for maturation of the axonal cytoskeleton. J Neurosci. 1999 Sep 1;19(17):7278-88.

Bungard, D, BJ Fuerth, PY Zeng, B Faubert, NL Maas, B Viollet, D Carling, CB Thompson, RG Jones, SL Berger. Signaling kinase AMPK activates stress-promoted transcription via histone H2B phosphorylation. Science. 2010 Sep 3;329(5996):1201-5.

Canela, N, A Rodriguez-Vilarrupla, JM Estanyol, C Diaz, MJ Pujol, N Agell, O Bachs. The SET protein regulates G2/M transition by modulating cyclin B-cyclin-dependent kinase 1 activity. J Biol Chem. 2003 Jan 10;278(2):1158-64. Epub 2002 Oct 28.

Cappadona, S, PR Baker, PR Cutillas, AJ Heck, B van Breukelen. Current challenges in software solutions for mass spectrometry-based quantitative proteomics. Amino Acids. 2012 Sep;43(3):1087-108.

Correa, SAL and KL Eales. The Role of p38 MAPK and Its Substrates in Neuronal Plasticity and Neurodegenerative Disease. J Signal Transduct. 2012;2012:649079.

Cottrell, JS. Protein identification using MS/MS data. J Proteomics. 2011 Sep 6;74(10):1842-51.

Croft, D, G O'Kelly, G Wu, R Haw, M Gillespie, L Matthews, M Caudy, P Garapati, G Gopinath, B Jassal, S Jupe, I Kalatskaya, S Mahajan, B May, N Ndegwa, E Schmidt, V Shamovsky, C Yung, E Birney, H Hermjakob, P D'Eustachio, L Stein. Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res. 2011 Jan;39(Database issue):D691-7.

Dantzer, R. Cytokine, Sickness Behavior, and Depression. Immunol Allergy Clin North Am. 2009 May; 29(2): 247–264.

Dıaz-Ramos, A, A Roig-Borrellas, A García-Melero, R López-Alemany. α-Enolase, a multifunctional protein: its role on pathophysiological situations. J Biomed Biotechnol. 2012;2012:156795.

Elias, JA, RJ Homer, Q Hamid, CG Lee. Chitinases and chitinase-like proteins in TH2 inflammation and asthma. J Allergy Clin Immunol. 2005 Sep;116(3):497-500.

Ellrichmann, G, C Reick, C Saft, RA Linker. The role of the immune system in Huntington's disease. Clin Dev Immunol. 2013;2013:541259.

Geer, LY, A Marchler-Bauer, RC Geer, L Han, J He, S He, C Liu, W Shi, SH Bryant. The NCBI BioSystems database. Nucleic Acids Res. 2010 Jan; 38(Database issue):D492-6.

Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. 2004 Jan 1;32(Database issue):D258-61.

Gilthorpe, JD, F Oozeer, J Nash, M Calvo, DLH Bennett, A Lumsden, A Pini. Extracellular histone H1 is neurotoxic and drives a pro-inflammatory response in microglia. F1000Res. 2013 Jul 8;2:148.

Gimeno-Bayon, J, A Lopez-Lopez, MJ Rodrıguez, N Mahy. Glucose pathways adaptation supports acquisition of activated microglia phenotype. J Neurosci Res. 2014 Jun;92(6):723-31.

Grebe, SKG and RJ Singh. LC-MS/MS in the Clinical Laboratory – Where to From Here?. Clin Biochem Rev. 2011 Feb;32(1):5-31.

Harry, GJ and AD Kraft. Microglia in the developing brain: a potential target with lifetime effects. Neurotoxicology. 2012 Mar;33(2):191-206.

Harry, GJ. Microglia During Development and Aging. Pharmacol Ther. 2013 Sep;139(3):313-26.

Huang, DW, BT Sherman, RA Lempicki. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44-57.

Huang, E, D Qu, Y Zhang, K Venderova, ME Haque, MWC Rousseaux, RS Slack, JM Woulfe, DS Park. The role of Cdk5-mediated apurinic/apyrimidinic endonuclease 1 phosphorylation in neuronal death. Nat Cell Biol. 2010 Jun;12(6):563-71.

Hummon, AB, SR Lim, MJ Difilippantonio, T Ried. Isolation and solubilization of proteins after TRIzol extraction of RNA and DNA from patient material following prolonged storage. Biotechniques. 2007 Apr;42(4):467-70, 472.

Jafari, M, V Primo, GB Smejkal, EV Moskovets, WP Kuo, AR Ivanov. Comparison of in-gel protein separation techniques commonly used for fractionation in mass spectrometry-based proteomic profiling. Electrophoresis. 2012 Aug;33(16):2516-26.

Kanehisa, M and S Goto. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000 Jan 1;28(1):27-30.

Kanehisa, M, S Goto, Y Sato, M Furumichi, M Tanabe. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 2012 Jan;40(Database issue):D109-14.

Kelley, KW, RM Bluthe, R Dantzer, JH Zhou, WH Shen, RW Johnson, SR Broussard. Cytokine-induced sickness behavior. Brain Behav Immun. 2003 Feb;17 Suppl 1:S112-8.

Kelley, KW, JC O'Connor, MA Lawson, R Dantzer, SL Rodriguez-Zas, RH McCusker. Aging leads to prolonged duration of inflammation-induced depression-like behavior caused by Bacillus Calmette-Guérin. Brain Behav Immun. 2013 Aug;32:63-9.

Kiecolt-Glaser, JK, R Glaser, S Gravenstein, WB Malarkey, J Sheridan. Chronic stress alters the immune response to influenza virus vaccine in older adults. Proc Natl Acad Sci U S A. 1996 Apr 2;93(7):3043-7.

Kusonmano, K, M Netzer, C Baumgartner, M Dehmer, KR Liedl, A Graber. Effects of Pooling Samples on the Performance of Classification Algorithms: A Comparative Study. Scientific World Journal. 2012;2012:278352.

Lacan, G, H Dang, B Middleton, MA Horwitz, J Tian, WP Melega, DL Kaufman. Bacillus Calmette-Guerin vaccine-mediated neuroprotection is associated with regulatory T-cell induction in the 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine mouse model of Parkinson's disease. J Neurosci Res. 2013 Oct;91(10):1292-302.

Lee, CG, CA Da Silva, CS Dela Cruz, F Ahangari, B Ma, MJ Kang, CH He, S Takyar, JA Elias. Role of Chitin and Chitinase/Chitinase-Like Proteins in Inflammation, Tissue Remodeling, and Injury. Annu Rev Physiol. 2011;73:479-501.

Loane, DJ, and KR Byrnes. Role of Microglia in Neurotrauma. Neurotherapeutics. 2010 Oct;7(4):366-77.

Ma, B, K Zhang, C Hendrie, C Liang, M Li, A Doherty-Kirby, G Lajoie. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. Rapid Commun Mass Spectrom. 2003;17(20):2337-42.

Magrane, M. UniProt Consortium. UniProt Knowledgebase: a hub of integrated protein data. Database (Oxford). 2011 Mar 29;2011:bar009.

Mann, M and ON Jensen. Proteomic analysis of post-translational modifications. Nat Biotechnol. 2003 Mar;21(3):255-61.

Martinez, FO, L Helming, S Gordon. Alternative activation of macrophages: an immunologic functional perspective. Annu Rev Immunol. 2009;27:451-83.

Matthiesen, R, and J Bunkenborg. Introduction to mass spectrometry-based proteomics. Methods Mol Biol. 2013;1007:1-45.

Matyszak, MK. Inflammation in the CNS: balance between immunological privilege and immune responses. Prog Neurobiol. 1998 Oct;56(1):19-35.

McGeer, EG and PL McGeer. Inflammatory processes in Alzheimer's disease. Prog Neuropsychopharmacol Biol Psychiatry. 2003 Aug;27(5):741-9.

Medzhitov, R. Recognition of microorganisms and activation of the immune response. Nature. 2007 Oct 18;449(7164):819-26.

Méndez-Samperio, P, A Trejo, A Pérez. Mycobacterium bovis Bacillus Calmette-Guérin (BCG) stimulates IL-10 production via the PI3K/Akt and p38 MAPK pathways in human lung epithelial cells. Cell Immunol. 2008 Jan;251(1):37-42.

Mi, H, A Muruganujan, PD Thomas. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res. 2013 Jan;41(Database issue):D377-86.

Möller, T. Neuroinflammation in Huntington's disease. J Neural Transm. 2010 Aug;117(8):1001-8.

Moreau, M, C André , JC O'Connor, SA Dumich , JA Woods , KW Kelley, R Dantzer, J Lestage, N Castanon. Inoculation of Bacillus Calmette-Guerin to mice induces an acute episode of sickness behavior followed by chronic depressive-like behavior. Brain Behav Immun. 2008 Oct;22(7):1087-95.

Nakajima, K and S Kohsaka. Microglia: activation and their significance in the central nervous system. J Biochem. 2001 Aug;130(2):169-75.

Nesvizhskii, AI, A Keller, E Kolker, R Aebersold. A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry. Anal Chem. 2003 Sep 1;75(17):4646-58.

Nikodemova, M and JJ Watters. Efficient isolation of live microglia with preserved phenotypes from adult mouse brain. J Neuroinflammation. 2012 Jun 28;9:147.

Nikolsky, Y and J Bryant. Protein Networks and Pathway Analysis. Methods Mol Biol. 2009;563.

O'Connor, JC, C Andre, Y Wang, MA Lawson, SS Szegedi, J Lestage, N Castanon, KW Kelley, R Dantzer. Interferon-γ and Tumor Necrosis Factor-α Mediate the Upregulation of Indoleamine 2,3-Dioxygenase and the Induction of Depressive-Like Behavior in Mice in Response to Bacillus Calmette-Guerin. J Neurosci. 2009 Apr 1;29(13):4200-9.

O'Neill, LAJ and DG Hardie. Metabolism of inflammation limited by AMPK and pseudo-starvation. Nature. 2013 Jan 17;493(7432):346-55.

Pancholi, V. Multifunctional alpha-enolase: its role in diseases. Cell Mol Life Sci. 2001 Jun;58(7):902-20.

Prinz, M and J Priller. Microglia and brain macrophages in the molecular age: from origin to neuropsychiatric disease. Nat Rev Neurosci. 2014 May;15(5):300-12.

Reading, SAJ, AC Dziorny, LA Peroutka, M Schreiber, LM Gourley, V Yallapragada, A Rosenblatt, RL Margolis, JJ Pekar, GD Pearlson, E Aylward, J Brandt, SS Bassett , CA Ross. Functional brain changes in presymptomatic Huntington's disease. Ann Neurol. 2004 Jun;55(6):879-83.

Rodriguez-Zas, SL, SE Nixon, MA Lawson, RH Mccusker, BR Southey, JC O'Connor, R Dantzer, KW Kelley. Advancing the understanding of behaviors associated with Bacille Calmette Guérin infection using multivariate analysis. Brain Behav Immun. 2014 Oct 7. pii: S0889-1591(14)00471-1.

Rubio-Perez, JM and and JM Morillas-Ruiz. A Review: Inflammatory Process in Alzheimer's Disease, Role of Cytokines. The Scientific World Journal 2012;2012:756357.

SAS Institute Inc. Base SAS® 9.3 Procedures Guide:Statistical Procedures. Cary, NC: SAS Institute Inc. 2011.

Setzu, A, JD Lathia, C Zhao, K Wells, MS Rao, C French-Constant, RJM Franklin. Inflammation stimulates myelination by transplanted oligodendrocyte precursor cells. Glia. 2006 Sep;54(4):297-303.

Smith, AM, FZ Rahman, BH Hayee, SJ Graham, DJB Marks, GW Sewell, CD Palmer, J Wilde, BMJ Foxwell, IS Gloger, T Sweeting, M Marsh, AP Walker, SL Bloom, AW Segal. Disordered macrophage cytokine secretion underlies impaired acute inflammation and bacterial clearance in Crohn's disease. J Exp Med. 2009 Aug 31;206(9):1883-97.

Suszynska-Zajczyk, J, M Łuczak, Ł Marczak, H Jakubowski. Hyperhomocysteinemia and bleomycin hydrolase modulate the expression of mouse brain proteins involved in neurodegeneration. J Alzheimers Dis. 2014;40(3):713-26.

Tai, YF, N Pavese, A Gerhard, SJ Tabrizi, RA Barker, DJ Brooks, P Piccini. Microglial activation in presymptomatic Huntington's disease gene carriers. Brain. 2007 Jul;130(Pt 7):1759-66.

Takeuchi, O and S Akira. Pattern recognition receptors and inflammation. Cell. 2010 Mar 19;140(6):805-20.

Verdoni, AM, N Aoyama, A Ikeda, S Ikeda. Effect of destrin mutations on the gene expression profile in vivo. Physiol Genomics. 2008 Jun 12;34(1):9-21.

Voloboueva, LA, JF Emery, X Sun, RG Giffard. Inflammatory response of microglial BV-2 cells includes a glycolytic shift and is modulated by mitochondrial glucose-regulated protein 75/mortalin. FEBS Lett. 2013 Mar 18;587(6):756-62.

Wang, Y, MA Lawson, R Dantzer, KW Kelley. LPS-induced indoleamine 2,3-dioxygenase is regulated in an interferon-γ-independent manner by a JNK signaling pathway in primary murine microglia. Brain Behav Immun. 2010 Feb;24(2):201-9.

Xu, D. Protein databases on the internet. Curr Protoc Protein Sci. 2012 Nov;Chapter 2:Unit2.6.

Yong, J, G Lacan, H Dang, T Hsieh, B Middleton, C Wasserfall, J Tian, WP Melega, DL Kaufman. BCG vaccine-induced neuroprotection in a mouse model of Parkinson's disease. PLoS One. 2011 Jan 31;6(1):e16610.

Zhang, JM and J An. Cytokines, Inflammation and Pain. Int Anesthesiol Clin. 2007; 45(2): 27–37.

Zhang, J, L Xin, B Shan, W Chen, M Xie, D Yuen, W Zhang, Z Zhang, GA. Lajoie, B Ma. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. Mol Cell Proteomics. 2012 Apr;11(4):M111.010587.

Zhang, Y, BR Fonslow, B Shan, MC Baek, JR Yates. Protein Analysis by Shotgun/Bottom-up Proteomics. Chem Rev. 2013 April 10; 113(4): 2343–2394.

Zhang, Z, S Wu, DL Stenoien, L Pasa-Tolic. High-Throughput Proteomics. Annu Rev Anal Chem (Palo Alto Calif). 2014;7:427-54.

Zuccato, C, and E Cattaneo. Role of brain-derived neurotrophic factor in Huntington's disease. Prog Neurobiol. 2007 Apr;81(5-6):294-330.