STUDY OF PHENOTYPIC AND GENE EXPRESSION RESPONSE TO BACTERIAL CHALLENGE

BY

SCOTT E. NIXON

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Informatics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2015

Urbana, Illinois

Doctoral Committee:

Professor Sandra Rodriguez-Zas, Chair
Professor Gustavo Caetano-Anolles
Assistant Professor Xiaohui Chen
Professor Emeritus Keith Kelley

Abstract

In the mouse model of Bacillus Calmette-Guérin (BCG)-induced inflammation, focus is placed upon phenotypic and gene-expression responses to the original challenge. Studies of the phenotypic response to BCG involves several behavior indicators, categorized as representing an initial period of sickness response, and a set of depression-like behaviors that can endure for several weeks. Little is known about the impact of relationships among indicators on these studies, as most analyses have treated indicators independently. Gene-expression studies have also been small in scale, limiting what is known about gene-expression during these behavioral changes. With the availability of Next-Generation Sequencing platforms, the scale of transcriptome analysis can be greatly increased. This study aims to address these previous limitations, characterizing behavioral and gene-expression responses to BCG-induced inflammation.

Acknowledgements

A project like this may only have one author, but is only possible through the support of an entire network of people. I first want to thank my wife, Alyson Nixon, for all of her help and support throughout this project. We made the decision together to take this research opportunity, and she has been a constant source of help throughout the entire program. I would also like to thank my family who have supported the both of us before and through the duration of this program.

The opportunity to work on these projects, much less their completion, are thanks to my advisor, Sandra Rodriguez-Zas. Entering into a multi-disciplinary project is a daunting task, with the constant struggle between depth and breadth of knowledge while summarizing appropriately. Her experience and constant efforts not only gave me the chance to practice, but pushed me to work harder and never accept an answer without understanding it. Her suggestions and direction were invaluable, with their true benefit often explained only as time showed the wisdom of the choices. Sandra's advice does not end at this research, but will be an ongoing reminder to continue learning.

I would also like to thank my committee members for their expertise, suggestions, and support throughout. I would like to thank Keith Kelley both for the chance to have worked in his lab and exposing me to the field of Psychoneuroimmunology, and the reminder to always ground the bioinformatics in the model it examines. My involvement in the bioinformatics program is in part thanks to the studies at his lab, and how they prepared me for this project. Gustavo Caetano-Anolles, for his explanations of systems biology and network biology, has and continues to demonstrate how they benefit analyses of all kinds. I would also like to thank Xiaohui Chen, whose

advice on statistics at basic and advanced levels has been invaluable. Regardless of the data and the explanation, statistics are what anchor the results.

My collaborators in this project have been resources of knowledge and support. Marcus Lawson has been an invaluable resource in neuroscience and this behavioral model since I first entered the field. Robert McCusker, whose financial support and intellectual guidance resulted in constant improvements to the methods involved. Jason O'Connor, whose early work in BCG lead to this study and current expertise helped give direction to understanding the results. I would also like to thank Robert Dantzer, whose insights both in the behavioral model and their explanation were repeatedly beneficial in interpreting and reporting the results of this study.

I also would like to thank all of Sandra's lab throughout these years, for helping make all these days at the lab a fun learning experience. It's just as important to laugh as to learn when years of effort are involved. Former members including Nicholas Serão, Kristin Delfino, Nadeem Malik Akhtar, and Cynthia Zavala, all provided insight by their differing viewpoints and a reminder that there is a light at the end of the research tunnel. Dianelys González-Peña, Robmay Garcia, and Kelsey Caetano-Anolles, for our shared struggle in exploring how to sift through the scale of data.

Finally, I would like to acknowledge the generous funding that made these projects possible. I would like to thank the National Institute of Health, The United States Department of Agriculture for the grants and funding to Sandra Rodriguez-Zas, Keith Kelley, and Robert McCusker that made these projects possible.

Table of Contents

Page

List of Figures

Figure **Page**

Chapter II

Chapter III

Chapter IV

List of Tables

Table **Page**

Chapter I

Chapter II

Chapter III

Chapter IV

CHAPTER I: Literature review

1.1 Inflammation and depression

Impact of depression on the population

Depression is a large burden upon society, as the leading cause of disability and more than 350 million cases worldwide (WHO, 2012). The 4th edition *Diagnostic and Statistical Manual of Mental Disorders* defines major depressive disorder as expression of at least five of nine symptoms for greater than two weeks (Dedic et al., 2011). The burden of depression extends beyond the disease itself. Successful treatment often still leaves residual symptoms and depression shares co-morbidity with other diseases including an increased mortality risk (Lépine and Briley, 2011). A potential linkage between depression and co-morbid diseases may be inflammation, due to the extensive interactions with inflammatory pathways (Raison and Miller, 2011). The limited effectiveness of current options motivates continued searching for treatment options. While the most widely used class of drugs for the treatment of depression are described as selective-serotonin-reuptake inhibitors, the antidepressant effects appear to operate less directly than originally believed (Gartside et al., 1995; Angoa-Pérez et al., 2014). A greater understanding of the mechanisms underlying depression is needed to identify and develop treatments with better performance.

Inflammation-induced depression and microglia

The immune system, and specifically dysregulation of the inflammatory response, has been linked to several diseases and neurodegenerative disorders (Frank-Cannon et al., 2009; Wyss-Coray and Mucke, 2002). A common factor within these disorders is neuroinflammation, when an inflammatory response develops within the central nervous system (CNS). Several

neurodegenerative disorders are themselves comorbid with depression (Rosenblatt, 2007; Aarsland et al., 2012; Aznar and Knudsen, 2011), which itself has been linked to inflammation (Dantzer et al., 2011). However while the link between depression and inflammation is established, the understanding of their complex relationship continues to evolve. Neuroinflammation itself involves the activation of microglia in the CNS, a mediator of both the physiological and behavioral responses of inflammatory inducers (McGeer and McGeer, 2011). As the principal resident immune cells of the brain, microglia are studied for their relation to other mononuclear phagocytes like macrophages and their involvement in immune signaling within the CNS. Profile comparisons between microglia and other cell types including macrophages are available (Hickman et al., 2013; Gautier et al., 2012), yet their varied response profiles have not been fully explored. Microglia constantly survey their surrounding environment, resulting in a variety of priming states and phenotypes that can have neuroprotective or neurotoxic effects. These response profiles can not only be unique to the inducer, but also adapt over time as a part of the total immune response. The comparison of gene-expression profiles between microglia and related macrophages during an inflammatory response may identify unique characteristics of the role of microglia in neuroinflammation.

The interactions between the immune system and the central nervous system are major areas of focus, based upon the impact of neuroinflammatory signaling centrally and peripherally (Eyre and Baune, 2012). Microglia, the resident macrophage of the brain, is particularly interesting for its rapid activation and responsive release of inflammatory mediators in response to challenge (García-Bueno et al., 2008). Constantly dynamically active when the CNS is healthy, microglia shift from a state of scanning to an active response once a challenge is detected (Hanisch and

Kettenmann, 2007). Microglia provide a unique expression profile of interest to researchers, and are known for strong involvement in cytokine signaling (Beutner et al., 2013; Hickman et al., 2013; Hanisch, 2002). The interaction between the immune system and brain, and the role of cytokines therein, makes microglia a group of interest for investigations of inflammation-induced depression (Miller et al., 2009; Kreisel et al., 2014). For instance, stress is another inducer utilized for stress-induced models of depression (Lee et al., 2013). The stress model can cause proliferation, activation, and apoptosis in microglia prior to depression-like behaviors, yet the behaviors can be reversed by lipopolysaccharide (LPS) exposure despite its function as an inflammatory-inducer of depression (Kreisel et al., 2014; O'Connor et al., 2009b). However, unpredictable chronic mild stress can also serve as an inducer of neuroinflammation (Farooq et al., 2013). To summarize, microglia are capable of diverse responses which can result in protective or harmful responses for the CNS environment, with both signal directions involving cytokines (Hanisch and Kettenmann, 2007; Hanisch, 2002).

Microglia are typically isolated from other cell populations in the brain by taking advantage of cell markers dotting their exterior, or separating them via a density gradient with centrifugation (Nikodemova and Watters, 2012). Methods utilizing cell markers bind an antigen to the cell surface marker CD-11b. The anti-CD-11b antigen can conjugate to a magnetic bead for separation in a magnetic field (Nikodemova and Watters, 2012). Microglia can also be dual-marked with antigens for CD-11b and CD-45, and both antigens conjugating to fluorescent markers for profiling via flow cytometry ($CD\text{-}11b^{+}/CD\text{-}45^{low}$) and isolated by fluorescence-activated cell sorting (Hickman et al., 2013).

Association between transcriptome profile and depression

Gene expression is clearly involved in the interactions between inflammation and depression, yet exploration has been limited in scale (Blume et al., 2011). The expression of individual inflammatory markers overlap between depressed and non-depressed populations, and functionally operate in a networked capacity to other depression-associated systems (Raison and Miller, 2011). Modeling depression within animals provides more direct opportunities to study the biological underpinnings of the phenotype (Dedic et al., 2011). Models inducting depression after immunological challenge have identified key players in the progression of the phenotype (Moreau et al., 2005; O'Connor et al., 2009a). The scale of studies limits the understanding of the mechanisms coinciding with depressive-like behaviors.

Mycobacterium and macrophages

Studies on the induction of depression-like behaviors following cytokine triggering often use LPS as an acute inflammatory inducer (Lestage et al., 2002). As bacterial cell wall components, LPS triggers the initial immune responses to bacterial challenge with detectable depressive-like behaviors following (Yirmiya, 1996). However, the entire episode of the immune response and depressive-like behaviors are within a 48 hours period, limiting the utility for modeling chronic immune activation (Frenois et al., 2007). Bacille Calmette-Guérin (BCG), an attenuated *Mycobacterium* strain, is useful as an inducer for chronic models of immune activation and depressive-like behaviors (Moreau et al., 2008). As a first-responder to infection, macrophage activation and signaling initiates the response tailored to the form of attack involving transcriptomic modifications (Ehrt et al., 2001; Keller et al., 2004). Macrophages and *Mycobacterium* strains interact in a complex manner, with impacts upon the particular signaling

of both populations (Schnappinger et al., 2006). When *Mycobacterium* successfully infects a macrophage cell, gene expression in the macrophage is modified to promote apoptosis and occurs at a higher rate when infected by an avirulent strain compared to a virulent one (Danelishvili et al., 2003). As macrophages kill *Mycobacterium* released from an apoptotic cell, apoptosis appears to limit the growth of infection while uncontrolled necrosis correlates with infection progression (Fratazzi et al., 1997). In comparison to the process of necrosis, the programmed cell death of apoptosis is an energy-intensive process that is limited by ATP availability, with a critical importance of energy metabolism for host defense (Buttgereit et al., 2000; Danelishvili et al., 2003; Danelishvili et al., 2010). However host defense involving apoptosis can be interrupted if the pathogen hijacks particular regulators of cell-cycling, a capability attributed to intracellular pathogens including some *Mycobacteria* (Iwai et al., 2007; Oswald et al., 2005). Appropriately named "mitotic catastrophe," the simultaneous signaling to a cell for mitosis and apoptosis can result in a failed combination of the two pathways (Castedo et al., 2004; Svoboda et al., 2007). Along with the direct impacts of mitotic catastrophe on host defense against a pathogen, dysregulation of cell-cycling can also be associated with neurodegenerative conditions when it occurs in microglia (Bonda et al., 2010).

1.2 Transcriptomics

Gene expression and transcript levels

Gene expression is the synthesis of gene products from a gene within the DNA. The DNA genome is identical among all cells in an organism, making gene expression the first step towards differentiating actions and roles of cells and their responses (Morozova et al., 2009). Gene expression can follow many routes, the two most common resulting in functional RNA or protein

with RNA mediation (Ball, 2013). The genes in DNA are ‘written’ in the nucleotide alphabet of Adenosines, Cytosines, Guanines, and Thymines. As a first step towards expression, a gene in the DNA is transcribed by using the gene’s complementary DNA (cDNA) strand as a template to produce an RNA copy of the strand containing the gene. The sequence of nucleotides should thus be identical between the DNA template and the RNA copy, excepting that Thymines are instead replaced by Uracil in RNA strands (Krebs et al., 2009). A single gene can contain multiple regions of coding separated by non-coding sections that must be spliced out. Assembly of the remaining coding regions can either result in a single product or several combinations, termed alternative splicing. The first step to deciding the end-product for an RNA copy is processing by a spliceosome, which removes the non-coding sections (introns) while assembling the coding sections (exons) in the appropriate order. This reassembled RNA strand, once modified with a 5’ cap and 3’ poly-adenosine tail, comprises a messenger-RNA (mRNA) strand, an individual transcript from the DNA gene. These mRNA strands are the target of gene expression analysis, and the following workflow.

Before analyzing RNA samples for transcriptomic studies, it is important for the input samples to be uncontaminated, pure, and high quality (Nagalakshmi et al., 2010). Methods for extraction of RNA are varied, including methods based on phenol extraction, spin columns, and hybridization of the two (Chomczynski and Sacchi, 2006; Morse et al., 2006; Reno et al., 1997; Tan and Yiap, 2009). Regardless of the extraction procedure, high quality RNA benefits downstream methods and is quantified by the RIN quality which scores samples based upon degradation in the form of RNA fragmentation (Schroeder et al., 2006).

Platforms available for transcriptomic studies

Reverse Transcription quantitative Polymerase Chain Reaction (RT-qPCR) is not itself a high-throughput method. Examining the expression levels of individual genes and validating high-throughput methodologies is a popular use of RT-qPCR thanks to high sensitivity with well-established protocols (VanGuilder et al., 2008; Wang et al., 2014). Measurement of gene expression with RT-qPCR first requires producing cDNA from the available RNA. The cDNA is then exponentially amplified by several cycles of PCR, with a fluorescent probe-sequence that is detectable once bound to the cDNA target (VanGuilder et al., 2008). Measurements using RT-qPCR are thus based on the count of cycles before fluorescence reaches a threshold, specifically the difference in count between conditions for relative abundance. However, counting the difference in cycles assumes a doubling of produced cDNA targets each cycle, requiring PCR cycles are 100% efficient and not limited by available substrates (Livak and Schmittgen, 2001). House-keeping genes and measured additions of non-native control genes can measure the count stability across samples. The design of RT-qPCR also requires specific primer and probe sequences for every tested gene, limiting exploratory analysis (VanGuilder et al., 2008). It is thus mentioned here due to its use in many gene expression studies and popularity as a standard by which to compare other transcriptomic platforms.

Microarrays are a common high-throughput platform for the identification and quantification of labeled nucleic acids in samples. A DNA microarray simultaneously measures thousands of pre-selected oligonucleotides or cDNAs of interest as probes to see if they bind to complementary sequences within a fluorescently-labeled sample, profiling samples for all of the oligonucleotide sequences. Studies are limited to annotated strains and closely homogenous sequences, as the pre-

selected nature of the microarray probes is strongly dependent upon annotation quality (Levy et al., 2007; Wang et al., 2009). Additionally when using this platform to examine alternative splicing, there is a demonstrated technical bias from sequence features of the probes (Gaidatzis et al., 2009).

The direct sequencing of nucleic acids on a large scale has developed into a high-throughput platform, RNA-Seq (Wang et al., 2009). Although multiple platforms are available, all are termed flow cell sequencers based upon the use of a multi-lane flow cell for massive parallelization and increased capacity (Holt and Jones, 2008). Samples are applied to the RNA-Seq flow cell for sequencing to occur. Instead of labeling sample nucleic acid and identifying which probes will hybridize, a library of cDNA fragments is produced from the RNA samples prior to sequencing. Several methods are available for library preparation with continual focus upon high-throughput methods to avoid a choke-point in the sample pipeline (Quail et al., 2008; Nagalakshmi et al., 2010; Wilhelm et al., 2010; Zhong et al., 2011; Wang et al., 2011; Kumar et al., 2012). Fragments are mixed with random short adapters which can attach as a primer for sequencing (Wang et al., 2009). Sequencing of the original RNA is determined via complementary pairing to the cDNA once primers attach. Only one nucleotide attaches at a time by utilizing nucleotides with reversible blockers at the 3' end. The four DNA nucleotides are uniquely labeled with a fluorescent dye. The dye associated with the attached nucleotide is identified by the system before removing the blocker and repeating this cycle until the appropriate sequencing length is achieved, producing millions of reads. The output reads from RNA-Seq are dependent upon the specific sequencing technology, but can range from 30-400 base pairs in length (Wang et al., 2009). Experiments in RNA-Seq are further organized into those using single-end sequencing, or paired-end sequencing. During cDNA

template sequencing, single-end sequencing attaches primers to one end of each template to produce reads while paired-end methods attach primers to both ends and sequence in towards the middle. While paired-end experiments cause some additional complexity to downstream tool choices, paired-end data provides more information for exploring isoform- and splicing-level expression (Salzman et al., 2011). For a read to carry any information on alternative splicing within a gene, the read must overlap an alternatively spliced exon or the junctions between exons. So if the produced reads are length *r* while investigating a gene of length *L*, the probability *P* of the read providing information about an alternatively spliced exon of length *l* is (Salzman et al., 2011):

$$P = \frac{l + r}{L - r}$$

While paired-end experiments do not increase read length, an increase in the number of reads and their sequencing from both ends results in more reads providing information at the isoform-level (Salzman et al., 2011). Along with avoiding dependence upon prior knowledge of the sequences to be examined, this also avoids non-biological biases associated with sequence features of the probes found in arrays when studies involve splicing used (Gaidatzis et al., 2009). Careful utilization of data processing can result in high concordance between microarrays and RNA-Seq platforms (Bottomly et al., 2011). Normalized, log-transformed RNA-Seq data correlates well with the current 'gold standard' of RT-qPCR and log-transformed microarray data when comparison does not expose differences in low-expression accuracy between platforms (Mortazavi et al., 2008; Wang et al., 2014).

RNA-Seq data processing and normalization

Gene expression data from RNA-Seq involves converting the set of short sequence reads from a sample into measurable mRNA expression. Several platforms are currently available for the

production of RNA-Seq data (Liu et al., 2012). This review of processing steps centers on the Illumina HiSeq2000 platform (Illumina, San Diego, CA).

Reads from sample sequencing are stored as FASTQ files, a FASTA-derived format containing the raw nucleotide sequence and a paired quality score for each nucleotide site (Cock et al., 2010). The FASTQ file format is not yet standardized, encompassing three separate standards based upon encoding and quality score. The quality score is most often calculated as a Q_{PHRED} value based upon the log-transformed estimated probability of error (P_e) where a greater value is a more trusted estimate (Cock et al., 2010). As an example, Q_{PHRED} values of 10, 20, and 30 would result in error probabilities of 0.01, 0.001, and 0.0001 respectively. A similar Q_{SOLEXA} score is also used, with the two quality scores calculated as:

$$Q_{PHRED} = -10\ x\ log_{10}(P_e)$$

$$Q_{SOLEXA} = -10\ x\ log_{10}\left(\frac{P_e}{1 - P_e}\right)$$

Three related but unique formats share the FASTQ file suffix: FASTQ-Sanger (Q_{PHRED}, stored in ASCII characters 33-126), FASTQ-Solexa (Q_{SOLEXA}, stored in ASCII characters 59-126), and FASTQ-Illumina (Q_{PHRED}, stored in ASCII characters 64-126), which are incompatible (Cock et al., 2010). The incompatibilities between the FASTQ formats requires downstream programs are manually or automatically capable of format identification. The quantitative value associated with each base position is important to identify potential substitution errors, a common issue with Illumina sequencing data (Yang et al., 2013). The substitution errors and associated poor quality values tend to concentrate at the 3' end of reads (Liu et al., 2012). The accumulation of lower quality values and associated errors at the 3' end are suggested to be the result of dephasing (Metzker, 2010). Dephasing occurs when the several copies of the original cDNA template fail to

remain synchronized for each sequencing cycle, resulting in more than one site being sequenced during a cycle. Dephasing consequently results in unclear signaling during a sequence cycle because multiple sites are being reported simultaneously. Substitution errors occur when the signal intensities of more than one nucleotide are similar, correlating with lower Q_{PHRED} values (Nakamura et al., 2011). To avoid the effects of poor quality data on later analysis, pre-processing the raw reads uses the quality value to filter data.

Prior to any analytical steps, pre-processing reads involves the removal of the adaptor sequences, evaluation of read quality, and filtering reads based upon quality. Downstream analysis can result in incorrect and missing alignments, incorrect construction, and an overall increase in computational resources to handle the increased complexity when poor quality data is not removed beforehand (Del Fabbro et al., 2013). Read filtering can include complete removal of a read from the dataset, or trimming of poor quality sections. While trimming a read section for single-end reads is comparatively straightforward, any modifications to a read in paired-end data requires complementary modification of the mate read so they remain complementary. Read trimming tools are classified based on: the algorithm utilized; capable of working with compressed files; can work with paired-end data; automatic detection of the quality score format; and if one or both ends of a read are trimmed (Del Fabbro et al., 2013). However, there is debate whether trimming reads is beneficial to the sequenced transcriptome. Read trimming improves precision and reduces computational load during later alignment (Del Fabbro et al., 2013). However, trimming can severely impact coverage when combined with typically stringent ($Q_{PHRED} > 20$) cutoff levels (MacManes, 2014). Whole removal of reads can involve filtering based upon the overall quality

and read length (Minoche et al., 2011). As with all filtering methods, care must be taken as overly-stringent processing can result in information loss.

When the genome and annotation information is available for a source organism, the first step performed towards analyzing data is mapping the reads. Because of the short length of the reads output by the RNA-Seq platform, mapping reads against a reference sequence provides positional data for reassembly into the originating biologically-relevant transcripts. However, large numbers (millions of reads) of comparatively short length reads searched against the large space of the genome and transcriptome causes alignment and computational problems. Efficiency thus becomes a necessity, and solutions found for when reads map ambiguously to multiple sites. Alignment algorithms are categorized as seed-based alignment or a Burrows-Wheeler transformation based upon their source of search efficiency (Burrows and Wheeler, 1994; Grant et al., 2011; Garber et al., 2011). Seed methods work by taking a small sequence of the read, the seed, and attempting a match to the reference which reduces the search space for full-length matching (Garber et al., 2011). Aligners like BWA and Bowtie use Burrows-Wheeler transformation as part of a Full-text, Minute space index called the FM-index (Li and Durbin, 2009; Langmead et al., 2009; Burrows and Wheeler, 1994; Ferragina and Manzini, 2001). As an example of Burrows-Wheeler transformation, a string (‘preliminary proposal’) to be transformed receives a character prefix ($) that sorts ahead of the sequence characters. A set of rows equal to the length of the string is produced, with each row being one cyclic step ahead of the above row. Finally, the rows are sorted alphabetically to produce a matrix *M* with the last column (L) being the transformed sequence:

Cycle						
$	b	a	n	a	n	a
a	$	b	a	n	a	n
n	a	$	b	a	n	a
a	n	a	$	b	a	n
n	a	n	a	$	b	a
a	n	a	n	a	$	b
b	a	n	a	n	a	$

Sort (*M* matrix)						
F						L
$	b	a	n	a	n	a
a	$	b	a	n	a	n
a	n	a	$	b	a	n
a	n	a	n	a	$	b
b	a	n	a	n	a	$
n	a	$	b	a	n	a
n	a	n	a	$	b	a

The transformation is easily reversible, and the strings of the same letter make for a more efficiently compressed design than the original sequence when there are several letters as is the case for nucleotide sequences. Importantly for the FM-index, the output transformation follows particular traits based upon using the First (F_i) and Last (L_i) columns. If *c* is an element of the sequence; *C[c]* is the count of characters alphabetically smaller than *c*; *Occ*[*c*,*k*] is the count of characters alphabetically smaller than *c* in *L* to position *k*; *T[i']* is the original position of character at position *L*(*i*); *LF*(*i*) is the Last-to-First column mapping (Ferragina and Manzini, 2001):

$$LF(i) = C\big[L[i]\big] + Occ(L[i], i)$$

$$T[i' - 1] = L[LF(i)]$$

The *Occ*[*c*,*k*] function becomes an important part of the FM-index. When checking large references like a genome, the entire string is transformed and then partitioned into superbuckets that are further partitioned into buckets. Each superbucket and bucket presents a header describing the occurrence count for each character until the start of the current partition. Instead of compressing the entirety of the reference in a single unit, the portion of the reference within each bucket is compressed independently. A *Occ*[*c*,*k*] call would search the headers of superbuckets and buckets to find where the position *k* is stored, uncompress only the bucket containing *L*[*k*], count the occurrences *c* within this bucket, add the sum of the character count from preceding buckets within the superbucket, and finally add the character count from preceding superbuckets

(Ferragina and Manzini, 2001). The index can then be searched for a query sequence because of strong relations between the matrix used to transform the original string and the suffix array (Ferragina and Manzini, 2000).

Beyond efficiently searching a massive search-space on the scale of a genome, a transcriptome assembler often requires the capability to handle the splicing characteristics of mRNA strands, specifically the gaps and reorganized nature of mRNA when compared to the DNA genome. These sequence inconsistencies when comparing mature mRNA to the DNA genome complicates alignment, especially for Burrows-Wheeler methods that search for perfect matches. To deal with splicing concerns, programs such as Tophat are designed (Kim et al., 2013). Tophat operates by using Bowtie for aligning reads without splicing concerns. Tophat then fragments reads that failed to map and repeats the search to test if the fragments align near junctions. Regardless of splicing concerns for alignment, reads must afterwards be assembled into units of biological transcription. Comparing expression levels in RNA-Seq data can be performed at the level of the genes, gene isoforms, or even individual exons (Rapaport et al., 2013; Trapnell et al., 2010; Anders and Huber, 2010; Trapnell et al., 2012; Anders et al., 2012). Knowledge of the genome and annotations can guide all levels of read assembly from mapping; if annotations are less robust assembly can occur independently (Garber et al., 2011). This review will use the assembly unit of ‘transcript’ to discuss methods.

Statistical analysis of any RNA-Seq data first requires correcting for variations introduced from the platform method and identifying an effective, quantifiable unit of measure for comparisons. Raw read counts can reflect technical bias based upon variations between overall sample

expression and the behavior of individual lanes on the flow cell. Normalizing read count data by scaling to the library size, or a factor therein, can account for these technical biases (Dillies et al., 2013). One of the common normalization methods is converting read count to Fragments Per Kilobase of exon per Million fragments mapped (FPKM) which accounts for differences in transcript length (Trapnell et al., 2010). Upper-quartile normalization orders the population of transcripts by the count of reads aligning to each transcript, then splits the population into quartiles. To calculate the median, upper-quartile methods use only the upper-quartile transcripts, those with high numbers of alignments, instead of the total population including the potentially large number of transcripts with no alignment data from the experiment. Upper-quartile normalization corrects the uninformative nature of the transcript median when including transcripts with zero expression data (Bullard et al., 2010). Combining the FPKM and upper-quartile normalization methods avoids variations in precision for transcripts with lower expression values (http://cufflinks.cbcb.umd.edu/manual.html; Dillies et al., 2013). The combination of FPKM and upper-quartile normalization modifies the denominator of FPKM, as the rate of fragments per million fragments mapped would instead be counted based only on the upper-quartile of the population instead of the total count of all fragments. Once the read values are appropriately normalized for a sample, comparison between samples requires additional consideration. Comparison across profiles also involves comparisons across sequencing runs, necessitating library-size normalization beyond the previous within-library steps taken. High-expression transcripts can account for a large portion of the total read population, and any difference in their expression between libraries of equal size would correlate with changes to the read counts for other transcripts (Trapnell et al., 2013). The impact on the total population can be rescued by removing the upper-quartile of transcripts from library size calculations, not to be confused with normalizing

by the upper-quartile within libraries (Bullard et al., 2010). Adjusting the data by the median of the ratio of observed read counts between samples is a popular method suggested during the development of DESeq and edgeR; this adjustment was subsequently adopted by other programs as a standard (Anders and Huber, 2010; Robinson and Oshlack, 2010; Trapnell et al., 2013).

The statistical calculation of differences between transcripts from RNA-Seq data evolves as the underlying complexity gains appreciation (Soneson and Delorenzi, 2013). When considering the number of reads that can align (count) towards a transcript, the value is bounded by zero and an upper limit as the total number of reads sequenced. The goal of these methods is to output the parameter which represents the mean read count aligning to a gene, so that log-likelihood or similar tests can compare the mean between conditions. The difficulty in finding the mean is described below, as the model and resultant estimates of variance become more complex (Robinson et al., 2010; Soneson and Delorenzi, 2013; Trapnell et al., 2013). The discrete nature of read counts and low variance among technical replicates first lead to the discrete Poisson distribution for modeling purposes (Marioni et al., 2008). Given a single parameter λ as the expected rate of reads mapping *p* to the individual transcript by the total population of reads produced *n*; *k* number of reads counted for a transcript, the Poisson distribution is (Hogg and Tanis, 2009):

$$P(X = k) = \frac{(c\lambda)^k e^{-(c\lambda)}}{k!}$$

Where *P* is the probability, and *e* is Euler's number. In actual modeling of RNA-Seq data, the null hypothesis is λ remains constant across samples and conditions, and the *c* represents the total read production for a sample (Marioni et al., 2008). As $\lambda c=\mu$ for Poisson distributions, the null hypothesis H_o of differential expression between conditions A and B, which would have means μ_A and μ_B respectively, can also be written as:

$$H_O: \mu_A = \mu_B$$

The alternative hypothesis:

$$H_A: \mu_A \neq \mu_B$$

The hypothesis can be tested with a X^2 test (Srivastava and Chen, 2010). Despite the Poisson distribution effectively modeling data from early studies, later discussion suggests it was based on the low variance among technical replicates (Soneson and Delorenzi, 2013). However, if the λ parameter of the Poisson is treated as a random variable which follows a Gamma distribution, then the count data instead follows a Negative Binomial distribution (Robles et al., 2012). The greater variance range among biological replicates motivated future differential expression programs to utilize the Negative Binomial distribution (Anders and Huber, 2010; Robinson and Oshlack, 2010; Trapnell et al., 2013). Given the *K* counted number of reads aligning to a transcript, and the abundance of the transcript in a particular library is *R* (Robles et al., 2012):

$$E(R) = q, Var(R) = v$$

Given the expectation that the observed read count *r* represents an accurate observation of *R*, then the distribution of K is proportional to a Poisson (Robles et al., 2012):

$$K|(R = r) \sim Pois(\lambda r)$$

$$E(K|R = r) = \lambda r, \quad Var(K|R = r) = \lambda r$$

Here in comparison to the Marioni et al. Poisson example (2008), λ would be a normalization factor based upon the read production of a lane. Also, because of the relationship between *R*, *r*, and *K* and allowing for a dispersion factor ϕ based upon higher technical variation or variation among biological replicates (Robles et al., 2012):

$$E(K) = \mu, Var(K) = \mu(1 + \phi\mu)$$

Where

$$\mu = \lambda q, \qquad \phi = \frac{v}{q^2}$$

Finally, if R was a random variable that follows a Gamma distribution (Robles et al., 2012):

$$R \sim Gamma(q, v) \text{ and } \lambda R \sim Gamma\big(\mu, \mu(1 + \phi\mu)\big)$$

Then

$$K \sim Negative\ Binomial(\mu, \mu(1 + \phi\mu))$$

As the expected dispersion ϕ approaches zero, the mean and variance become equal, and the Negative Binomial collapses back down to a Poisson. While methods such as DESeq and edgeR both use the Negative Binomial to model overdispersion in the read count data, the methods differ in generation of an overdispersion estimate (Robles et al., 2012; Soneson and Delorenzi, 2013). EdgeR attempts to estimate a singular common dispersion parameter among samples for each gene, and estimate raw variance with quantile-adjusted conditional maximum likelihood, (Robinson et al., 2010). DESeq instead attempts to model a mean-dependent local regression to estimate the dispersion parameter and raw variance through linear scaling to correct for differing library size (Anders and Huber, 2010). Both edgeR and DESeq then look for differential expression for each gene with a test analogous to Fisher's Exact test which uses read counts and overdispersed-tolerant Negative Binomial probabilities instead of Hypergeometric (Robinson and Smyth, 2007). As models increase in complexity to consider uncertainty in fragment counts as well as biological variability, the mixing of several Negative Binomial distributions can combine into a singular Beta Negative Binomial model as with Cuffdiff2 (Trapnell et al., 2013). Cuffdiff2 considers fragments instead of reads, but utilizes the same design as DESeq for modeling the count for an individual transcript with a Negative Binomial and a local regression to estimate dispersion. The test statistic used to determine significance of differential expression is based upon the log-transformed ratio of expression, divided by its variance of the transformed ratio to follow a normal

distribution (Trapnell et al., 2010; Trapnell et al., 2013). Where Cuffdiff2 more drastically departs from the other differential expression designs is utilizing an additional Negative Binomial to consider the uncertainty in fragment counts, which is based on a fragment ambiguously mapping to several transcripts based upon high similarity (Trapnell et al., 2013). This additional level of complexity is considered necessary by the authors to explore beyond the scale of genes to their isoforms, the differential splicing between isoforms, and variation in transcriptional start sites (Trapnell et al., 2013). The Beta distribution with parameters α and β is described as having (Hogg and Tanis, 2009):

$$\mu = \frac{\alpha}{\alpha + \beta}$$

Modified for notational simplicity in Cuffdiff2 descriptions, the term changes to (Trapnell et al., 2013):

$$\mu = \frac{\alpha - 1}{\alpha + \beta - 1}$$

And a Beta Negative Binomial is described with parameters r, α, and β such that:

$$X \sim BNB(r, \alpha, \beta)$$

Where

$$X|p \sim NB(r, p) \text{ and } p \sim B(\alpha, \beta),$$

Then if $V_E(N)$ is an obtained function that outputs predicted variance for condition E when given the mean counts N; X_g which is the abundance of transcripts at the locus containing the tested transcript; $\hat{\gamma}_t$is the estimated probability that a fragment from the locus originated from transcript t, the parameters for the Beta Negative Binomial can be solved by the equations (Trapnell et al., 2013):

$$A = X_g \hat{\gamma}_t; \; B = V_E(X_g)\hat{\gamma}_t$$

$$\frac{1(1-p)}{p} = A; \frac{1(1-p)}{p^2} = B; r = \frac{A^2}{B-A}; p = \frac{A}{B}$$

So that *r* must be greater than zero, and assuming that B > A since it implies that there is overdispersion in the data (Trapnell et al., 2013). If B = A, then variance across replicates equals the mean which collapses the distribution back to a singular Negative Binomial. If there is no uncertainty in the read counting, then the model also collapses to a Negative Binomial, akin to the model design of edgeR and DESeq. While this additional model term increases the complexity, it is also needed to address the ambiguous fragment counting so that quantification can occur at the sub-gene level (Trapnell et al., 2013). The scale of output produced from Cuffdiff2 analysis produces the more common analysis of differential expression at the read level, while including tests for differential expression of gene isoforms, differences in the how often conditions are observed to use transcription start sites for those isoforms, and thus allows analysis of alternative splicing based upon considering differential expression among the population of isoforms that represent a gene (Trapnell et al., 2012; Trapnell et al., 2013).

Like any other platform, RNA-Seq data can contain technical and biological bias not addressed by normalization. Variation from RNA-Seq platforms can involve unequal distribution of available binding sites combined with the potential for primers to bind unevenly to the sites (Hansen et al., 2010). Reads originating from regions with high similarity may be indiscernible and result in multiple mapping sites. If the mappings sites split between transcripts that express at different levels, assigning reads to the assembled transcripts for quantification purposes must account for these differences (Mortazavi et al., 2008). Reads can be mapped uniformly before re-estimating the counts. The probability of observing *C* reads on an *L* length transcript with *X* number of copies in *N* tries over a transcriptome of length *T* is calculated as (Mortazavi et al., 2008):

$$\frac{C}{N} = \frac{XL}{T}$$

By distributing multi-mapped reads based upon the uniformly-mapped count to get the read estimate *R*, a new distribution can be calculated:

$$X = \frac{C}{NL}T = \frac{R}{10^9}T$$

Available analytical pipelines include methods to address these biases based upon the above calculations (http://cufflinks.cbcb.umd.edu/howitworks.html). In my studies, corrections were implemented with Tophat (splice-tolerant alignment using genome and transcriptome annotations using Bowtie with the FM-Index) at the mapping level (Kim et al., 2013). Transcript assembly, as well as the normalizations and corrections at that stage (transcript assembly guided by reference annotations while implementing fragment bias corrections, multi-read corrections, and upper-quartile normalization to FPKM calculation), were handled by Cufflinks (Trapnell et al., 2010). Finally, normalization among libraries (DESeq method) and differential tests were handled by Cuffdiff (Trapnell et al., 2013).

1.3 Functional and network analysis

Categories and gene sets

Comparing transcriptomic profiles involving thousands of differentially expressed genes demands efficient interpretation that scales beyond the individual gene to evaluate functional roles and network interaction (Dopazo, 2009). Mining particular genes are still effective, but the importance of individual genes becomes a singular part of examining the interactivity and sum effect of groups. Functional analysis focuses upon analyzing categories and gene sets constructed from knowledge of shared characteristics. The gene ontology catalogue is widely accessed as a source of functional information (Ashburner et al., 2000). Designed as a directed acyclic graph, individual terms are

organized with broader terms serving as parents to more specific terms, with terms capable of belonging to more than one parent term. The Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathways is an integrated database of 15 main databases with systems, genomic, chemical, and health information (Kanehisa et al., 2002; Kanehisa et al., 2014). KEGG maps and modules are generalized so that information entered on a specific organism can be adapted for analysis across other species. Reactome is somewhat related to KEGG by design, and is a manually curated database that classifies biological units based upon their reactions, which are then networked by pathways and assembled into a biological process-based hierarchy (Croft et al., 2014). Shared characteristics can be based upon function, pathway, or other traits; many of the category designs are interconnected by integrated knowledgebases and linked database IDs (Huang et al., 2009b; Drăghici et al., 2006). However despite the relationships between databases, consensus is limited due to differences in structure and curation (Stobbe et al., 2011). Yet until the databases can be effectively integrated, parallel analyses using several databases can provide a broader, albeit fragmented, understanding of functional relationships.

Cutoff and non-cutoff based approaches

When testing a subset gene list of the transcriptome profile such as those meeting a particular significance level, the subset selection method is termed cutoff-based approach. Also described as Class I tools, many utilize similarly performing tests (Huang et al., 2009a; Rivals et al., 2007). As a representative test of cutoff-based tools, the hypergeometric test is a popular and computationally straightforward method of determining whether a gene set is enriched. The total number of genes tested is N. Within the number of tested genes, the number in the significant list is K. The hypergeometric distribution describes in a subsample of n genes belonging to a gene set, the

probability that *k* number of genes $P(X=k)$ would be in the significant list. Given a size to test without replacement, there are: $\binom{N}{n}$ are the possible subsamples; $\binom{K}{k}$ possible subsamples of significant genes in a gene set; and $\binom{N-K}{n-k}$ ways to fill the remaining spaces in the subsample where$\binom{K}{k} = \frac{K!}{k!-(K-k)!}$. The hypergeometric probability is the likelihood that the level of enrichment observed could occur due to random choice based upon the size of the gene set and genes tested in the experiment (Rodriguez-Zas, 2013):

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

The application of the hypergeomtric distribution for gene set analysis can be found in (Boyle et al., 2004). A popular derivative of the hypergeometric distribution is the scoring method originally from Expression Analysis Systematic Explorer (EASE), a one-tailed jackknifed Fisher exact test which provides a conservative adjustment of the standard probability (Hosack et al., 2003). Jackknifing removes a single point from *k* to penalize gene sets. The Database for Annotation, Visualization and Integrated Discovery (DAVID) Knowledgebase is a popular online set of bioinformatics tools that utilizes the EASE score (http://david.abcc.ncifcrf.gov; Huang et al., 2009b). The DAVID knowledgebase is built upon several non-redundant database sources (**Table 1.1**), web-accessible, and tests themes from Gene Ontology (GO) categories, the Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathways, and several others (**Table 1.2**) (Maglott et al., 2005; Apweiler et al., 2004; Ashburner et al., 2000; Kanehisa et al., 2002; Huang et al., 2009b). Enriched functional categories can be clustered in DAVID by using a fuzzy heuristic partition algorithm on the gene list to cluster categories based upon highly related genes (http://david.abcc.ncifcrf.gov/helps/functional_classification.html). Fuzzy heuristic partition uses

the underlying list of genes enriching the categories. Genes are mapped with proximity related to shared annotation terms. The pairwise relationship between all genes is calculated as a kappa score, based upon the degree of relationship in sharing categories. Each gene can initiate a cluster, filtered for sufficient neighbors with a high enough kappa score. Clusters are then merged based upon sharing more than half of their member genes, until only less similar clusters remain. The sensitivity of cutoff-based approaches, combined with a user-defined list of significant genes based upon an arbitrary cutoff level can result in artifacts. These artifacts, in the genes and gene sets deemed significant based upon the inclusion or exclusion of genes near the cutoff, can complicate downstream interpretation. A false discovery rate adjustment to the individual significance values helps limit the impact upon results.

Non-cutoff based approaches including Gene Set Enrichment Analysis (GSEA) are an alternative to cutoff based approaches, using the entire gene list output by a platform (Subramanian et al., 2005). Instead of categorically splitting the list, genes are ranked based upon their correlation to the compared conditions; highly correlated genes receive high absolute values. Gene sets receive Enrichment Scores (ES), a Kolmogorov-Smirnov based statistic that can be normalized for gene set size among all gene sets tested. The ES determines if the gene set is preferentially overrepresented by genes that correlate to one condition. The ES is based on working stepwise through the ranked list of genes in order, with the ES becoming the greatest absolute value reached. If the rank order of N genes tested in the experiment to form a list L of genes. The j^{th} gene in L has a correlation score $r(g_j) = r_j$ in relation to a particular condition. The g_j is considered a "hit" if it is a member of the gene set S which has the number of genes N_H, and is elsewise considered a "miss". The ES score is calculated as (Subramanian et al., 2005):

$$(N)ES = supremum|P_{hit}(S,i) - P_{miss}(S,i)|$$

$$P_{hit}(S,i) = \sum_{\substack{g_j \in S \\ j \leq i}} \frac{|r_j|^p}{N_R}$$

$$N_R = \sum_{g_j \in S} |r_j|^p$$

$$P_{miss}(S,i) = \sum_{\substack{g_j \notin S \\ j \leq i}} \frac{1}{(N - N_H)}$$

The ES can then be normalized (NES) by permutation to account for differences in gene set sizes and correlation between gene sets and the gene list data (Subramanian et al., 2005). The difference in the above calculations between the ES and the original Kolmogorov-Smirnov equation is the binary parameter $p \in \{0,1\}$. When $p=0$, the ES collapses to Kolmogorov-Smirnov, while when $p=1$, genes are weighted by their correlations with a normalization based upon the sum of gene set correlations (Subramanian et al., 2005). The GSEA method detects gene sets where member genes express at a lower level in comparison to the relatively higher individual expression required for detection by cutoff-based approaches (Subramanian et al., 2005).

When analyzing gene expression data at the level of gene sets, testing methods are separated based upon their null hypothesis definition, and the sampling model (Goeman and Bühlmann, 2007). The most popular null hypotheses are described as (Goeman and Bühlmann, 2007; Nam and Kim, 2008):

1. Competitive null hypothesis (Q1): *The genes in the gene-set are at most as often differentially expressed as the complementary set*;
2. Self-contained null hypothesis (Q2): *No genes in the gene-set are differentially expressed*;

3. and (Q3): *None of the gene sets considered is associated with the phenotype.*

The sampling model provides two possible categories (Goeman and Bühlmann, 2007):

1. Subject-sampling: Given *n* number of sample subjects; each subject is measured for the fixed set of genes *m*; each subject belongs to a tested condition *Y*. The i^{th} sample is thus represented by X_i, the *m*–dimensional vector of expression measurements, and Y_i as the condition such that :

$$(X_1, Y_1), \cdots, (X_n, Y_n)$$

2. Gene-sampling: Each subject sample consists of a set of observations of the *m* genes; if the i^{th} gene is part of the gene set A is recorded as A_i and whether it is a member of the differentially expressed gene list B is recorded as B_i; each subject sample provides thus provides the observation set:

$$(A_1, B_1), \cdots, (A_m, B_m)$$

Subject-sampling assumes independence between subjects with the possibility of gene expression correlations; gene-sampling assumes the independence of all gene expression observations (Goeman and Bühlmann, 2007). The correlation among genes within a gene set violates the assumptions of independence among genes in the gene-sampling approach. Hypergeometric approaches categorize as using a competitive null hypothesis and a gene-sampling model. The GSEA method will at times be uniquely categorized as Q3 (Nam and Kim, 2008; Wang et al., 2011). However, GSEA is typically categorized as using a competitive null hypothesis and a subject-sampling model (Goeman and Bühlmann, 2007; Dinu et al., 2008; Tarca et al., 2013; Maciejewski, 2014). Gene set analysis methods violating the assumption of sampling independence are expected to suffer from a high false-positive rate (Goeman and Bühlmann, 2007). However, the hypergeometric-based methods have demonstrated effective control of false-

positives (Tarca et al., 2013). Resampling methods such as GSEA are more debatable, with reports that resampling can or cannot control the rate of false-positives (Gatti et al., 2010; Tarca et al., 2013).

Network construction and visualizations

Genes do not impact the phenotypic level individually in a vacuum, nor as a listed category, but as a unit of complex networks (Wu and Chan, 2012). To increase our understanding and consequently identify more appropriate routes of diagnosis and treatment, it is potentially more beneficial to represent the transcriptomic data in a more biologically-valid network design. Basic network construction represents members of a set (genes; gene sets; pathways; etc.) as nodes and their relationships as edges. Networks can be parameterized based upon the number of nodes, edges, and their distribution within the network (Doncheva et al., 2012). For example, categorization by degree distribution is defined by how edges are distributed among nodes, or the degree distributions of the nodes (Barabási and Oltvai, 2004). One of the degree distribution categories is scale-free, a design that models many biological networks (Barabási and Oltvai, 2004; Albert, 2005). If the number of edges connecting to a particular node is *k*, then the probability of a node in a scale-free network possessing being of degree *k* is (Barabási and Oltvai, 2004):

$$P(k) \sim k^{-\gamma}$$

Where γ is the degree exponent and (Doncheva et al., 2012):

$$2 < \gamma < 3$$

Networks can also be parameterized based upon the shortest path between nodes as the number of edges traveled to connect two nodes, and evaluation of the connectedness within "neighborhoods" or sub-networks of a network based upon the average *k* for all neighbor nodes of a particular node

(Doncheva et al., 2012). A parameterized network provides quantifiable traits that can thus be compared among networks, and used to compare networks based upon their metrics (Assenov et al., 2008).

Cytoscape (http://www.cytoscape.org/) is an open-source platform for calculating and visualizing networks, capable of providing publication-ready visualizations (Killcoyne et al., 2009). Designed to accept multiple data input formats, the platform can read in data from standard network formats, XML based formats, delimited text files or Excel workbook files, and the output files from annotation programs to construct networks from the relationship data (http://www.cytoscape.org/what_is_cytoscape.html). As an openly accessible Java application, Cytoscape handles diverse data and interaction networks. Several plugins are available for customized functions, the most popular being designed to work with functional analysis data (Saito et al., 2012). The most popular plugin, BiNGO displays enriched terms from the GO database in a hierarchical fashion (Maere et al., 2005). The Enrichment Map plugin is designed to display enriched gene sets as nodes in a network linked by an overlap of enriching genes (Merico et al., 2010). The overlap in enriching genes is quantified from zero (no overlapping genes) to one when there is a complete overlap of enriching genes. This overlap is quantified by the Jaccard coefficient (JC) if gene sets are similar in size (Merico et al., 2010):

$$JC = \frac{|A \cap B|}{|A \cup B|}$$

or the overlap coefficient (OC) when dissimilar in size (Merico et al., 2010):

$$OC = \frac{|A \cap B|}{Min(|A|, |B|)}$$

When the goal is to integrate several networks together, Cytoscape comes pre-installed with AdvancedNetworkMerge and NetworkAnalyzer. While AdvancedNetworkMerge allows for basic set operations (union, intersection, difference) on multiple networks, available plugins offer additional utility (Saito et al., 2012). NetworkAnalyzer calculates topological network parameters for multiple networks so that they can be quantifiably compared (Assenov et al., 2008). Multiple plugins are available for producing Venn or Euler diagrams of shared network traits. Integrating networks and exploring the resultant design is possible with CABIN, albeit primarily designed for protein data (Singhal and Domico, 2007). The open-accessibility for plugin-design and support of several import/export file formats are designed to customize the network visualization aspects as needed (Smoot et al., 2011).

Thesis motivation

The goal of my studies is to characterize behavioral and unique transcriptomic characteristics in the BCG model of inflammation. Statistical modeling of the behavioral indicators that considers relationships among the indicators can identify relationships not apparent when analyzed independently. Large-scale gene expression analysis utilizing high-throughput platforms and network construction from the functional categories enriched by the gene expression profiles can identify the unique pathways for the biological conditions. Shared pathways and branching points to unique network traits can be examined for changes in expressed genes or gene expression levels for further analysis and exploration.

1.4 References

Aarsland, D., Påhlhagen, S., Ballard, C.G., Ehrt, U., Svenningsson, P. (2011). Depression in Parkinson disease—epidemiology, mechanisms, and management, *Nature Reviews Neurology*, **8**(1): 35-47. http://dx.doi.org/10.1038/nrneurol.2011.189

Albert, R. (2005). Scale-free networks in cell biology, *Journal of Cell Science*, **118**(Pt 21): 4947-4957. http://dx.doi.org/10.1242/jcs.02714

Anders, S., Reyes, A., Huber, W. (2012). Detecting differential usage of exons from RNA-seq data, *Genome Research*, **22**: 4025. http://dx.doi.org/10.1101/gr.133744.111

Anders, S., Huber, W. (2010). Differential expression analysis for sequence count data, *Genome Biology*, **11**: R106. http://dx.doi.org/10.1186/gb-2010-11-10-r106

Angoa-Pérez, M., Kane, M.J., Briggs, D.I., Herrera-Mundo, N., Sykes, C.E., Francescutti, D.M., Kuhn, D.M. (2014). Mice genetically depleted of brain serotonin do not display a depression-like behavioral phenotype, *ACS Chemical Neuroscience*, **5**(10): 908-919. http://dx.doi.org/10.1021/cn500096g

Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M.J., Natale, D.A., O'Donovan, C., Redaschi, N., Yek, L.S. (2004). UniProt: the universal protein knowledgebase, *Nucleic Acids Research*, **32**(Database issue): D115-119. http://dx.doi.org/10.1093/nar/gkh131

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G. (2000). Gene Ontology: tool for the unification of biology, *Nature Genetics*, **25**(1): 25-29. http://dx.doi.org/10.1038/75556

Assenov, Y., Ramírez, F., Schelhorn, S.E., Lengauer, T., Albrecht, M. (2008). Computing topological parameters of biological networks, *Bioinformatics*, **24**(2): 282-284. http://dx.doi.org/10.1093/bioinformatics/btm554

Aznar, S., Knudsen, G.M. (2011). Depression and Alzheimer's disease: is stress the initiating factor in a common neuropathological cascade? *Journal of Alzheimer's Disease*, **23**(2): 177-193. http://dx.doi.org/10.3233/JAD-2010-100390

Ball, P. (2013). DNA: celebrate the unknowns, *Nature*, **496**: 419-420. http://dx.doi.org/10.1038/496419a

Barabási, A.L., Oltvai, Z.N. (2004). Network biology: understanding the cell's functional organization, *Nature Reviews Genetics*, **5**(2): 101-113. http://dx.doi.org/10.1038nrg1272

Beutner, C., Linnartz-Gerlach, B., Schmidt, S.V., Beyer, M., Mallmann, M.R., Staratschek-Jox, A., Schultze, J.L., Neumann, H. (2013). Unique transcriptome signature of mouse microglia, *Glia*, **61**(9): 1429-1442. http://dx.doi.org/10.1002/glia.22524

Blume, J., Douglas, S.D., Evans, D.L. (2011). Immune suppression and immune activation in depression, *Brain Behavior and Immunity*, **25**(2): 221-229. http://dx.doi.org/10.1016/j.bbi.2010.10.008

Bonda, D.J., Lee, H.P., Kudo, W., Zhu, X., Smith, M.A., Lee, H.G. (2010). Pathological implications of cell cycle re-entry in Alzheimer disease, *Expert Reviews in Molecular Medicine*, **12**: e19. http://dx.doi.org/10.1017/S146239941000150X

Bottomly, D., Walter, N.A., Hunter, J.E., Darakjian, P., Kawane, S., Buck, K.J., Searles, R.P., Mooney, M., McWeeney, S.K., Hitzemann, R. (2011). Evaluating gene expression in C57Bl/6J and DBA/2J mouse striatum using RNA-seq and microarrays, *PLoS One*, **6**(3): e17820. http://dx.doi.org/10.1371/journal.pone.0017820

Boyle, E.I., Weng, S., Gollub, J., Jin, H., Botstein, D., Cherry, J.M., Sherlock, G. (2004). GO::TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes, *Bioinformatics*, **20**(18): 3710-3715. http://dx.doi.org/10.1093/bioinformatics/bth456

Bullard, J.H., Purdom, E., Hansen, K.D., Dudoit, S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-seq experiments, *BMC Bioinformatics*, **11**: 94. http://dx.doi.org/10.1186/1471-2105-11-94

Burrows, M., Wheeler, D.J. (1994). *A block-sorting lossless data compression algorithm*, Technical Report 124, Digital Equipment Corporation.

Buttgereit, F., Burmester, G.R., Brand, M.D. (2000). Bioenergetics of immune functions: fundamental and therapeutic aspects, *Immunology Today*, **21**(4): 192-199. http://dx.doi.org/10.1016/S0167-5699(00)01593-0

Castedo, M., Perfettini, J.L., Roumier, T., Andreau, K., Medema, R., Kroemer, G. (2004). Cell death by mitotic catastrophe: a molecular definition, *Oncogene*, **23**(16): 2825-2837. http://dx.doi.org/10.1038/sj.onc.1207528

Chomczynski, P., Sacchi, N. (2006). The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on, *Nature Protocols*, **1**(2): 581-585. http://dx.doi.org/10.1038/nprot.2006.83

Cock, P.J., Fields, C.J., Goto, N., Heuer, M.L., Rice, P.M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants, *Nucleic Acids Research*, **38**(6): 1767-1771. http://dx.doi.org/10.1093/nar/gkp1137

Croft, D., Mundo, A.F., Haw, R., Milacic, M., Weiser, J., Wu, G., Caudy, M., Garapati, P., Gillespie, M., Kamdar, M.R., Jassal, B., Jupe, S., Matthews, L., May, B., Palatnik, S., Rothfels, K., Shamovsky, V., Song, H., Williams, M., Birney, E., Hermjakob, H., Stein, L., D'Eustachio, P. (2014). The Reactome pathway knowledgebase, *Nucleic Acids Research*, **42**(D1): D472-D477. http://dx.doi.org/10.1093/nar/gkt1102

Danelishvili, L., McGarvey, J., Li, Y.J., Bermudez, L.E. (2003). *Mycobacterium tuberculosis* infection causes different levels of apoptosis and necrosis in human macrophages and alveolar

epithelial cells, *Cellular Microbiology*, **5**(9): 649-660. http://dx.doi.org/10.1046/j.1462-5822.2003.00312.x

Danelishvili, L., Yamazaki, Y., Selker, J., Bermudez, L.E. (2010). Secreted *Mycobacterium tuberculosis* Rv3654c and Rv3655c proteins participate in the suppression of macrophage apoptosis, *PLoS One*, **5**(5): e10474. http://dx.doi.org/10.1371/journal.pone.0010474

Dantzer, R., O'Connor, J.C., Lawson, M.A., Kelley, K.W. (2011). Inflammation-associated depression: from serotonin to kynurenine, *Psychoneuroendocrinology*, **36**(3): 426-436. http://dx.doi.org/10.1016/j.psyneuen.2010.09.012

Dedic N., Walser, S.M., Deussing, J.M. (2011). Mouse Models of Depression, Psychiatric Disorders - Trends and Developments, Dr. Toru Uehara (Ed.), ISBN: 978-953-307-745-1, InTech. http://dx.doi.org/10.5772/27498

Del Fabbro, C., Scalabrin, S., Morgante, M., Giorgi, F.M. (2013). An extensive evaluation of read trimming effects on Illumina NGS data analysis, *PLoS One*, **8**(12): e85024. http://dx.doi.org/10.1371/journal.pone.0085024

Dillies, M.A.., Rau, A., Aubert, J., Henneguet-Antier, C., Jeanmougin, M., Servant, N., Keime, C., Marot, G., Castel, D., Estelle, J., Guernec, G., Jagla, B., Jouneau, L., Laloë, D., Le Gall, C., Schaëffer, B., Le Crom, S., Guedj, M., Jaffrézic, F. (2013). A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis, *Brief Bioinformatics*, **14**(6): 671-683. http://dx.doi.org/10.1093/bib/bbs046

Dinu, I., Liu, Q., Potter, J.D., Adewale, A.J., Jhangri, G.S., Mueller, T., Einecke, G., Famulsky, K., Halloran, P., Yasui, Y. (2008). A biological evaluation of six gene set analysis methods for identification of differentially expressed pathways in microarray data, *Cancer Informatics*, **6**: 357-368. http://dx.doi.org/10.1093/bib/bbn042

Doncheva, N.T., Assenov, Y., Domingues, F.S., Albrecht, M. (2012). Topological analysis and interactive visualization of biological networks and protein structures, *Nature Protocols*, **7**(4): 670-685. http://dx.doi.org/10.1038/nprot.2012.004

Dopazo, J. (2009). Formulating and testing hypotheses in functional genomics, *Artificial Intelligence in Medicine*, **45**(2-3): 97-107. http://dx.doi.org/10.1016/j.artmed.2008.08.003

Drăghici, S., Sellamuthu, S., Khatri, P. (2006). Babel's tower revisited: a universal resource for cross-referencing across annotation databases, *Bioinformatics*, **22**(23): 2934-2939. http://dx.doi.org/10.1093/bioinformatics/btl372

Ehrt, S., Schnappinger, D., Bekiranov, S., Drenkow, J., Shi, S., Gingeras, T.R., Gaasterland, T., Schoolnik, G., Nathan, G. (2001). Reprogramming of the macrophage transcriptome in response to interferon-gamma and mycobacterium tuberculosis: signaling roles of nitric oxide synthase-2 and phagocyte oxidase, *Journal of Experimental Medicine*, **194**(8): 1123-1140. http://dx.doi.org/10.1084/jem.194.8.1123

Eyre, H., Baune, B.T. (2012). Neuroplastic changes in depression: a role for the immune system, *Psychoneuroendocrinology*, **37**(9): 1397-1416. http://dx.doi.org/10.1016/j.psyneuen.2012.03.019

Farooq, R.K., Isingrini, E., Tanti, A., Le Guisguet, A.M., Arlicot, N., Minier, F., Leman, S., Chalon, S., Belzung, C., Camus, V. (2012). Is unpredictable chronic mild stress (UCMS) a reliable model to study depression-induced neuroinflammation, *Behavioural Brain Research*, **231**(1): 130-137. http://dx.doi.org/10.1016/j.bbr.2012.03.020

Ferragina, P., Manzini, G. (2001). An experimental study of a compressed index, *Information Sciences*, **35**(1-2): 13-28. http://dx.doi.org/10.1016/S0020-0255(01)00098-6

Ferragina, P., Manzini, G. (2000). Opportunistic data structures and applications, *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, 390-398, 2000. http://dx.doi.org/10.1109/SFCS.2000.892127

Frank-Cannon, T.C., Alto, L.T., McAlpine, F.E., Tansey, M.G. (2009). Does neuroinflammation fan the flame in neurodegenerative diseases? *Molecular Neurodegeneration*, **4**: 47. http://dx.doi.org/10.1186/1750-1326-4-47

Fratazzi, C., Arbeit, R.D., Carini, C., Remold, H.G. (1997). Programmed cell death of *Mycobacterium avium* serovar 4-infected human macrophages prevents the mycobacteria from spreading and induces mycobacterial growth inhibition by freshly added, uninfect macrophages, *Journal of Immunology*, **158**(9): 4320-4327.

Frenois, F., Moreau, M., O'Connor, J., Lawson, M., Micon, C., Lestage, J., Kelley, K.W., Dantzer, R., Castanon, N. (2007). Lipopolysaccharide induces delayed FosB/DeltaFosB immunostaining within the mouse extended amygdala, hippocampus and hypothalamus, that

parallel the expression of depressive-like behavior, *Pyschoneroendocrinology*, **32**(5): 516-531. http://dx.doi.org/10.1016/j.psyneuen.2007.03.005

Gaidatzis, D., Jacobeit, K., Oakeley, E.J., Stadler, M.B. (2009). Overestimation of alternative splicing caused by variable probe characteristics in exon arrays, *Nucleic Acids Research*, **37**(16): e107. http://dx.doi.org/10.1093/nar/gkp508

Garber, M., Grabherr, M.G., Guttman, M., Trapnell, C. (2011). Computational methods for transcriptome annotation and quantification using RNA-seq, *Nature Methods*, **8**(6): 469-477. http://dx.doi.org/10.1038/nmeth.1613

García-Bueno, B., Caso, J.R., Leza, J.C. (2008). Stress as a neuroinflammatory condition in brain: damaging and protective mechanisms, *Neuroscience and Biobehavioral Reviews*, **32**(6): 1136-1151. http://dx.doi.org/10.1016/j.neubiorev.2008.04.001

Gartside, S.E., Umbers, V., Hajós, M., Sharp, T. (1995). Interaction between a selective 5-HT1A receptor antagonist and an SSRI in vivo: effects on 5-HT cell firing and extracellular 5-HT, *British Journal of Pharmacology*, **115**(6): 1064-1070. http://dx.doi.org/10.1111/j.14776-5381.1995.tb15919.x

Gatti, D.M., Barry, W.T., Nobel, A.B., Rusyn, I., Wright, F.A. (2010). Heading down the wrong pathway: on the influence of correlation within gene sets, *BMC Genomics*, **11**: 574. http://dx.doi.org/10.1186/1471-2164-11-574

Goeman, J.J., Bühlmann, P. (2007). Analyzing gene expression data in terms of gene sets: methodological issues, *Bioinformatics*, **23**(8): 980-987. http://dx.doi.org/10.1093/bioinformatics/btm051

Grant, G.R., Garkas, M.H., Pizarro, A.D., Lahens, N.F., Schug, J., Brunk, B.P., Stoeckert, C.J., Hogenesch, J.B., Pierce, E.A. (2011). Comparative analysis of RNA-seq alignment algorithms and the RNA-seq unified mapper (RUM), *Bioinformatics*, **27**(18): 2518-2528. http://dx.doi.org/10.1093/bioinformatics/btr427

Hanisch, U.K., Kettenmann, H. (2007). Microglia: active sensore and versatile effector cells in the normal and pathologic brain, *Nature Neuroscience*, **10**(11): 1387-1394. http://dx.doi.org/10.1038/nn1997

Hanisch, U.K. (2002). Microglia as a source and target of cytokines, *Glia*, **40**(2): 140-155. http://dx.doi.org/10.1002/glia.10161

Hansen, K.D., Brenner, S.E., Dudoit, S. (2010). Biases in Illumina transcriptome sequencing caused by random hexamer priming, *Nucleic Acids Research*, **38**(12): e131. http://dx.doi.org/10.1093/nar/gkq224

Hickman, S.E., Kingery, N.D., Ohsumi, T.K., Borowsky, M.L., Wang, L.C., Means, T.K., El Khoury, J. (2013). The microglial sensome revealed by direct RNA sequencing, *Nature Neuroscience*, **16**(12): 1896-1905. http://dx.doi.org/10.1038/nn.3554

Hogg and Tanis, *Probability and Statistical Inference: 8th ed.*, Prentice Hall, 2009.

Holt, R.A., Jones, S.J. (2008). The new paradigm of flow cell sequencing, *Genome Research*, **18**(6): 839-846, 2008. http://dx.doi.org/10.1101/gr.073262.107

Hosack, D.A., Dennis, G. Jr, Sherman, B.T., Lane, H.C., Lempicki, R.A. (2003). Identifying biological themes within lists of genes with EASE, *Genome Biology*, **4**(10): R70. http://dx.doi.org/10.1186/gb-2003-4-10-r70

Huang, da W., Sherman, B.T., Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths towards the comprehensive functional analysis of large gene lists, *Nucleic Acids Research*, **37**(1): 1-13. http://dx.doi.org/10.1093/nar/gkn923

Huang, da W., Sherman, B.T., Lempicki, R.A. (2009b) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, *Nature Protocols*, **4**(1): 44-57. http://dx.doi.org/10.1038/nprot.2008.211

Iwai, H., Kim, M., Yoshikawa, Y., Ashida, H., Ogawa, M., Fujita, Y., Muller, D., Kirikae, T., Jackson, P.K., Kotani, S., Sasakawa C. (2007). A bacterial effector targets Mad2L2, an APC inhibitor, to modulate host cell cycling, *Cell*, **130**(4): 611-623. http://dx.doi.org/10.1016/j.cell.2007.06.043

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG, *Nucleic Acids Research*, **42**(D1): D199-D205. http://dx.doi.org/10.1093/nar/gkt1076

Kanehisa, M., Goto, S., Kawashima, S., Nakaya, A. (2002). The KEGG databases at GenomeNet, *Nucleic Acids Research*, **30**(1): 42-46. http://dx.doi.org/10.1093/nar/30.1.42

Keller, C., Lauber, J., Blumenthal, A., Buer, J., Ehlers, S. (2004). Resistance and susceptibility to tuberculosis analysed at the transcriptome level: lessons from mouse macrophages, *Tuberculosis*, **84**(3-4): 144-158. http://dx.doi.org/10.1016/j.tube.2003.12.003

Killcoyne, S., Carter, G.W., Smith, J., Boyle, J. (2009). Cytoscape: A community-based framework for network modeling, *Methods in Molecular Biology*, **563**: 219-239. http://dx.doi.org/10.1007/978-1-60761-175-2_12

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., Salzberg, S.L. (2013). Tophat2: accurate alignment of transcriptomes in the presence of insertions, deletions, and gene fusions, *Genome Biology*, **14**(4): R36. http://dx.doi.org/10.1186/gb-2013-14-4-r36

Krebs, J.E., Goldstein, E.S., Kilpatrick, T. (2009). *Lewin's genes X: 10th ed.*, Sudbury, MA: Jones & Bartlett Publishers.

Kreisel, T., Frank, M.G., Licht, T., Reshef, R., Ben-Menachem-Zidon, O., Baratta, M.V., Maier, S.F., Yirmiya, R. (2014). Dynamic microglial alterations underlie stress-induced depress-like behavior and suppressed neurogenesis, *Molecular Psychiatry*, **19**: 699-709. http://dx.doi.org/10.1038/mp.2013.155

Kumar, R., Ichihashi, Y., Kimura, S., Chitwood, D.H., Headland, L.R., Peng, J., Maloof, J.N., Sinha, N.R. (2012). A high-throughput method for Illumina RNA-seq library preparation, *Frontiers in Plant Science*, **3**: 202. http://dx.doi.org/10.3389/fpls.2012.00202

Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, *Genome Biology*, **10**(3): R25. http://dx.doi.org/10.1186/gb-2009-10-3-r25

Lee, H., Ohno, M., Ohta, S., Mikami, T. (2013). Regular moderate or intense exercise prevents depression-like behavior without change of hippocampal tryptophan content in chronically tryptophan-deficient and stressed mice, *PLoS One*, **8**(7): e66996. http://dx.doi.org/10.1371/journal.pone.0066996

Lépine, J.P., Briley, M. (2011). The increasing burden of depression, *Neuropsychiatric Disease and Treatment*, **7**(Suppl 1): 3-7. http://dx.doi.org/10.2147/NDT.S19617

Lestage, J., Verrier, D., Palin, K., Dantzer, R. (2002). The enzyme indoleamine 2,3-dioxygenase is induced in the mouse brain in response to peripheral administration of lipopolysaccharide and superantigen, *Brain Behavior and Immunity*, **16**(5): 596-601. http://dx.doi.org/10.1016/S0889-1591(02)00014-4

Levy, S., Sutton, G., Ng, P.C., Feuk, L., Halpern, A.L., Walenz, B.P., Axelrod, N., Huang, J., Kirkness, E.F., Denisov, G., Lin, Y., MacDonald, J.R., Pang, A.W., Shago, M., Stockwell, T.B., Tsiamouri, A., Bafna, V., Bansal, V., Kravitz, S.A., Busam, D.A., Beeson, K.Y., McIntosh, T.C., Remington, K.A., Abril, J.F., Gill, J., Borman, J., Rogers, Y.H., Frazier,

M.E., Scherer, S.W., Strausberg, R.L., Venter, J.C. (2007). The diploid genome sequence of an individual human, PLoS Biology, **5**(10): e254. http://dx.doi.org/10.1371/journal.pbio.0050254

Li, H., Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform, *Bioinformatics*, **25**(14): 1754-1760. http://dx.doi.org/10.1093/bioinformatics/btp324

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., Law, M. (2012). Comparison of next-generation sequencing systems, *Journal of Biomedicine and Biotechnology*, **2012**: 251364. http://dx.doi.org/10.1155/2012/251364

Livak, K.J., Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method, *Methods*, **25**(4): 402-408. http://dx.doi.org/10.1006/meth.2001.1262

Maciejewski, H. (2014). Gene set analysis methods: statistical models and methodological differences, *Briefings in Bioinformatics*, **15**(4): 504-518. http://dx.doi.org/10.1093/bib/bbt002

MacManes, M.D. (2014). On the optimal trimming of high-throughput mRNA sequence data, *Frontiers in Genetics*, **5**:13. http://dx.doi.org/10.3389/fgene.2014.00013

Maere, S., Heymans, K., Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks, *Bioinformatics*, **21**(16): 3448-3449. http://dx.doi.org/10.1093/bioinformatics/bti551

Maglott, D., Ostell, J., Pruitt, K.D., Tatusova, T. (2005). Entrez Gene: gene-centered information at NCBI, *Nucleic Acids Research*, **33**(Database issue): D54-58. http://dx.doi8.org/10.1093/nar/gki031

Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays, *Genome Research*, **18**(9): 1509-1517. http://dx.doi.org/10.1101/gr.079558.108

McGeer, P.L., McGeer, E.G. (2011). History of innate immunity in neurodegenerative disorders, *Frontiers in Pharmacology*, **2**: 77. http://dx.doi.org/10.3389/fphar.2011.00077

Merico, D., Isserlin, R., Stueker, O., Emili, A., Bader, G.D. (2010). Visualizing gene-set enrichment results using Cytoscape plug-in enrichment map, *PLoS One*, **5**(11): e13984. http://dx.doi.org/10.1371/journal.pone.0013984

Metzker, M.L. (2010). Sequencing technologies – the next generation, *Nature Reviews* Genetics, **11**(1): 31-46. http://dx.doi.org/10.1038/nrg2626

Miller, A.H., Maletic, V., Raison, C.L. (2009). Inflammation and its discontents: the role of cytokines in the pathophysiology of major depression, *Biological Psychiatry*, **65**(9): 732-741. http://dx.doi.org/10.1016/j.biopsych.2008.11.029

Minoche, A.E., Dohm, J.C., Himmelbauer, H. (2011). Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems, *Genome Biology*, **12**(11): R112. http://dx.doi.org/10.1186/gb-2011-12-11-r112

Moreau, M., André, C., O'Connor, J.C., Dumich, S.A., Woods, J.A., Kelley, K.W., Dantzer, R., Lestage, J., Castanon, N. (2008). Inoculation of Bacillus Calmette-Guérin to mice induces an acute episode of sickness behavior followed by chronic-depressive like behavior, *Brain, Behavior, and Immunity*, 192(3): **22**(7): 1087-1095. http://dx.doi.org/10.1016/j.bbi.2008.04.001

Moreau, M., Lestage, J., Verrier, D., Mormede, C., Kelley, K.W., Dantzer, R., Castanon, N. (2005). Bacille Calmette-Guérin inoculation induces chronic activation of peripheral and brain indoleamine 2,3-dioxygenase in mice, *Journal of Infectious Diseases*, **192**(3): 537-544. http://dx.doi.org/10.1086/431603

Morozova, O., Hirst, M., Marra, M.A. (2009). Applications of new sequencing technologies for transcriptome analysis, *Annual Review of Genomics and Human Genetics*, **10**: 135-151. http://dx.doi.org/10.1146/annurev-genom-082908-145957

Morse, S.M., Shaw, G., Larner, S.F. (2006). Concurrent mRNA and protein extraction from the same experimental sample using a commercially available column-based RNA preparation kit, *BioTechniques*, **40**(1): 54-58.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-seq, *Nature* Methods, **5**(7): 621-628. http://dx.doi.org/10.1038/nmeth.1226

Nagalakshmi, U., Waern, K., Snyder, M. (2010). RNA-seq: a method for comprehensive transcriptome analysis, *Current Protocols in Molecular Biology*, **4**(4): 4.11.1-4.11.13. http://dx.doi.org/10.1002/0471142727.mb0411s89

Nakamura, K., Oshima, T., Morimoto, T., Ikeda, S., Yoshikawa, H., Shiwa, Y., Ishikawa, S., Linak, M.C., Hirai, A., Takahashi, H., Altaf-Ul-Amin, M., Ogasawara, N., Kanaya, S. (2011). Sequence-specific error profile of Illumina sequencers, *Nucleic Acids Research*, **39**(13): e90. http://dx.doi.org/10.1093/nar/gkr344

Nam, D., Kim, S.Y. (2008). Gene-set approach for expression pattern analysis, *Briefings in Bioinformatics*, **9**(3): 189-197. http://dx.doi.org/10.1093/bib/bbn001

Nikodemova, M., Watters, J.J. (2012). Efficient isolation of live microglia with preserved phenotypes from adult mouse brain, *Journal of Neuroinflammation*, **9**: 147. http://dx.doi.org/10.1186/1742-2094-9-147

O'Connor, J.C., Lawson, M.A., André, C., Briley, E.M., Szegedi, S.S., Lestage, J., Castanon, N., Herkenham, M., Dantzer, R., Kelley, K.W. (2009a). Induction of IDO by Bacille Calmette-Guerin is responsible for development of murine depressive-like behavior, *Journal of Immunology*, **182**(5): 3202-3212, 2009a. http://dx.doi.org/10.4049/jimmunol.0802722

O'Connor, J.C., Lawson, M.A., André, C., Moreau, M., Lestage, J., Castanon, N., Kelley, K.W., Dantzer, R. (2009b). Lipopolysaccharide-induced depressive-like behavior is mediated by indoleamine 2,3-dioxygenase activation in mice, *Molecular Psychiatry*, **14**(5): 511-522. http://dx.doi.org/10.1038/sj.mp.4002148

Oswald, E., Nougayrède, J.P., Taieb, F., Sugai, M. (2005). Bacterial toxins that modulate host cell-cycle progression, *Current Opinion in Microbiology*, **8**(1): 83-91. http://dx.doi.org/10.1016/j.mib.2004.12.011

Quail, M.A., Kozarewa, I., Smith, F., Scally, A., Stephens, P.J., Durbin, R., Swerdlow, H., Turner, D.J. (2008). A large genome center's improvements to the Illumina sequencing system, *Nature Methods*, **5**(12): 1005-1010. http://dx.doi.org/10.1038/nmeth.1270

Raison, C.L., Miller, A.H. (2011). Is depression an inflammatory disorder, *Current Psychiatry Reports*, **13**(6): 467-475. http://dx.doi.org/10.1007/s11920-011-0232-0

Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., Mason, C.E., Socci, N.D., Betel, D. (2013). Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data, *Genome Biology*, **14**(9): R95. http://dx.doi.org/10.1186/gb-2013-14-9-r95

Reno, C., Marchuk, L., Sciore, P., Frank, C.B., Hart, D.A. (1997). Rapid isolation of total RNA from small samples of hypocellular, dense connective tissues, *BioTechniques*, **22**(6): 1082-1086.

Rivals, I., Personnaz, L., Taing, L., Potier, M.C. (2007). Enrichment or depletion of a GO category within a class of genes: which test? *Bioinformatics*, **23**(4): 401-407. http://dx.doi.org/10.1093/bioinformatics/btl633

Robinson, M.D., Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data, *Genome Biology*, **11**(3): R25. http://dx.doi.org/10.1186/gb-2010-11-3-r25

Robinson, M.D., Smyth, G.K. (2007). Moderated statistical tests for assessing differences in tag abundance, *Bioinformatics*, **23**(21): 2881-2887. http://dx.doi.org/10.1093/bioinformatics/btm453

Robinson, M.D., McCarthy, D.J., Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, *Bioinformatics*, **26**(1):139-140. http://dx.doi.org/10.1093/bioinformatics/btp616

Robles, J.A., Qureshi, S.E., Stephen, S.J., Wilson, S.R., Burden, C.J., Taylor, J.M. (2012). Efficient experimental design and analysis strategies for the detection of differential expression using RNA-sequencing, *BMC Genomics*, **13**:484. http://dx.doi.org/10.1186/1471-2164-13-484

Rodriguez-Zas, S.L. (2013). Functional or pathway analysis and network visualization. Adapted from class notes of ANSC 545 – Statistical Genomics.

Rosenblatt, A. (2007). Neuropsychiatry of Huntington's disease, *Dialogues in Clinical Neuroscience*, **9**(2): 191-197.

Saito, R., Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.L., Lotia, S., Pico, A.R., Bader, G.D., Ideker, T. (2012). A travel guide to Cytoscape plugins, *Nature Methods*, **9**(11): 1069-1076. http://dx.doi.org/10.1038/nmeth.2212

Salzman, J., Jiang, H., Wong, W.H. (2011). Statistical modeling of RNA-seq data, *Statistical Science*, **26**(1): 62-83. http://dx.doi.org/10.1214/10-STS343

Schnappinger, D., Schoolnik, G.K., Ehrt, S. (2006). Expression profiling of host pathogen interactions: how Mycobacterium tuberculosis and the macrophage adapt to one another, *Microbes and Infection*, **8**(4): 1132-1140. http://dx.doi.org/10.1016/j.micinf.2005.10.027

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

Singhal, M., Domico, K. (2007). CABIN: collective analysis of biological interaction networks, *Computational Biology and Chemistry*, **31**(3): 222-225. http://dx.doi.org/10.1016/j.compbiolchem.2007.03.006

Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.L., Ideker, T. (2011). Cytoscape 2.8: new features for data integration and network visualization, *Bioinformatics*, **27**(3): 431-432. http://dx.doi.org/10.1093/bioinformatics/btq675

Soneson, C., Delorenzi, M. (2013). A comparison of methods for differential expression analysis of RNA-seq data, *BMC Bioinformatics*, **14**:91. http://dx.doi.org/10.1186/1471-2105-14-91

Stobbe, M.D., Houten, S.M., Jansen, G.A., van Kampen, A.H., Moerland, P.D. (2011). Critical assessment of human metabolic pathway databases: a stepping stone for future integration, *BMC Systems Biology*, **5**: 165. http://dx.doi.org/10.1186/1752-0509-5-165

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, *Proceedings of the National Academy of Sciences*, **102**(43): 15545-15550. http://dx.doi.org/10.1073/pnas.0506580102

Svoboda, N., Zierler, S., Kerschbaum, H.H. (2007). cAMP mediates ammonia-induced programmed cell death in microglial cell line BV-2, *The European Journal of Neuroscience*, **25**(8): 2285-2295. http://dx.doi.org/10.1111/j.1460-9568.2007.05452.x

Tan, S.C., Yiap, B.C. (2009). DNA, RNA, and protein extraction: the past and the present, *Journal of Biomedicine and Biotechnology*, **2009**: 574398. http://dx.doi.org/10.1155/2009/574398

Tarca, A.L., Bhatti, G., Romero, R. (2013). A comparison of gene set analysis methods in terms of sensitivity, prioritization and specificity, *PLoS One*, **8**(11): e79217. http://dx.doi.org/10.1037/journal.pone.0079217

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq, *Nature Biotechnology*, **31**(1): 46-53. http://dx.doi.org/10.1038/nbt.2450

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with Tophat and Cufflinks, *Nature Protocols*, **7**(3): 562-578. http://dx.doi.org/10.1038/nprot.2012.016

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., Pachter, L. (2010). Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation, *Nature Biotechnology*, **28**(5): 511-515.

VanGuilder, H.D., Vrana, K.E., Freeman, W.M. (2008). Twenty-five years of quantitative PCR for gene expression analysis, *Biotechniques*, **44**(5): 619-626. http://dx.doi.org/10.2144/000112776

Wang, L., Dedow, L.K., Shao, Y., Liu, P., Brutnell, T.P. (2011). A low-cost library construction protocol and data analysis pipeline for Illumina-based strand-specific multiplex RNA-seq, *PLoS One*, **6**(10): e26426. http://dx.doi.org/10.1371/journal.pone.0026426

Wang, Z., Gerstein, M., Snyder, M. (2009). RNA-seq: a revolutionary tool for transcriptomics, *Nature Review Genetics*, **10**(1): 57-63. http://dx.doi.org/10.1038/nrg2484

Wang, C., Gong, B., Bushel, P.R., Thierry-Mieg, J., Thierry-Mieg, D., Xu, J., Fang, H., Hong, H., Shen, J., Su, Z., Meehan, J., Li, X.,Yang, L., Li, H., Łabaj, P.P., Kreil, D.P., Megherbi, D., Gaj, S., Caiment, F., van Delft, J., Kleinjans, J., Scherer, A., Devanarayan, V., Wang, J., Yang, Y., Qian, H.R., Lancashire, L.J., Bessarabova, M., Nikolsky, Y., Furlanello, C., Chierici, M., Albanese, D., Jurman, G., Riccadonna, S., Filosi, M., Visintainer, R., Zhang, K.K., Li, J., Hsieh, J.H., Svoboda, D.L., Fuscoe, J.C., Deng, Y., Shi, L., Paules, R.S., Auerbach, S.S., Tong, W. (2014). The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance, *Nature Biotechnology*, **32**(9): 926-932. http://dx.doi.org/10.1038/nbt.3001

Wilhelm, B.T., Marguerat, S., Goodhead, I., Bähler, J. (2010). Defining transcribed regions using RNA-seq, *Nature Protocols*, **5**(2): 255-266. http://dx.doi.org/10.1038/nprot.2009.229

World Health Organization, 2012, *Depression- Fact Sheet*, viewed Sept 2014, http://www.who.int/mediacentre/factsheets/fs369/en/

Wu, M., Chan, C. (2012). Learning transcriptional regulation on a genome scale: A theoretical analysis based on gene expression data, *Briefings in Bioinformatics*, **13**(2): 150-161. http://dx.doi.org/10.1093/bib/bbr029

Wyss-Coray, T,. Mucke, L. (2002). Inflammation in neurodegenerative disease—a double-edged sword, *Neuron*, **35**(3): 419-432. http://dx.doi.org/10.1016/S0896-6273(02)00794-8

Yang, X., Chockalingam, S.P., Aluru, S. (2013). A survey of error-correction methods for next-generation sequencing, *Briefings in Bioinformatics*, **14**(1): 56-66. http://dx.doi.org/10.1093/bib/bbs0115

Yirmiya, R. (1996). Endotoxin produces a depressive-like episode in rats, *Brain Research*, **711**(1-2): 163-174. http://dx.doi.org/10.1016/0006-8993(95)01415-2

Zhong, S., Joung, J.G., Zheng, Y., Chen, Y.R., Liu, B., Shao, Y., Xiang, J.Z., Fei, Z., Giovannoni, J.J. (2011). High-throughput Illumina strand-specific RNA sequencing library preparation, *Cold Spring Harbor Protocols*, **2011**(8): 940-949. http://dx.doi.org/10.1101/pdb.prot5652

1.5 Tables

Table 1.1 DAVID knowledgebase database sources
(from list at http://david.abcc.ncifcrf.gov/content.jsp?file=Acknowledgement.htm)

Primary – gene clustering	
NCBI Entrez Gene	http://www.ncbi.nlm.nih.gov/Entrez
UniProt and UniRef100	http://www.uniprot.org
PIR NREF and iProClass	http://pir.georgetown.edu
Secondary – additional annotation information	
Affymetrix Probeset Mapping: NetAffx; TIGR	http://www.netaffx.com; http://www.tigr.org/tdb
Ontologies Entrez Gene; PIR iProClass; UniProt; GOA; Panther Ontology; Gene Ontology	http://www.ncbi.nlm.nih.gov/Entrez; http://pir.georgetown.edu/iproclass; http://www.uniprot.org; http://www.ebi.ac.uk/GOA; http://www.pantherdb.org; http://www.geneontology.org
Protein Domains Entrez Gene; PIR iProClass; UniProt; InterPro; Pfam; COG/KOG; Blocks; SMART; PDB; ProDom; PROSITE; TIGRFAMS; PRINTS; Panther Family; SCOP	http://www.ncbi.nlm.nih.gov/Entrez; http://pir.georgetown.edu/iproclass; http://www.uniprot.org; http://www.ebi.ac.uk/interpro; http://www.sanger.ac.uk/Software/Pfam; http://www.ncbi.nlm.nih.gov/COG/new; http://blocks.fhcrc.org; http://smart.embl-heidelberg.de; http://www.rcsb.org/pdb; http://protein.toulouse.inra.fr/prodom/current/html/home.php; http://au.expasy.org/prosite; http://www.tigr.org/TIGRFAMs/index/shtml; http://www.bioinf.man.ac.uk/dbbrowser/PRINTS; http://www.pantherdb.org; http://scop.mrc-lmb.cam.ac.uk/scop
Pathways KEGG Pathways, Reaction, and Compound; CGAP BioCarta; Pathways; Biocarta; BBID; PID; Panther Pathway; REACTOME	http://www.genome.jp/kegg; http://cgap.nci.nih.gov/Pathways/BioCart_Pathways; http://www.biocarta.com/genes/index.asp; http://bbid.grc.nia.nih.gov; http://pid.nci.nih.gov; http://www.pantherdb.org; http://www.reactome.org
General Annotation Entrez Gene; PIR iProClass; UniProt	http://www.ncbi.nlm.nih.gov/Entrez; http://pir.georgetown.edu/iproclass; http://www.uniprot.org
Functional Categories Entrez Gene; PIR iProClass; UniProt; COG/KOG	http://www.ncbi.nlm.nih.gov/Entrez; http://pir.georgetown.edu/iproclass; http://www.uniprot.org; http://www.ncbi.nlm.nih.gov/COG/new
Protein Interactions HIV Interactions (Entrez and RefSeq); NCICB caPathway; Bind; Mint; HPRD; DIP; REACTOME	http://www.ncbi.nlm.nih.gov/Entrez; http://www.ncbi.nlm.nih.gov/RefSeq/HIVInteractions; http://ncicb.nci.nih.gov; http://www.clueprint.org/bind/bind.php; http://mint.bio.uniroma2.it/mint; http://www.hprd.org; http://dip.doe-mbi.ucla.edu; http://www.reactome.org
Literature PubMedIDs (Entrez Gene, PIR iProClass, UniProt) GeneRIF (Entrez Gene)	http://www.ncbi.nlm.nih.gov/Entrez; http://pir.georgetown.edu/iproclass; http://www.uniprot.org;
Disease Genetic Association Database OMIM Phenotype (Entrez Gene and OMIM)	http://geneticassociationdb.nih.gov; http://www.ncbi.nlm.nih.gov/Entrez; http://www.ncbi.nlm.nih.gov/omim
Tissue Expression GNF Microarray; CGAP SAGE; CGAP Tissue EST; NCBI Unigene EST Profile	http://wombat.gnf.org/index.html; http://cgap.nci.nih.gov/SAGE; http://cgap.nci.nih.gov/Tissues; ftp://ftp.ncbi.nih.gov/repository/UniGene/Homo_sapiens
Other Data Sources Homologene	http://www.ncbi.nlm.nih.gov/homologene

Table 1.2 Annotation Categories in DAVID

Type of Category	Categories
Ontology	GO_BIOLOGICAL PROCESS GO_MOLEDCULAR FUNCTION GO_CELLULAR COMPONENT PANTHER_BIOLOGICAL PROCESS PANTHER_MOLECULAR FUNCTION COG_KOG_ONTOLOGY
Protein Domain/Family	BLOCKS_ID COG_KOG_NAME INTERPRO_NAME PDB_ID PFAM_NAME PIR_ALN PIR_HOMOLOGY_DOMAIN PIR_SUPERFAMILY_NAME PRINTS_NAME PRODOM_NAME PROSITE_NAME SCOP_ID SMART_NAME TIGRFAMS_NAME PANTHER_SUBFAMILY PATHER_FAMILY
Sequence Features	ALIAS_GENE_SYMBOL CHROMOSOME CYTOBAND GENE_NAME GENE_SYMBOL HOMOLOGOUS_GENE ENTREZ_GENE_SUMMARY OMIM_ID PIR_SUMMARY PROTEIN_MW REFSEQ_PRODUCT SEQUENCE_LENGTH SP_COMMENT
P-P Interaction	BIND DPI MINT NCICB_CAPATHWAY TRANSFAC_ID HIV_INTERACTION HIV_INTERACTION_CATEGORY HPRD_INTERACTION REACTOME_INTERACTION
Disease Association	GENETIC_ASSOCIATION_DB OMIM_DISEASE
Literature	GENERIF_SUMMARY PUBMED_ID HIV_INTERACTION_PUBMED_ID
Pathways	BioCarta KEGG_PATHWAY PANTHER_PATHWAY PID BBID KEGG_REACTION
Functional Category	PIR_SEQ_FEATURE SP_COMMENT_TYPE SP_PIR_KEYWORDS UP_SEQ_FEATURE
Gene Tissue Expression	GNF Microarray UNIGENE EST CGAP SAGE CGAP EST

(Source: Huang et al., 2009b supplement)

CHAPTER II: Exploring the effect of covariates on univariate modeling of murine behavior after Bacillus Calmette-Guérin challenge

2.1 Abstract

In the mouse model of Bacille Calmette Guérin (BCG)-induced inflammation, individual and independent analysis of behavioral indicators is common. Model design can benefit from considering the established relationships among indicators in the model design, but no study has explored the improvement upon univariate analysis of these behavioral indicators. The objectives of this study were: (1) to characterize the individual response of behavioral indicators in the BCG-induced inflammation model, and (2) to characterize the impact of covariates on the univariate analysis of behavior indicators. Adult male mice were injected with BCG at a level of 10 mg, 5 mg, or a saline control. Daily body weights were measured with weight loss from Day 0 to Day 2 (p-value < 0.01) followed by weight recovery that became non-significant between Day 4 and Day 5 (p-value < 0.01). On Day 6, the sickness indicators of locomotor activity and rearing were measured as well as the depression-like indicators of increased immobility forced swim test and tail suspension test, and finally sucrose preference testing on Day 7. The modeling of all behavior indicators was determined by significant (X^2 < 0.05) reduction in log likelihood ratios, combined with the consideration of Akaike's and Bayesian Information Criteria. A significant improvement in log likelihood was found for sucrose preference and tail suspension test data by including covariates in the model when compared to a model without covariates. Including age as a covariate with a cubic relationship in sucrose preference identified a significant treatment effect (p-value < 0.05). Similarly, including body weight recovery from Day 2 to Day 5 as a covariate with a cubic relationship when modeling tail suspension test immobility identified a borderline significant

treatment effect (p-value = 0.055) that was not identified by models lacking covariates. Body weight differences between treatment groups were identified when the model included no covariates, while including Day -1 body weight as a linear covariate identified a significant difference between the two BCG treatment levels at Day 2 (p-value = 0.0007). Rearing and locomotor activity were sensitive to covariate inclusion with significant treatment effects arising, while forced swim results showed no significant treatment effect regardless of covariate inclusion in the model. Univariate analysis can effectively capture the treatment level effects on body weight, mobility-based indicators, and sucrose preference with the inclusion of covariates.

2.2 Introduction

The immune system response to stimuli involves the ability to ellicit symptoms of sickness response, and a less understood association to depression (Jones and Thomsen, 2013). As an animal model of chronic immune activation, Bacille Calmette-Guérin (BCG) administration in mice results in persistent mycobacterial dissemination throughout the body (Tsenova et al., 1999) and a long-lived inflammatory response with detectable difference found four weeks post-challenge (Moreau et al., 2005). Systemic challenges induced by BCG include an initial sickness response involving anorexia, weight-loss, and lethargy, and depression-like behaviors including anhedonia and learned-helplessness which can last for up to 3 weeks post-challenge (Moreau et al., 2008). However studies on the behavioral response following BCG-treatment remain focused on a measurement subset which includes body weight, locomotor activity, anhedonia, and despair (Moreau et al., 2008; O'Connor et al., 2009; Kelley et al., 2013; Painsipp et al., 2013; Saleh et al., 2014; Rodriguez-Zas et al., 2015). Previous studies of the behavioral response to infection have analyzed response metrics with no or minimal covariates (Moreau et al., 2008; Painsipp et al.,

2011; Schmuckermair et al., 2013). The objectives of this study were: (1) to characterize the individual response of behavioral indicators in the BCG-induced inflammation model, and (2) to characterize the impact of covariates on the univariate analysis of behavior indicators.

2.3 Materials and Methods

Experiment

Male adult C57Bl/6J mice (n=19) were selected for the experiment, ranging from 91 – 107 days old with a median age of 96 days. Access to water and chow (Teklad 8640, Harlan Laboratories, Indianapolis, IN, USA) were provided *ad libitum*. Mice were individually housed in a temperature- (23° C) and humidity-controlled (45%) room maintained on a 12-hour reverse light/dark cycle, and handled daily for at least 7 days prior to experimentation for acclimation purposes. Only male adult mice were used to avoid behavioral variations associated with hormonal changes from the estrous cycle (Ellacott et al., 2010).

For immune challenge, fresh solutions were prepared by dispersing lyophilized cultures of BCG (TICE, Organon USA Inc., USA) into sterile endotoxin-free isotonic saline on the day of injection. Lyophilized cultures were 50 mg wet weight, containing 1 to 8 x 10^8 colony forming units (cfu). Three treatment levels were evaluated: 1) intraperitoneal (i.p.) administration of a 10 mg dose of BCG, between 2 x 10^7 and 1.6 x 10^8 cfu (High BCG, n=6); 2) a 5 mg dose, between 1 x 10^7 and 8 x 10^7 cfu (Low BCG, n=6); and 3) i.p. administration of saline (control, n=7). All administrations were a standardized 0.3 mL volume. Measurements of sickness and depression-like behaviors were taken across time in an immune-response experiment that spanned a week after the injection (**Figure 2.1**). All experiments were conducted in accordance with the Animal Care and Use of

Laboratory Program approved by the Institutional Animal Care and Use Committee at the University of Illinois.

Measurements of Sickness Behaviors

Measuring body weight encompasses a broad spectrum of sickness effects, involving anorexia and modifications to metabolic homeostasis (Ellacott et al., 2010). Body weight was recorded daily at the beginning of the dark cycle, with all records in grams (g). Recordings began the day before treatment (Day -1) for a baseline, immediately before treatment (Day 0), and continued afterwards for one week (Day 7). On Day 6 immediately following the daily body weight measurement, locomotor activity was used to measure the impact of sickness on exploratory behavior and mobility (Gould et al., 2009). Following O'Connor et al. (2009), each mouse was place in a standard acrylic housing cage with opaque walls. A 5-minute test session was video recorded and analyzed by a trained observer blind to treatment. The cage was divided into virtual sectors and locomotor activity measured as the number of line-crossings between quadrants while rearing was measured as the number of times the mice stood on their hind paws. Four mice could be tested simultaneously and the group order was recorded.

Measurements of Depression-like Behaviors

Three measures of depression-like behavior were measured: the forced swim and tail suspension tests measured behavioral despair while sucrose preference served as an anhedonic measure (Castagné et al., 2011; Dedic et al., 2011). Forced swim testing was performed after locomotor activity testing on Day 6 (O'Connor et al., 2009). Mice were placed into an opaque cylinder with water maintained at approximately 23°C, with activity recorded for 6 minutes. Time spent

immobile was recorded during the last five minutes by a trained observer blind to treatment, with tail suspension following after the forced swim testing (O'Connor et al., 2009). During the Tail Suspension Test (TST), mice were suspended by their tails to a computerized strain gauge that collects the movement data for each mouse. Depression was interpreted as learned helplessness when the duration of immobility increases compared to the control group of mice (Steru et al., 1985). Following O'Connor et al (2009), the test was conducted for 10 minutes with time spent immobile recorded using the Mouse Tail Suspension Package software (MED-TSS-MS; Med Associates). Program settings were:

1) integration on;
2) resolution (data acquisition rate) 10 milliseconds;
3) gain (voltage signal amplification for animal weight): 4 input/output ratio;
4) start trigger (necessary load to begin recording): 20% of maximum (50 g load cell).

Settings were determined according to the manufacturer's instructions with consideration of the weight of the mouse strain. The lower threshold setting for the load cell, which was the user-defined load-cell detection cutoff determining if time was recorded as mobile activity or immobility, was adjusted manually based upon measurement of immobile periods for each mouse individually. Time spent motionless between 3 and 8 minutes after original suspension were analyzed. Similar to the locomotor activity recording, only 4 mice could be recorded at a time in the TST and the testing order was recorded. Consumption of a sucrose solution was used as a measurement of anhedonic behavior, a popular non-invasive method of measurement (Moreau et al., 2008; Steensland et al., 2010; Dedic et al., 2011; Kelley et al., 2013). Following Lawson et al. (2013), mice were provided two bottles in the home cage, containing water and 1% sucrose solution, respectively. The bottles were weighed after a 24-hour period, and the difference in

weight was attributed to consumption. Sucrose and water were preliminarily measured before treatment (Day -1), and after activity measurements (Day 6). Timing of measurement was immediately following the body weight measurement, at the beginning of the dark cycle. The position of the two bottles was switched throughout training and during the experiment to avoid positional preferences. Sucrose preference was expressed as sucrose / (sucrose + water) * 100. Switching the bottle positions and use of the sucrose preference measurement controlled for the potential confounds of side preference and high variability among mice on liquid intake volume, respectively (Strekalova et al., 2011).

Statistical Analysis

Univariate linear models describe the response of a single dependent variable, here the specific behavioral indicator to be modeled, as it relates to one or more independent variables. Prior investigations include the fixed effect of treatment (BCG versus control), time (days) when repeated measures are performed, and testing for interaction between the effects (Moreau et al., 2008; Painsipp et al., 2011; Kelley et al., 2013; Schmuckermair et al., 2013). Treatment, time, and their interactions were treated as fixed effects due to interest in examining specific levels.

Body weight (Day 0 to Day 2; Day 2 to Day 5) models

Preliminary analysis determined that there were no differences from Day 5 to Day 7, within or between treatment groups (p-value > 0.1). As the maximum weight loss following BCG-treatment occurred on Day 2, analysis of body weights was split to more specifically model the body weight loss from Day 0 to Day 2, and body weight recovery from Day 2 to Day 5.

The model for body weight measurements from Day 0 to Day 2 and Day 2 to Day 5 without covariates was:

$$Y_{ijk} = \mu + T_i + D_j + TD_{ij} + e_{ijk} \quad \text{I}$$

Y_{ijk} = Body weight measurement of the kth mouse in the ith treatment (T_i); μ = The overall body weight mean across all mice; T_i = The ith fixed effect of treatment; D_k = The fixed effect of the jth day over which body weight was measured; TD_{ij} = The fixed effect of the interaction between ith treatment and jth day; e_{ijk} = The residual associated with the measurement Y_{ijk}.

The full model, combining all tested covariates was:

$$Y_{ijk} = \mu + T_i + D_j + TD_{ij} + b_1\left(B_{ijk} - \bar{B}\right) + b_2\left(F_{ijk} - \bar{F}\right) + b_3\left(U_{ijk} - \bar{U}\right) + b_4\left(S_{ijk} - \bar{S}\right) + b_5\left(A_{ijk} - \bar{A}\right) + e_{ijk} \quad \text{II}$$

Y_{ijk} = Body weight measurement of the kth mouse in the ith treatment (T_i); μ = The overall body weight mean across all mice; T_i = The ith fixed effect of treatment; D_k = The fixed effect of the jth day over which body weight was measured; TD_{ij} = The fixed effect of the interaction between ith treatment and jth day; b_1 = The regression coefficient for the linear covariate of body weight on Day -1 (B_{ijk}); b_2 = The regression coefficient for the linear covariate of forced swim measurement (F_{ijk}); b_3 = The regression coefficient for the linear covariate of tail suspension test immobility measurement (U_{ijk}); b_4 = The regression coefficient for the linear covariate of sucrose preference measurement (S_{ijk}); b_5 = The regression coefficient for the linear covariate of age (A_{ijk}); e_{ijk} = The residual associated with the measurement Y_{ijk}.

Locomotor activity models

The model for locomotor activity without covariates was:

$$Y_{ij} = \mu + T_i + e_{ij} \quad \text{III}$$

Y_{ij} = The number of line-crossings in the j[th] mouse in the i[th] treatment (T_i); μ = The overall mean of line-crossings across all mice; T_i = The fixed effect of treatment; e_{ij} = The residual associated with the measurement Y_{ij}.

The full locomotor activity model, combining all tested covariates was:

$$Y_{ij} = \mu + T_i + b_1(O_{ij} - \bar{O}) + b_2(B_{ij} - \bar{B}) + b_3(F_{ij} - \bar{F}) + b_4(F_{ij} - \bar{F})^2 + b_5(U_{ij} - \bar{U}) + b_6(S_{ij} - \bar{S}) + b_7(A_{ij} - \bar{A}) + e_{ij} \quad \text{IV}$$

Y_{ij} = Number of line-crossings from the j[th] mouse in the i[th] treatment (T_i); μ = The overall mean of line-crossings across all mice; T_i = The fixed effect of treatment; b_1 = The regression coefficient for the covariate of the testing order (O_{ij}); b_2 = The regression coefficient for the covariate of body weight measurement on Day -1 (B_{ij}); b_3, b_4 = The regression coefficients for the linear and quadratic covariates of forced swim test measurement (F_{ij}), respectively; b_5 = The regression coefficient for the covariate of tail suspension test immobility measurement (U_{ij}); b_6 = The regression coefficient for the covariate of the sucrose preference measurement (S_{ij}); b_7 = The regression coefficient for the covariate of age (A_{ij}); e_{ij} = The residual associated with the measurement Y_{ij}.

Rearing models

The model for rearing without covariates was:

$$Y_{ij} = \mu + T_i + e_{ij} \quad \text{V}$$

Y_{ij} = Number of rearings in the jth mouse in the ith treatment (T_i); μ = The overall mean of rearings across all mice; T_i = The fixed effect of treatment; e_{ij} = The residual associated with the measurement Y_{ij}.

The full rearing model, combining all tested covariates was:

$$Y_{ij} = \mu + T_i + b_1(O_{ij} - \bar{O}) + b_2(B_{ij} - \bar{B}) + b_3(F_{ij} - \bar{F}) + b_4(U_{ij} - \bar{U}) + b_5(S_{ij} - \bar{S}) + b_6(A_{ij} - \bar{A}) + b_7(A_{ij} - \bar{A})^2 + e_{ij} \quad \text{VI}$$

Y_{ij} = Number of rearings from the jth mouse in the ith treatment (T_i); μ = The overall mean of rearings across all mice; T_i = The fixed effect of treatment; b_1 = The regression coefficient for the covariate of the testing order (O_{ij}); b_2 = The regression coefficient for the covariate of body weight measurement on Day -1 (B_{ij}); b_3 = The regression coefficient for the covariate of forced swim test measurement (F_{ij}); b_4 = The regression coefficient for the covariate of tail suspension test immobility measurement (U_{ij}); b_5 = The regression coefficient for the covariate of the sucrose preference measurement (S_{ij}); b_6, b_7 =The regression coefficients for the linear and quadratic covariates of age (A_{ij}), respectively; e_{ij} = The residual associated with the measurement Y_{ij}.

Forced swim test models

The model for forced swim test without covariates was:

$$Y_{ij} = \mu + T_i + e_{ij} \quad \text{VII}$$

Y_{ij} = Forced swim test immobility measurement in the jth mouse in the ith treatment (T_i); μ = The overall mean of forced swim test immobility across all mice; T_i = The fixed effect of treatment; e_{ij} = The residual associated with the measurement Y_{ij}.

The full forced swim test model, combining all tested covariates was:

$$Y_{ij} = \mu + T_i + b_1(B_{ij} - \bar{B}) + b_2(E_{ij} - \bar{E}) + b_3(P_{ij} - \bar{P}) + b_4(L_{ij} - \bar{L}) + b_5(R_{ij} - \bar{R}) + b_6(A_{ij} - \bar{A}) + e_{ij} \quad \text{VIII}$$

Y_{ij} = Forced swim test immobility measurement from the jth mouse in the ith treatment (T_i); μ = The overall mean; T_i = The ith fixed effect of treatment; b_1 = the regression coefficient for the covariate of body weight measurement on Day -1 (B_{ij}); b_2 = the regression coefficient for the covariate of body weight loss between Day 0 and Day 2 (E_{ij}); b_3 = the regression coefficient for the covariate of recovering body weight between Day 2 and Day 5 (P_{ij}); b_4 = the regression coefficient for the covariate of the locomotor activity measurement (L_{ij}); b_5 = the regression coefficient for the covariate of rearing measurement (R_{ij}); b_6 = the regression coefficient for the covariate of age (A_{ij}); e_{ij} = the residual associated with the measurement Y_{ij}.

Tail suspension test models

The model for tail suspension test without covariates was:

$$Y_{ij} = \mu + T_i + e_{ij} \quad \text{XI}$$

Y_{ij} = Tail suspension test immobility measurement in the jth mouse in the ith treatment (T_i); μ = The overall mean of tail suspension test immobility across all mice; T_i = The fixed effect of treatment; e_{ij} = The residual associated with the measurement Y_{ij}.

The full tail suspension test model, combining all tested covariates was:

$$Y_{ij} = \mu + T_i + b_1(B_{ij} - \bar{B}) + b_2(E_{ij} - \bar{E}) + b_3(P_{ij} - \bar{R}) + b_4(P_{ij} - \bar{P})^2 + b_5(P_{ij} - \bar{P})^3 + b_6(L_{ij} - \bar{L}) + b_7(R_{ij} - \bar{R}) + b_8(A_{ij} - \bar{A}) + e_{ij} \quad \text{X}$$

Y_{ij} = Tail suspension immobility measurement from the jth mouse in the ith treatment (T_i); μ = The overall mean of tail suspension test immobility across all mice; T_i = the ith fixed effect of treatment; b_1 = the regression coefficient for the covariate of body weight measurement on Day -1 (B_{ij}); b_2 = the regression coefficient for the covariate of body weight loss between Day 0 and Day 2 (E_{ij}); b_3, b_4, b_5 = the regression coefficients for the linear, quadratic, and cubic covariates of recovering body weight between Day 2 and Day 5 (P_{ij}), respectively; b_6 = the regression coefficient for the covariate of the locomotor activity measurement (L_{ij}); b_7 = the regression coefficient for the covariate of rearing measurement (R_{ij}); b_8 = the regression coefficient for the covariate of age (A_{ij}); e_{ij} = the residual associated with the measurement Y_{ij}.

Sucrose preference models

The model for sucrose preference test without covariates was:

$$Y_{ij} = \mu + T_i + e_{ij} \quad \text{XI}$$

Y_{ij} = Sucrose preference test measurement in the jth mouse in the ith treatment (T_i); μ = The overall mean of sucrose preference test measurement across all mice; T_i = The fixed effect of treatment; e_{ij} = The residual associated with the measurement Y_{ij}.

The full sucrose preference test model, combining all tested covariates was:

$$Y_{ij} = \mu + T_i + b_1(B_{ij} - \bar{B}) + b_2(I_{ij} - \bar{I}) + b_3(E_{ij} - \bar{E}) + b_4(P_{ij} - \bar{P}) + b_5(L_{ij} - \bar{L}) + b_6(R_{ij} - \bar{R}) + b_7(A_{ij} - \bar{A}) + b_8(A_{ij} - \bar{A})^2 + b_9(A_{ij} - \bar{A})^3 + e_{ij} \quad \text{XII}$$

Y_{ij} = Sucrose preference measurement from the jth mouse in the ith treatment (Trt_i); μ = The overall mean for sucrose preference measurements across all mice; T_i = The fixed effect of treatment; b_1 = the regression coefficient for the covariate of body weight measurement on Day -1 (B_{ij}); b_2 = the regression coefficient for the covariate of sucrose preference measurements on Day -1 (I_{ij}); b_3 = the regression coefficient for the covariate of body weight loss between Day 0 and Day 2 (E_{ij}); b_4 = the regression coefficient for the covariate of recovering body weight between Day 2 and Day 5 (P_{ij}); b_5 = the regression coefficient for the covariate of the locomotor activity measurement (L_{ij}); b_6 = the regression coefficient for the covariate of rearing measurement (R_{ij}); b_7, b_8, b_9 = The regression coefficients for the linear, quadratic, and cubic regression coefficients for the covariate of age (A_{ij}), respectively; e_{ij} = The residual associated with the measurement Y_{ij}.

Model Comparisons

For each indicator, a model without covariates was first tested (**Eq. I**, **III**, **V**, **VII**, **IX**, **XI**). To determine the order of the relationship between a tested indicator and each covariate, covariate terms were preliminarily tested for a linear, quadratic, or cubic relationship. The full models listed (**Eq. II**, **IV**, **VI**, **VIII**, **X**, and **XII**) include all covariate terms at the order of their best performing relationship. Residuals were calculated from each model. Comparison of nested models relied upon the log likelihood ratio, Akaike's Information Criteria (AIC; Akaike, 1974), the corrected AIC (AICC), and the Bayesian Information Criterion (BIC; Schwarz, 1978) in combination (Anderson, 2008). Improvement was determined as the reduction in value of the likelihood ratio and criteria. Significance of a likelihood ratio reduction was tested with a Chi-square distribution, with the degrees of freedom calculated as the difference in parameterization between the models. While differences in the information criteria cannot be tested against a known distribution for significance, their comparisons are important to avoiding the tendency in log likelihood comparisons for over-parameterization (Anderson, 2008). Instead, the comparing nested models by the values of their metric criteria followed qualitative descriptions (Raftery, 1996). All combinations were tested, although no combination of more than three covariates demonstrated model improvement for an indicator. All covariates were nested within the BCG-treatment group, with differences between observed and predicted values used for tests of normality and homogeneity of variance. Model designs for each indicator were compared in PROC MIXED using maximum likelihood estimates, with REML estimates used when testing for significant terms within models (SAS Institute, Cary, NC). Model residuals (from the REML estimates) were checked for normality with a Kolmogorov-Smirnov test using PROC UNIVARIATE, and for homogeneity of variance using the Brown-Forsythe test with PROC GLM. For the repeated

measurements of body weight, the repeated statement in PROC MIXED was used. For repeated measures, the non-zero covariance among measures can be represented and thus modeled with the inclusion of an appropriate structure (Kaps and Lamberson, 2004). To model the covariance structure in the repeated measurements, structures studied included: Variance Components; Banded; Unstructured; Autoregressive 1; Heterogeneous Autoregressive 1; and Compound Symmetry. A Chi-squared test was used to test for homogeneity of variance in repeated structures once an appropriate model and structure were selected.

2.4 Results

Initial measurements of Body weight and sucrose preference on Day -1 were not significantly different between treatment groups (BW: p-value = 0.78; SP: p-value = 0.39). The same set of covariates were tested for preliminary measurements of body weight and sucrose preference as in modeling of later measurements of body weight and sucrose preference, which also found no significant treatment effect. The repeated structure of the body weight measurements was adequately modeled by an autoregressive order 1 structure.

Sickness indicators

Body weight loss results from Day 0 to Day 2 remained stable in regards to treatment effect across the tested models (**Eq. I**; **Eq. II**). The inclusion of body weight at Day -1 as a linear covariate (Eq. XIII) significantly improved log likelihood ($X_1^2 < 0.01$), with a strong improvement in all three criteria (**Table 2.1**). No fuller model including additional covariates demonstrated significant likelihood ratio improvement over the Day -1 body weight model. All other models without Day

-1 body weight, whether a single covariate or in combinations, showed no significant improvement in likelihood ratio improvement over the basic model (**Eq. I**). The model including Day -1 body weights satisfied homogeneity of variance ($X_4^2 > 0.1$). There was a significant interaction between treatment and time (approximate $R^2 = 0.96$; p-value < 0.0001) with High BCG treated mice losing the most weight from Day 0 to Day 2 at 3.01 g ± 0.22 g. Over the same time, Low BCG treated mice lost 1.62 g ± 0.20 g, while the control treatment showed a non-significant *gain* of 0.35 g ± 0.19 g. One notable change with the inclusion of covariates into the model, was the detection of a significant post-hoc difference between High BCG and Low BCG body weights at Day 2 (p-value = 0.0007) which was not detectable in the basic model (p-value = 0.68).

$$Y_{ijk} = \mu + T_i + D_j + TD_{ij} + b_1\left(B_{ijk} - \bar{B}\right) + e_{ijk} \qquad \text{XIII}$$

Y_{ijk} = Body weight measurement of the kth mouse in the ith treatment (T_i); μ = The overall body weight mean across all mice; T_i = The ith fixed effect of treatment; D_k = The fixed effect of the jth day over which body weight was measured; TD_{ij} = The fixed effect of the interaction between ith treatment and jth day; b_1 = The regression coefficient for the linear covariate of body weight on Day -1 (B_{ijk}); e_{ijk} = The residual associated with the measurement Y_{ijk}.

When modeling later body weights from Day 2 to Day 5 (**Table 2.2**), the time by treatment interaction was found significant regardless of covariates (**Eq. I**; p-value < 0.05). Modeling also followed a similar trend for fit improvement, with the inclusion of Day -1 body weights as a linear covariate (**Eq. XIII**) providing a significant improvement in likelihood ratio ($X_4^2 > 0.1$) and strong improvement across all three information criteria (**Table 2.2**). As with measuring body weight from Day 0 to Day 2, no other covariates individually or in combination showed an improvement

over their nested reduced models. Where likelihood ratios and the information criteria did not agree, however, was in determination of homogeneity of variance. The model for body weights from Day 2 to Day 5 (**Eq. XIII**) showed significant improvement in likelihood ratio by considering homogeneity of variance by treatment (approximate R^2 = 0.82; X_4^2 < 0.05) and a positive improvement in the AIC. However, the BIC showed a very weak improvement with a difference of only 0.5 while the AICC showed a weak worsening of the model. Interaction of time and treatment was significant under both structures (p-value < 0.01), and both BCG treatment levels showed significant (p-value < 0.01) increases in weight between days with High BCG mice gaining 0.68 g between Day 2 and Day 3. Body weight differences became non-significant by Day 4 to Day 5 (p-value > 0.1). Over the same period, no changes to body weight were found in the Control mice (p-value > 0.1).

Locomotor activity showed no significant effect of BCG treatment when modeled without covariates (**Eq. III**; p-value > 0.1). Model fit significantly improved as determined by log likelihood ratio with the addition of forced swim test in a quadratic relationship (X_2^2 < 0.05), and was improved further by including age in a linear relationship (X_1^2 < 0.01; **Eq. XIV**; **Table 2.3**). This model identified a significant effect of treatment (approximate R^2 = 0.71; p-value < 0.01) based upon a reduction of 15 line-crossings in locomotor activity in the High BCG treatment group compared to the control group over the test period (p-value < 0.01).

$$Y_{ij} = \mu + T_i + b_1\left(F_{ij} - \bar{F}\right) + b_2\left(F_{ij} - \bar{F}\right)^2 + b_3\left(A_{ij} - \bar{A}\right) + e_{ij} \qquad \text{XIV}$$

Y_{ij} = Number of line-crossings from the j^{th} mouse in the i^{th} treatment (T_i); μ = The overall mean of line-crossings across all mice; T_i = The fixed effect of treatment; b_1, b_2 = The regression

coefficients for the linear and quadratic covariates of forced swim test measurement (F_{ij}), respectively; b_3 = The regression coefficient for the covariate of age (A_{ij}); e_{ij} = The residual associated with the measurement Y_{ij}.

Rearing showed a borderline treatment effect when modeled without any covariates (p-value = 0.06). Including a quadratic covariate of age was the best fitting model with a single covariate, strongly improving all three criteria as well as significant difference in likelihood ratio ($X_2^2 < 0.01$; **Table 2.4**). Including the age covariate resulted in a significant treatment effect (p-value < 0.05). Larger models continued to improve up to including the quadratic age, forced swim test, and sucrose preference test covariates (approximate $R^2 = 0.77$; **Eq. XV**). Using this model, control mice on average performed 11 more rearings than High BCG mice over the test period (p-value < 0.01). The AICC criterion did show a weak penalty for the additional terms, but weak improvement for both the AIC and BIC, as well as a significant difference in likelihood ratio. All of the models including covariates maintained the significant treatment effect (p-value < 0.05).

$$Y_{ij} = \mu + T_i + b_1(F_{ij} - \bar{F}) + b_2(S_{ij} - \bar{S}) + b_3(A_{ij} - \bar{A}) + b_4(A_{ij} - \bar{A})^2 + e_{ij} \qquad \text{XV}$$

Y_{ij} = Number of rearings from the jth mouse in the ith treatment (T_i); μ = The overall mean of rearings across all mice; T_i = The fixed effect of treatment; b_1 = The regression coefficient for the covariate of forced swim test measurement (F_{ij}); b_2 = The regression coefficient for the covariate of the sucrose preference measurement (S_{ij}); b_3, b_4 =The regression coefficients for the linear and quadratic covariates of age (A_{ij}), respectively; e_{ij} = The residual associated with the measurement Y_{ij}.

Depression-like behavior indicators

Regardless of the exclusion (**Eq. VII**) or inclusion (**Eq. VIII**) of covariates in the model, no significant effect of treatment was found in forced swim testing (p-value > 0.1). Including rearing data in a linear relationship was the best single covariate for the model by likelihood ratio (**Eq. XVI**; X_1^2 < 0.05), with no significant improvements over this model from including additional covariates (**Table 2.5**). However the criteria scores for the model were only moderately improved, with only weak change to AICC and positive improvements to both the AIC and BIC (approximate $R^2 = 0.04$).

$$Y_{ij} = \mu + T_i + b_1\left(R_{ij} - \bar{R}\right) + e_{ij} \qquad \text{XVI}$$

Y_{ij} = Forced swim test immobility measurement from the jth mouse in the ith treatment (T_i); μ = The overall mean of forced swim test immobility; T_i = The ith fixed effect of treatment; b_1 = the regression coefficient for the covariate of rearing measurement (R_{ij}); e_{ij} = the residual associated with the measurement Y_{ij}.

Tail suspension data also showed no significant treatment effect when the model did not include covariates (**Eq. XI**; p-value > 0.1), while model residuals also indicated a borderline deviation from normality in the residuals (Kolmogorov-Smirnov p-value = 0.0605). However including body weight from Day 2 to Day 5 as a cubic covariate (**Eq. XVII**) improved the model by all measurements against the model without covariates (**Eq. XI**), with a significant reduction in the log likelihood ratio (X_3^2 < 0.05) while reporting the lowest criteria of all tested models (**Table 2.6**). By including body weights from Day 2 to Day 5 as a cubic covariate (**Eq. XVII**), the model detected a borderline significant effect of treatment while rescuing residual normality, with High

BCG mice spending an additional 58 seconds immobile than control mice over the test period (approximate $R^2 = 0.53$; p-value = 0.055; KS p-value > 0.1).

$$Y_{ij} = \mu + T_i + b_1\left(P_{ij} - \bar{R}\right) + b_2\left(P_{ij} - \bar{P}\right)^2 + b_3\left(P_{ij} - \bar{P}\right)^3 + e_{ij} \quad \text{XVII}$$

Y_{ij} = Tail suspension immobility measurement from the jth mouse in the ith treatment (T_i); μ = The overall mean of tail suspension test immobility across all mice; T_i = the ith fixed effect of treatment; b_1, b_2, b_3 = the regression coefficients for the linear, quadratic, and cubic covariates of recovering body weight between Day 2 and Day 5 (P_{ij}), respectively; e_{ij} = the residual associated with the measurement Y_{ij}.

No treatment effect was found in sucrose preference when modeled without covariates (**Eq. XI**; p-value > 0.1). However when mouse age was included as a cubic covariate (**Eq. XVIII**), likelihood ratio was significantly improved ($X_3^2 < 0.01$). While all criteria agreed that there was some improvement by including age as a cubic covariate, it was a very weak improvement to AICC compared to the model without covariates (**Eq. XI**), a strong improvement in AIC, and a positive improvement to BIC (**Table 2.7)**. The inclusion of the cubic age covariate also identified a significant effect of treatment (approximate $R^2 = 0.59$; p-value < 0.05). Among the treatments in the model including the cubic covariate of age, there was a greater difference between the High BCG and control groups, a difference of 0.12 in sucrose preference (p-value < 0.03) with a smaller difference when comparing Low BCG to control (p-value < 0.05). Residuals from all reported models of sucrose preference displayed normality and homogeneity of variance (data not shown).

$$Y_{ij} = \mu + T_i + b_1\left(A_{ij} - \bar{A}\right) + b_2\left(A_{ij} - \bar{A}\right)^2 + b_3\left(A_{ij} - \bar{A}\right)^3 + e_{ij} \quad \text{XVIII}$$

Y_{ij} = Sucrose preference measurement from the jth mouse in the ith treatment (Trt_i); μ = The overall mean for sucrose preference measurements across all mice; T_i = The fixed effect of treatment; b_1, b_2, b_3 = The regression coefficients for the linear, quadratic, and cubic regression coefficients for the covariate of age (A_{ij}), respectively; e_{ij} = The residual associated with the measurement Y_{ij}.

2.5 Discussion

When modeling body weight changes from Day 0 to Day 2 by treatment and time, the estimates of treatment effect were not greatly impacted by inclusion of covariates in the model. Still, the criteria of model fit were improved with a subsequent sensitivity that enabled identification of differences between the BCG treatment levels at Day 2. For modeling later body weight changes, the inclusion of Day -1 body weights is recommended (**Table 2.2**). This model reinforced that body weights increased in both BCG treatment levels from Day 2 to Day 5, as opposed to no change over the same period when given saline (p-value > 0.1). Univariate modeling of locomotor activity and rearing presents some complexities, as previous studies have indicated no significant (Kwon et al., 2012; Kelley et al., 2013), significant (O'Connor et al., 2009), or borderline significant (Painsipp et al., 2013) treatment effects in locomotory measurements at Day 7. The inclusion of covariates appears to expose several underlying relationships between locomotor measurements and the other indicators, with implications to the use of locomotor activity as a proof of separation between sickness and depression behaviors (Vijaya Kumar et al., 2014). No significant effect of treatment was found in forced swim testing, with or without covariates included in the model (p-value > 0.1). Including rearing in a linear relationship was the best model by likelihood ratio with varying improvements to criteria (**Table 2.5**). The effect of BCG treatment

upon forced swim testing has varied among univariate studies (Moreau et al., 2008; O'Connor et al., 2009). However the exploration of the underlying relationships between locomotor activity, rearing, and other indicators is perhaps better suited to multivariate studies.

Tail suspension data was perhaps more affected by the inclusion of covariates, improving model fit while also identifying a previously unidentified borderline significant effect of treatment (**Table 2.6**; p-value = 0.055). Despite the forced swim and tail suspension tests measuring the same behavior (Dedic et al., 2011) and clustering together in multivariate studies (Rodriguez-Zas et al., 2015), their respective optimum models did not share the same covariates. In fact, the tail suspension test was the only mobility-based indicators that did not benefit from including other mobility-based indicators as covariates (**Table 2.6**). Thus, including mobility-based indicators as covariates during the modeling of tail suspension is discouraged, while including weight recovery data is instead recommended.

Sucrose preference was clearly benefitted by the inclusion of covariates (**Eq. XVIII**), significantly improving the log likelihood ratio the criteria, although the AICC showed only very weak improvement (**Table 2.7**). Additionally, including age as a covariate with a cubic relationship detected a significant treatment effect (p-value < 0.05) that was not found without the age covariate. Previous univariate studies did not identify significant treatment effects in sucrose preference following BCG treatment, although covariation was not examined or focused upon the impact of behavioral indicators (Moreau et al., 2008; Rodriguez-Zas et al., 2015). Age can significantly impact changes in sucrose preference (Kelley et al., 2013), a behavioral measurement

sensitive to disruption without sufficient internal control (Strekalova et al., 2011). Age-matching is a common control in behavioral studies, yet the strength of this covariation suggests controlling for even small age differences among mice can greatly improve the sensitivity of sucrose preference testing. In addition to the consideration of covariates, multivariate analysis is an additional approach that can provide complementary information to the univariate analysis presented in this study (Rodriguez-Zas et al., 2015).

In conclusion, univariate modeling of behavioral indicators is a historically popular method of analysis. Mice treated with BCG displayed a reduction in body weight by Day 2 post-treatment, with changes in body weight between days becoming non-significant by Day 5. There was a borderline significant effect of BCG treatment on tail suspension immobility at Day 6, and a significant effect of BCG on sucrose preference at Day 7. Several sickness and depression indicators have a detectable treatment effect when univariate modeling is improved by including covariates. Modeling sucrose preference can be improved by including mouse age, while a covariate of body weight recovery after treatment can improve the modeling of tail suspension data. For measurements that have much stronger signaling, the benefit of covariates was limited to model fit rather than a change in significance. The interaction between time and treatment in body weight modeling were an example, with the same set of effects found significant across all model designs. Locomotor and rearing were not strongly benefited by inclusion of the other tested indicators as covariates. Univariate models can benefit from the inclusion of covariate terms.

2.6 References

Akaike, H. (1974). A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, **19**(6): 716-723. http://dx.doi.org/10.1109/TAC.1974.1100705

Anderson, D.R., *Model based inference in the life sciences: a primer on evidence*, New York: Springer; 2008.

Castagné, V., Moser, P., Roux, S., Porsolt, R.D. (2011). Rodent models of depression: Forced swim and tail suspension behavioral despair tests in rats and mice. *Current Protocols in Neuroscience*, **55**(8.10A): 8.10A.1-8.10A.14. http://dx.doi.org/10.1002/0471142301.ns0810as55

Committee for the Update of the Guide for the Care and Use of Laboratory Animals, & National Research Council. (2011). *Guide for the care and use of laboratory animals: Eighth edition* The National Academies Press.

Dedic N., Walser, S.M., Deussing, J.M. (2011). Mouse Models of Depression, Psychiatric Disorders - Trends and Developments, Dr. Toru Uehara (Ed.), ISBN: 978-953-307-745-1, InTech. http://dx.doi.org/10.5772/27498

Ellacott, K.L.J., Morton, G.J., Woods, S.C., Tso, P., Schwartz, M.W. (2010). Assessment of feeding behavior in laboratory mice. *Cell Metabolism*, **12**(1): 10-17. http://dx.doi.org/10.1016/j.cmet.2010.06.001

Gould, T., Dao, D., Kovacsics, C. (2009). The open field test. In T. D. Gould (Ed.), (pp. 1-20) Humana Press.

Jones, K.A., Thomsen, C. (2013). The role of the innate immune system in psychiatric disorders. *Molecular and Cellular Neurosciences*. **53**: 52-62.

Kelley, K.W., O'Connor, J.C., Lawson, M.A., Dantzer, R., Rodriguez-Zas, S.L., McCusker, R.H., 2013. Aging leads to prolonged duration of inflammation-induced depression-like behavior caused by Bacillus Calmette-Guérin. *Brain Behavior and Immunity*, **32:** 63-69. http://dx.doi.org/10.1016/j.bbi.2013.02.003

Lawson, M.A., Parrott, J.M., McCusker, R.H., Dantzer, R., Kelley, K.W., O'Connor, J.C. (2013). Intracerebroventricular administration of lipopolysaccharide induces indoleamine-2,3-dioxygenase-dependent depression-like behaviors. *Journal of Neuroinflammation*, **10**: 87. http://dx.doi.org/10.1186/1742-2094-10-87

Moreau, M., André, C., O'Connor, J.C., Dumich, S.A., Woods, J.A., Kelley, K.W., Dantzer, R., Lestage, J., Castanon, N. (2008). Inoculation of Bacillus Calmette-Guerin to mice induces an acute episode of sickness behavior followed by chronic depressive-like behavior. *Brain, Behavior, and Immunity*, **22**(7): 1087-1095. http://dx.doi.org/10.1016/j.bbi.2008.04.001

Moreau, M., Lestage, J., Verrier, D., Mormède, C., Kelley, K.W., Dantzer, R., Castanon, N. (2005). Bacille Calmette-Guérin inoculation induces chronic activation of peripheral and brain indoleamine 2,3-dioxygenase in mice. *Journal of Infectious Diseases*, **192**(3): 537-544. http://dx.doi.org/10.1086/431603

O'Connor, J.C., André, C., Wang, Y., Lawson, M.A., Szegedi, S.S., Lestage, J., Castanon, N., Kelley, K.W., Dantzer, R. (2009). 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-Guérin. *Journal of Neuroscience*, **29**(13): 4200-4209. http://dx.doi.org/10.1523/JNEUROSCI.5032-08.2009

Painsipp, E., Köfer, M.J., Farzi, A., Dischinger, U.S., Sinner, F., Herzog, H., Holzer P. (2013). Neuropeptide Y and peptide YY protect from weight loss caused by Bacille Calmette-Guérin in mice. *British Journal of Pharmacology*, **170**(5): 1014-1026. http://dx.doi.org/10.1111/bph.12354

Painsipp, E., Köfer, M.J., Sinner, F., Holzer P. (2011). Prolonged depression-like behavior caused by immune challenge: influence of mouse strain and social environment. *PLoS One*, **6**(6): e20719. http://dx.doi.org/10.1371/journal.pone.0020719

Raftery, A.E. (1996). Approximate Bayes factors and accounting for model uncertainty in generalized linear models. *Biometrika*, **83**(2): 251-266. http://dx.doi.org/10.1093/biomet/83.2.251

Rodriguez-Zas, S.L., Nixon, S.E., Lawson, M.A., McCusker, R.H., Southey, B.R., O'Connor, J.C., Dantzer, R., Kelley, K.W. (2015). Advancing the understanding of behaviors associated with Bacillus Calmette Guérin infection using multivariate analysis. *Brain, Behavior, and Immunity*, **44**: 176-186. http://dx.doi.org/10.1016/j.bbi.2014.09.018

Saleh, L.A., Hamza, M., El Gayar, N.H., Abd El-Samad, A.A., Nasr, E.A., Masoud, S.I. (2014). Ibuprofen suppresses depressive like behavior induced by BCG inoculation in mice: role of nitric oxide and prostaglandin. *Pharmacology, Biochemistry, and Behavior*, **125**: 29-39. http://dx.doi.org/10.1016/j.pbb.2014.07.013

Schmuckermair C., Gaburro S, Sah A., Landgraf R., Sartori S.B., Singewald N. (2013). Behavioral and neurobiological effects of deep brain stimulation in a mouse model of high anxiety- and depression-like behavior. *Neuropsychopharmacology*, **38**(7): 1234-1244. http://dx.doi.org/10-1038/npp.2013.21

Schwarz, G (1978). Estimating the dimension of a model, *Annals of Statistics*, **42**(3): 872-917. http://dx.doi.org/10.1214/aos/1176344136

Srivastava, S., Chen, L. (2010). A two-parameter generalized Poisson model to improve the analysis of RNA-seq data, *Nucleic Acids Research*, **38**(17): e170. http://dx.doi.org/10.1093/nar/gkq670

Steensland, P., Simms, J.A., Nielsen, C.K., Holgate, J., Bito-Onon, J.J., Bartlett, S.E. (2010). The neurokinin 1 receptor antagonist, Ezlopitant, reduces appetitive responding for sucrose and ethanol. *PLoS One*. **5**(9): e12527. http://dx.doi.org/10.1371/journal.pone.0012527

Steru, L., Chermat, R., Thierry, B., Simon, P. (1985). The tail suspension test: A new method for screening antidepressants in mice. *Psychopharmacology*, **85**(3): 367-370. http://dx.doi.org/10.1007/BF00428203

Strekalova, T., Couch, Y., Kholod, N., Boyks, M., Malin, D., Leprince, P., Steinbusch, H.M. (2011). Update in the methodology of the chronic stress paradigm: internal control matters. *Behavioral and Brain Functions: BBF*, **7**: 9. http://dx.doi.org/10.1186/1744-9081-7-9

Tsenova, L., Bergtold, A., Freedman, V.H., Young, R.A., Kaplan, G. (1999). Tumor necrosis factor α is a determinant of pathogenesis and disease progression in mycobacterial infection in the central nervous system. *Proceedings of the National Academy of Sciences*, **96**(10): 5657-5662. http://dx.doi.org/10.1073/pnas.96.10.5657

Vijaya Kumar, K., Rudra, A., Sreedhara, M.V., Siva Subramani, T., Prasad, D.S., Das, M.L., Murugesan, S., Yadav, R., Trivedi, R.K., Louis, J.V., Li, Y.W., Bristow, L.J., Naidu, P.S., Vikramadithyan, R.K. (2014). Bacillus Calmette-Guérin vaccine induces a selective serotonin reuptake inhibitor (SSRI)-resistant depression like phenotype in mice. *Brain, Behavior, and Immunity*, **42**: 204-211. http://dx.doi.org/10.1016/j.bbi.2014.06.205

2.7 Figures

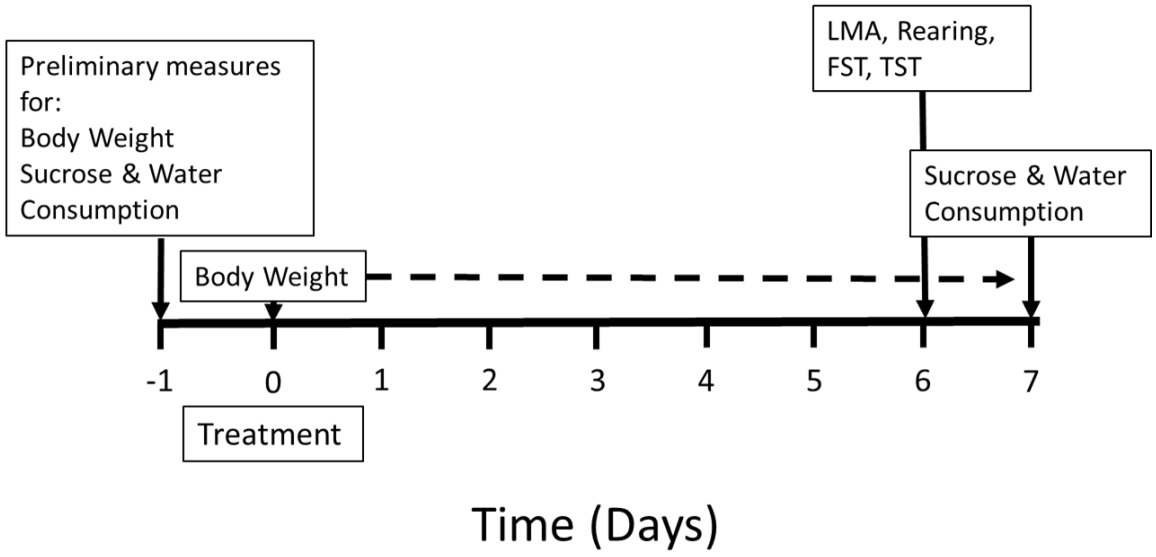


Figure 2.1. Timeline of measurement for behavioral indicators and body weight
Timeline indicating the order of experimental measurements in relation to treatment. Body Weight was measured daily at the beginning of the dark cycle, starting with one day before treatment (Day -1). A preliminary measurement of sucrose and water consumption was also measured on Day -1. The mobility-based sickness and depression indicators of locomotor activity (LMA), rearing, tail suspension test (TST), and forced swim test (FST) were measured on Day 6. For the schedule of Day 6, mice were weighed at the beginning of the dark cycle as usual, with LMA and rearing measured immediately after. Upon their completion, FST was performed on all mice followed shortly thereafter by TST. Consumption of sucrose and water were measured on Day 7 with animal weights for the sucrose preference indicator.

2.8 Tables

Table 2.1 Models of mouse body weight from Day 0 to Day 2, from simplest (no covariates included) to fullest. Higher order relationships (quadratic or cubic) for covariates included all lower order relationships. Model rows demonstrating improvement over lower complexity designs are in white, with the final model selection in bold.

Covariates included in the model						**Model criteria**			
IBW	**FST**	**TST**	**SP**	**Age**	**DF**	**AIC**	**AICC**	**BIC**	**-2 LL**
-	-	-	-	-	40	151.1	156.9	161.5	129.1
IBW	**-**	**-**	**-**	**-**	**39**	**118.7**	**125.8**	**130.1**	**94.7**
IBW^2	-	-	-	-	38	120.5	129	132.8	94.5
IBW^3	-	-	-	-	37	122.1	132.1	135.3	94.1
-	FST	-	-	-	39	153	160.1	164.3	129
-	FST^2	-	-	-	38	154	162.4	166.2	128
-	FST^3	-	-	-	37	155.8	165.8	169	127.8
-	-	TST	-	-	39	152	159.1	163.3	128
-	-	TST^2	-	-	38	152.7	161.2	165	126.7
-	-	TST^3	-	-	37	152	162	165.2	124
-	-	-	SP	-	39	153.1	160.2	164.4	129.1
-	-	-	SP^2	-	38	154.7	163.2	167	128.7
-	-	-	SP^3	-	37	156.5	166.5	169.7	128.5
-	-	-	-	Age	39	150.7	157.8	162	126.7
-	-	-	-	Age^2	38	151.6	160	163.8	125.6
-	-	-	-	Age^3	37	153.5	163.5	166.8	125.5
IBW	FST	-	-	-	38	120.7	129.2	133	94.7
IBW	-	TST	-	-	38	120.7	129.2	133	94.7
IBW	-	-	SP	-	38	120.1	128.6	132.4	94.1
IBW	-	-	-	Age	38	120.7	129.2	133	94.7
-	FST	TST	-	-	38	154	162.5	166.3	128
-	FST	-	SP	-	38	155	163.5	167.3	129
-	FST	-	-	Age	38	152.7	161.1	165	126.7
-	-	TST	SP	-	38	154	162.4	166.3	128
-	-	TST	-	Age	38	150.8	159.2	163	124.8
-	-	-	SP	Age	38	152.6	161	164.9	126.6

IniBW: body weight at Day -1; FST: Forced swim test; TST: Tail suspension test; SP: Sucrose preference at Day 7; DF: Error degrees of freedom; AIC: Akaike Information Criterion; AICC: corrected AIC; BIC: Bayesian Information Criterion; -2 LL: -2 Log Likelihood.

Table 2.2 Models of mouse body weight from Day 2 to Day 5, from simplest (no covariates included) to fullest. Higher order relationships (quadratic or cubic) for covariates included all lower order relationships. Model rows demonstrating improvement over lower complexity designs are in white, with the final model selection in bold.

Covariates included in the model						**Model criteria**			
IBW	**FST**	**TST**	**SP**	**Age**	**DF**	**AIC**	**AICC**	**BIC**	**-2 LL**
-	-	-	-	-	55	133.1	140	146.3	105.1
IBW	**-**	**-**	**-**	**-**	**54**	**114.3**	**122.3**	**128.4**	**84.3**
IBW^2	-	-	-	-	53	116.2	125.5	131.4	84.2
IBW^3	-	-	-	-	52	118.2	128.7	134.2	84.2
-	FST	-	-	-	54	135.1	143.1	149.2	105.1
-	FST^2	-	-	-	53	134.8	144	149.9	102.8
-	FST^3	-	-	-	52	136.6	147.2	152.7	102.6
-	-	TST	-	-	54	133.7	141.7	147.8	103.7
-	-	TST^2	-	-	53	132.9	142.1	148	100.9
-	-	TST^3	-	-	52	133	143.5	149	99
-	-	-	SP	-	54	135.1	143.1	149.2	105.1
-	-	-	SP^2	-	53	137	146.2	152.1	105
-	-	-	SP^3	-	52	138.9	149.5	155	104.9
-	-	-	-	Age	54	133.6	141.6	147.8	103.6
-	-	-	-	Age^2	53	135.3	144.5	150.4	103.3
-	-	-	-	Age^3	52	137.3	147.8	153.3	103.3
IBW	FST	-	-	-	53	116.1	125.3	131.2	84.1
IBW	-	TST	-	-	53	116	125.2	131.1	84
IBW	-	-	SP	-	53	115.5	124.7	130.6	83.5
IBW	-	-	-	Age	53	116.2	125.5	131.4	84.2
-	FST	TST	-	-	53	135.6	144.8	150.7	103.6
-	FST	-	SP	-	53	137.1	146.3	152.2	105.1
-	FST	-	-	Age	53	135.6	144.8	150.7	103.6
-	-	TST	SP	-	53	135.7	144.9	150.8	103.7
-	-	TST	-	Age	53	133.4	142.6	148.5	101.4
-	-	-	SP	Age	53	135.6	144.8	150.7	103.6

IBW: body weight at Day -1; FST: Forced swim test; TST: Tail suspension test; SP: Sucrose preference at Day 7; DF: Error degrees of freedom; AIC: Akaike Information Criterion; AICC: corrected AIC; BIC: Bayesian Information Criterion; -2 LL: -2 Log Likelihood.

Table 2.3 Models of mouse locomotor activity, from simplest (no covariates included) to fullest. Higher order relationships (quadratic or cubic) for covariates included all lower order relationships. Model rows demonstrating improvement over lower complexity designs are in white, with the final model selection in bold.

Covariates included in the model							**Model criteria**			
O	**IBW**	**FST**	**TST**	**SP**	**Age**	**DF**	**AIC**	**AICC**	**BIC**	**-2 LL**
-	-	-	-	-	-	15	150.3	153.2	154.1	142.3
O	-	-	-	-	-	14	151.7	156.3	156.4	141.7
-	IBW	-	-	-	-	14	149.9	154.5	154.6	139.9
-	IBW^2	-	-	-	-	13	151.7	158.7	157.4	139.7
-	IBW^3	-	-	-	-	12	153.1	163.3	159.7	139.1
-	-	FST	-	-	-	14	151	155.6	155.8	141
-	-	FST^2	-	-	-	13	146.1	153.1	151.7	134.1
-	-	FST^3	-	-	-	12	146.3	156.5	152.9	132.3
-	-	-	TST	-	-	14	152.1	156.7	156.8	142.1
-	-	-	TST^2	-	-	13	153.8	160.8	159.4	141.8
-	-	-	TST^3	-	-	12	151.2	161.4	157.8	137.2
-	-	-	-	SP	-	14	152	156.6	156.8	142
-	-	-	-	SP^2	-	13	153.9	160.9	159.6	141.9
-	-	-	-	SP^3	-	12	155.9	166.1	162.5	141.9
-	-	-	-	-	Age	14	150.5	155.1	155.2	140.5
-	-	-	-	-	Age^2	13	152.4	159.4	158.1	140.4
-	-	-	-	-	Age^3	12	154.4	164.6	161	140.4
O	IBW	-	-	-	-	13	151.7	158.7	157.3	139.7
O	-	FST^2	-	-	-	12	138.1	148.3	144.7	124.1
O	-	-	TST	-	-	13	153.6	160.6	159.3	141.6
O	-	-	-	SP	-	13	153.5	160.5	159.2	141.5
O	-	-	-	-	Age	13	151.9	158.9	157.5	139.9
-	IBW	FST^2	-	-	-	12	144.9	155.1	151.5	130.9
-	IBW	-	TST	-	-	13	151	158	156.7	139
-	IBW	-	-	SP	-	13	151.4	158.4	157.1	139.4
-	IBW	-	-	-	Age	13	151.1	158.1	156.8	139.1
-	-	FST^2	TST	-	-	12	148	158.1	154.6	134
-	-	FST^2	-	SP	-	12	147.7	157.9	154.3	133.7
-	**-**	**FST^2**	**-**	**-**	**Age**	**12**	**138.3**	**148.5**	**144.9**	**124.3**
-	-	-	TST	SP	-	13	153.8	160.8	159.4	141.8
-	-	-	TST	-	Age	13	152.4	159.4	158.1	140.4
-	-	-	-	SP	Age	13	152.4	159.4	158	140.4

O: Order of testing; IBW: Body weight at Day -1; FST: Forced swim test; TST: Tail suspension test; SP: Sucrose preference at Day 7; DF: Error degrees of freedom; AIC: Akaike Information Criterion; AICC: corrected AIC; BIC: Bayesian Information Criterion; -2 LL: -2 Log Likelihood.

Table 2.4 Models of mouse rearing, from simplest (no covariates included) to fullest. Higher order relationships (quadratic or cubic) for covariates included all lower order relationships. Model rows demonstrating improvement over lower complexity designs are in white, with the final model selection in bold.

Covariates included in the model							**Model criteria**			
O	**IBW**	**FST**	**TST**	**SP**	**Age**	**DF**	**AIC**	**AICC**	**BIC**	**-2 LL**
-	-	-	-	-	-	15	150.3	153.2	154.1	142.3
O	-	-	-	-	-	14	151.7	156.3	156.4	141.7
-	IBW	-	-	-	-	14	149.9	154.5	154.6	139.9
-	IBW^2	-	-	-	-	13	151.7	158.7	157.4	139.7
-	IBW^3	-	-	-	-	12	153.1	163.3	159.7	139.1
-	-	FST	-	-	-	14	123.7	128.4	128.5	113.7
-	-	FST^2	-	-	-	13	125.7	132.7	131.4	113.7
-	-	FST^3	-	-	-	12	126.5	136.7	133.2	112.5
-	-	-	TST	-	-	14	128.9	133.5	133.6	118.9
-	-	-	TST^2	-	-	13	130.5	137.5	136.2	118.5
-	-	-	TST^3	-	-	12	130.5	140.7	137.1	116.5
-	-	-	-	SP	-	14	125.3	130	130.1	115.3
-	-	-	-	SP^2	-	13	126	133	131.7	114
-	-	-	-	SP^3	-	12	128	138.1	134.6	114
-	-	-	-	-	Age	14	128.7	133.3	133.5	118.7
-	-	-	-	-	Age^2	13	121.7	128.7	127.3	109.7
-	-	-	-	-	Age^3	12	123	133.2	129.6	109
O	IBW	-	-	-	-	13	130.4	137.4	136.1	118.4
O	-	FST	-	-	-	13	125.7	132.7	131.4	113.7
O	-	-	TST	-	-	13	130.9	137.9	136.5	118.9
O	-	-	-	SP	-	13	127.3	134.3	132.9	115.3
O	-	-	-	-	Age^2	12	123.5	133.7	130.1	109.5
-	IBW	FST	-	-	-	13	125.5	132.5	131.1	113.5
-	IBW	-	TST	-	-	13	130.4	137.4	136	118.4
-	IBW	-	-	SP	-	13	127	134	132.7	115
-	IBW	-	-	-	Age^2	12	123	133.2	129.6	109
-	-	FST	TST	-	-	13	125.1	132.1	130.7	113.1
-	-	FST	-	SP	-	13	119.1	126.1	124.7	107.1
-	-	FST	-	-	Age^2	12	117.6	127.8	124.2	103.6
-	-	-	TST	SP	-	13	127.3	134.3	133	115.3
-	-	-	TST	-	Age^2	12	123.6	133.8	130.2	109.6
-	-	-	-	SP	Age^2	12	121.9	132.1	128.5	107.9
-	**-**	**FST**	**-**	**SP**	**Age^2**	**11**	**115.2**	**129.6**	**122.7**	**99.2**

O: Order of testing; IBW: Body weight at Day -1; FST: Forced swim test; TST: Tail suspension test; SP: Sucrose preference at Day 7; DF: Error degrees of freedom; AIC: Akaike Information Criterion; AICC: corrected AIC; BIC: Bayesian Information Criterion; -2 LL: -2 Log Likelihood.

Table 2.5 Models of forced swim test immobility, from simplest (no covariates) to fullest. Higher order relationships (quadratic, cubic) for covariates include all lower order relationships. Models demonstrating improvement are in white, final model is in bold.

Covariates included in the model							Model criteria			
IBW	**BW0-2**	**BW2-5**	**LMA**	**Rear**	**Age**	**DF**	**AIC**	**AICC**	**BIC**	**-2 LL**
-	-	-	-	-	-	15	199.2	202.1	203	191.2
IBW	-	-	-	-	-	14	201	205.6	205.8	191
IBW^2	-	-	-	-	-	13	202.5	209.5	208.2	190.5
IBW^3	-	-	-	-	-	12	204.5	214.7	211.1	190.5
-	BW0-2	-	-	-	-	14	201.2	205.8	205.9	191.2
-	$BW0\text{-}2^2$	-	-	-	-	13	202.5	209.5	208.2	190.5
-	$BW0\text{-}2^3$	-	-	-	-	12	202.4	212.6	209	188.4
-	-	BW2-5	-	-	-	14	200.2	204.8	204.9	190.2
-	-	$BW2\text{-}5^2$	-	-	-	13	201.5	208.5	207.1	189.5
-	-	$BW2\text{-}5^3$	-	-	-	12	202.4	212.6	209	188.4
-	-	-	LMA	-	-	14	199.9	204.6	204.7	189.9
-	-	-	LMA^2	-	-	13	201.7	208.7	207.3	189.7
-	-	-	LMA^3	-	-	12	198.9	209.1	205.5	184.9
-	**-**	**-**	**-**	**Rear**	**-**	**14**	**196.1**	**200.7**	**200.8**	**186.1**
-	-	-	-	$Rear^2$	-	13	195.9	202.9	201.6	183.9
-	-	-	-	$Rear^3$	-	12	197.5	207.7	204.1	183.5
-	-	-	-	-	Age	14	200.8	205.4	205.5	190.8
-	-	-	-	-	Age^2	13	202.3	209.3	207.9	190.3
-	-	-	-	-	Age^3	12	201.8	212	208.4	187.8
IBW	BW0-2	-	-	-	-	11	203	210	208.7	191
IBW	-	BW2-5	-	-	-	11	202.2	209.2	207.9	190.2
IBW	-	-	LMA	-	-	11	201.1	208.1	206.8	189.1
IBW	-	-	-	Rear	-	11	198.1	205.1	203.7	186.1
IBW	-	-	-	-	Age	11	202.7	209.7	208.4	190.7
-	BW0-2	BW2-5	-	-	-	11	202.1	209.1	207.8	190.1
-	BW0-2	-	LMA	-	-	11	201.9	208.9	207.6	189.9
-	BW0-2	-	-	Rear	-	11	198.1	205.1	203.8	186.1
-	BW0-2	-	-	-	Age	11	202.7	209.7	208.4	190.7
-	-	BW2-5	LMA	-	-	11	201.4	208.4	207.1	189.4
-	-	BW2-5	-	Rear	-	11	198.1	205.1	203.7	186.1
-	-	BW2-5	-	-	Age	11	201.6	208.6	207.3	189.6
-	-	-	LMA	Rear		11	197.8	204.8	203.4	185.8
-	-	-	LMA	-	Age	11	200.8	207.8	206.4	188.8
-	-	-	-	Rear	Age	11	197.2	204.2	202.8	185.2

IBW: Body weight at Day -1; BW0-2: Body weight change from Day 0 to Day 2; BW2-5: Body weight change from Day 2 to Day 5; LMA: Locomotor activity; Rear: Rearing; DF: Error degrees of freedom; AIC: Akaike Information Criterion; AICC: corrected AIC; BIC: Bayesian Information Criterion; -2 LL: -2 Log Likelihood.

Table 2.6 Models of tail suspension test immobility, from simplest (no covariates) to fullest. Higher order relationships (quadratic, cubic) for covariates include all lower order relationships. Models demonstrating improvement are in white, final model is in bold.

Covariates included in the model							Model criteria			
IBW	**BW0-2**	**BW2-5**	**LMA**	**Rear**	**Age**	**DF**	**AIC**	**AICC**	**BIC**	**-2 LL**
-	-	-	-	-	-	15	218.4	221.3	222.2	210.4
IBW	-	-	-	-	-	14	219.2	223.8	223.9	209.2
IBW^2	-	-	-	-	-	13	219.6	226.6	225.3	207.6
IBW^3	-	-	-	-	-	12	219.1	229.3	225.7	205.1
-	BW0-2	-	-	-	-	14	220.4	225.1	225.2	210.4
-	$BW0\text{-}2^2$	-	-	-	-	13	221.2	228.2	226.9	209.2
-	$BW0\text{-}2^3$	-	-	-	-	12	221.4	231.6	228	207.4
-	-	BW2-5	-	-	-	14	219.7	224.3	224.4	209.7
-	-	$BW2\text{-}5^2$	-	-	-	13	221.3	228.3	227	209.3
-	**-**	**$BW2\text{-}5^3$**	**-**	**-**	**-**	**12**	**215.3**	**225.5**	**221.9**	**201.3**
-	-	-	LMA	-	-	14	220.2	224.8	224.9	210.2
-	-	-	LMA^2	-	-	13	222.1	229.1	227.8	210.1
-	-	-	LMA^3	-	-	12	217.4	227.6	224.1	203.4
-	-	-	-	Rear	-	14	220.4	225.1	225.2	210.4
-	-	-	-	$Rear^2$	-	13	222.3	229.3	228	210.3
-	-	-	-	$Rear^3$	-	12	220.8	231	227.4	206.8
-	-	-	-	-	Age	14	220	224.6	224.7	210
-	-	-	-	-	Age^2	13	221.9	228.9	227.6	209.9
-	-	-	-	-	Age^3	12	223.8	234	230.4	209.8
IBW	BW0-2	-	-	-	-	13	221.1	228.1	226.8	209.1
IBW	-	$BW2\text{-}5^3$	-	-	-	11	215.6	230	223.1	199.6
IBW	-	-	LMA	-	-	13	220.3	227.3	225.9	208.3
IBW	-	-	-	Rear	-	13	221.1	228.1	226.8	209.1
IBW	-	-	-	-	Age	13	219.7	226.7	225.3	207.7
-	BW0-2	$BW2\text{-}5^3$	-	-	-	11	217.2	231.6	224.8	201.2
-	BW0-2	-	LMA	-	-	13	222.2	229.2	227.9	210.2
-	BW0-2	-	-	Rear	-	13	222.4	229.4	228.1	210.4
-	BW0-2	-	-	-	Age	13	221.9	228.9	227.6	209.9
-	-	$BW2\text{-}5^3$	LMA	-	-	11	217.2	231.6	224.7	201.2
-	-	$BW2\text{-}5^3$	-	Rear	-	11	217.3	231.7	224.8	201.3
-	-	$BW2\text{-}5^3$	-	-	Age	11	217.2	231.6	224.7	201.2
-	-	-	LMA	Rear	-	13	222.2	229.2	227.9	210.2
-	-	-	LMA	-	Age	13	221.9	228.9	227.5	209.9
-	-	-	-	Rear	Age	13	221.9	228.9	227.6	209.9

IBW: Body weight at Day -1; BW0-2: Body weight change from Day 0 to Day 2; BW2-5: Body weight change from Day 2 to Day 5; LMA: Locomotor activity; Rear: Rearing; DF: Error degrees of freedom; AIC: Akaike Information Criterion; AICC: corrected AIC; BIC: Bayesian Information Criterion; -2 LL: -2 Log Likelihood.

Table 2.7 Sucrose preference, from simplest (no covariates) to fullest. Higher order relationships (quadratic, cubic) for covariates include all lower order relationships. Models demonstrating improvement are in white, final model is in bold.

Covariates included in the model								Model criteria			
IBW	**ISP**	**BW0-2**	**BW2-5**	**LMA**	**Rear**	**Age**	**DF**	**AIC**	**AICC**	**BIC**	**-2 LL**
-	-	-	-	-	-	-	15	-26.1	-23.2	-22.3	-34.1
IBW	-	-	-	-	-	-	14	-24.2	-19.6	-19.5	-34.2
IBW^2	-	-	-	-	-	-	13	-22.3	-15.3	-16.6	-34.3
IBW^3	-	-	-	-	-	-	12	-20.7	-10.5	-14.1	-34.7
-	ISP	-	-	-	-	-	14	-24.1	-19.5	-19.4	-34.1
-	ISP^2	-	-	-	-	-	13	-22.1	-15.1	-16.5	-34.1
-	ISP^3	-	-	-	-	-	12	-20.1	-10	-13.5	-34.1
-	-	BW0-2	-	-	-	-	14	-24.1	-19.5	-19.3	-34.1
-	-	$BW0\text{-}2^2$	-	-	-	-	13	-22.1	-15.1	-16.4	-34.1
-	-	$BW0\text{-}2^3$	-	-	-	-	12	-20.1	-9.9	-13.5	-34.1
-	-	-	BW2-5	-	-	-	14	-25.9	-21.3	-21.2	-35.9
-	-	-	$BW2\text{-}5^2$	-	-	-	13	-24.1	-17.1	-18.4	-36.1
-	-	-	$BW2\text{-}5^3$	-	-	-	12	-22.1	-11.9	-15.5	-36.1
-	-	-	-	LMA	-	-	14	-24.3	-19.7	-19.6	-34.3
-	-	-	-	LMA^2	-	-	13	-22.6	-15.6	-16.9	-34.6
-	-	-	-	LMA^3	-	-	12	-20.8	-10.6	-14.2	-34.8
-	-	-	-	-	Rear	-	14	-27.6	-23	-22.9	-37.6
-	-	-	-	-	$Rear^2$	-	13	-27.6	-20.6	-21.9	-39.6
-	-	-	-	-	$Rear^3$	-	12	-27	-16.8	-20.4	-41
-	-	-	-	-	-	Age	14	-24.5	-19.8	-19.7	-34.5
-	-	-	-	-	-	Age^2	13	-24.1	-17.1	-18.5	-36.1
-	**-**	**-**	**-**	**-**	**-**	**Age^3**	**12**	**-33.7**	**-23.5**	**-27**	**-47.7**

IBW: Body weight at Day -1; ISP: Sucrose preference at Day -1; BW0-2: Body weight change from Day 0 to Day 2; BW2-5: Body weight change from Day 2 to Day 5; LMA: Locomotor activity; Rear: Rearing; DF: Error degrees of freedom; AIC: Akaike Information Criterion; AICC: corrected AIC; BIC: Bayesian Information Criterion; -2 LL: -2 Log Likelihood.

CHAPTER III: Analytical workflow profiling gene expression in murine macrophages[1]

S. E. Nixon, D. González-Peña, M. A. Lawson, R. H. McCusker, A. G. Hernandez, J. C. O'Connor, R. Dantzer, K. W. Kelley, and S. L. Rodriguez-Zas

3.1 Abstract

Comprehensive and simultaneous analysis of all genes in a biological sample is a capability of RNA-Seq technology. Analysis of the entire transcriptome benefits from summarization of genes at the functional level. As a cellular response of interest not previously explored with RNA-Seq, peritoneal macrophages from mice under two conditions (control and immunologically challenged) were analyzed for gene expression differences. Quantification of individual transcripts modeled RNA-Seq read distribution and uncertainty (using a Beta Negative Binomial distribution), then tested for differential transcript expression (False Discovery Rate-adjusted p-value < 0.05). Enrichment of functional categories utilized the list of differentially expressed genes. A total of 2,079 differentially expressed transcripts representing 1,884 genes were detected. Enrichment of 92 categories from Gene Ontology Biological Processes and Molecular Functions, and KEGG pathways were grouped into 6 clusters. Clusters included defense and inflammatory response (Enrichment Score = 11.24) and ribosomal activity (Enrichment Score = 17.89). Our work provides a context to the fine detail of individual gene expression differences in murine peritoneal macrophages during immunological challenge with high throughput RNA-Seq.

[1] This chapter has been published as an open-access manuscript in the *Journal of Bioinformatics and Computational Biology* (13). The rights to reprint were retained by the authors.
Nixon, S.E., González-Peña, D., Lawson, M.A., McCusker, R.H., Hernandez, A.G., O'Connor, J.C., Dantzer, R., Kelley, K.W., Rodriguez-Zas, S.L. (2015). Analytical workflow profiling gene expression in murine macrophages. *Journal of Bioinformatics and Computational Biology*, **13**(2): 1550010.
http://dx.doi.org/10.1142/S0219720015500109

3.2 Introduction

Identification and analysis of an individual gene may offer limited insights. While genes serve as one of the smallest units by which biological change can be measured, critical information comes from considering the sum of their individual effects. Expanding the "snapshot view" available for differential expression motivates a drive towards the enlargement of analyses from single gene studies with quantitative real-time PCR to microarrays, and more recently RNA-Seq (Oshlack et al., 2010).

The range of tools available for RNA-Seq analysis, as well as the tools themselves, undergoes a rapid pace of modification. These changes demand a thorough understanding of how the tools operate to choose appropriate settings for a particular experiment. Without a singular accepted method or settings to address all applications, transcriptomics relies upon the validation of data quality and controls (Van Verk et al., 2013). TopHat, Cufflinks, and Cuffdiff comprise a set of tools for analyzing RNA-Seq datasets (Trapnell et al., 2012). These tools have gained popularity for the capability to handle intron-spanning reads, and options to address various biological- and technical-biases that are of concern during analysis (Trapnell et al., 2013).

Mapping RNA reads to an annotated genome is one of the popular and well-established methods for differential expression testing between treatments in model organisms (Garber et al., 2011). With the potential to detect thousands of differentially expressed genes, organizing these differences into more interpretable groups becomes the purpose of downstream tools. One possibility that is explored here involves grouping the gene information into groups based upon

their functional actions, a form of gene set enrichment (Huang et al., 2009). This study examines the capability of an RNA-Seq based workflow to evaluate transcriptomic changes. Efficient identification of differentially expressed genes and the functions they impact elucidates their modification of the biological state between treatments. The novelty of this experiment is in the application of RNA-Seq and the associated algorithms to a particular biological model, the analysis of peritoneal macrophages from Bacille Calmette-Guérin (BCG)-challenged mice compared to those receiving a saline control (Moreau et al., 2005). This challenge has been associated with substantial changes in sickness and depression-like indicators (Rodriguez-Zas et al., 2015). The characterization of the transcriptome during immunological resolution and behavioral transition 7 days after initial challenge is of interest. RNA-Seq has yet to be applied to characterize the transcriptome at this time point (Moreau et al., 2008; O'Connor et al., 2009; Platt et al., 2013). The application of RNA-Seq and downstream methods to analyze changes in transcriptomics in this model has not been reported, providing a new level of capability in constructing an inflammation-induced immunological response profile. Transcript profiles were further studied and interpreted using functional enrichment analyses to uncover categories that may be over-represented among particular profiles.

3.3 Materials and Methods

RNA-Seq technology was used to study changes in gene expression in macrophages taken from mice following a previously established immune-challenge model (O'Connor et al., 2009). Male adult (~22 weeks of age; n=6/group) C57BL/6J mice were injected into the peritonium with TICE strain BCG (Organon USA Inc., USA) or equal volume (10mg) physiological saline (Control). Utilizing the same inbred strain as used for the Mouse Genome Project minimizes genetic

variations that could hinder mapping (Levy et al., 2007). RNA was isolated from macrophages collected from the peritoneal cavity 7 d post-challenge (Reno et al., 1997; Zhang et al., 2008). This timing of collection was selected to capture transcriptome changes during a period of immunological and behavioral transitions (Platt et al., 2013; Rodriguez-Zas et al., 2015).

Library Sequencing and Abundance Quantification

The workflow of RNA-Seq data analysis is presented in **Fig. 3.1**. Transcriptomic analysis with RNA-Seq involves producing libraries of reads that represent gene transcripts from the samples for quantitative comparison. Individual mouse RNA-Seq libraries were sequenced using Illumina HiSeq2000 (Illumina, San Diego, CA) to produce paired-end 100-bp reads, summarized as "left" and "right" reads. One library of reads per biological sample was examined for sequencing errors prior to mapping to genome and transcriptome features. Quality control of sequence reads used FastQC (Babraham Institute, 2013; **Fig. 3.2**). Quality was determined by the reported score at each base position (> 30), a Qphred quality value which is the negative logarithmic transformation of the estimated probability of error (Eq. 1; Cock et al., 2010).

$$Q_{PHRED} = -10\ x\ log_{10}(P_e)$$

Reads were mapped to the mouse genome (GRCm38) and assembled using TopHat2 (TopHat v2.0.9) and Cufflinks and analyzed using Cuffmerge, and Cuffdiff 2 (v2.1.1; Trapnell et al., 2012; Fig. 1). TopHat2 maps reads via the use of Bowtie2, the core read-alignment program, while TopHat2 deals with splicing concerns from mapping intron-spanning RNA reads to a DNA genome (Trapnell et al., 2012). Due to the computational scale of mapping millions of reads to large genomes, Bowtie2 implements Burrows-Wheeler transformation to efficiently scan the

genome during mapping (Langmead and Salzberg, 2012). TopHat2 was chosen for its two-step method to deal with spliced alignments and preferential alignment of reads onto real genes from an annotation (Kim et al., 2013).

Reads were assembled based upon mapping information into gene transcripts, with transcripts quantified by condition for differential comparison as elaborated in Trapnell et al. (2012). The Cufflinks program (http://cufflinks.cbcb.umd.edu/) takes the mapping information from TopHat2, and assembles the reads back into the biologically-relevant transcripts that would have produced them. Cufflinks offers optional assembly methods that correct for biological and technical biases, including biases in Illumina's read-creation process (Hansen et al., 2010). Options to correct for fragment bias during transcription priming with random hexamers and estimation of appropriate counting for those reads that can map to multiple sites were used (Mortazavi et al., 2008; Roberts et al., 2011). Upper Quartile normalization was enabled for its superior performance compared to the default Total Count method available in Cufflinks (Dillies et al., 2013).

Cuffdiff 2 (referred to from here on simply as "Cuffdiff") performs differential expression testing between conditions by checking if each gene follows a beta negative binomial distribution. The beta negative binomial distribution can account for potential overdispersion between groups or uncertainty in read counts that may otherwise be ignored by simpler models (Trapnell et al., 2013). Before any testing for significance, all loci in the genome first needed a minimum number of fragment alignments (10 fragments; Test Status "OK"). Genes within a locus could be analyzed for significance after this minimum alignment (MA) within Cuffdiff was satisfied. Of those genes

in locations with > 10 fragment alignments, a list of genes exhibiting significant differential expression between conditions (False Discovery Rate or FDR-adjusted p-value < 0.05) was obtained. The genes were named based upon annotation available from the UCSC database (Karolchik et al., 2014; www.genome.ucsc.edu).

Functional Analysis

Two complementary approaches were used to identify functional categories among transcript profiles. Enrichment based on the hypergeometric test applied to a list of differentially abundant transcript isoforms and gene set enrichment analysis (GSEA) of all transcript isoforms based on the Kolmogorov-Smirnov statistics were evaluated (Subramanian et al., 2005). Gene Ontological (GO; Ashburner et al., 2000; www.geneontology.org) terms related to Biological Process (BP) and Molecular Function (MF) were tested, along with the Kyoto Encyclopedia of Genes and Genomes (KEGG)-Pathway database (Kanehisa et al., 2008; http://www.genome.jp/kegg). For the hypergeometric test, functional category enrichment and functional annotation clustering were performed in the Database for Annotation, Visualization, and Integrated Discovery (DAVID; Huang et al., 2009). Specifically the GO FAT categories within DAVID were tested, a filter of GO categories to minimize repetition of general categories and to focus on more specific term identification. Individual categories in DAVID are deemed enriched by using a one-tailed jackknifed Fisher exact test, the EASE score (Hosack et al., 2003). The downstream functional annotation clustering of these categories used Enrichment Score, calculated as the –log scale geometric mean of the EASE scores of member categories (Serão et al., 2013).

For the purposes of clustering, DAVID by default considers categories individually by their EASE score (EASE ≤ 0.1) without concern for experiment-wide false-detection (Hosack et al., 2003). To avoid errors related to multiple tests, categories were only considered enriched if they were significant at FDR-adjusted EASE score based p-value < 0.1 (Delfino and Rodriguez-Zas, 2013). Cluster Enrichment Scores were re-calculated to reflect the remaining member categories (Enrichment Score > 4). The GSEA methodology was implemented using the software package GSEA-P and enrichment was tested against the functional categories present in the Molecular Signature Database (MSigDB, (Subramanian et al., 2005). The recommended GSEA FDR-adjusted p-value < 0.25 threshold was used in agreement with the statistical testing implemented (Subramanian et al., 2005). Categories consistent between the hypergeometric and GSEA approaches are reported and discussed. These results are robust to differences in assumptions and methodologies between the approaches.

Network Visualization

Network designs can offer information on relationships between categories, within and across clusters. Using the categories enriched in both DAVID and GSEA methods as input, networks were created with the Enrichment Map plugin (Merico et al., 2010) for Cytoscape (http://www.cytoscape.org; Killcoyne et al., 2009). In the networks, enriched categories are represented by nodes and relationships between categories as edges. An edge between two node categories required at least a 50% overlap in enriching member genes from the Cuffdiff output gene list (similarity coefficient > 0.5). Node size represented a greater number of enriching member genes, while edge thickness represented the range of overlap coefficients from 0.5 to 1. Finally, nodes were colored by the direction of enrichment for member genes. Red nodes had at

least 60% of enriching member genes over-expressed in the BCG treatment versus control. Blue nodes had at least 60% of enriching member genes under-expressed in the BCG treatment versus control, and nodes were grey if enriching member genes were split between over-expression and under-expression.

3.4 Results

The quality control was evaluated for every sample. No evidence of low quality reads was observed within the samples, with quality scores greater than 30 across the entire length of the reads. Quality scores were similarly high across both Control and BCG groups (**Fig 3.2**). Scores ranged between 30 and 40, indicating accuracies between 99.9% and 99.99% for the bases at those positions. Based upon the observed quality of the sample data as well as the read filtering internal to TopHat2, trimming was not needed.

The RNA-Seq reads produced 54 ± 8.5 million reads and 64 ± 6 million paired-end reads of 100 bp in length per sample for Control and BCG, respectively. On average 91% of total reads mapped to the genome for both Control and BCG. The percentage of reads per sample that successfully mapped to the genome ranged from 74% to 95%, the same range as the percentage of reads that produced aligning pairs (**Table 3.1**).

Following evaluation of read quality, their assembly into transcripts produced over 60,000 transcripts among all samples. Prior to differential testing between the groups, these transcripts

were filtered based upon sufficient alignment coverage and experiment-wide significance cut-offs. The number of differentially expressed transcripts between Control and BCG groups was 2,079 (**Table 3.2**; 1,373: FDR p-value < 0.01; 706: 0.01 < FDR p-value < 0.05), representing 1,884 genes.

Among the differentially expressed genes, 802 were under-expressed in BCG versus control, indicating similar quantity of up- and down-regulated genes post-challenge. However, there was a predominance of genes over-expressed in BCG relative to Control among the most significant profiles. The most significantly differentially expressed genes (FDR p-value < 0.01) are listed separately for those over- (**Table 3.3**) and under- (**Table 3.4**) expressed in BCG versus Control, together with supporting references when previously associated with macrophage populations and their immunological response profile.

Functional categorization of the gene list resulted in 92 significantly enriched terms (BP: 69 terms; MF: 20 terms; KEGG: 3 terms; FDR p-value < 0.1; not listed). Clustering the enriched terms further reduced the list to 6 highly enriched clusters (Enrichment Score > 4), listed by score in **Table 3.5**. Clusters were dominated by GO Biological Process terms as they were the majority of the significantly enriched term list, with terms in the clusters underscoring the activation and regulation of the immune system following challenge. These clusters accounted for 24 of the significantly enriched terms. Constructing a network from the clustered terms and any connected (overlap coefficient > 0.5) enriched terms, neither Clusters 1 or 4 had any external connections

while the other clusters were directly or indirectly connected along with 8 additional non-clustered terms (**Figure 3.3**).

3.5 Discussion

Quality control of the input reads is an important step to successful downstream mapping. Once the reads were determined to be of high quality, the filtering controls implemented by TopHat2 prior to mapping made additional trimming of the reads unnecessary (**Table 3.1**; Trapnell et al., 2009). Percentages of mapped reads were similar to those reported in previous high-stringency methods, and approached the percentages seen when previously tested on simulated error-free data (Mende et al., 2012; Kim et al., 2013). The mapping capability of aligners like TopHat2 is dependent upon the genome and annotation, meaning unmapped reads may include those associated with transcripts not yet represented in the annotation. Findings from these RNA-Seq confirmed several results from previous studies that used similar models and quantitative real-time PCR or microarray technologies and uncovered additional profiles and enriched categories. This study centered on one type of peripheral macrophage, collected at one time point and using a specific collection method on macrophage activation status. A longitudinal study of additional macrophage populations using alternative collection methods is necessary for extrapolation of our findings to wider conditions.

The workflow described here effectively identifies genes that are differentially expressed during an immunological challenge and clusters these results based upon functionality. Significantly differentially expressed genes illustrated the extended expression response after BCG-challenge.

Among the over-expressed genes, the most over-expressed gene S100a9 works as a heterodimer with S100a8, also found to be in the over-expressed list (**Table 3.3**). As both are associated with inflammatory events and are inducible in mature macrophages, their presence after BCG-challenge is expected (Ehrchen et al., 2009). The over-expression of CCL5 and CXCL10 (**Table 3.3**) was also unsurprising, considering the inhibitory action of IL-10 upon both, and that reduced IL-10 levels were associated with increased resistance to intracellular pathogens (Moore et al., 2001) such as BCG (Il10, **Table 3.4**). The overexpression of Arg1 in the BCG group is consistent with previous work studying the effect of this challenge in macrophages (Speranza et al., 2010). Along with the most under-expressed gene in BCG compared to Control, Retnla, these indicate under-expression of Th2-associated genes due to the classic Th1-response to BCG (Gordon and Martinez, 2010; Pesce et al., 2009). It is interesting to find Mt1 and Il10 together in the under-expressed category. Although studies were previously performed in T cells, Mt1/Mt2-deficient mice were found to produce increased levels of Il10 following an immune challenge with anti-CD3/CD28 (Wu et al., 2013). Still, the role of metallothionein genes during immune challenges and inflammation are not fully elucidated, and low expression of Mt1 supports the proinflammatory nature of the response at the measured time-point (Inoue et al., 2009). Direct association in the literature between macrophages and the under-expressed gene Ptprcap was less clear, although it has been found in the monocyte precursors to macrophages (Thamilarasan et al., 2013). However, Ptprcap is known as a CD45-associate, regulating the interaction of CD45 with other proteins. As CD45 regulates apoptosis, this may explain the relationship to immune-challenge (Dupéré-Minier et al., 2010). Although the number of differentially expressed genes were similarly split between over-expressed and under-expressed in the BCG relative to the control group, a more stringent significance cutoff found a predominance of genes over-expressed in the BCG group. These results

are consistent with other reports of over-expression in the microglia of genes associated with inflammation response in response to an inflammatory challenge (Lee et al., 2014; Przanowski et al., 2014).

Table 3.5 summarizes the enriched functional categories consistently detected by the hypergeometric test and GSEA approaches. Enrichment analysis highlighted the biological response of macrophages to an immunological challenge (Inflammation-based defense responses; clusters 2 and 3 in Table 5). Clustering was effective at identifying cytokine and chemokine activity in immune cells that are typically associated with activation of macrophages (Mantovani et al., 2004) and apoptosis as a part of the response to intracellular pathogens including Mycobacteria (Danelishvili et al., 2010). Additional terms related to the regulation of cytokine production (GO:0001817) and their binding (GO:0019955), the interaction of cytokines with their receptors (KEGG mmu04060) and chemokine signaling (KEGG mmu04062) were not clustered, but enriched and connected to clustered terms (**Figure 3.3**). While not identified by the cluster method, GO categories related to the immune response (GO:0006955) and adaptive immune response (GO:0002250; GO:0002460) were connected in the network visualization and served as a link between apoptotic and inflammation-based cateogires. Categories previously associated with similar immune challenges (Marquis et al., 2011) were clustered to better clarify the transcriptomic differences between experimental groups. Several genes that were over-expressed in BCG relative to Control (**Table 3.3**) are affiliated to regulation of locomotion (**Table 3.5**) including Xcl1, Cxcl13, Cxcr2, Cxcl10, and Ccl5. These associations could be related to the typical amelioration of sickness behaviors and higher activity observed in mice 7 days post-challenge (O'Connor et al., 2009; Rodriguez-Zas et al., 2015). A ribosomal cluster (cluster 1) dominating

the list is expected, as protein regulation is at the core of immunological response (Mantovani et al., 2004). However network visualization showed that both the ribosomal cluster and carbohydrate-based binding clusters possessed enriching member genes that were unique when compared to the other immunologically based categories (**Figure 3.3**). The gene lists and resulting clusters from RNA-Seq technology allows for analysis based upon the shared and unique genes. In future studies, this response profile of immunologically-challenged peritoneal macrophages can be compared to similar constructed profiles of other cell populations or challenges to identify profile characteristics unique to each combination.

3.6 References

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Mates,e J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G. (2000). Gene ontology: tool for the unification of biology, *Nature Genetics*, **25**(1): 25-29. http://dx.doi.org/10.1038/75556

The Babraham Institute. FastQC. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/, 2013.

Beutner, C., Linnartz-Gerlach, B., Schmidt, S.V., Beyer, M., Mallmann, M.R., Staratschek-Jox, A., Schultze, J.L., Neumann, H. (2013). Unique transcriptome signature of mouse microglia, *Glia*, **61**(9): 1429-1442. http://dx.doi.org/10.1002/glia.22524

Cock, P.J., Fields, C.J., Goto, N., Heuer, M.L., Rice, P.M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants, *Nucleic Acids Research*, **38**(6): 1767-1771. http://dx.doi.org/10.1093/nar/gkp1137

Danelishvili, L., Yamazaki, Y., Selker, J., Bermudez, L.E. (2010). Secreted Mycobacterium tuberculosis Rv3654c and Rv3655c proteins participate in the suppression of macrophage apoptosis, *PLoS One*, **4**(5): e10474. http://dx.doi.org/10.1371/journal.pone.0010474

Delfino, K., Rodriguez-Zas, S.L. (2013). Transcription factor-microRNA-target gene networks associated with ovarian cancer survival and recurrence, *PLoS One*, **8**(3): e58608. http://dx.doi.org/10.1371/journal.pone.0058608

Dillies, M.A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., Keime, C., Marot, G., Castel, D., Estelle, J., Guernec, G., Jagla, B., Jouneau, L., Laloë, D., Le Gall, C., Schaëffer, B., Le Crom, S., Guedj, M., Jaffrézic, F. (2013). A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis, *Briefings in Bioinformatics*, **14**(6):671-683. http://dx.doi.org/10.1093/bib/bbs046

Dupéré-Minier, G., Desharnais, P., Bernier, J. (2010). Involvement of tyrosine phosphatase CD45 in apoptosis, *Apoptosis*, **15**(1): 1-13. http://dx.doi.org/10.1007/s10495-009-0413-z

Ehrchen, J.M., Sunderkötter, C., Foell, D., Vogl, T., Roth, J. (2009). The endogenous Toll-like receptor 4 agonist S100A8/S100A9 (calprotectin) as innate amplifier of infection, autoimmunity, and cancer, *Journal of Leukocyte Biology*, **86**(3): 557-566. http://dx.doi.org/10.1189/jlb.1008647

Ehrt, S., Schnappinger, D., Bekiranov, S., Drenkow, J., Shi, S., Gingeras, T.R., Gaasterland, T., Schoolnik, G., Nathan, C. (2001). Reprogramming of the macrophage transcriptome in response to interferon-γ and Mycobacterium tuberculosis: signaling roles of nitric oxide synthase-2 and phagocyte oxidase, *Journal of Experimental Medicine*, **194**(8): 1123-1140. http://dx.doi.org/10.1084/jem.194.8.1123

Espejo, C., Penkowa, M., Demestre, M., Montalban, X., Martínez-Cáceres, E.M. (2005). Time-course expression of CNS inflammatory, neurodegenerative tissue repair markers and metallothioneins during experimental autoimmune encephalomyelitis, *Neuroscience*, **132**(4): 1135-1149. http://dx.doi.org/10.1016/j.neuroscience.2005.01.057

Garber, M., Grabherr, M.G., Guttman, M., Trapnell, C. (2011). Computational methods for transcriptome annotation and quantification using RNA-Seq, *Nature Methods*, **8**(6): 469-477. http://dx.doi.org/10.1038/nmeth.1613

Gladue, D.P., Zhu, J., Holinka, L.G., Fernandez-Sainz, I., Carrillo, C., Prarat, M.V., O'Donnell, V., Borca, M.V. (2010). Patterns of gene expression in swine macrophages infected with classical swine fever virus detected by microarray, *Virus Research*, **151**(1):10-18. http://dx.doi.org/10.1016/j.virusres.2010.03.007

Gordon, S., Martinez, F.O. (2010). Alternative activation of macrophages: mechanism and functions, *Immunity*, **32**(5): 593-604. http://dx.doi.org/10.1016/j.immuni.2010.05.007

Gundra, U.M., Girgis, N.M., Ruckerl, D., Jenkins, S., Ward, L.N., Kurtz, Z.D., Wiens, K.E., Tang, M.S., Basu-Roy, U., Mansukhani, A., Allen, J.E., Loke, P. (2014). Alternatively activated macrophages derived from monocytes and tissue macrophages are phenotypically and functionally distinct, *Blood*, **123**(20): e110-122. http://dx.doi.org/10.1182/blood-2013-08-520619

Hansen, K.D., Brenner, S.E., Dudoit, S. (2010). Biases in Illumina transcriptome sequencing caused by random hexamer priming, *Nucleic Acids Research*, **38**(12): e131. http://dx.doi.org/10.1093/nar/gkq224

Hickman, S.E., Kingery, N.D., Ohsumi, T.K., Borowsky, M.L., Wan, L., Means, T.K., Khoury, J.E. (2013). The microglial sensome revealed by direct RNA sequencing, *Nature Neuroscience*, **16**(12): 1896-1905. http://dx.doi.org/10.1038/nn.3554

Hosack, D.A., Dennis, G. Jr, Sherman, B.T., Lane, H.C., Lempicki, R.A. (2003). Identifying biological themes within lists of genes with EASE, *Genome Biology*, **4**(10): R70. http://dx.doi.org/10.1186/gb-2003-4-10-r70

Huang, D.W., Sherman, B.T., Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, *Nature Protocols*, **4**(1): 44-57. http://dx.doi.org/10.1038/nprot.2008.211

Inoue, K., Takano, H., Shimada, A., Satoh, M. (2009). Metallothionein as an anti-inflammatory mediator, *Mediators of Inflammation*, **2009**: 101659. http://dx.doi.org/10.1155/2009/101659

Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T., Yamanishi, Y. (2008). KEGG for linking genomes to life and the environment, *Nucleic Acids Research*, **36**(S1): D480-484. http://dx.doi.org/10.1093/nar/gkm882

Karlstetter, M., Walczak, Y., Weigelt, K., Ebert, S., Van den Brulle, J., Schwer, H., Fuchshofer, R., Langmann, T. (2010). The novel activated microglia/macrophage WAP domain protein, AMWAP, acts as a counter-regulator of proinflammatory response, *Journal of Immunology*, **185**(6): 3379-3390. http://dx.doi.org/10.4049/jimmunol.0903300

Karolchik, D., Barber, G.P., Casper, J., Clawson, H., Cline, M.S., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M., Harte, R.A., Heitner, S., Hinrichs, A.S., Learned, K., Lee, B.T., Li, C.H, Raney, B.J., Rhead, B., Rosenbloom, K.R., Sloan, C.A., Speir, M.L., Zweig, A.S., Haussler, D., Kuhn, R.M., Kent, W.J. (2014). The UCSC genome browser

database: 2014 update, *Nucleic Acids Research*, **42**(Database issue): D764-770. http://dx.doi.org/10.1093/nar/gkt1168

Killcoyne, S., Carter, G.W., Smith, J., Boyle, J. (2009). Cytoscape: A community-based framework for network modeling, *Methods in Molecular Biology*, **563**:219-239. http://dx.doi.org/10.1007/978-1-60761-175-2_12

Kim, B.O., Liu, Y., Zhou, B.Y., He, J.J. (2004). Induction of C chemokine XCL1 (lymphotactin/single C motif-1α/activation-induced, T cell-derived and chemokine-related cytokine) expression by HIV-1 tat protein, *Journal of Immunology*, **172**(3): 1888-1895. http://dx.doi.org/10.4049/jimmunol.172.3.1888

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions, *Genome Biology*, **14**(4): R36. http://dx.doi.org/10.1186/gb-2013-14-4-r36

Kota, R.S., Rutledge, J.C., Gohil, K., Kumar, A., Enelow, R.I., Ramana, C.V. (2006). Regulation of gene expression in RAW 264.7 macrophage cell line by interferon-gamma, *Biochemical and Biophysical Research Communications*, **342**(4): 1137-1146. http://dx.doi.org/10.1016/j.bbrc.2006.02.087

Langmead, B., Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie2, *Nature Methods*, **9**(4): 357-359. http://dx.doi.org/10.1038/nmeth.1923

Lee, H.T., Kim, S.K., Kim, S.H., Kim, K., Lim, C.H., Park, J., Roh, T.Y., Kim, N., Chai, Y.G. (2014). Transcription-related element gene expression pattern differs between microglia and macrophages during inflammation, *Inflammation Research*, **63**(5): 389-397. http://dx.doi.org/10.1007/s00011-014-0711-y

Levy, S., Sutton, G., Ng, P.C., Feuk, L., Halpern, A.L., Walenz, B.P., Axelrod, N., Huang, J., Kirkness, E.F., Denisov, G., Lin, Y., MacDonald, J.R., Pang, A.W.C., Shago, M., Stockwell, T.B., Tsiamouri, A., Bafna, V., Bansal, V., Kravitz, S.A., Busam, D.A., Beeson, K.Y., McIntosh, T.C., Remington, K.A., Abril, J.F., Gill, J., Borman, J., Rogers, Y.H., Frazier, M.E., Scherer, S.W., Strausberg, R.L., Ventner, J.C. (2007). The diploid genome sequence of an individual human, *PLoS Biology*, **5**(10): e254. http://dx.doi.org/10.1371/journal.pbio.0050254

Liu, H., Liu, Z., Chen, J., Chen, L., He, X., Zheng, R., Yang, H., Song, P., Weng, D., Hu, H., Fan, L., Xiao, H., Kaufmann, S.H.E., Ernst, J., Ge, B. (2013). Induction of CCL8/MCP-2 by Mycobacteria through the activation of TLR2/PI3K/Akt signaling pathway, *PLoS One*, **8**(2): e56815. http://dx.doi.org/10.1371/journal.pone.0056815

Liu, Q., Zheng, J., Yin, D.D., Xiang, J., He, F., Wang, Y.C., Liang, L., Qin, H.Y., Liu, L., Liang, Y.M., Han, H. (2012). Monocyte to macrophage differentiation-associated (MMD) positively regulates ERK and Akt activation and TNF-α and NO production in macrophages, *Molecular Biology Reports*, **39**(5): 5643-5650. http://dx.doi.org/10.1007/s11033-011-1370-5

MacMicking, J., Xie, Q.W., Nathan, C. (1997). Nitric oxide and macrophage function, *Annual Review of Immunology*, **15**(1): 323-350. http://dx.doi.org/10.1146/annurev.immunol.15.1.323

Magee, D.A., Taraktsoglou, M., Killick, K.E., Nalpas, N.C., Browne, J.A., Park, S.D., Conlon, K.M., Lynn, D.J., Hokamp, K., Gordon, S.V., Gormley, E., MacHugh, D.E. (2012). Global gene expression and systems biology analysis of bovine monocyte-dervied macrophages in response to in vitro challenge with Mycobacterium bovis, *PLoS One*, **7**(2): e32034. http://dx.doi.org/10.1371/journal.pone.0032034

Mantovani, A., Sica, A., Sozzani, S., Allavena, P., Vecchi, A., Locati, M. (2004). The chemokine system in diverse forms of macrophage activation and polarization, *Trends in Immunology*, **25**(12): 677-686. http://dx.doi.org/10.1016/j.it.2004.09.015

Marquis, J.F., Kapoustina, O., Langlais, D., Ruddy, R., Dufour, C.R., Kim, B.H., MacMicking, J.D., Giguère, V., Gros, P. (2011). Interferon regulatory factor 8 regulates pathways for antigen presentation in myeloid cells and during tuberculosis, *PLoS Genetics*, **7**(6):e1002097. http://dx.doi.org/10.13/journal.pgen.1002097

Martinez, A.N., Mehra, S., Kaushal, D. (2013). Role of Interleukin 6 in innate immunity to Mycobacterium tuberculosis infection, *Journal of Infectious Diseases*, **207**(8): 1253-1261. http://dx.doi.org/10.1093/infdis/jit037

Mende, D.R., Waller, A.S., Sunagawa, S., Järvelin, A.I., Chan, M.M., Arumugam, M., Raes, J., Bork, P. (2012). Assessment of metagenomic assembly using simulated next generation sequencing data, *PLoS One*, **7**(2): e31386. http://dx.doi.org/10.1371/journal.pone.0031386

Merico, D., Isserlin, R., Stueker, O., Emili, A., Bader, G.D. (2010). Enrichment map: a network-based method for gene-set enrichment visualization and interpretation, *PLoS One*, **5**(11): e13984. http://dx.doi.org/10.1371/journal.pone.0013984

Moore, K.W., de Waal Malefyt, R., Coffman, R.L., O'Garra, A. (2001). Interleukin-10 and the interleukin-10 receptor, *Annual Review of Immunology*, **19**: 683-765. http://dx.doi.org/10.1146/annurev.immunol.19.1683

Moreau, M., André, C., O'Connor, J.C., Dumich, S.A., Woods, J.A., Kelley, K.W., Dantzer, R., Lestage, J., Castanon, N. (2008). Inoculation of Bacillus Calmette-Guérin to mice induces an acute episode of sickness behavior followed by chronic depressive-like behavior, *Brain, Behavior, and Immunity*, **22**(7): 1087-1095. http://dx.doi.org/10.1016/j.bbi.2008.04.001

Moreau, M., Lestage, J., Verrier, D., Mormède, C., Kelley, K.W., Dantzer, R., Castanon, N. (2005). Bacille Calmette-Guérin inoculation induces chronic activation of peripheral and brain indoleamine 2,3-dioxygenase in mice, *Journal of Infectious Diseases*, **192**(3): 537-544. http://dx.doi.org/10.1086/431603

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq, *Nature Methods*, **5**: 621-628. http://dx.doi.org10.1038/nmeth.1226

Mulcahy, H., O'Rourke, K.P., Adams, C., Molloy, M.G., O'Gara, F. (2006). LST1 and NCR3 expression in autoimmune inflammation and in response to IFN-γ, LPS and microbial infection, *Immunogenetics*, **57**(12): 893-903. http://dx.doi.org/10.1007/s00251-005-0057-2

de Oliveira, C.C., de Oliveira, S.M., Goes, V.M., Probst, C.M., Krieger, M.A., Buchi Dde, F. (2008). Gene expression profiling of macrophages following mice treatment with an immunomodulator medication, *Journal of Cellular Biochemistry*, **104**(4): 1364-1377. http://dx.doi.org/10.1002/jcb.21713

Oshlack, A., Robinson, M.D., Young, M.D. (2010). From RNA-Seq reads to differential expression results, *Genome Biology*, **11**(12): 220. http://dx.doi.org/10.1186/gb-2010-11-12-220

Osorio y Fortéa, J., de La Llave, E., Regnault, B., Coppée, J.Y., Milon, G., Lang, T., Prina, E. (2009). Transcriptional signatures of BALB/c mouse macrophages housing multiplying Leishmania amazonensis amastigotes, *BMC Genomics*, **10**: 119. http://dx.doi.org/10.1186/1471-2164-10-119

O'Connor, J.C., Lawson, M.A., André, C., Briley, E.M., Szegedi, S.S., Lestage, J., Castanon, N., Herkenham, M., Dantzer, R., Kelley, K.W. (2009). Induction of IDO by Bacille Calmette-Guérin is responsible for development of murine depressive-like behavior. *Journal of Immunology*, **182**(5): 3202-3212. http://dx.doi.org/10.4049/jimmunol.0802722

Pesce, J.T., Ramalingam, T.R., Wilson, M.S., Mentink-Kane, M.M., Thompson, R.W., Cheever, A.W., Urban, J.F. Jr., Thomas, A., Wynn, T.A. (2009). Retnla (Relma/Fizz1) suppresses Helminth-induced Th2-type-immunity, *PLoS Pathogens*, **5**(4): e1000393. http://dx.doi.org/10.1371/journal.ppat.1000393

Platt, B., Schulenberg, J., Klee, N., Nizami, M., Clark, J.A. (2013). A depressive phenotype induced by Bacille Calmette Guérin in 'susceptible' animals: sensitivity to antidepressants, *Psychopharmacology*, **226**(3): 501-513. http://dx.doi.org/10.1007/s00213-012-2923-6

Przanowski, P., Dabrowski, M., Ellert-Miklaszewska, A., Kloss, M., Mieczkowski, J., Kaza, B., Ronowicz, A., Hu, F., Piotrowski, A., Kettenmann, H., Komorowski, J., Kaminska, B. (2014). The signal transducers Stat1 and Stat3 and their novel target Jmjd3 drive the expression of inflammatory genes in microglia, *Journal of Molecular Medicine (Berlin, Germany)*, **92**(3): 239-254. http://dx.doi.org/10.1007/s00109-013-1090-5

Ramadas, R.A., Ewart, S.L., Medoff, B.D., LeVine, A.M. (2011). Interleukin-1 family member 9 stimulates chemokine production and neutrophil influx in mouse lungs, *American Journal of Respiratory Cell and Molecular Biology*, **44**(2): 134-145. http://dx.doi.org/10.1165/rcmb.2009-0315OC

Reno, C., Marchuk, L., Sciore, P., Frank, C.B., Hart, D.A. (1997). Rapid isolation of total RNA from small samples of hypocellular, dense connective tissues, *BioTechniques*, **22**: 1082-1086, 1997.

Roberts, A., Trapnell, C., Donaghey, J., Rinn, J.L., Pachter, L. (2011). Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biology*, **12**(3): R22. http://dx.doi.org/10.1186/gb-2011-12-3-r22

Rodriguez-Zas, S.L., Nixon, S.E., Lawson, M.A., Mccusker, R.H., Southey, B.R., O'Connor, J.C., Dantzer, R., Kelley, K.W. (2015). Advancing the understanding of behaviors associated with

Bacille Calmette Guérin infection using multivariate analysis. *Brain, Behavior, and Immunity*, **44**: 176-186. http://dx.doi.org/10.1016/j.bbi.2014.09.018

Schmieder, A., Michel, J., Schönhaar, K., Goerdt, S., Schledzewski, K. (2012). Differentiation and gene expression profile of tumor-associated macrophages, *Seminars in Cancer Biology*, **22**(4): 289-297. http://dx.doi.org/10.1016/j.semcancer.2012.02.002

Serão, N.V., González-Peña, D., Beever, J.E., Faulkner, D.B., Southey, B.R., Rodriguez-Zas, S.L. (2013). Single nucleotide polymorphisms and haplotypes associated with feed efficiency in beef cattle, *BMC Genetics*, **14**: 94. http://dx.doi.org/10.1186/1471-2156-14-94

Sigruener, A., Buechler, C., Bared, S.M., Grandl, M., Aslanidis, C., Ugocsai, P., Gehrmann, M., Schmitz, G. (2007). E-LDL upregulates TOSO expression and enhances the survival of human macrophages, *Biochemical and Biophysical Research Communications*, **359**(3):723-728. http://dx.doi.org/10.1016/j.bbrc.2007.05.169

Sohn, S.H., Ko, E., Kim, S.H., Kim, Y., Shin, M., Hong, M., Bae, H. (2009). Genome wide expression profile of Agrimonia pilosa in LPS-stimulated BV-2 microglial cells, *Molecular and Cellular Toxicology*, **5**(1): 1-6, 2009.

Speranza, V., Colone, A., Cicconi, R., Palmieri, G., Giovanni, D., Grassi, M., Mattei, M., Sali, M., Delogu, G., Andreola, F., Colizzi, V., Mariani, F. (2010). Recombinant BCG-Rv1767 amount determines, in vivo, antigen-specific T cells location, frequency, and protective outcome, *Microbial Pathogenesis*, **48**(5):150-159. http://dx.doi.org/10.1016/j.micpath.2010.02.003

Stables, M.J., Shah, S., Camon, E.B., Lovering, R.C., Newson, J., Bystrom, J., Farrow, S., Gilroy, D.W. (2011). Transcriptomic analyses of murine resolution-phase macrophages, *Blood*, **118**(26): e192-208. http://dx.doi.org/10.1182/blood-2011-04-345330

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, *Proceedings of the National Academy of Sciences of the United States of America*, **102**(43): 15545-15550. http://dx.doi.org/10.1073/pnas.0506580102

Thamilarasan, M., Hecker, M., Goertsches, R.H., Paap, B.K., Schröder, I., Koczan, D., Thiesen, H.J., Zettl, U.K. (2013). Glatiramer acetate treatment effects on gene expression in monocytes of multiple sclerosis patients, *Journal of Neuroinflammation*, **10**:126. http://dx.doi.org/10.1186/1742-2094-10-126

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, R.L., Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq, *Nature Biotechnology*, **31**(1): 46-53. http://dx.doi.org/10.1038/nbt.2450

Trapnell, C., Pachter, L., Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq, *Bioinformatics*, **25**(9): 1105-1111. http://dx.doi.org/10.1093/bioinformatics/btp120

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq

experiments with TopHat and Cufflinks, *Nature Protocols*, **7**(3): 562-578. http://dx.doi.org/10.1038/nprot.2012.016

Uhrin, P., Perkmann, T., Binder, B., Schabbauer, G. (2013). ISG12 is a critical modulator of innate immune responses in murine models of sepsis, *Immunobiology*, **218**(9): 1207-1206. http://dx.doi.org/10.1016/j.imbio.2013.04.009

Van Verk, M.C., Hickman, R., Pieterse, C.M., Van Wees, S.C. (2013). RNA-Seq: revelation of the messengers, *Trends in Plant Science*, **18**(4): 175-179, 2013. http://dx.doi.org/10.1016/j.tplants.2013.02.001

Vogt, L., Schmitz, N., Kurrer, M.O., Bauer, M., Hinton, H.I., Behnke, S., Gatto, D., Sebbel, P., Beerli, R.R., Sonderegger, I., Kopf, M., Saudan, P., Bachmann, M.F. (2006). VSIG4, a B7 family-related protein, is a negative regulator of T cell activation, *Journal of Clinical Investigation*, **116**(10): 2817-2826. http://dx.doi.org/10.1172/JCI25673

Wu, C., Pot, C., Apetoh, L., Thalhamer, T., Zhu, B., Murugaiyan, G., Xiao, S., Lee, Y., Rangachari, M., Yosef, N., Kuchroo, V.K. (2013). Metallothioneins negatively regulate IL-27-induced type 1 regulatory T-cell differentiation, *Proceedings of the National Academy of Sciences of the United States of America*, **110**(19): 7802-7807. http://dx.doi.org/10.1073/pnas.1211776110

Zhang, X., Goncalves, R., Mosser, D.M. (2008). The isolation and characterization of murine macrophages, *Current Protocols in Immunology*, **83**(14.1): 14.1.1-14.1.114. http://dx.doi.org/10.1002/0471142735.im1401s83

3.7 Figures

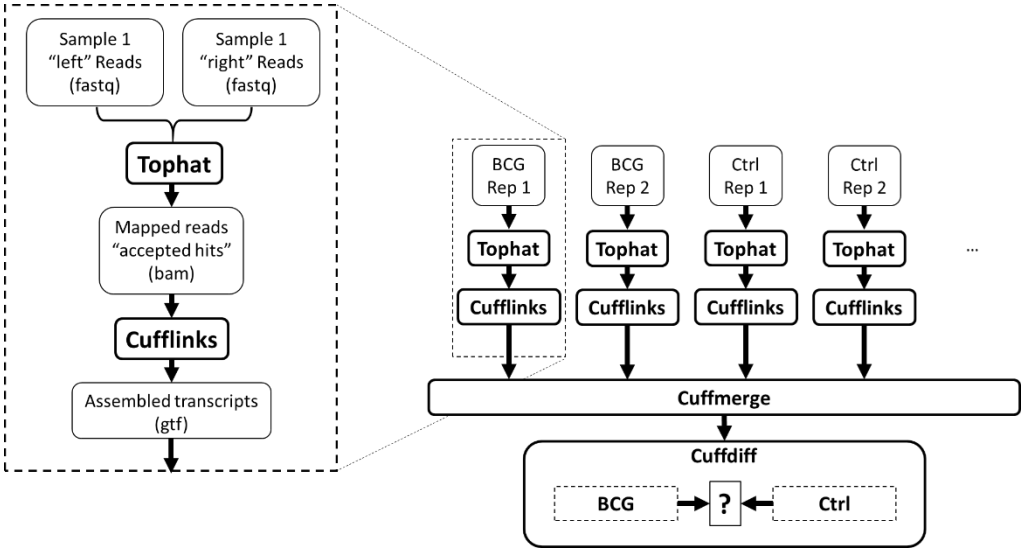


Figure 3.1 RNA-Seq workflow, showing the analysis of each sample individually by Tophat and Cufflinks (inset) before the collective analysis of all samples in Cuffdiff to test for differential expression between conditions

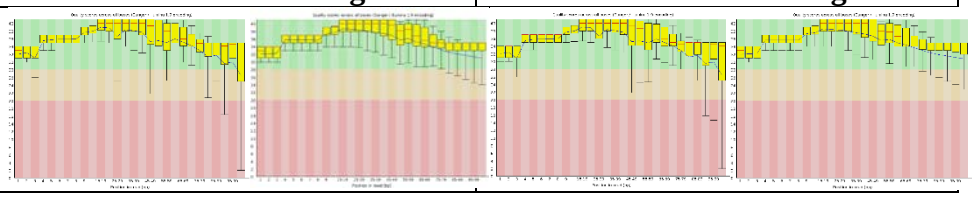
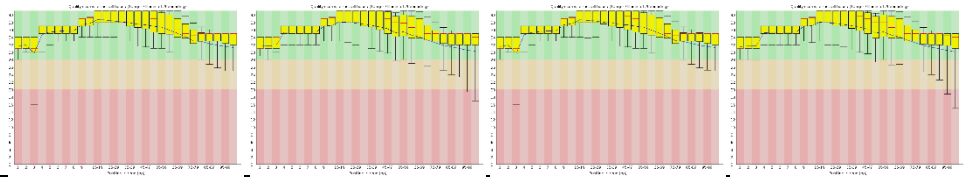
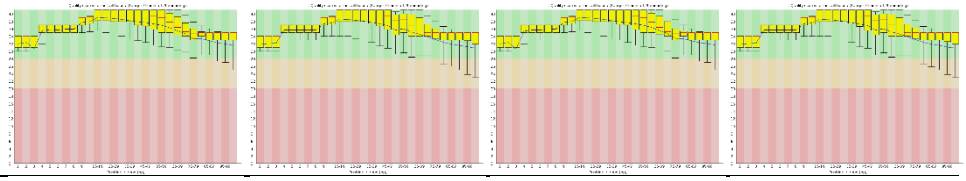
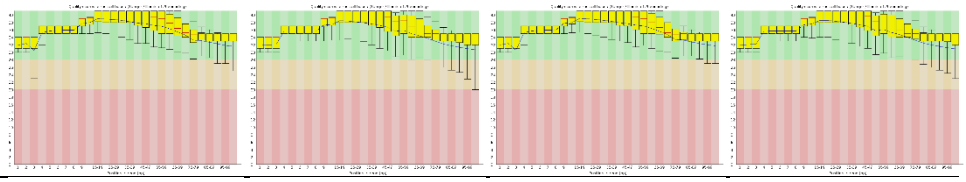
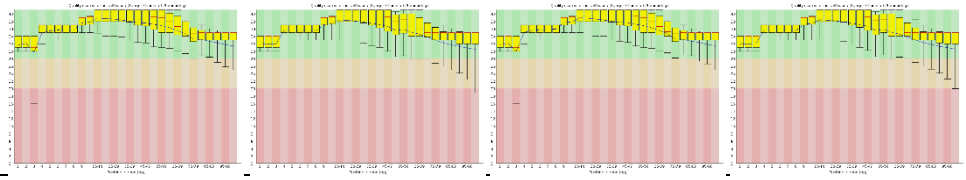
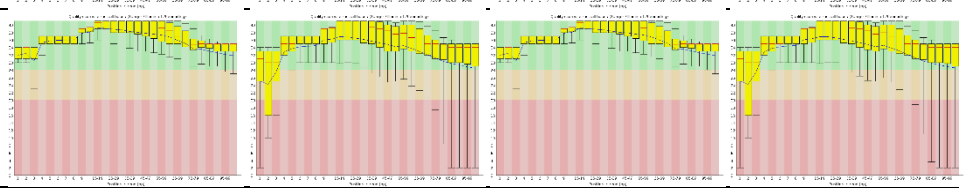
	Ctrl		BCG	
Rep	**Left**	**Right**	**Left**	**Right**
1				
2				
3				
4				
5				
6				

Figure 3.2 Sample quality box-and-whisker graphs via FastQC illustrating quality scores across the read length from left to right

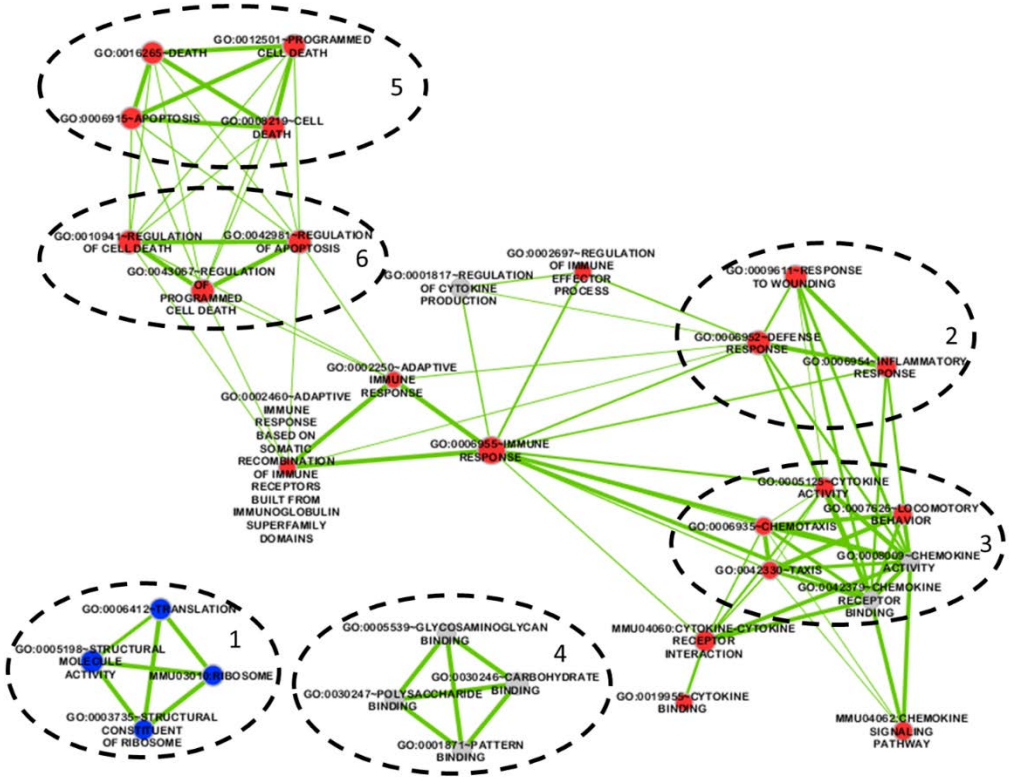


Figure 3.3 Network constructed from clustered terms and any connected (overlap coefficient > 0.5) enriched terms, with clusters from Table 3.5 circled and numbered

3.8 Tables

Table 3.1 Read count of the paired-end (left/right) sample data

Group	Sample	Reads[1]	Direction	Mapped (%)	Aligned Pairs (%)
Ctrl	1	48137202	Left	78	77
		48137202	Right	81	
	2	54465049	Left	94	92
		54465049	Right	93	
	3	50384929	Left	95	94
		50384929	Right	95	
	4	58747069	Left	95	94
		58747069	Right	95	
	5	68739003	Left	95	93
		68739003	Right	94	
	6	45314712	Left	92	88
		45314712	Right	89	
BCG	1	64092232	Left	75	74
		64092232	Right	77	
	2	72364837	Left	94	92
		72364837	Right	94	
	3	67524327	Left	95	94
		67524327	Right	95	
	4	61335634	Left	95	94
		61335634	Right	95	
	5	54744239	Left	94	93
		54744239	Right	94	
	6	62302008	Left	92	89
		62302008	Right	90	

[1] Number of reads for the sample, for each of the paired runs (direction)

Table 3.2 Transcript and gene counts within Cuffdiff

	Total Tested	MA > 10[1]	Significant (FDR p-value < 0.05)	Named Genes[2]
Transcript	62490	29844	2258	2079
Gene	23274	12009	1885	1884

[1] MA: minimum alignment; a locus (ie transcripts) needs at least this many fragments aligned before significance testing will be performed

[2] Named genes were determined using the UCSC database (http://genome.ucsc.edu)

Table 3.3 Twenty five genes showing the greatest differential over-expression in BCG relative to control (FDR p-value < 0.01)

Gene Name[1]	Gene ID	Log_2 (BCG/Control)	Ref*
S100a9	20202	10.08	Hickman
Ly6i	57248	9.50	Martinez
Asprv1	67855	9.44	Sohn, Stables
Il1f9	215257	8.19	Ramadas, Stables
Spon1	233744	8.02	Magee
Nos2	18126	8.01	Ehrt, MacMicking
S100a8	20201	7.69	Ehrt, Hickman
Ccl8	20307	7.58	Liu 2013, Hickman
Cxcr2	12765	7.50	Hickman
Fcgr1	14129	6.03	Hickman
Xcl1	16963	5.49	Stables, Kim 2004
AW112010	107350	4.98	de Oliveira
Lst1	16988	4.80	Mulcahy
Oas3	246727	4.77	Kota
Rsad2	58185	4.72	Magee
Smpdl3b	100340	4.52	Osorio y Fortea
Gbp2	14469	4.38	Ehrt, Beutner
AA467197	433470	4.26	Gundra
Ccl5	20304	3.99	Stables, Liu 2013
Cxcl10	15945	3.81	Hickman
Chi3l3	12655	3.81	Hickman
Acsl1	14081	3.59	Magee
Isg15	100038882	3.38	Magee
Ifi27l2a	76933	2.65	Uhrin
Arg1	11846	3.35	Hickman, Stables

[1] AA467197: expressed sequence AA467197; Acsl1: Acyl-CoA synthetase long-chain family member 1; Arg1: Arginase 1; Asprv1: Aspartic peptidase, retroviral-like 1; AW112010: Expressed sequence AW112010; Ccl5: Chemokine (C-C motif) ligand 5; Ccl8: Chemokine (C-C motif) ligand 8; Chi3l3: Chitinase-like 3; Cxcr2: Chemokine (C-X-C motif) receptor 2; Cxcl10: Chemokine (C-X-C motif) ligand 10; Fcgr1: Fc receptor, IgG, high affinity I; Gbp2: Guanylate binding protein 2; Ifi27l2a: Interferon, alpha-inducible protein 27 like 2A; Il1f9: Interleukin 1 family, member 9; Isg15: ISG15 ubiquitin-like modifier; Lst1: Leukocyte specific transcript 1; Ly6i: Lymphocyte antigen 6 complex, locus I; Nos2: Nitric oxide synthase 2, inducible; Oas3: 2'-5' oligoadenylate synthetase 3; Rsad2: Radical S-adenosyl methionine domain containing 2; Smpdl3b: Sphingomyelin phosphodiesterase, acid-like 3B; Spon1: Spondin 1; S100a8: S100 calcium binding protein 8; S100a9: S100 calcium binding protein 9; Xcl1: Chemokine (C motif) ligand 1.

*Literature associating the listed gene with a macrophage population

Table 3.4 Twenty five genes showing the greatest differential under-expression in BCG relative to control (FDR p-value < 0.01)

Gene Name[1]	Gene ID	Log_2 (BCG/Control)	Ref*
Retnla	57262	-4.90	Pesce, Hickman
Cxcl13	55985	-4.67	Magee, Stables, Hickman
Cd209a	170786	-4.42	Stables
Pf4	56744	-4.07	Hickman, Magee
Fcrls	80891	-3.90	Hickman
Adm	11535	-3.51	Hickman, Magee
Lyve1	114332	-3.41	Schmieder
Vsig4	278180	-3.19	Magee, Vogt
Bank1	242248	-3.07	Gladue
Il10	16153	-3.01	Hickman, Magee, Moore, Stables
Cd83	12522	-2.87	Magee
Faim3	69169	-2.79	Sigruener
Blk	12143	-2.79	Magee
Pou2af1	18985	-2.78	Magee
Mmd	67468	-2.71	Liu 2012, Magee
Cd79b	15985	-2.66	Hickman
Bcar3	29815	-2.62	Osorio y Fortea
Cd2	12481	-2.55	Magee
Fabp4	11770	-2.43	Hickman, Magee
Gimap6	231931	-2.36	Magee
F13a1	74145	-2.35	Hickman
Ptprcap	19265	-2.34	-
Phgdh	236539	-1.94	Gundra
Mt1	17748	-1.70	Espejo et al., 2005
Wfdc17	100034251	-1.22	Karlstetter

[1] Adm: Adrenomedullin; Bank1: B cell scaffold protein with ankyrin repeats 1; Bcar3: Breast cancer anti-estrogen resistance 3; Blk: B lymphoid kinase; Cd2: Cd2 antigen; Cd209a: Cd209a antigen; Cd79b: Cd79b antigen; Cd83: CD83 antigen; Cxcl13: Chemokine (C-X-C motif) ligand 13; Fabp4: Fatty acid binding protein 4, adipocyte; Faim3: Fas apoptotic inhibitory molecule 3; Fcrls: Fc receptor-like S, scavenger receptor; F13a1: Coagulation factor XIII, A1 subunit; Gimap6: GTPase, IMAP family member 6; Il10: Interleukin 10; Lyve1: Lymphatic vessel endothelial hyaluronan receptor 1; Mmd: Monocyte to macrophage differentiation-associated; Mt1: Metallothionein 1; Pf4: Platelet factor 4; Phgdh: 3-phosphoglycerate dehydrogenase; Pou2af1: POU domain, class 2, associating factor 1; Ptprcap: Protein tyrosine phosphatase, receptor type, C polypeptide-associated protein; Retnla: Resistin like alpha; Wfdc17: WAP four-disulfide core domain 17; Vsig4: V-set and immunoglobulin domain containing 4.

*Literature associating the listed gene with a macrophage population

Table 3.5 List of member terms for each functional cluster (Enrichment Score > 4)

Cluster Identifier	ES (Genes)[1]	Identifier (Genes)[2]	Term Name
1	17.89 (98)	GO:0006412 (69)	Translation
		GO:0003735 (59)	Structural Constituent of Ribosome
		GO:0005198 (85)	Structural Molecule Activity
		Mmu03010 (60)	Ribosome
2	11.24 (118)	GO:0009611 (80)	Response to Wounding
		GO:0006952 (92)	Defense Response
		GO:0006954 (54)	Inflammatory Response
3	7.15 (66)	GO:0042330 (35)	Taxis
		GO:0006935 (35)	Chemotaxis
		GO:0007626 (46)	Locomotory Behavior
		GO:0008009 (17)	Chemokine Activity
		GO:0042379 (17)	Chemokine Receptor Binding
		GO:0005125 (38)	Cytokine Activity
4	5.89 (67)	GO:0030246 (67)	Carbohydrate Binding
		GO:0030247 (30)	Polysaccharide Binding
		GO:0001871 (30)	Pattern Binding
		GO:0005539 (27)	Glycosaminoglycan Binding
5	5.83 (83)	GO:0006915 (79)	Apoptosis
		GO:0012501 (79)	Programmed Cell Death
		GO:0008219 (83)	Cell Death
		GO:0016265 (83)	Death
6	5.45 (88)	GO:0042981 (88)	Regulation of Apoptosis
		GO:0043067 (88)	Regulation of Programmed Cell Death
		GO:0010941 (88)	Regulation of Cell Death

[1] ES: Enrichment Score; Listed below the ES is the number of genes enriching the cluster

[2] Member terms of the cluster, with the number of genes enriching that term in parentheses

CHAPTER IV: Differential gene expression and networks in murine microglia and peritoneal macrophages challenged with Bacillus Calmette-Guérin

4.1 Abstract

Transcriptomic studies have compared the resting profiles of several cell types, including microglia and peritoneal macrophages. However the exploration of transcriptome responses to immune-challenge are limited, despite being the primary role of immune cells. Bacille Calmette Guérin (BCG) is an inducer in mice for modeling chronic inflammation, resulting in acute sickness behaviors and chronic depressive-like behaviors. Brain microglia mediate behavioral changes in response to BCG, however the relationship between microglia and peripheral macrophage responses is incompletely understood. The objective of this study was to characterize the transcriptome profile of microglia and peritoneal macrophages one week after a peripheral BCG challenge, coinciding with depressive-like behavior changes, relative to saline-challenged controls. Profiles were then compared between microglia and peritoneal macrophages for unique and shared patterns of gene set enrichment which were visualized in networks. Using Cuffdiff, adequate read coverage (minimum read alignment > 10) was identified for 14,046 genes in microglia and 12,352 genes in peritoneal macrophages. The genes with adequate read coverage were tested for differential expression (False Discovery Rate-adjusted p-value < 0.05, $\log_2$ fold-change > 1.5) between BCG and saline control treatments, with 11,738 of the genes tested in both cell types. In microglia, 147 genes showed significant differential expression, all over-expressed in the BCG versus control group. Peritoneal macrophages had 655 over-expresseed and 355 under-expressed genes. The top 5 significantly over-expressed genes, BCG versus control, in microglia were Serum amyloid A3 (Saa3; fold-change 5.90), Cell adhesion molecule 3 (Cadm3; fold-change

5.14), Carbonic anhydrase 6 (Car6; fold-change 5.01), STEAP family member 4 (Steap4; fold-change: 4.89), and Selectin, endothelial cell (Sele; fold-change: 4.80). While no genes met the fold-change cutoff for under-expression in microglia, the most under-expressed genes were Coiled-coil domain containing 162 (Ccdc162; fold-change: -1.21), Titin-cap (Tcap; fold-change: -1.15), Hemoglobin alpha, adult chain 1 (Hba-a1; fold-change: -1.10) and 2 (Hba-a2; fold-change: -1.09), and Resistin like alpha (Retnla; fold-change: -1.00). The most over-expressed genes for peritoneal macrophages were S100 calcium binding protein A9 (S100A9; fold-change: 10.11), MAS-related GPR, member A2a (Mrgpra2a; fold-change: 1.10), Stefin A2 like 1 (Stfa2l1; fold-change: 10.03), Lymphocyte antigen 6 complex, locus I (Ly6i; fold-change: 9.42), and Aspartic peptidase, retroviral-like 1 (Asprv1; fold-change: 9.15). Top under-expressed genes in peritoneal macrophages were Protease, serine 37 (Prss37; fold-chagne: -3.49), Mucin-like 1 (Mucl1; fold-change: -3.09), Reprimo, TP53 dependent G2 arrest mediator candidate (rprm; fold-chagne: -3.07), SRY (sex determining region Y)-box 7 (Sox7; fold-change: -3.01), with Retnla being the fifth most under-expressed gene for peritoneal macrophages (fold-change: -2.89) as well as microglia. All genes with adequate coverage were then entered into GSEA, testing for enriched gene sets in the Molecular Signature Database. Categories were tested separately, resulting in 3 analyses: GO Biological Processes and Molecular Functions, KEGG pathways, and REACTOME pathways. Across the tested categories, 60 gene sets (including 2,407 genes) were enriched (FDR-adjusted p-value < 0.05) in both microglia and peritoneal macrophages. Of the gene sets enriched in both cell types, 27 gene sets (including 287 genes) were enriched but exhibited opposing differential expression in microglia and peritoneal macrophages. Of the 27 gene sets, 15 over-expressed the BCG treatment group compared to the control group in microglia and under-expressed in macrophages. These 15 gene sets (including 123 genes) had a singular gene set (REACTOME

GPCR Downstream Signaling, REACT_14797) that did not share enriching member genes with the other 14, which were GO and REACTOME gene sets related to ribosomal activity, RNA metabolism, and protein synthesis sharing several genes from the Ribosomal protein S (*Rps)* and Ribosomal protein L (*Rpl*) families. The remaining 12 gene sets (including 164 genes) were over-expressed in the BCG treatment versus control in peritoneal macrophages, included respiratory electron transport chain gene sets (REACTOME) sharing several enriching gene members with KEGG gene sets related to neurodegenerative disorders, and REACTOME gene sets related to the regulation of cell-cycle, Ornithine Decarboxylase, APOBEC3G, and cross-presentation. This indicates a simultaneous over-expression of genes in BCG treatment versus control associated with apoptosis, energy metabolism, and cell-cycle regulation, primary macrophage methods of controlling intracellular pathogens that are also under-expressed in microglia to maintain a neuroprotective state.

4.2 Introduction

Macrophages are a heterogeneous group of resident immune cell in tissues (Gordon and Plűddemann, 2013). Among these groups, macrophage-like brain microglia and peritoneal macrophages share the common role as first-responders and communicators during immunological challenge (Davies et al., 2013). However microglia are considered to be primarily tasked with maintenance of normal functions in the central nervous system, greatly impacting their ability to respond to inflammatory events compared to macrophages (Prinz and Priller, 2014; London et al., 2013). This perhaps helps explain the identification of a core gene-expression profile signature shared among macrophage populations and microglia, as well as unique expression characteristics to each cell type (Gautier et al., 2012; Gautier et al., 2014). Shared and differentiating

transcriptome profiles between macrophage populations enable the understanding of common and unique activities (Hickmann et al., 2013).

Bacille Calmette Guérin (BCG) is an attenuated *Mycobacterium* strain and an effective inducer for chronic modeling of immune activation and behavioral impacts in mice (Moreau et al., 2005; Moreau et al., 2008; Rodriguez-Zas et al., 2015). Studies of the profile of individual genes in response to BCG have demonstrated that the proinflammatory response of macrophages is dependent upon Toll-like receptor (TLR)-2, with TLR2 and TLR4 both regulating the adaptive immune response (Heldwein et al., 2003) while BCG itself can inhibit parts of the immune response (Master et al., 2008; Danelishvili et al., 2010). While the murine depressive-like behavioral response to BCG is dependent upon interferon (IFN)-γ signaling, which in combination with tumor necrosis factor (TNF)-α induces indoleamine 2,3-dioxygenase (IDO) expression in microglia, the proinflammatory cytokine and immune response are independent of IFN-γ (O'Connor et al., 2009). However, the profile of the murine transcriptome in response to BCG relative to resting state within and across macrophage cell groups has yet to be completely characterized.

Functional Gene Set Enrichment Analysis (GSEA), which considers the profiles of all transcripts simultaneously regardless of significance level, enabled the identification of gene sets enriched within and across profiles (Subramanian et al., 2005). As the gene sets indicate relationships among genes, genes shared between gene sets can also be treated as relationships to construct gene set networks (Saito et al., 2012). Visualization of relationships between enriched terms based on

treatment groups within and between cell types is expected to provide a more complete profile of immune response bacterial challenge (Merico et al., 2010).

The objective of this study was to characterize the gene expression response of microglia and peritoneal macrophagesto BCG challenge relative to a saline control, coinciding with behavioral changes in the murine host (Rodriguez-Zas et al., 2015). Genes that differ in expression between BCG and control groups were identified and compared between microglial and macrophage cells. Functional analysis identified enriched categories within cell types and networks of categories were identified. Comparison of networks and gene expression profiles enables the identification of unique and shared patterns between microglia and peritoneal macrophages.

4.3 Materials and Methods

All animal experiments were done in accordance with the NIH Policies for Animal Care and Use of Laboratory Animals, and approved by the Institutional Animal Care and Use Committee at the University of Illinois. Microglia and peritoneal macrophages were collected from adult (~22 weeks old) male C57Bl/6J mice sourced from Charles River Laboratories. Mice were housed individually in cages under a 12:12 hr light/dark cycle with controlled temperature and humidity at 23° C and 45%, respectively. Water and food (Teklad 8640 chow, Harlan Laboratories, Indianapolis, IN, USA) were available *ad libitum*.

Mice received a dose of TICE BCG (10 mg; Organon USA Inc., USA) or saline solution as a peritoneal injection and each treatment included 12 mice. Prior to treatment, a BCG vial (50 mg wet weight of lyophilized culture, $1x10^8$ colony forming units per vial) was reconstituted in sterile

saline, with both BCG and saline (Ctrl) treatments a 0.3 ml/mouse injection volume. Brains and peritoneal cells were collected from mice one week after injection for the isolation of microglia and peritoneal macrophages, respectively. The isolations of peritoneal macrophages (Zhang et al., 2008) and microglia (Nikodemova and Watters, 2012) were modified from previously published methods.

Peritoneal cells were harvested by peritoneal lavage with 10 mL of cold HBSS (without Ca^{2+} or Mg^{2+}) before spinning them down at 400g at room temperature for 10 minutes. The peritoneal cells were then resuspended in DMEM supplemented with 10 mM glutamine, 100 µg/mL of penicillin and streptomycin, plated and incubated at 37° C under 5% CO_2 for 2 hours. Pre-plating examination with flow cytometry verified the population distributions previously indicated in the literature for peritoneal cells (Zhang et al., 2008). To isolate the peritoneal macrophages, the plates were washed gently with warm PBS, leaving macrophages as the surviving cells attached to the plate for further extraction. The macrophages were stored in Trizol at -80° C until ready for RNA extraction.

Microglial collection immediately followed peritoneal lavage. Mice were perfused with ice-cold PBS + 2 mM EDTA, and the brain dissected. Brains were digested enzymatically using the Miltenyi Neural Tissue Dissociation Kit (Miltenyi Biotec, Germany). Cell debris was removed by passage of the resulting solution through a 40 µm filter, followed by myelin removal via resuspension and centrifugation in 30% Percoll (GE Healthcare, Princeton, NJ). Cells were then magnetically labeled with anti-CD11b Miltenyi MicroBeads and separated in a magnetic field with MS columns (Miltenyi Biotec, Germany). The resulting Cd-$11b^{+}$ fraction was spun down and

resuspended in Trizol for storage at -80° C until ready for RNA extraction. Separate tissue samples were used to verify the enrichment of Cd-11b^{+} cells after column separation by flow cytometry (Fenn et al., 2014). No study has shown monocytes to migrate past the blood-brain barrier at this time-point following a peripheral BCG challenge, minimizing concern that migrating monocytes are responsible for the microglial responses listed here (Sirén et al., 2001).

Flow Cytometry

Cells were stained with primary fluorescent antibodies for two primary markers for macrophages and microglia: CD-11b and CD-45 (Fenn et al., 2014; Henry et al., 2009). Briefly, Fc receptors were blocked by incubation with anti-CD16/CD32 antibody before incubation with anti-CD-11b and anti-CD-45 antibodies (eBioscience, CA). Surface receptor expression was identified using a Biosciences LSR II Flow Cytometry Analyzer with BD FACSDiva software. Antibody gating was determined using isotype-stained controls.

RNA Extraction

Following tissue storage in Trizol, RNA was extracted based upon the Tri-spin method (Reno et al., 1997) with modifications. Briefly, the Trizol-immersed samples were thawed to room temperature and and homogenized before being mixed with chloroform (0.2 ml chloroform : 1 ml Trizol) and incubated for 3 min. A phase separation was achieved by centrifuging for 15 min at 15000 g at room temperature, then transferring the aqueous layer to mix with acid phenol (0.6 ml acid phenol : 1 ml Trizol) to repeat the phase separation. The aqueous layer was then mixed with ethanol (1 : 1 ratio) before spinning through an extraction column from the E.Z.N.A. Total RNA Kit (Omega Biotek, Norcross, GA). The optional DNase step was included to eliminate potential

DNA contamination. Purity and quantity of RNA was measured on a Nanodrop (Nanodrop Products, Wilmington, DE, USA). RNA integrity was determined with the Agilent 2100 Bioanalyzer with RNA Pico chip (Agilent Technologies, Palo Alto, CA) for RNA Integrity Numbers (RIN). All RIN values were > 7, with 90% of all samples > 9.

Library Sequencing and Abundance Quantification

Individual samples from each mouse were used to obtain libraries that were sequenced using an Illumina HiSeq 2000 (Illumina, San Diego, CA, USA), producing paired reads 100 base-pairs in length. Quality determination of reads and processing for testing of differential abundance between treatments was performed as previously published (Nixon et al., 2015). Briefly, reads were quality-checked using FastQC (Babraham Institute, 2013). Reads were then mapped to the mouse genome and annotations as available through the Illumina iGenomes package (mm10; http://support.illumina.com/sequencing/sequencing_software/igenome.html). Options designed to address biological and technical biases were enabled throughout the workflow (Nixon et al., 2015). Mapping utilized the gene annotations as a guide (argument –G) in Tophat2 (v 2.0.8) before assembly of reads into transcripts in Cufflinks (v2.1.1). The calculation of the transcript median in Cufflinks utilized Upper Quartile Normalization (argument –N) to avoid skewed results based on transcripts not expressed in the dataset (Dillies et al., 2013). Biases from reads mapping to multiple sites and unequal site availability for read creation were corrected with multi-read correction (argument –u) and fragment bias correction (argument –b), respectively (Roberts et al., 2011; Mortazavi et al., 2008). Individual sample assemblies were then merged with Cuffmerge before differential expression between both treatment groups within cell type was tested in Cuffdiff2 by modeling read distribution and uncertainty with a Beta Negative Binomial distribution (Trapnell

et al., 2013). Multi-read and fragment bias corrections were enabled in Cuffdiff as well. Differential expression was evaluated on all genes with minimum alignment count > 10. The Cuffdiff output test-statistic was calculated, accounting for expression variance while relating greater differential expression between groups in terms of a larger absolute value. Genes exhibiting False Discovery Rate or FDR-adjusted p-value < 0.05 between the BCG and control groups within cell type and a $\log_2$ fold-change greater than 1.5 were considered differentially expressed.

Functional Analysis

Within cell type, functional analysis considered all genes using GSEA-PreRanked analysis (Subramanian et al., 2005; http://www.broadinstitute.org/gsea/index.jsp). To identify enriched functional categories within the gene expression profiles of the tested treatments (BCG or control), the genes were entered with their Cuffdiff test statistic. The list of genes was sorted by their Cuffdiff test statistic, resulting in an ordered list from those genes most over-expressed to those most under-expressed in the BCG treatment compared to Control. The gene list was tested against gene sets, determining enrichment as member genes of the gene set being overrepresented near the top or bottom of the gene list. Each gene set is given a normalized enrichment score (NES), normalizing the enrichment score by permutation to account for differences in gene set sizes and correlation between gene sets and the gene list data (Subramanian et al., 2005; http://www.broadinstitute.org/gsea/doc/GSEAUserGuideTEXT.htm). The ranked list of genes were tested against the Molecular Signature Database (MSigDB; Liberzon et al., 2011) that includes the following terms: Gene Ontology (GO) Biological Processes and Molecular Functions (1,221 tested gene sets; Ashburner et al., 2000), Kyoto Encyclopedia of Genes and Genomes pathways (KEGG; 186 tested gene sets; Kanehisa et al., 2002), and pathways from the

REACTOME knowledgebase (674 tested gene sets; Matthews et al., 2009). Despite KEGG and REACTOME overlapping in their focus, their designs and methods of annotations produce limited consensus with benefits to using both databases (Stobbe et al., 2011).

Visualization

Typically, relationships among terms enriched by a dataset are defined by overlapping members from the gene list used in enrichment testing (Huang et al., 2009; Merico et al., 2011), or the cumulative functional relationships of member genes (Wang et al., 2011). Clustering methods identify relationships among all member terms above the established cutoff, assembling independent groups of terms. Network designs offer information on relationships between terms within and across such groupings to further relate groups of terms and improve identification of relationships (Newman, 2012). Using the GSEA results as input, networks were created with the Enrichment Map plugin (Merico et al., 2010) for Cytoscape (http://www.cytoscape.org; Killcoyne et al., 2009) as a visualization plugin specifically designed to work with funcational analysis data and GSEA results. In the networks, enriched gene sets were represented by nodes with relationships between gene sets defined by sufficient overlap in member genes from the Cuffdiff output gene list (similarity coefficient > 0.5) as edges. Networks visualized within-cell type (microglia or peritoneal macrophages) comparisons between BCG and control groups.

4.4 Results

Adequate read coverage as per Cuffdiff default settings of >10 reads was tested for 26,773 genes in microglia and 25,767 genes in peritoneal macrophages. Differential expression was evaluated for 14,046 and 12,352 genes with adequate read coverage in the respective cell types. Among the

genes tested for differential expression, 11,738 genes were tested in both cell types; 2,307 were tested only in microglia; and 613 were tested only in peritoneal macrophages. Over-expressed genes in BCG versus control dominated the list of significantly differentially expressed genes in peritoneal macrophages (FDR-adjusted p-value < 0.05; $\log_2$ fold-change > 1.5), with a greater number of genes and greater observed fold changes in expression level relative to under-expressed genes. In microglia when comparing the BCG treatment group versus control, 147 genes showed significant differential expression, and all were over-expressed (FDR-adjusted p-value < 0.05; $\log_2$ fold-change > 1.5). In peritoneal macrophages 655 over-expressed and 355 under-expressed genes (FDR-adjusted p-value < 0.05; $\log_2$ fold-change > 1.5) were identified. The top 20 genes are listed for over-expressed (BCG versus control) in microglia are listed in **Table 4.1**, with fold-changes ranging from 5.90 – 2.99. As no under-expressed genes in microglia met both significance requirements, the top 20 genes meeting only the p-value requirement (FDR-adjusted p-value < 0.05) were considered (**Table 4.2**) with fold-changes ranging from 1.21 – 0.69. Over-expressed genes (**Table 4.3**, fold-changes from 10.11 – 6.27) and under-expressed genes (**Table 4.4**, fold-changes from 3.49 – 2.49) in peritoneal macrophages were listed based upon meeting both requirements (FDR-adjusted p-value < 0.05; $\log_2$ fold-change > 1.5).

Of the genes differentially over-expressed in microglia from BCG versus control mice, the largest difference in expression was observed in Serum amyloid A3 (Saa3; fold-change: 5.90; FDR-adjusted p-value < 0.01), followed by Cell adhesion molecule 3 (Cadm3; fold-change: 5.14), Carbonic anhydrase 6 (Car6; fold-change: 5.01), STEAP family member 4 (Steap4; fold-change: 4.89), and Selectin, endothelial cell (Sele; fold-change: 4.80; **Table 4.1**). When considering genes under-expressed in BCG versus control groups in microglia, the most differentially under-

expressed gene was Coiled-coil domain containing 162 (Ccdc162; fold-change: -1.21) followed by Titin-cap (Tcap; fold-change: -1.15), Hemoglobin alpha, adult chain 1 (Hba-a1; fold-change: -1.10), Hemoglobin alpha, adult chain 2 (Hba-a2; fold-change: -1.09), and Resistin like alpha (Retnla; fold-change: -1.00; **Table 4.2**). The S100 calcium binding protein A9 (S100A9) gene was the most over-expressed gene in peritoneal macrophages (fold-change: 10.11; FDR-adjusted p-value < 0.01), followed by MAS-related GPR, member A2A (Mrgpra2a; fold-change: 1.10), Stefin A2 like 1 (Stfa2l1; fold-change: 10.03), Lymphocyte antigen 6 complex, locus I (Ly6i; fold-change: 9.42), and Aspartic peptidase, retroviral-like 1 (Asprv1; fold-change: 9.15; **Table 4.3**). The S100A9 heterodimer partner, S100A8, was the 14th most over-expressed genes in peritoneal macrophages (fold-change: 7.15). S100A9 and S100A8 were not significant in microglia when considering fold-change requirements (Fold-change: 1.49). The IFN-γ gene, Ifng (6.45 fold-change; FDR-adjusted p-value < 0.01; **Table 4.3**), was significantly over-expressed in peritoneal macrophage, although no differential expression was detected in microglia (0.92 fold-change; FDR-adjusted p-value = 0.06). IL-1β was significantly over-expressed in both microglia (fold-change: 1.60) and peritoneal macrophages (fold-change: 1.97). TNFα did not meet the fold-change requirements for significant differential expression in peritoneal macrophages (1.29 fold-change; FDR-adjusted p-value < 0.01) or microglia (0.14 fold-change; FDR-adjusted p-value > 0.1). The most under-expressed gene in the BCG group versus control comparison in peritoneal macrophages was Protease, serine 37 (Prss37; fold-change: -3.49; FDR-adjusted p-value < 0.05), followed by Mucin-like 1 (Mucl1; fold-change: -3.09), Reprimo, TP53 dependent G2 arrest mediator candidate (Rprm; fold-change: -3.07), SRY (sex determining region Y)-box 7 (Sox7; fold-change: -3.01), with the fifth most under-expressed gene being Retnla as in microglia (fold-change: -2.89; **Table 4.4**). Krüppel-like factor 15 (klf15) was also under-expressed in peritoneal

macrophages when comparing the BCG treatment group versus control (-2.68 fold-change; FDR-adjusted p-value < 0.01; **Table 4.4**). The klf15 gene was previously found to be expressed in macrophages, though no study has involved challenge states to determine the expression response in an inflammation model (Nagare et al., 2011). Of note, the classic M2 indicator IL-10 (Martinez et al., 2008) was also found to be under-expressed in peritoneal macrophages (fold-change: -1.56; FDR-adjusted p-value < 0.05).

Within the GO BP and MF analysis, 113 gene sets were enriched in microglia or peritoneal macrophages (FDR-adjusted p-value < 0.05), with 104 of these gene sets connected to at least one other gene set (overlap coefficient > 0.5), and 21 total (regardless of connectivity) gene sets enriched in both cell types. For instance, the GO BP gene sets Apoptosis (GO:0006915) and Programmed Cell Death (GO:0012501) were enriched in both microglia and peritoneal macrophages, with enriching member genes over-expressed in the BCG treatment group versus control (**Table 4.5**). These gene sets were linked (overlap coefficient > 0.5) to gene sets that were enriched in microglia or peritoneal macrophages. For example, the GO BP gene set enriched in peritoneal macrophages with member genes over-expressed in the BCG treatment versus control, Apoptotic Program (GO:0008632) was linked to the shared Apoptosis gene set and involves the signaling cascade for cells triggered to undergo apoptosis. Meanwhile in microglia with member genes over-expressed in BCG treatment versus control, a link connected the GO BP gene sets Apoptosis (GO:0006915) and Anti-Apoptosis (GO:0006916), which were both further connected to Negative Regulation of Apoptosis (GO:0043066) and negative Regulation of Programmed Cell death (GO:0043069) (**Figure 4.1**). The KEGG pathway Toll-like Receptor Signaling Pathway (hsa04620) was enriched in both microglia and peritoneal macrophages, with enriching member

genes over-expressed in the BCG treatment versus control. Only enriched in microglia with member genes over-expressed in the BCG group versus control, the NOD-like receptor signaling pathway (hsa04621, FDR-adjusted p-value < 0.01) involves the inflammasome response. The associated REACTOME pathways of Activated TLR4 Signaling (REACT_6890) and Toll Receptor Cascades (REACT_6966) were also found to be enriched with member genes over-expressed in BCG treatment versus control, although only in peritoneal macrophages (**Figure 4.3**). GSEA identified 13 gene sets with negative enrichment scores in microglia, meaning their enrichment was based upon genes under-expressed in BCG versus control (**Table 4.4**; **Table 4.5**). The largest cluster of networked GO gene sets involved 14 gene sets enriched in both cell types connected to a unique microglia response of 46 gene sets and a comparatively smaller unique peritoneal macrophage response of 4 gene sets (**Figure 4.1**). Three of the gene sets enriched in both cell types exhibited opposing direction of enrichment: Structural Constituent of Ribosome (GO MF, GO:0003735, 76 enriched member genes), Structural Molecule Activity (GO MF, GO:0005198, 133 enriched member genes), and Translation (GO BP, GO:0006412, 130 enriched member genes). The majority of enriching member genes in microglia were not significant when analyzed at the gene level, with less than 5% of genes in any category found significant (FDR-adjusted p-value < 0.05). The enriching member genes in peritoneal macrophages were better represented, with 50-75% of genes identified as significant. These 3 gene sets were directly and highly connected to each other (Overlap Coefficient > 0.98), with all enriching member genes over-expressing genes in microglia, while all enriching member genes were under-expressing genes in peritoneal macrophages. Across all enriched gene sets in the GO BP and MF analysis, 65 gene sets were connected (Overlap Coefficient > 0.5) together in a cluster, with 8 additional clusters ranging in size from 13 to 2 gene sets (**Figure 4.1**). A portion of the large cluster included

a well-connected group of gene sets related to defense responses (Cellular Defense Response; Defense Response; Response to Wounding; Inflammatory Response; Response to External Stimulus) which included a strong connection (Overlap Coefficient = 1) to Locomotory Behavior and Behavior (**Figure 4.1**). The analysis of KEGG gene sets across both microglia and peritoneal macrophages produced 15 enriched gene sets (**Table 4.6**). Of the 8 gene sets that were enriched in both cell types, 4 enriched in both cell populations by member genes over-expressed in the BCG treatment versus control, involving Extracellular matrix receptor interactions (hsa04512), the TLR signaling pathway (hsa04620), as well as T-cell receptor signaling pathways (hsa04660) and natural killer cell mediated cytotoxicity (hsa04650) (**Table 4.6**). The remaining 4 gene sets enriched in both microglia and peritoneal macrophages had member genes exhibiting opposing direction of enrichment between cell types: in BCG treatment versus control, the genes were under-expressed in microglia while over-expressed genes in peritoneal macrophages (**Figure 4.2**). These gene sets were related to neurodegenerative diseases, and oxidative phosphorylation. The REACTOME analysis produced 111 enriched gene sets with 31 enriched in both cell types (**Table 4.7**). Additionally, REACTOME pathway gene sets did not all display the same direction of change between cell types (12 over-expressed in BCG treatment versus control in microglia versus under-expressed in BCG versus control in peritoneal macrophages; 8 gene sets under-expressed in microglia versus over-expressed in peritoneal macrophages). A single large cluster was produced containing 89 gene sets, with 5 more clusters ranging between 8 to 2 connected gene sets (Overlap Coefficient > 0.5; **Figure 4.3**). Regarding the top 5 most differentially over- and under-expressed genes in microglia and peritoneal macrophages, it is interesting to note that only S100a9 was an enriching member of any of the enriched gene sets across GO, KEGG, and REACTOME. S100a9 was an enriching member of the GO BP terms Defense Response (GO:0006952), Inflammatory

Response (GO:0006954), Response to External Stimulus (GO:0009605), Response to Wounding (GO:0009611), and the GO MF term Cation Binding (GO:0043169). If considering the top 20 genes for over- and under-expression in each cell type (**Tables 4.1-4.4**), 13 genes were members of 38 enriched gene sets.

Gene sets demonstrating opposing direction of enrichment between cell-types were found in all tested collections. However, only the REACTOME collection possessed gene sets showing both (BCG treatment versus control) over-expression in microglia and under-expression in peritoneal macrophages, as well as the opposite (**Table 4.8**). Across the analyses of GO, Kegg, and REACTOME, the 27 gene sets split into 15 over-expressed in BCG treatment versus control in microglia and under-expressed in peritoneal macrophages with the remaining 12 reversed in the expression pattern of their enriching member genes. Across the 15 gene sets over-expressed in the BCG treatment in microglia while the BCG treatment was under-expressed in peritoneal macrophages, 123 enriched member genes were present in both cell types with 88 of those shared among more than one gene set. Overlap Coefficients were calculated between all 15 gene sets (**Table 4.9**). Almost all of the gene sets were connected (Overlap Coefficient > 0.5), but for "GPCR Downstream Signaling" (REACT_14797) which shared no enriching member genes with any of the other gene sets. There was also a low Overlap Coefficient (< 0.5) between Translation (GO BP, GO:0006412) and the two REACTOME pathways REACT_1258 or REACT_1079 (**Table 4.9**). Ribosomal protein L (Rpl) and Ribosomal protein S (Rps) were the primarily shared gene families among these gene sets. The Rps family genes were members of all the gene sets (**Table 4.9**) but the REACTOME pathway "GPCR Downstream Signaling" (REACT_14797). The Rpl family genes possessed similar membership as Rps genes, additionally not members of

REACT_1258, and REACT_1079. The 12 gene sets under-expressed in BCG treatment versus control in microglia and over-expressed in peritoneal macrophages represented a total of 164 enriching member genes across both cell types, with 102 genes shared by more than one gene set. The gene sets were sharply divided based upon enriching member genes, creating two completely separate groups of gene sets (Overlap Coefficient =0) with each group containing connected (Overlap Coefficient > 0.5) gene sets (**Table 4.10**).

4.5 Discussion

It is interesting to note that none of the top 5 genes over-expressed in microglia when comparing the BCG treatment group versus control have been previously associated with the microglial BCG model (**Table 4.1**). The most over-expressed gene, Saa3, is an acute-phase reactant that is over-expressed in the brain during peripheral inflammatory events (Thomson et al., 2014). Previous studies have demonstrated that Saa3 is expressed in microglia (Hickman et al., 2013), can be induced 12-fold in the inflammatory profile of microglia (Baker and Manuelidis, 2003), and that its expression can be induced by Mycobacterium in macrophage populations (Keller et al., 2004). The expression of S100A8 and S100A9 have been found to induce Saa3 expression in an inflammation-like state in studies of tumor recruitment (Hiratsuka et al., 2008). Reports on Cadm3 indicate that this gene is associated other inflammatory models in the brain (Kyan et al., 2014) and with BCG response in the lung of guinea pigs (Jain et al., 2012). The coded protein for the Car6 gene is expressed in the murine macrophage 264.7 cell line, yet was under-expressed 2 hours after exposure to lipopolysaccharide (LPS) *in vitro* (Das et al., 2013) in comparison to over-expression in this study (**Table 4.1**). Steap4 is expressed in microglia (Chiu et al., 2013) and inducible in macrophages by more than 8-fold with LPS (Fleming, 2014). Sele was over-expressed in BCG

versus control groups in microglia (4.80 fold-change; **Table 4.1**) and is over-expressed during the resolution-phase of murine macrophages compared to M1 activated macrophages (Stables et al., 2011). However as the previous study compared resolution-phase and activated macrophages, there is limited value in comparison of fold-changes. The most under-expressed gene in microglia, Ccdc162, has only recently been characterized in some gene studies with no relationship to the current model (Lin et al., 2013; Zhang et al., 2014; Hudson et al., 2014). Tcap, as the second most under-expressed gene (BCG versus control) in microglia, has previously found in microglia to be over-expressed 13-fold in an Alzheimer's Disease model (Orre et al., 2014). Both Hba-a1 and Hba-a2 can be expressed in activated macrophages, although in brain they are more typically associated with the activation of astrocytes (Orre et al., 2014). However several markers for other brain cell types are expressed in microglia (Solga et al., 2015), which may motivate new exploration of these genes in microglia. The under-expression of Retnla in both cell types, a marker of M2 activation in macrophages (Pesce et al., 2009) and microglia (Tada et al., 2014), matched fold-change reductions in the previous studies and suggests an M1 state of activation. As IFN-γ and IL-1β are classic indicators of M1 activated macrophages (Martinez et al., 2008) induced by BCG exposure, a pro-inflammatory state would be expected based upon their over-expression in peritoneal macrophages in this study (Gordon and Martinez, 2010; O'Connor et al., 2009). Mrgpra2a and Stfa2l1 are both typically associated with neutrophils, although some non-neutrophil populations can also express them (Ericson et al., 2014). However, their relationship to macrophages is otherwise unexplored. In comparison, the relationship of Ly6i over-expression to the current study is well-established, as Ly6 genes are suggested to impact host defense against *Mycobacterium tuberculosis* infection (Martinez et al., 2013). Asprv1 and Interleukin 1 family member 9 (Il1f9) are both over-expressed in the BCG group versus control in peritoneal

macrophages (fold-changes: 9.15, 7.66; **Table 4.3**) and associated with pro-inflammatory events (Stables et al., 2011). These genes were also over-expressed (BCG vs control groups) in microglia, though with smaller fold-changes (fold-changes: 2.94, 2.37; FDR-adjusted p-value < 0.01). Little information is available regarding Prss37, the most under-expressed gene in BCG versus control groups in peritoneal macrophages, and has been suggested to have tumor suppressor activity based on loss-of-function truncated variants in colorectal cancer studies (Gylfe et al., 2013). Mucl1 is shown to over-express in macrophages upon exposure to a combination of IFN-γ and LPS, albeit during *in vitro* studies meant to examine more immediate responses (Zollbrecht et al., 2013). The effect of under-expressed Rprm in macrophages after BCG challenge is less clear, although the Myocyte enhance factor-2 family of transcription factors are linked to host response to bacterial pathogens (McKinsey et al., 2002) and also represses Rprm (Yu et al., 2015). However, a more focused analysis would be necessary to clarify the relationship. Sox7 also has limited association with macrophages, although macrophage-derived cells have been shown to express the gene (Hall et al., 2012) as well as microglia (Solga et al., 2015). Information is lacking on klf15 response in macrophages to inflammatory challenge, but the loss of klf15 in insulin-resistance models suggest an anti-inflammatory role (Jung et al., 2013). The inflammatory cytokine interleukin (IL)-17 is also known to inhibit klf15 (Ahmed and Gaffen, 2013). Although IL-17 is induced by BCG in mice (Gopal et al., 2012), it did not meet significance thresholds within this study 7 days after challenge. Further study within macrophages would be necessary to determine the contextual role of klf15 within the macrophage response. While the Complement component 1 s subcomponent itself has not previously been associated with macrophage expression, other C1 components are known to be inducible in macrophages (Martinez and Gordon, 2014). The gene expression patterns of peritoneal macrophages supported previously reported response to BCG challenge , including

Resistin like alpha (Retnla), the CD209a antigen, Platelet factor 4 (Nixon et al., 2015). Of interest was Mucl1, one of the most under-expressed gene in BCG groups versus control, in peritoneal macrophages (-3.09 fold-change; **Table 4.4**). Previous *in vitro* studies indicated macrophage up-regulation of Mucl1 following treatment with inflammatory stimulators, suggesting a need for further study of this relationship (Zollbrecht et al., 2013).

The most differentially-expressed genes between treatments, as determined by fold-change, appear to have a minimal contribution to findings at the gene set and network level. Of the top 20 genes over- or under-expressed in BCG versus control groups in microglia and peritoneal macrophages respectively, 3 genes (**Table 4.1**), no genes (**Table 4.2**), 6 genes (**Table 4.3**) and 4 genes (**Table 4.4**) were members of enriched gene sets. The focus of GSEA analysis is towards the discovery of gene sets that are enriched through a broad number of member genes with consistent profiles (Subramanian et al., 2005).

Enrichment by over-expressing genes across both microglia and peritoneal macrophages involved immunological response, apoptosis, and intracellular signaling cascades (**Tables 4.5-4.7**). However where the apoptosis-related gene sets enriched in peritoneal macrophages were related to inducing apoptosis, those enriched in microglia were associated with the negative regulation and control of apoptosis. (**Figure 4.1**). The 4 gene sets with switching enrichment across cell types from KEGG involved Alzheimer's Disease (hsa05010), Huntington's Disease (hsa05016), Parkinson's Disease (hsa05012), and Oxidative Phosphorylation (hsa00190). The induction of the inflammatory immune system response has been coupled to oxidative stress, with associations to

depression and neurodegenerative disorders (Anderson et al., 2014). The connections between them primarily highlights the metabolic dysfunctions connecting these neurodegenerative disorders (Federico et al., 2012). In the remaining 11 gene sets, immunological signaling exhibited over-expression of genes across cell types (**Table 4.6**) and identified relationships like the NOD-like receptor signaling pathway's inflammasome activity working in concert with the Toll-like receptor signaling pathway as a proinflammatory response (**Figure 4.2**; Kumar et al., 2009). While inflammasome-related genes like Il1f9, S100a9, and S100a8 (Novikov et al., 2011) were enriched in peritoneal macrophages (**Table 4.3**), the inflammasome related pathway (hsa04621) was only enriched in microglia (**Figure 4.2**). The lack of inflammasome-related gene set enrichment in peritoneal macrophages may be the result of inhibition by BCG, a method of avoiding host defenses utilized by Mycobacterium and other pathogens (Master et al., 2008; Taxman et al., 2010). The larger number of gene sets with switching enrichment across cell types in REACTOME involved several clusters involving metabolic regulation, (**Figure 4.3**). The REACTOME gene sets enriched in the same direction in both cell types involved cell cycling and DNA replication, as well as interferon and cytokine signaling (**Table 4.7**) which are to be expected following BCG challenge (Castedo et al., 2004; Moreau et al., 2005; Svoboda et al., 2007). The differences in scale of and limited consensus between KEGG and REACTOME can be partially attributed to the difference in scale between their curated gene set lists in MSigDB (186 gene sets in KEGG compared to 674 gene sets listed for REACTOME) as well as differences between their annotation methods resulting in limited consensus (Stobbe et al., 2011).

When considering the gene sets that were over-expressed in BCG treated groups versus Control in microglia while under-expressed in peritoneal macrophages (**Table 4.8**), their shared ribosomal genes have previously been associated with the human innate immune system's inflammatory

response (Calvano et al., 2005). In considering how member genes were shared among the gene sets with enriched member genes under-expressed in microglia and over-expressed in peritoneal macrophages in BCG treatment groups versus control (**Table 4.10**), two profiles developed across the KEGG and REACTOME pathways. The first profile of shared genes were among the 4 KEGG gene sets, and REACTOME gene sets for respiratory electron transport (REACT_22393, REACT_6305) and TCA cycle (REACT_111083) (49 genes; **Table 4.10**). This first profile included Cox genes and mitochondrial complex units associated with ATP production. Previous studies indicated mitochondrial oxidative phosphorylation and protein synthesis are important to the ongoing inflammatory response of macrophages, and in their transition towards resolution as part of the overall immune response (Hall et al., 2014). Such cellular resource allocation is part of the larger immunological response, potentially involving reallocation of energy resources throughout the organism (Straub et al., 2010). Early reallocation of energy from other tissues including the brain to immune functions may include the modification of behaviors (Straub et al., 2010). The opposing direction of enrichment between translation and protein synthesis associated gene sets compared to the ATP synthesizing gene sets suggests an alteration of energetic trends, which may suggest cellular phase modification (Hall et al., 2014) and is possibly reflected in similarly timed changing behavioral patterns (Rodriguez-Zas et al., 2015). The second profile of shared member genes among gene sets under-expressed in microglia and over-expressed in peritoneal macrophages in BCG treatment groups versus control involved several proteasome genes from the alpha, beta, and 26S subunits. These genes were shared among the REACTOME pathways associated with cell-cycle regulation (REACT_6785, REACT_6821), cross-presentation (REACT_111056), Ornithine Decarboxylase regulation (REACT_13565, and degradation of Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G (APOBEC3G;

REACT_9453) (**Table 4.10**). An enriched gene set for cross-presentation of antigens (REACT_111056) is important to the macrophage defense response against intracellular bacteria (Rock and Shen, 2005), although Several enriched gene sets were associated with cellular control of mitotic stages including Early mitotic inhibitor (Emi)-1 degradation by the complex of S-phase kinase-associated protein 1, Cullin 1, and the F-box protein β-transducin repeat-containing protein (SCF-βTrCP; Silverman et al., 2012) and Cadherin (Cdh)1 degradation by the anaphase-promoting complex/cytosome (APC/C) during G1 and G0 in several species including mouse (Margottin-Goguet et al., 2003; Guardavaccaro et al., 2003). There is also speculation that these mitotic regulators are causally involved in mitotic catastrophe in microglial cell lines (Castedo et al., 2004; Svoboda et al., 2007). Other macrophage-infecting pathogens have demonstrated methods of controlling the APC regulation system (Iwai et al., 2007), including a secreted toxin from a Mycobacterium impacting cell-cycle control and apoptosis (Oswald et al., 2005). The re-balancing of energy needs, as well as modifying and regulating cell stages are also associated with programmed cell death, with ATP availability impacting whether cells can proceed with the more energy-intensive apoptosis rather than necrosis (Buttgereit and Brand, 1995; Buttgereit et al., 2000). As apoptosis is necessary to macrophages successfully controlling Mycobacterium challenges while necrosis allows disease progression, regulation is critically important to host defense (Danelishvili et al., 2003; Danelishvili et al., 2010). The simultaneous under-expression of BCG treated groups compared to control in microglia could be a neuroprotective effect, prohibiting the cell-cycle reentry associated with neurodegenerative conditions like Alzheimer's Disease (Bonda et al., 2010).

In conclusion, RNA-seq analysis of murine microglia and peritoneal macrophages identified several highly differentially-expressed genes between BCG-treated and control groups 7 days after challenge. The response of peritoneal macrophages primarily involved over-expression of genes associated with the classic macrophage activation state. Microglia also showed high differential expression in some resolution-phase genes. The GSEA identified enrichment of several gene sets across cell-types. Visual comparison of enriched lists indicated that both cell types shared enrichment by over-expressed genes of terms relating to cell cycle, apoptosis, and immunological signaling. Gene sets similarly enriched in both cell types also connected to networks of gene sets unique to each. When gene sets were assembled into networks, those enriched by under-expressed genes in microglia were primarily related by cellular metabolism. Enrichment in peritoneal macrophages involved over-expressed genes and were only a subset of the gene sets enriched by opposing directions of expression between cell types. This switch in enrichment involved gene sets associated with cell cycle and protein synthesis regulation. Reflecting the shared immunological purpose of microglia and macrophages in their respective regions, these shared areas of switching enrichment may indicate points of gene expression regulating the divergent response to BCG of peritoneal macrophages and microglia. This primarily seems to translate as the direct defensive measures of peritoneal macrophages to BCG through apoptosis signaling and control of bacterial targets for cell-cycle regulators, while microglia negatively regulate the same to maintain a neuroprotective roles and avoid dysregulation associated with neurodegenerative scenarios.

4.6 References

Adami, C., Sorci, G., Blasi, E., Agneletti, A.L., Bistoni, F., Donato, R. (2001). S100B expression in and effects on microglia, *Glia*, **33**(2): 131-142. http://dx.doi.org/10.1002/1098-1136(200102)33:2<131::AID-GLIA1012>3.0.CO;2-D

Ahmed, M., Gaffen, S.L. (2013). IL-17 inhibits adipogenesis in part via C/EBPα, PPARγ and Krüppel-like factors, *Cytokine*, **61**(3): 898-905. http://dx.doi.org/10.1016/j.cyto.2012.12.007

Anderson, G., Berk, M., Dean, O., Moylan, S., Maes, M. (2014). Role of immune-inflammatory and oxidative and nitrosative stress pathways in the etiology of depression: therapeutic implications, *CNS Drugs*, **28**(1): 1-10. http://dx.doi.org/10.1007/s40263-013-0119-1

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, **25**(1): 25-29. http://dx.doi.org/10.1038/75556

Baker, C.A., Manuelidis, L. (2003). Unique inflammatory RNA profiles of microglia in Creutzfeldt-Jakob disease, *Proceedings of the National Academy of Sciences of the United States of America*, **100**(2): 675-679. http://dx.doi.org/10.1073/pnas.0237313100

Bonda, D.J., Lee, H.P., Kudo, W., Zhu, X., Smith, M.A., Lee, H.G. (2010). Pathological implications of cell cycle re-entry in Alzheimer disease, *Expert Reviews in Molecular Medicine*, **12**: e19. http://dx.doi.org/10.1017/S146239941000150X

Buttgereit, F., Brand, M.D. (1995). A hierarchy of ATP-consuming processes in mammalian cells, *The Biochemical Journal*, **312**(Pt 1): 163-167.

Buttgereit, F., Burmester, G.R., Brand, M.D. (2000). Bioenergetics of immune functions: fundamental and therapeutic aspects, *Immunology Today*, **21**(4): 192-199. http://dx.doi.org/10.1016/S0167-5699(00)01593-0

Castedo, M., Perfettini, J.L., Roumier, T., Andreau, K., Medema, R., Kroemer, G. (2004). Cell death by mitotic catastrophe: a molecular definition, *Oncogene*, **23**(16): 2825-2837. http://dx.doi.org/10.1038/sj.onc.1207528

Calvano, S.E., Xiao, W., Richards, D.R., Felciano, R.M., Baker, H.V., Cho, R.J., Chen, R.O., Brownstein, B.H., Cobb, J.P., Tschoeke, S.K., Miller-Graziano, C., Moldawer, L.L., Mindrinos, M.N., Davis, R.W., Tompkins, R.G., Lowry, S.F. (2005). A network-based analysis of systemic inflammation in humans, *Nature*, **437**(7061): 1032-1037. http://dx.doi.org/10.1038/nature03985

Cheong, C., Matos, I., Choi, J.H., Dandamudi, D.B., Shrestha, E., Longhi, M.P., Jeffrey, K.L., Anthony, R.M., Kluger, C., Nchinda, G., Koh, H., Rodriguez, A., Idoyaga, J., Pack, M., Velinzon, K., Park, C.G., Steinman, R.M. (2010). Microbial stimulation fully differentiates monocytes to DC-SIGN/CD209(+) dendritic cells for immune T cell areas, *Cell*, **143**(3): 416-429. http://dx.doi.org/10.1016/j.cell.2010.09.039

Chiu, I.M., Morimoto, E.T., Goodarzi, H., Liao, J.T., O'Keeffe, S., Phatnani, H.P., Muratet, M., Carroll, M.C., Levy, S., Tavazoie, S., Myers, R.M., Maniatis, T. (2013). A neurodegeneration-specific gene-expression signature of acutely isolated microglia from an amyotrophic lateral sclerosis mouse model, *Cell Reports*, **4**(2): 385-401. http://dx.doi.org/10.1016/j.celrep.2013.06.018

Danelishvili, L., McGarvey, J., Li, Y.J., Bermudez, L.E. (2003). Mycobacterium tuberculosis infection causes different levels of apoptosis and necrosis in human macrophages and alveolar

epithelial cells, *Cellular Microbiology*, **5**(9): 649-660. http://dx.doi.org/10.1046/j.1462-5822.2003.00312.x

Danelishvili, L., Yamazaki, Y., Selker, J., Bermudez, L.E. (2010). Secreted Mycobacterium tuberculosis Rv3654c and Rv3655c proteins participate in the suppression of macrophage apoptosis, *PLoS One*, **4**(5): e10474. http://dx.doi.org/10.1371/journal.pone.0010474

Das, A., Das, N.D., Park, J.H., Lee, H.T., Choi, M.R., Jung, K.H., Chai, Y.G. (2013). Identification of survival factors in LPS-stimulated anthrax lethal toxin tolerant RAW 264.7 cells through proteomic approach, *BioChip Journal*, **7**(1): 75-84. http://dx.doi.org/10.1007/s13206-013-7112-0

Davies, L.C., Jenkins, S.J., Allen, J.E., Taylor, P.R. (2013). Tissue-resident macrophages, *Nature Immunology*, **14**(10): 986-995. http://dx.doi.org/10.1038/ni.2705

Ebert, S., Walczak, Y., Remé, C., Langmann, T. (2012). Microglial activation and transcriptomic changes in the blue light-exposed mouse retina, *Advances in Experimental Medicine and Biology*, **723**: 619-632. http://dx.doi.org/10.1007/978-1-4614-0631-0_79

Ehrt, S., Schnappinger, D., Bekiranov, S., Drenkow, J., Shi, S., Gingeras, T.R., Gaasterland, T., Schoolnik, G., Nathan, C. (2001). Reprogramming of the macrophage transcriptome in response to interferon-γ and *Mycobacterium tuberculosis*: signaling roles of nitric oxide synthase-2 and phagocyte oxidase, *The Journal of Experimental Medicine*, **194**(8): 1123-1140. http://dx.doi.org/10.1084/jem.194.8.1123

Federico, A., Cardaioli, E., Da Pozzo, P., Formichi, P., Gallus, G.N., Radi, E. (2012). Mitochondria, oxidative stress and neurodegeneration, *Journal of Neurological Sciences*, **322**(1-2): 254-262. http://dx.doi.org/10.1016/j.jns.2012.05.030

Fenn, A.M., Hall, J.C., Gensel, J.C., Popovich, P.G., Godbout, J.P. (2014). IL-4 signaling drives a unique Arginase^{+}/IL-1β^{+} microglia phenotype and recruits macrophages to the inflammatory CNS: consequences of age-related deficits in IL-4Rα after traumatic spinal cord injury, *The Journal of Neuroscience*, **34**(26): 8904-8917. http://dx.doi.org/10.1523/jneurosci.1146-14.2014

Fensterl, V., Sen, G.C. (2015). Interferon-induced Ifit proteins: their role in viral pathogenesis, *Journal of Virology*, **89**(5): 2462-2468. http://dx.doi.org/10.1128/JVI.02744-14

Fleming, B.D. (2014). *Expanding our understanding of the regulatory macrophage* (Doctoral Dissertation). Retrieved from ProQuest Dissertations and Theses. (Accession Order No. AAT 3682597)

Flynn, G., Maru, S., Loughlin, J., Romero, I.A., Male, D. (2003). Regulation of chemokine receptor expression in human microglia and astrocytes, *Journal of Neuroimmunology*, **136**(1-2): 84-93. http://dx.doi.org/10.1016/S0165-5728(03)00009-2

Gautier, E.L., Shay, T., Miller, J., Greter, M., Jakubzick, C., Ivanov, S., Helft, J., Chow, A., Elpek, K.G., Godonov, S., Mazloom, A.R., Ma'ayan, A., Chua, W.J., Hansen, T.H., Turley, S.J., Merad, M., Randolph, G.J. (2012). Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages, *Nature Immunology*, **13**(11): 1118-1128. http://dx.doi.org/10.1038/ni.2419

Gautier, E.L., Yvan-Charvet, L. (2014). Understanding macrophage diversity at the ontogenic and transcriptomic levels, *Immunological Review*, **262**(1): 85-95. http://dx.doi.org/10.1111/imr.12231

Gopal, R., Lin, Y., Obermajer, N., Slight, S., Nuthalapati, N., Ahmed, M., Kalinski, P., Khader, S.A. (2012). IL-23-dependent IL-17 drives Th1-cell responses following Mycobacterium bovis

BCG vaccination, *European Journal of Immunology*, **42**(2): 364-373. http://dx.doi.org/10.1002/eji.201141569

Gordon, S., Martinez, F.O. (2010). Alternative activation of macrophages: mechanism and functions, *Immunity*, **32**(5): 593-604. http://dx.doi.org/10.1016/j.immuni.2010.05.007

Gordon, S., Plűddemann, A. (2013). Tissue macrophage heterogeneity: issues and prospects, *Seminars in Immunopathology*, **35**(5): 533-540. http://dx.doi.org/10.1007/s00281-013-0386-4

Guardavaccaro, D., Kudo, Y., Boulaire, J., Barchi, M., Busino, L., Donzelli, M., Margottin-Goguet, F., Jackson, P.K., Yamasaki, L., Pagano, M. (2003). Control of meiotic and mitotic progression by the F box protein beta-Trcp1 in vivo, *Developmental Cell*, **4**(6): 799-812. http://dx.doi.org/10.1016/S1534-5807(03)00154-0

Gylfe, A.E., Katainen, R., Kondelin, J., Tanskanen, T., Cajuso, T., Hänninen, U., Taipale, J., Taipale, M., Renkonen-Sinisalo, L., Järvinen, H., Mecklin, J.P., Kilpivaara, O., Pitkänen, E., Vahteristo, P., Tuupanen, S., Karhu, A., Aaltonen, L.A. (2013). Eleven candidate susceptibility genes for common familial colorectal cancer, *PLoS Genetics*, **9**(10): e1003876. http://dx.doi.org/10.1371/journal.pgen.1003876

Hall, C.J., Sanderson, L.E., Crosier, K.E., Crosier, P.S. (2014). Mitochondrial metabolism, reactive oxygen species, and macrophage function-fishing for insights, *Journal of Molecular Medicine*, **92**: 1119-1128. http://dx.doi.org/10.1007/s00109-014-1186-6

Hall, K.L., Volk-Draper, L.D., Flister, M.J., Ran, S. (2012). New model of macrophage acquisition of the lymphatic endothelial phenotype, *PLoS One*, **7**(3): e31794. http://dx.doi.org/10.1371/journal.pone.0031794

Heldwein, K.A., Liang, M.D., Andresen, T.K., Thomas, K.E., Marty, A.M., Cuesta, N., Vogel, S.N., Fenton, M.J. (2003). TLR2 and TLR4 serve distinct roles int eh host immune response

against Mycobacterium bovis BCG, *Journal of Leukocyte Biology*, **74**(2): 277-286. http://dx.doi.org/10.1189/jlb.0103026

Hickman, S.E., Kingery, N.D., Ohsumi, T.K., Borowsky, M.L., Wang, L.C., Means, T.K., El Khoury, J. (2013). The microglial sensome revealed by direct RNA sequencing, *Nature Neuroscience*, **16**(12): 1896-1905. http://dx.doi.org/10.1038/nn.3554

Hiratsuka, S., Watanabe, A., Sakurai, Y., Akashi-Takamura, S., Ishibashi, S., Miyake, K., Shibuya, M., Akira, S., Aburatani, H., Maru, Y. (2008). The S100A8-serum amyloid A3-TLR4 paracrine cascade establishes a pre-metastatic phase, *Nature Cell Biology*, **10**(11): 1349-1355. http://dx.doi.org/10.1038/ncb1794

Huang, da W., Sherman, B.T., Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, *Nature Protocols*, **4**(1): 44-57. http://dx.doi.org/10.1038/nprot.2008.211

Hudson, C., Schwanke, C., Johnson, J.P., Elias, A.F., Phillips, S., Schwalbe, T., Tunby, M., Xu, D. (2014). Confimation of 6q21-6q22.1 deletion in acro-cardio-facial syndrome and further delineation of this contiguous gene deletion syndrome, *American Journal of Medical Genetics*, **164A**(8): 2109-2113. http://dx.doi.org/10.1002/ajmg.a.36548

Iwai, H., Kim, M., Yoshikawa, Y., Ashida, H., Ogawa, M., Fujita, Y., Muller, D., Kirikae, T., Jackson, P.K., Kotani, S., Sasakawa C. (2007). A bacterial effector targets Mad2L2, an APC inhibitor, to modulate host cell cycling, *Cell*, **130**(4): 611-623. http://dx.doi.org/10.1016/j.cell.2007.06.043

Jain, R., Dey, B., Tyagi, A.K. (2012). Development of the first oligonucleotide microarray for global gene expression profiling in guinea pigs: defining the transcription signature of infectious diseases, *BMC Genomics*, **13**: 520. http://dx.doi.org/10.1186/1471-2164-13-520

Kanehisa, M., Goto, S., Kawashima, S., Nakaya, A. (2002). The KEGG databases at GenomeNet, *Nucleic Acids Research*, **30**(1): 42-46. http://dx.doi.org/10.1093/nar/30.1.42

Keller, C., Lauber, J., Blumenthal, A., Buer, J., Ehlers, S. (2004). Resistance and susceptibility to tuberculosis analysed at the transcriptome level: lessons from mouse macrophages, *Tuberculosis*, **84**(3-4): 144-158. http://dx.doi.org/10.1016/j.tube.2003.12.003

Killcoyne, S., Carter, G.W., Smith, J., Boyle, J. (2009). Cytoscape: A community-based framework for network modeling, *Methods in Molecular Biology*, **563**:219-239. http://dx.doi.org/10.1007/978-1-60761-175-2_12

Kota, R.S., Rutledge, J.C., Gohil, K., Kumar, A., Enelow, R.I., Ramana, C.V. (2006). Regulation of gene expression in RAW 264.7 macrophage cell line by interferon-γ, *Biochemical and Biophysical Research Communications*, **342**(4): 1137-1146. http://dx.doi.org/10.1016/j.bbrc.2006.02.087

Kumar, H., Kawai, T., Akira, S. (2009). Toll-like receptors and innate immunity, *Biochemical and Biophysical Research Communications*, **388**(4): 621-625. http://dx.doi.org/10.1016/j.bbrc.2009.08.062

Kyan, Y., Ueda, Y., Yoshida, M., Sasahara, K., Shinya, K. (2014). Transcriptome profiling of brain edemas caused by influenza infection and lipopolysaccharide treatment, *Journal of Medical Virology*, **86**(5): 905-911. http://dx.doi.org/10.1002/jmv.23801

Lin, L., Wang, D., Cao, N., Lin, Y., Jin, Y., Zheng, C. (2013). Whole-transcriptome analysis of hepatocellular carcinoma, *Medical Oncology*, **30**(4): 736. http://dx.doi.org/10.1007/s12032-013-0736-z

Liu, H., Liu, Z., Chen, J., Chen, L., He, X., Zheng, R., Yang, H., Song, P., Weng, D., Hu, H., Fan, L., Xiao, H., Kaufmann, S.H.E., Ernst, J., Ge, B. (2013). Induction of CCL8/MCP-2 by

Mycobacteria through the activation of TLR2/PI3K/Akt signaling pathway, *PLoS One*, **8**(2): e56815. http://dx.doi.org/10.1371/journal.pone.0056815

London, A., Cohen, M., Schwartz, M. (2013). Microglia and monocyte-derived macrophages: functionally distinct populations that act in concert in CNS plasticity and repair, *Frontiers in Cellular Neuroscience*, **7**: 34. http://dx.doi.org/10.3389/fncel.2013.00034

MacMicking, J., Xie, Q.W., Nathan, C. (1997). Nitric oxide and macrophage function, *Annual Review of Immunology*, **15**: 323-350. http://dx.doi.org/10.1146/annurev.immunol.15.1.323

Magee, D.A., Taraktsoglou, M., Killick, K.E., Nalpas, N.C., Browne, J.A., Park, S.D., Conlon, K.M., Lynn, D.J., Hokamp, K., Gordon, S.V., Gormley, E., MacHugh, D.E. (2012). Global gene expression and systems biology analysis of bovine monocyte-derived macrophages in response to *in vitro* challenge with *Mycobacterium bovis*, *PLoS One*, **7**(2): e32034. http://dx.doi.org/10.1371/journal.pone.0032034

Margottin-Goguet, F., Hsu, J.Y., Loktev, A., Hsieh, H.M., Reimann, J.D., Jackson, P.K. (2003). Prophase destruction of Emi1 by the SCF(betaTrCP/Slimb) ubiquitin ligase activates the anaphase promoting complex to allow progression beyond prometaphase, *Developmental Cell*, **4**(6): 813-826. http://dx.doi.org/10.1016/S1534-5807(03)00153-9

Martinez, A.N., Mehra, S., Kaushal, D. (2013). Role of interleukin 6 in innate immunity to Mycobacterium tuberculosis infection, *Journal of Infectious Diseases*, **207**(8): 1253-1261. http://dx.doi.org/10.1093/infdis/jit037

Martinez, F.O., Gordon, S. (2014). The M1 and M2 paradigm of macrophage activation: time for reassessment, *F1000Prime Reports*, **6**: 13. http://dx.doi.org/10.12703/P6-13

Martinez, F.O., Sica, A., Mantovani, A., Locati, M. (2008). Macrophage activation and polarization, *Frontiers in Bioscience*, **13**: 453-461. http://dx.doi.org/10.2741/2692

Master, S.S., Rampini, S.K., Davis, A.S., Keller, C., Ehlers, S., Springer, B., Timmins, G.S., Sander, P., Deretic, V. (2008). Mycobacterium tuberculosis prevents inflammasome activation, *Cell Host & Microbe*, **3**(4): 224-232. http://dx.doi.org/10.1016/j.chom.2008.03.003

Matthews, L., Gopinath, G., Gillespie, M., Caudy, M., Croft, D., de Bono, B., Garapati, P., Hemish, J., Hermjakob, H., Jassal, B., Kanapin, A., Lewis, S., Mahajan, S., May, B., Schmidt, E., Vastrik, I., Wu, G., Birney, E., Stein, L., D'Eustachio, P. (2009). Reactome knowledgebase of human biological pathways and processes, *Nucleic Acids Research*, **37**(Database issue): D619-622. http://dx.doi.org/10.1093/nar/gkn863

McKinsey, T.A., Zhang, C.L., Olson, E.N. (2002). MEF2: a calcium-dependent regulator of cell division, differentiation and death, *Trends in Biochemical Sciences*, **27**(1): 40-47. http://dx.doi.org/10.1016/S0968-0004(01)02031-X

Merico, D., Isserlin, R., Stueker, O., Emili, A., Bader, G.D. (2010). Enrichment map: a network-based method for gene-set enrichment visualization and interpretation, *PLoS One*, **5**(11): e13984. http://dx.doi.org/10.1371/journal.pone.0013984

Moreau, M., André, C., O'Connor, J.C., Dumich, S.A., Woods, J.A., Kelley, K.W., Dantzer, R., Lestage, J., Castanon, N. (2008). Inoculation of Bacillus Calmette-Guerin to mice induces an acute episode of sickness behavior followed by chronic depressive-like behavior. Brain, Behavior, and Immunity, **22**(7): 1087-1095. http://dx.doi.org/10.1016/j.bbi.2008.04.001

Moreau, M., Lestage, J., Verrier, D., Mormède, C., Kelley, K.W., Dantzer, R., Castanon, N. (2005). Bacille Calmette-Guérin inoculation induces chronic activation of peripheral and brain indoleamine 2,3-dioxygenase in mice. Journal of Infectious Diseases, **192**(3): 537-544. http://dx.doi.org/10.1086/431603

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-seq, *Nature Methods*, **5**(7): 621-628. http://dx.doi.org/10.1038/nmeth.1226

Nagare, T., Sakaue, H., Matsumoto, M., Cao, Y., Inagaki, K., Sakai, M., Takashima, Y., Nakamura, K., Mori, T., Okada, Y., Matsuki, Y., Watanabe, E., Ikeda, K., Taguchi, R., Kamimura, N., Ohta, S., Hiramatsu, R., Kasuga, M. (2011). Overexpression of KLF15 transcription factor in adipocytes of mice results in down-regulation of SCD1 protein expression in adipocytes and consequent enhancement of glucose-induced insulin secretion, *Journal of Biological Chemistry*, **286**(43): 37458-37469. http://dx.doi.org/10.1074/jbc.M111.242651

Newman, M.E.J. (2012). Communities, modules and large-scale structure in networks, *Nature Physics*, **8**: 25-31. http://dx.doi.org/10.1038/nphys2162

Nikodemova, M., Watters, J.J. (2012). Efficient isolation of live microglia with preserved phenotypes from adult mouse brain, *Journal of Neuroinflammation*, **9**: 147. http://dx.doi.org/10.1186/1742-2094-9-147

Nixon, S.E., González-Peña, D., Lawson, M.A., McCusker, R.H., Hernandez, A.G., O'Connor, J.C., Dantzer, R., Kelley, K.W., Rodriguez-Zas, S.L. (2015). Analytical workflow profiling gene expression in murine macrophages. *Journal of Bioinformatics and Computational Biology*, **13**(2): 1550010. http://dx.doi.org/10.1142/S0219720015500109

Nonaka, H., Niidome, T., Shinozuka, Y., Akaike, A., Kihara, T., Sugimoto, H. (2009). A role for SOX2 in the generation of microtubule-associated protein 2-positive cells from microglia, *Biochemical and Biophysical Research Communications*, **380**(1): 60-64. http://dx.doi.org/10.1016/j.bbrc.2009.01.027

Novikov, A., Cardone, M., Thompson, R., Shenderov, K., Kirshchman, K.D., Mayer-Barber, K.D., Myers, T.G., Rabin, R.L., Trinchieri, G., Sher, A., Feng, C.G. (2011). Mycobacterium tuberculosis triggers host type I IFN signaling to regulate IL-1β production in human macrophages, *Journal of Immunology*, **187**(5): 2540-2547. http://dx.doi.org/10.4049/jimmunol.1100926

O'Connor, J.C., André, C., Wang, Y., Lawson, M.A., Szegedi, S.S., Lestage, J., Castanon, N., Kelley, K.W., Dantzer, R. (2009). 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-Guérin. *Journal of Neuroscience*, **29**(13): 4200-4209. http://dx.doi.org/10.1523/JNEUROSCI.5032-08.2009

de Oliveira, C.C., de Oliveira, S.M., Goes, V.M., Probst, C.M., Krieger, M.A., Buchi Dde, F. (2008). Gene expression profiling of macrophages following mice treatment with an immunomodulator medication, *Journal of Cellular Biochemistry*, **104**(4): 1364-1377. http://dx.doi.org/10.1002/jcb.21713

Orre, M., Kamphuis, W., Osborn, L.M., Jansen, A.H.P., Kooijman, L., Bossers, K., Hol, E.M. (2014). Isolation of glia from Alzheimer's mice reveals inflammation and dysfunction, Neurobiology of Aging, 35(12): 2746-2760. http://dx.doi.org/10.1016/j.neurobiolaging.2014.06.004

Oswald, E., Nougayrède, J.P., Taieb, F., Sugai, M. (2005). Bacterial toxins that modulate host cell-cycle progression, *Current Opinion in Microbiology*, **8**(1): 83-91. http://dx.doi.org/10.1016/j.mib.2004.12.011

Pesce, J.T., Ramalingam, T.R., Wilson, M.S., Mentink-Kane, M.M., Thompson, R.W., Cheever, A.W., Urban, J.F. Jr., Wynn, T.A. (2009). Retnla (relmalpha/fizz1) suppresses helminth-

induced Th2-type immunity, *PLoS Pathogens*, **5**(4): e1000393. http://dx.doi.org/10.1371/journal.ppat.1000393

Prinz, M., Priller, J. (2014). Microglia and brain macrophages in the molecular age: from origin to neuropsychiatric disease, *Nature Reviews Neuroscience*, **15**(5): 300-312. http://dx.doi.org/10.1038/nrn3722

Reno, C., Marchuk, L., Sciore, P., Frank, C.B., Hart, D.A. (1997). Rapid isolation of total RNA from small samples of hypocellular, dense connective tissues, *BioTechniques*, **22**(6): 1082-1086.

Roberts, A., Pimentel, H., Trapnell, C., Pachter, L. (2011). Identification of novel transcripts in annotated genomes using RNA-seq, *Bioinformatics*, **27**(17): 2325-2329. http://dx.doi.org/10.1093/bioinformatics/btr355

Rock, K.L., Shen, L. (2005). Cross-presentation: underlying mechanisms and role in immune surveillance, *Immunological Reviews*, **207**(1): 166-183. http://dx.doi.org/10.1111/j.0105-2896.2005.00301.x

Rodriguez-Zas, S.L., Nixon, S.E., Lawson, M.A., McCusker, R.H., Southey, B.R., O'Connor, J.C., Danzter, R., Kelley, K.W. (2015). Advancing the understanding of behaviors associated with Bacille Calmette Guérin infection using multivariate analysis, *Brain Behavior and Immunity*, **44**: 176-186. http://dx.doi.org/10.1016/j.bbi.2014.09.018

Saito, R., Smooth, M.E., Ono, K., Ruscheinski, J., Wang, P.L., Lotia, S., Pico, A.R., Bader, G.D., Ideker, T. (2012). A travel guide to Cytoscape plugins, *Nature Methods*, **9**(11): 1069-1076. http://dx.doi.org/10.1038/nmeth.2212

Silverman, J.S., Skaar, J.R., Pagano, M. (2012). SCF ubiquitin ligases in the maintenance of genome stability, *Trends in Biochemical Sciences*, **37**(2): 66-73. http://dx.doi.org/10.1016/j.tibs.2011.10.004

Sirén, A.L., McCarron, R., Wang, L., Garcia-Pinto, P., Ruetzler, C., Martin, D., Hallenbeck, J.M. (2001). Proinflammatory cytokine expression contributes to brain injury provoked by chronic monocyte activation, *Molecular Medicine*, **7**(4): 219-229.

Solga, A.C., Pong, W.W., Walker, J., Wylie, T., Magrini, V., Apicelli, A.J., Griffith, M., Griffith, O.L., Kohsaka, S., Wu, G.F., Brody, D.L., Mardis, E.R., Gutmann, D.H. (2015). RNA-sequencing reveals oligodendrocyte and neuronal transcripts in microglia relevant to central nervous system disease, *Glia*, **63**(4): 531-548. http://dx.doi.org/10.1002/glia.22754

Stables, M.J., Shah, S., Camon, E.B., Lovering, R.C., Newson, J., Bystrom, J., Farrow, S., Gilroy, D.W. (2011). Transcriptomic analyses of murine resolution-phase macrophages, *Blood*, **118**(26): e192-208. http://dx.doi.org/10.1182/blood-2011-04-345330

Stobbe, M.D., Houten, S.M., Jansen, G.A., van Kampen, A.H., Moerland, P.D. (2011). Critical assessment of human metabolic pathway databases: a stepping stone for future integration, *BMC Systems Biology*, **5**: 165. http://dx.doi.org/10.1186/1752-0509-5-165

Straub, R.H., Cutolo, M., Buttgereit, F., Pongratz, G. (2010). Energy regulation and neuroendocrine-immune control in chronic inflammatory disease, *Journal of Internal Medicine*, **267**(6): 543-560. http://dx.doi.org/10.1111/j.1365-2796.2010.02218.x

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression

profiles, *Proceedings of the National Academy of Sciences of the United States of America*, **102**(43): 15545-15550. http://dx.doi.org/10.1073/pnas.0506580102

Svoboda, N., Zierler, S., Kerschbaum, H.H. (2007). cAMP mediates ammonia-induced programmed cell death in microglial cell line BV-2, *The European Journal of Neuroscience*, **25**(8): 2285-2295. http://dx.doi.org/10.1111/j.1460-9568.2007.05452.x

Tada, S., Okuno, T., Hitoshi, Y., Yasui, T., Honorat, J.A., Takata, K., Koda, T., Shimagami, H., Chi-Jing, C., Namba, A., Sugimoto, T., Sakoda, S., Mochizuki, H., Kikutani, H., Nakatsuiji, Y. (2014). Partial suppression of M1 microglia by Janus kinase 2 inhibtor does not protect against neurodegeneration in animal models of amyotrophic lateral sclerosis, *Jouranl of Neuroinflammation*, **11**: 179. http://dx.doi.org/10.1186/s12974-014-0179-2

Takikawa, O., Oku, T., Ito, N., Ushio, Y., Yamamoto, N., Yoneda, Y., Tsuji, J., Sanchez-Bueno, A., Verkhusha, V., Yoshida, R. (1996). Multiple expression of Ly-6C and accumulation of a Ly-6C pre-mRNA in activated macrophages involved in rejection of an allografted tumor, *Biochemical and Biophysical Research Communications*, **226**(1): 247-253. http://dx.doi.org/10.1006/bbrc.1996.1341

Taxman, D.J., Huang, M.T., Ting, J.P. (2010). Inflammasome inhibition as a pathogenic stealth mechanism, *Cell Host & Microbe*, **8**(1): 7-11. http://dx.doi.org/10.1016/j.chom.2010.06.005

Thomas, K.E., Galligan, C.L., Newman, R.D., Fish, E.N., Vogel, S.N. (2006). Contribution of interferon-beta to the murine macrophage response to the toll-like receptor 4 agonist, lipopolysaccharide, *Journal of Biological Chemistry*, **281**(41): 31119-31130. http://dx.doi.org/10.1074/jbc.M604958200

Thomson, C.A., McColl, A., Cavanagh, J., Graham, G.J. (2014). Peripheral inflammation is associated with remote global gene expression changes in the brain, *Journal of*

Neuroinflammation, **11**: 73. http://dx.doi.org/10.1186/1742-2094-11-73 Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq, *Nature Biotechnology*, **31**(1): 46-53. http://dx.doi.org/10.1038/nbt.2450

Wang, Q., Sun, J., Zhou, M., Yang, H., Li, Y., Li, X., Lv, S., Li, X., Li, Y. (2011). A novel network-based method for measuring the functional relationship between gene sets, *Bioinformatics*, **27**(11): 1521-1528. http://dx.doi.org/10.1093/bioinformatics/btr154

Yu, H., Sun, H., Bai, Y., Han, J., Liu, G., Liu, Y., Zhang, N. (2015). MEF2D overexpression contributes to the progression of osteosarcoma, *Gene*, **563**(2): 130-135. http://dx.doi.org/10.1016/j.gene.2015.03.046

Zanin, R.F., Braganhol, E., Bergamin, L.S., Campesato, L.F., Filho, A.Z., Moreira, J.C., Morrone, F.B., Sévigny, J., Schetinger, M.R., de Souza Wyse, A.T., Battastini, A.M. (2012). Differential macrophage activation alters the expression profile of NTPDase and ecto-5'-nucleotidase, *PLoS One*, **7**(2): e31205. http://dx.doi.org/10.1371/journal.pone.0031205

Zhang, X., Goncalves, R., Mosser, D.M. (2008). The isolation and characterization of murine macrophages, *Current Protocols in Immunology*, **Chapter 14**: Unit 14.1. http://dx.doi.org/10.1002/0471142735.im1401s83

Zhang, X., Johnson, A.D., Hendricks, A.E., Hwang, S.J., Tanriverdi, K., Ganesh, S.K., Smith, N.L., Peyser, P.A., Freedman, J.E., O'Donnell, C.J. (2014). Genetic assocations with expression for genes implicated in GWAS studies for atherosclerotic cardiovascular disease and blood phenotypes, *Human Molecular Genetics*, **23**(3): 782-795. http://dx.doi.org/10.1093/hmg/ddt461

Zollbrecht, C., Grassl, M., Fenk, S., Höcherl, R., Hubauer, U., Reinhard, W., Esslinger, U.B., Ebert, S., Langmann, T., Stark, K., Hengstenberg, C. (2013). Expression pattern in human macrophages dependent on 9p21.3 coronary artery disease risk locus, *Atherosclerosis*, **227**(2): 244-249. http://dx.doi.org/10.1016/j.atherosclerosis.2012.12.030

4.7 Figures

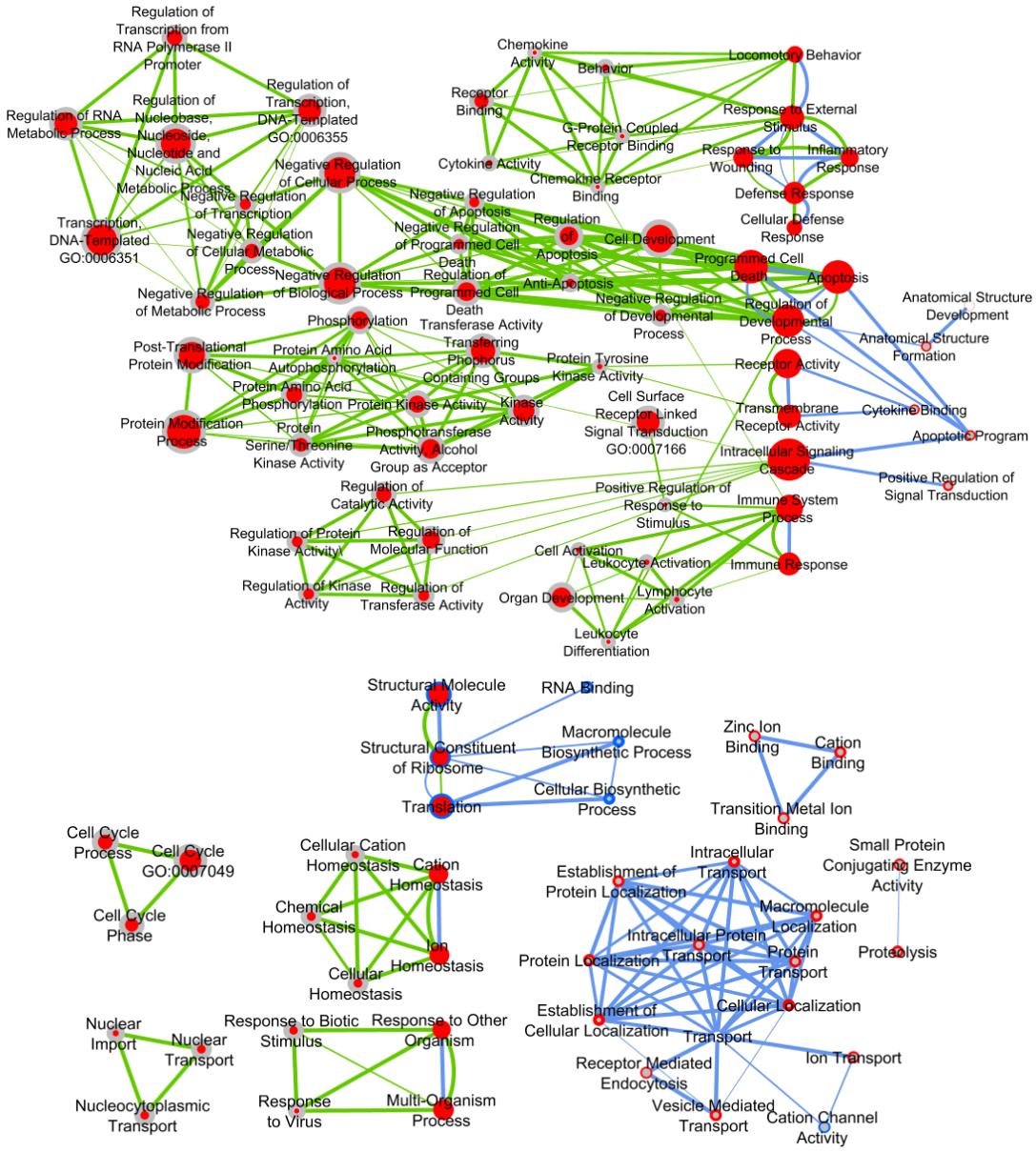


Figure 4.1 Immunologically enriched Gene Ontology gene set network in microglia and peritoneal macrophages

Network of gene sets from Gene Ontology Biological Process and Molecular Function. Node color corresponds to significant up- (Red) or down- (Blue) regulation in BCG compared to Ctrl (nominal p-value < 0.001; FDR-adjusted p-value < 0.05). Non-significant tissue results (Grey) are also indicated. The color of the node center indicates response in microglia while the node border indicates response in peritoneal macrophages. Edges indicate that the enriching genes for gene sets in microglia (Green) or peritoneal macrophages (Blue) are shared (similarity coefficient > 0.5). Node size reflects the number of enriching genes in the set, and edge width corresponds to the number of genes overlapping between the two nodes.

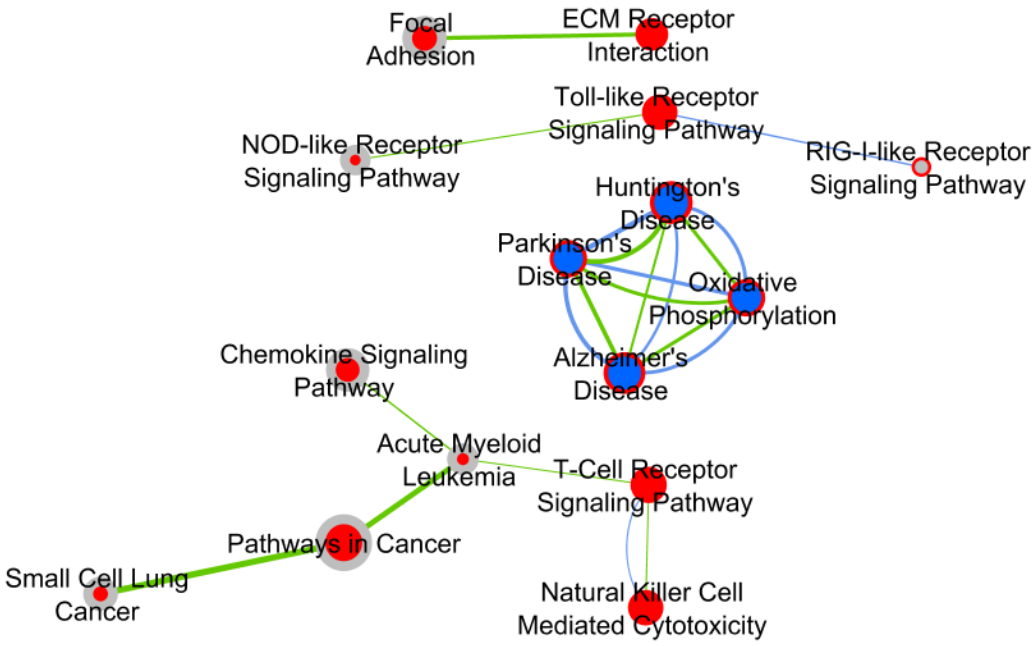


Figure 4.2 Immunologically enriched Kyoto Encyclopedia of Genes and Genomes gene set network in microglia and peritoneal macrophages
Network of gene sets from Kyoto Encyclopedia of Genes and Genomes. Node color corresponds to significant up- (Red) or down- (Blue) regulation in BCG compared to Ctrl (nominal p-value < 0.001; FDR-adjusted p-value < 0.05). Non-significant tissue results (Grey) are also indicated. The color of the node center indicates response in microglia while the node border indicates response in peritoneal macrophages. Edges indicate that the enriching genes for gene sets in microglia (Green) or peritoneal macrophages (Blue) are shared (similarity coefficient > 0.5). Node size reflects the number of enriching genes in the set, and edge width corresponds to the number of genes overlapping between the two nodes.

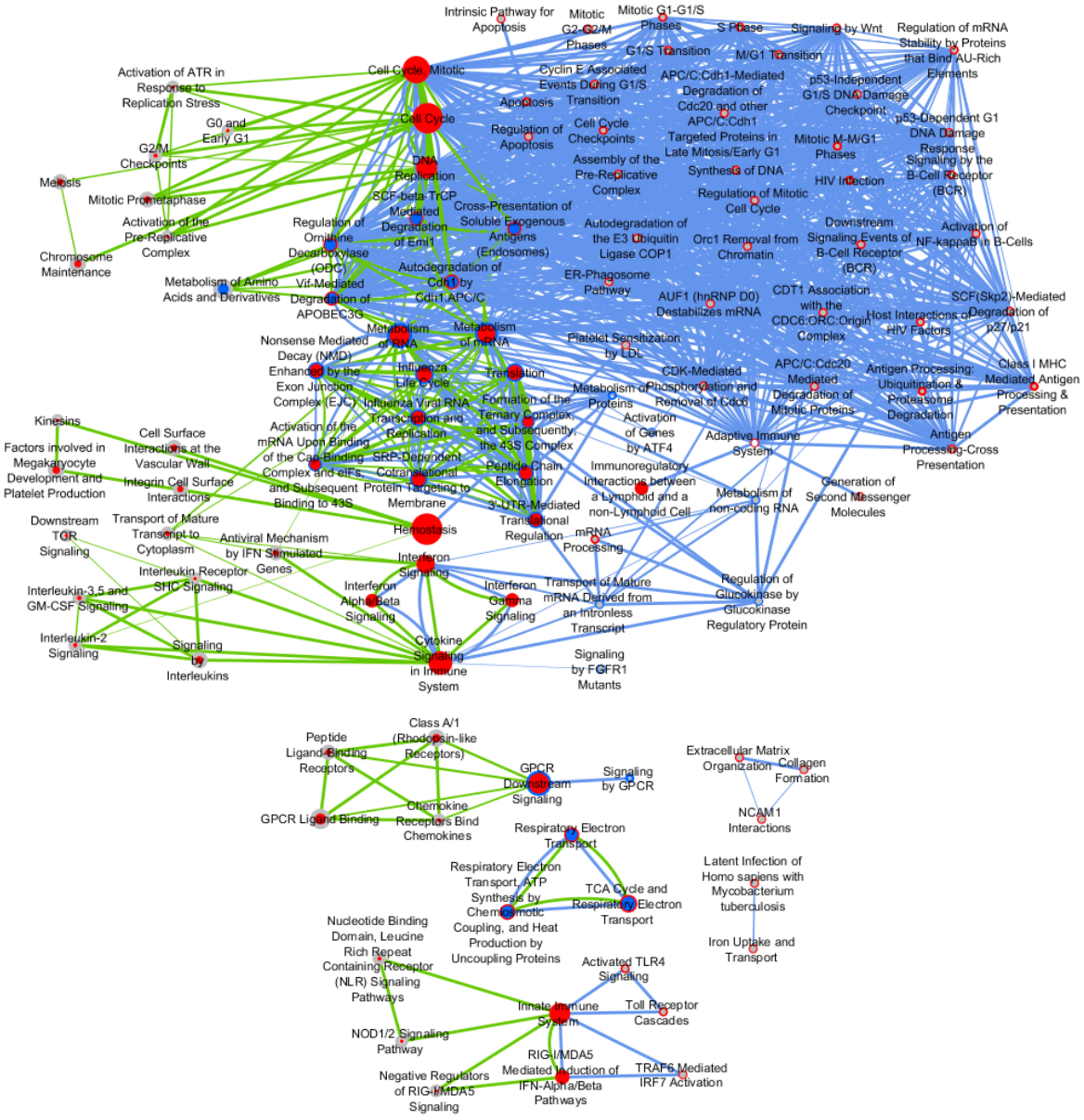


Figure 4.3 Immunologically enriched REACTOME gene set network in microglia and peritoneal macrophages

Network of gene sets from Kyoto Encyclopedia of Genes and Genomes. Node color corresponds to significant up- (Red) or down- (Blue) regulation in BCG compared to Ctrl (nominal p-value < 0.001; FDR-adjusted p-value < 0.05). Non-significant tissue results (Grey) are also indicated. The color of the node center indicates response in microglia while the node border indicates response in peritoneal macrophages. Edges indicate that the enriching genes for gene sets in microglia (Green) or peritoneal macrophages (Blue) are shared (similarity coefficient > 0.5). Node size reflects the number of enriching genes in the set, and edge width corresponds to the number of genes overlapping between the two nodes.

4.8 Tables

Table 4.1 Top 20 genes by fold-change, over-expressed in BCG relative to control groups in microglia (FDR-adjusted p-value < 0.05)

Gene Name[1]	Gene ID (NCBI)	Log_2 Fold-Change (BCG/Control)	P-value	Ref
Saa3	20210	5.90	0.002	Baker and Manuelidis, 2003; Hickman et al., 2013
Cadm3	94332	5.14	0.002	-
Car6	12353	5.01	0.044	Das et al., 2013
Steap4	117167	4.89	0.002	Chiu et al., 2013; Fleming, 2014
Sele	20339	4.80	0.002	Stables et al., 2011
Cxcr1	227288	4.03	0.002	Hickman et al., 2013; Flynn et al., 2003
Ifitm1	68713	3.72	0.002	Hickman et al., 2013
Irg1	16365	3.66	0.002	Thomas et al., 2006
Lrg1	76905	3.57	0.002	-
Prok2	50501	3.48	0.002	-
Cldn4	12740	3.44	0.030	Hickman et al., 2013
Cfb	14962	3.39	0.002	Chiu et al., 2013
Slfn4	20558	3.33	0.002	Chiu et al., 2013
Nxpe5	381680	3.29	0.002	-
Ly6i	57248	3.29	0.002	Martinez et al., 2013
Oas3	246727	3.27	0.002	Kota et al., 2006
Ifi205	226695	3.15	0.002	Thomas et al., 2006
Stfa1	20861	3.12	0.002	Stables et al., 2011
Plac8	231507	3.09	0.002	Zollbrecht et al., 2013
Tarm1	245126	2.99	0.002	-

[1] Cadm3: Cell adhesion molecule 3; Car6: Carbonic anhydrase 6; Cfb: Complement factor B; Cldn4: Claudin 4; Cxcr1: Chemokine (C-X-C motif) receptor 1; Ifi205: Interferon activated gene 205; Ifitm1: Interferon induced transmembrane protein 1; Irg1: Immunoresponsive gene 1; Lrg1: Leucine-rich alpha-2-glycoprotein 1; Ly6i: Lymphocyte antigen 6 complex, locus I; Nxpe5: Neurexophilin and PC-esterase domain family, member 5; Oas3: 2'-5' oligoadenylate synthetase 3; Plac8: Placenta-specific 8; Prok2: Prokineticin 2; Saa3: Serum amyloid A3; Sele: Selectin, endothelial cell; Slfn4: Schlafen 4; Steap4: STEAP family member 4; Stfa1: Stefin A1; Tarm1: T cell-interacting, activating receptor on myeloid cells 1.

Log_2 FC: Log_2-based Fold-change

P-value: False Detection Rate-adjusted p-value

References indicate previous identification in microglia or in a macrophage inflammation model

Table 4.2 Top 20 genes by fold-change, under-expressed in BCG relative to control groups in microglia (FDR-adjusted p-values < 0.05)

Gene Name[1]	Gene ID	Log_2 Fold-Change (BCG/Control)	P-value	Ref
Ccdc162	72002	-1.21	0.0021	-
Tcap	21393	-1.15	0.0302	Chiu et al., 2013; Orre et al., 2014
Hba-a1	15122	-1.10	0.0021	de Oliveira et al., 2008
Hba-a2	110257	-1.09	0.0021	de Oliveira et al., 2008
Retnla	57262	-1.00	0.0447	Pesce et al., 2009
Bcas1	76960	-0.96	0.0266	-
Mobp	17433	-0.95	0.0122	Solga et al., 2015
Tnfrsf17	21935	-0.88	0.0021	Chiu et al., 2011
D7Ertd443e	71007	-0.87	0.0166	-
Olig1	50914	-0.84	0.0202	Nonaka et al., 2009
Sema4b	20352	-0.84	0.0021	Chiu et al., 2011
Sec16b	89867	-0.82	0.0091	-
S100b	20203	-0.81	0.0190	Adami et al., 2001
Klhdc8b	78267	-0.81	0.0021	-
Plekhb1	27276	-0.77	0.0166	Solga et al., 2015
Cldn11	18417	-0.77	0.0091	Solga et al., 2015
Gpr37	14763	-0.77	0.0266	Ebert et al., 2012
Tmem88b	320587	-0.76	0.0290	-
Hbb-bs	100503605	-0.71	0.0021	-
Upk1b	22268	-0.69	0.0021	Orre et al., 2014

[1] Bcas1: Breast carcinoma amplified sequence 1; Ccdc162: Coiled-coil domain containing 162; Cldn11: Claudin 11; D7Ertd443e: DNA segment, Chr 7, ERATO doi 443, expressed; Gpr37: G protein-coupled receptor 37; Hba-a1: Hemoglobin alpha, adult chain 1; Hba-a2: Hemoglobin alpha, adult chain 2; Hbb-bs: Hemoglobin, beta adult S chain; Klhdc8b: Kelch domain containing 8B; Mobp: Myelin-associated oligodendrocytic basic protein; Olig1: Oligodendrocyte transcription factor 1; Plekhb1: Pleckstrin homology domain containing, family B (evectins) member 1; Retnla: Resistin like alpha; S100b: S100 protein, beta polypeptide, neural; Sec16b: SEC16 homolog B (S. cerevisiae); Sema4b: Sema domain, immunoglobulin domain (Ig), transmembrane domain (TM) and short cytoplasmic domain, (semaphoring) 4B; Tcap: Titin-cap; Tmem88b: Transmembrane protein 88B; Tnfrsf17: Tumor necrosis factor receptor superfamily, member 17; Upk1b: Uroplakin 1B.

Log_2 FC: Log_2-based Fold-change

P-value: False Detection Rate-adjusted p-value

References indicate previous identification in microglia or in a macrophage inflammation model

Table 4.3 Top 20 genes by fold-change, over-expressed in BCG relative to control groups in peritoneal macrophages (FDR-adjusted p-value < 0.05)

Gene Name[1]	Gene ID	Log_2 FC (BCG/Control)	P-value	Ref
S100a9	20202	10.11	0.0004	Hickman et al., 2013
Mrgpra2a	668727	10.10	0.0004	-
Stfa2l1	268885	10.03	0.0009	-
Ly6i	57248	9.42	0.0003	Martinez et al., 2013
Asprv1	67855	9.15	0.0003	Stables et al., 2011
Gm5483	433016	9.12	0.0025	Fleming, 2014
Wfdc21	66107	8.87	0.0025	Hickman et al., 2013
Ly6c2	100041546	7.92	0.0004	Takikawa et al., 1996
Nos2	18126	7.69	0.0004	Ehrt et al., 2001; MacMicking et al., 1997
Il1f9	215257	7.66	0.0004	Stables et al., 2011
Ccl8	20307	7.45	0.0004	Liu et al., 2013
2010002M12Rik	112419	7.32	0.0004	Fensterl and Sen, 2015
Spon1	233744	7.23	0.0004	Magee et al., 2012
S100a8	20201	7.15	0.0004	Ehrt et al., 2001; Hickman et al., 2013
Entpd3	215446	7.03	0.0078	Zanin et al., 2012
Cxcr2	12765	7.02	0.0004	Hickman et al., 2013
A530046M15Rik	328190	6.96	0.0012	-
Ifng	15978	6.45	0.0004	Stables et al., 2011
Cxcl9	17329	6.40	0.0004	Hickman et al., 2013
Col5a3	53867	6.27	0.0234	-

[1] 2010002M12Rik: RIKEN cDNA 2010002M12 gene; A530046M15Rik: RIKEN cDNA A530046M15 gene; Asprv1: Aspartic peptidase, retroviral-like 1; Ccl8: Chemokine (C-C motif) ligand 8; Col5a3: Collagen, type V, alpha 3; Cxcl9: Chemokine (C-X-C motif) ligand 9; Cxcr2: Chemokine (C-X-C motif) receptor 2; Entpd3: Ectonucleoside triphosphate diphosphohydrolase 3; Gm5483: Predicted gene 5483; Ifng: Interferon gamma; Il1f9: Interleukin 1 family, member 9; Ly6c2: Lymphocyte antigen 6 complex, locus C2; Ly6i: Lymphocyte antigen 6 complex, locus I; Mrgpra2a: MAS-related GPR, member A2A; Nos2: Nitric oxide synthase 2, inducible; S100a8: S100 calcium binding protein A8 (calgranulin A); S100a9: S100 calcium binding protein A9 (calgranulin B); Spon1: Spondin 1, (f-spondin) extracellular matrix protein; Stfa2l1: Stefin A2 like 1; Wfdc21: WAP four-disulfide core domain.
Log_2 FC: Log_2-based Fold-change
P-value: False Detection Rate-adjusted p-value

Table 4.4 Top 20 genes by fold-change, under-expressed in BCG relative to control groups in peritoneal macrophages (FDR-adjusted p-value < 0.05)

Gene Name[1]	Gene ID	Log_2 Fold-Change (BCG/Control)	P-value	Ref
Prss37	67690	-3.49	0.0341	-
Mucl1	20771	-3.09	0.0041	Zollbrecht et al., 2013
Rprm	67874	-3.07	0.0004	-
Sox7	20680	-3.01	0.0017	Hall et al., 2012; Solga et al., 2015
Retnla	57262	-2.89	0.0004	Pesce et al., 2009; Hickman et al., 2013
Cyp26a1	13082	-2.84	0.0004	-
Cd209b		-2.76	0.0106	Cheong et al., 2010
Cd209f	69142	-2.73	0.0020	Stables et al., 2011
A4galt	239559	-2.70	0.0004	-
Klf15	66277	-2.68	0.0015	Nagare et al., 2011
Cd209a	170786	-2.66	0.0004	Stables et al., 2011
1810046K07Rik	69809	-2.61	0.0004	Hickman et al., 2013
Fcrls	80891	-2.60	0.0004	Stables et al., 2011
Frmpd1	666060	-2.57	0.0385	-
Pf4	56744	-2.56	0.0004	Stables et al., 2011
Tpsab1	100503895	-2.54	0.0004	-
Ccl24	56221	-2.53	0.0004	Chiu et al., 2013
Gna14	14675	-2.53	0.0015	-
Kcne3	57442	-2.50	0.0004	-
Artn	11876	-2.49	0.0025	-

[1] 1810046K07Rik: RIKEN cDNA 1810046K07 gene; A4galt: Alpha 1,4-galactosyltransferase; Artn: Artemin; Ccl24: Chemokine (C-C motif) ligand 24; Cd209a: Cd209a antigen; Cd209b: Cd209b antigen; Cd209f: Cd209f antigen; Cyp26a1: Cytochrome P450, family 11, subfamily a, polypeptide 1; Fcrls: Fc receptor-like S, scavenger receptor; Frmpd1: FERM and PDZ domain containing 1; Gna14: Guanine nucleotide binding protein, alpha 14; Kcne3: Potassium voltage-gated channel, Isk-related subfamily, member 3; Klf15: Kruppel-like factor 15; Mucl1: Mucin-like 1; Pf4: Platelet factor 4; Prss37: Protease, serine 37; Retnla: Resistin like alpha; Rprm: Reprimo, TP53 dependent G2 arrest mediator candidate; Sox7: SRY (sex determining region Y)-box 7; Tpsab1: Tryptase alpha/beta 1.

Table 4.5 Gene Ontology Biological Process (BP) and Molecular Functions (MF) gene sets enriched in both microglia and peritoneal macrophages (FDR-adjusted p-value < 0.05)

GO Gene Set (GO Identifier)	Microglia			Macrophage		
	Genes	NES	FDR	Genes	NES	FDR
BP: Apoptosis (GO:0006915)	325	3.23	> 0.001	311	2.60	2.27E-03
BP: Cation Homeostasis (GO:0055080)	58	3.03	> 0.001	61	2.28	8.90E-03
BP: Cellular Defense Response (GO:0006968)	35	2.48	1.87E-03	38	3.13	> 0.001
BP: Defense Response (GO:0006952)	149	4.41	> 0.001	148	4.46	> 0.001
BP: Immune Response (GO:0006955)	153	4.05	> 0.001	155	3.19	> 0.001
BP: Immune System Process (GO:0002376)	221	4.31	> 0.001	218	3.19	> 0.001
BP: Inflammatory Response (GO:0006954)	79	3.55	> 0.001	74	2.63	1.99E-03
BP: Intracellular Signaling Cascade (GO:0007242)	478	3.63	> 0.001	439	2.26	9.46E-03
BP: Ion Homeostasis (GO:0050801)	65	2.88	> 0.001	68	2.34	6.75E-03
BP: Locomotory Behavior (GO:0007626)	58	3.18	> 0.001	59	2.02	0.030
BP: Multi-Organism Process (GO:0051704)	79	2.64	> 0.001	70	2.25	9.80E-03
BP: Programmed Cell Death (GO:0012501)	326	3.25	> 0.001	312	2.60	2.17E-03
MF: Receptor Activity (GO:0004872)	313	2.95	> 0.001	253	3.90	> 0.001
BP: Regulation of Developmental Process (GO:0050793)	320	2.72	> 0.001	299	2.22	0.011
BP: Response to External Stimulus (GO:0009605)	185	3.85	> 0.001	174	2.36	6.57E-03
BP: Response to Other Organism (GO:0051707)	43	2.81	> 0.001	42	2.79	> 0.001
BP: Response to Wounding (GO:0009611)	114	2.96	> 0.001	107	2.12	0.018
MF: Structural Constituent of Ribosome (GO:0003735)	77	5.55	> 0.001	76	-6.72	> 0.001
MF: Structural Molecule Activity (GO:0005198)	165	5.17	> 0.001	141	-4.97	> 0.001
BP: Translation (GO:0006412)	138	2.89	> 0.001	134	-3.52	> 0.001
MF: Transmembrane Receptor Activity (GO:0004888)	204	2.34	4.34E-03	161	4.16	> 0.001

Genes: # of genes in the gene set represented in the expression list; NES: Normalized Enrichment Score; FDR: The gene set enrichment FDR-adjusted p-value.

Table 4.6 Kyoto Encyclopedia of Genes and Genomes pathways enriched in both microglia or peritoneal macrophages (FDR-adjusted p-value < 0.05)

Gene Set (KEGG hsa Identifier)	Microglia			Macrophage		
	Genes	NES	FDR	Genes	NES	FDR
Acute Myeloid Leukemia (05221)	52	2.42	1.51E-03	NE	NE	NE
Alzheimer's Disease (05010)	132	-2.40	4.32E-03	124	3.68	> 0.001
Chemokine Signaling Pathway (04062)	145	3.23	> 0.001	NE	NE	NE
ECM-Receptor Interaction (04512)	61	2.64	> 0.001	45	2.64	> 0.001
Focal Adhesion (04510)	153	2.64	> 0.001	NE	NE	NE
Huntington's Disease (05016)	140	-3.17	> 0.001	137	3.83	> 0.001
Natural Killer Cell-Mediated Cytotoxicity (04650)	82	3.36	> 0.001	80	2.63	> 0.001
NOD-Like Receptor Signaling Pathway (04621)	43	2.47	1.02E-03	NE	NE	NE
Oxidative Phosphorylation (00190)	98	-3.81	> 0.001	95	3.99	> 0.001
Parkinson's Disease (05012)	97	-3.84	> 0.001	91	3.66	> 0.001
Pathways in Cancer (05200)	248	2.27	4.30E-03	NE	NE	NE
RigI-Like Receptor Signaling Pathway (04622)	NE	NE	NE	47	2.71	> 0.001
Small Cell Lung Cancer (05222)	75	2.22	4.88E-03	NE	NE	NE
T-Cell Receptor Signaling Pathway (04660)	89	2.90	> 0.001	97	2.50	> 0.001
Toll-Like Receptor Signaling Pathway (04620)	82	2.67	> 0.001	80	2.62	> 0.001

Genes: # of genes in the gene set represented in the expression list; NES: Normalized Enrichment Score; FDR: The gene set enrichment FDR-adjusted p-value;
NE: Not enriched, gene set is only enriched in microglia or peritoneal macrophages, not both.

Table 4.7 REACTOME pathways enriched in both microglia and peritoneal macrophages (FDR-adjusted p-value < 0.05)

Gene Set (REACT Identifier)	Microglia			Macrophage		
	Genes	NES	FDR	Genes	NES	FDR
3' UTR-Mediated Translational Regulation (1762)	86	7.18	> 0.001	84	-7.67	> 0.001
Activation of mRNA upon binding Cap-Binding Complex, eIFs, and to 43S (1258)	39	4.63	> 0.001	38	-4.53	> 0.001
Autodegradation of Cdh1 by Cdh1:APC/C (6785)	56	-2.31	0.019	56	4.04	> 0.001
Cell Cycle (115566)	319	2.64	> 0.001	306	3.20	> 0.001
Cell Cycle, Mitotic (152)	265	2.20	0.007	266	3.53	> 0.001
Cross-Presentation of Soluble Exogenous Antigens (Endosomes) (111056)	44	-2.30	0.017	43	4.14	> 0.001
Cytokine Signaling in Immune System (75790)	203	4.72	> 0.001	198	3.08	> 0.001
DNA Replication (383)	162	2.33	0.003	163	3.40	> 0.001
Formation of the Ternary Complex, and the 43S Complex (1079)	33	4.33	> 0.001	32	-5.11	> 0.001
GPCR Downstream Signaling (19184)	233	2.51	> 0.001	205	-2.39	0.003
Immunoregulatory Interactions, Lymphoid, non-Lymphoid Cells (11152)	34	3.62	> 0.001	36	3.58	> 0.001
Influenza Life Cycle (6145)	125	5.25	> 0.001	125	-5.76	> 0.001
Influenza Viral RNA Transcription and Replication (6152)	93	5.99	> 0.001	92	-7.14	> 0.001
Hemostasis	335	3.10	> 0.001	282	2.35	0.002
Innate Immune System (6802)	158	2.78	> 0.001	159	3.06	> 0.001
Interferon Alpha/Beta Signaling (25162)	37	3.69	> 0.001	36	4.09	> 0.001
Interferon Gamma Signaling (25078)	40	2.69	> 0.001	38	3.46	> 0.001
Interferon Signaling (25229)	112	3.94	> 0.001	108	3.56	> 0.001
Metabolism of mRNA (20605)	193	3.81	> 0.001	192	-3.93	> 0.001
Metabolism of RNA (21257)	234	3.13	> 0.001	232	-3.64	> 0.001
Nonsense-Mediated Decay, Enhanced by the Exon Junction Complex (75822)	98	6.31	> 0.001	95	-6.70	> 0.001
Peptide Chain Elongation (1404)	79	7.10	> 0.001	76	-8.11	> 0.001
Regulation of Ornithine Decarboxylase (ODC) (13565)	48	-2.41	0.013	47	4.16	> 0.001
Respiratory Electron Transport (22393)	60	-4.01	> 0.001	60	3.50	> 0.001
Respiratory Electon Transport, ATP Synthesis by Chemiosmotic Coupling, and Heat Production (6305)	72	-4.04	> 0.001	73	3.99	> 0.001
Rig-I/MDA5 Mediated Induction of IFN Alpha/Beta Pathways (25359)	48	2.69	> 0.001	47	2.62	> 0.001
SCF-beta-TrCP Mediated Degradation of Emi1 (6821)	48	-2.23	0.021	48	3.68	> 0.001
SRP-Dependent Cotranslational Protein Targeting to Membrane (115902)	100	5.73	> 0.001	98	-7.04	> 0.001
TCA Cycle and Respiratory Electron Transport (111083)	105	-3.72	> 0.001	104	3.53	> 0.001
Translation (1014)	122	5.54	> 0.001	119	-6.67	> 0.001
VIF-Mediated Degradation of APOBEC3G (9453)	48	-2.35	0.015	48	4.19	> 0.001

Genes: # of genes in the gene set represented in the expression list; NES: Normalized Enrichment Score; FDR: The gene set enrichment FDR-adjusted p-value.

Table 4.8 Gene sets with enriched gene members in opposite direction (BCG versus control groups) of enrichment between microglia and peritoneal macrophages

Collection	Gene Set	Direction of Enrichment in Microglia (BCG vs Ctrl)	Direction of Enrichment in Peritoneal Macrophages (BCG vs Ctrl)
GO MF	Structural Constituent of Ribosome (GO:0003735)	Over	Under
GO MF	Structural Molecule Activity (GO:0005198)	Over	Under
GO BP	Translation (GO:0006412)	Over	Under
KEGG	Alzheimer's Disease (05010)	Under	Over
KEGG	Huntington's Disease (05016)	Under	Over
KEGG	Oxidative Phosphorylation (00190)	Under	Over
KEGG	Parkinson's Disease (05012)	Under	Over
REACTOME	3'-UTR-Mediated Translational Regulation (1762)	Over	Under
REACTOME	Activation of mRNA Upon Binding Cap-Binding Complex, eIFs, and Subsequent Binding to 43S (1258)	Over	Under
REACTOME	Autodegradation of Cdh1 by Cdh1:APC/C (6785)	Under	Over
REACTOME	Cross-Presentation of Soluble Exogenous Antigens (Endosomes) (111056)	Under	Over
REACTOME	Formation of the Ternary Complex, and Subsequently, the 43S Complex (1079)	Over	Under
REACTOME	GPCR Downstream Signaling (19184)	Over	Under
REACTOME	Influenza Life Cycle (6145)	Over	Under
REACTOME	Influenza Viral RNA Transcription and Replication (6152)	Over	Under
REACTOME	Metabolism of mRNA (20605)	Over	Under
REACTOME	Metabolism of RNA (21257)	Over	Under
REACTOME	Nonsense-Mediated Decay Enhanced by the Exon Junction Complex (EJC) (75822)	Over	Under
REACTOME	Peptide Chain Elongation (1404)	Over	Under
REACTOME	Regulation of Ornithine Decarboxylase (ODC) (13565)	Under	Over
REACTOME	Respiratory Electron Transport (22393)	Under	Over
REACTOME	Respiratory Electon Transport, ATP Synthesis by Chemiosmotic Coupling, and Heat Production by Uncoupling Proteins (6305)	Under	Over
REACTOME	SCF-beta-TrCP Mediated Degradation of Emi1 (6821)	Under	Over
REACTOME	SRP-Dependent Cotranslational Protein Targeting to Membrane (115902)	Over	Under
REACTOME	TCA Cycle and Respiratory Electron Transport (111083)	Under	Over
REACTOME	Translation (1014)	Over	Under
REACTOME	VIF-Mediated Degradation of APOBEC3G (9453)	Under	Over

Table 4.9 Overlap Coefficient among enriched gene sets with enriching gene members over-expressed (BCG treatment versus control) in microglia, under-expressed in peritoneal macrophages

	GO			REACTOME										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
1	1.00	0.98	0.78	1.00	0.77	0.75	0	1.00	1.00	0.85	0.98	1.00	1.00	0.98	1.00
2		1.00	0.78	0.95	0.96	0.93	0	0.95	0.95	0.83	0.95	0.95	0.95	0.95	0.95
3			1.00	0.82	0.35	0.36	0	0.84	0.80	0.67	0.84	0.82	0.80	0.78	0.82
4				1.00	1.00	0.96	0	0.99	1.00	0.97	0.97	0.99	1.00	1.00	1.00
5					1.00	1.00	0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
6						1.00	0	1.00	0.96	0.96	0.96	0.96	0.96	0.96	0.96
7							1.00	0	0	0	0	0	0	0	0
8								1.00	1.00	0.97	0.92	1.00	1.00	1.00	0.96
9									1.00	0.97	0.97	1.00	1.00	1.00	1.00
10										1.00	1.00	0.97	0.97	0.97	0.97
11											1.00	0.97	0.97	1.00	0.95
12												1.00	1.00	1.00	0.99
13													1.00	1.00	1.00
14														1.00	1.00
15															1.00

Gene Sets
Gene Ontology
1) Structural Constituent of Ribosome (GO MF, GO:0003735);
2) Structural Molecule Activity (GO MF, GO:0005198);
3) Translation (GO BP, GO:0006412).
Reactome
4) 3'-UTR-Mediated Translational Regulation (REACT_1762);
5) Activation of mRNA Upon Binding of the Cap-Binding Complex and eIFs, and Subsequently, Binding to 43S (REACT_1258);
6) Formation of the Ternary Complex, and Subsequently, the 43S Complex (REACT_1079);
7) GPCR Downstream Signaling (REACT_14797);
8) Influenza Life Cycle (REACT_6145);
9) Influenza Viral RNA Transcription and Replication (REACT_6152);
10) Metabolism of mRNA (REACT_20605);
11) Metabolism of RNA (REACT_21257);
12) Nonsense Mediated Decay (NMD) Enhanced by the Exon Junction Complex (EJC) (REACT_75822);
13) Peptide Chain Elongation (REACT_1404);
14) SRP-Dependent Cotranslational Protein Targeting to Membrane (REACT_115902);
15) Translation (REACT_1014).

Table 4.10 Overlap Coefficient among enriched gene sets with enriching gene members under-expressed (BCG treatment versus control) in microglia, over-expressed in peritoneal macrophages

	KEGG				REACTOME							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	1.00	0.85	0.67	0.73	0	0	0	0.81	0.71	0	0.75	0
2		1.00	0.80	0.95	0	0	0	0.86	0.86	0	0.75	0
3			1.00	0.82	0	0	0	0.76	0.82	0	0.80	0
4				1.00	0	0	0	0.76	0.79	0	0.79	0
5					1.00	1.00	0.88	0	0	0.97	0	0.91
6						1.00	1.00	0	0	0.97	0	1.00
7							1.00	0	0	0.97	0	0.94
8								1.00	1.00	0	1	0
9									1.00	0	1	0
10										1.00	0	0.97
11											1.00	0
12												1.00

Gene Sets
KEGG
1) Alzheimer's Disease (hsa05010)
2) Huntington's Disease (hsa05016)
3) Oxidative Phosphorylation (hsa00190)
4) Parkinson's Disease (hsa05012)
REACTOME
5) Autodegradation of Cdh1 by Cdh1:APC/C (REACT_6785)
6) Cross-Presentation of Soluble Exogenous Antigens (Endosomes) (REACT_111056)
7) Regulation of Ornithine Decarboxylase (ODC) (REACT_13565)
8) Respiratory Electron Transport (REACT_22393)
9) Respiratory Electon Transport, ATP Synthesis by Chemiosmotic Coupling, and Heat Production by Uncoupling Proteins (REACT_6305)
10) SCF-beta-TrCP Mediated Degradation of Emi1 (REACT_6821)
11) TCA Cycle and Respiratory Electron Transport (REACT_111083)
12) VIF-Mediated Degradation of APOBEC3G (REACT_9453)