2017

Using RNA Seq to Define Genes and Pathways Affecting Salmonella Vaccine Response in Beef Cattle

Sara Awrahman
South Dakota State University

USING RNA SEQ TO DEFINE GENES AND PATHWAYS AFFECTING SALMONELLA VACCINE RESPONSE IN BEEF CATTLE

BY

SARA AWRAHMAN

A thesis submitted in partial fulfillment of the requirements for the

Master of Science□

Major in Animal Science

South Dakota State University

2017

USING RNA SEQ TO DEFINE GENES AND PATHWAYS AFFECTING SALMONELLA VACCINE RESPONSE IN BEEF CATTLE

This thesis is approved as a creditable and independent investigation by a candidate for the Master of Science in Animal Science degree and is acceptable for meeting the thesis requirements for this degree. Acceptance of this does not imply that the conclusions reached by the candidate are necessarily the conclusions of the major department.

Michael Gonda, Ph.D.
Thesis Advisor Date

Joseph Cassady, Ph.D.
Head, Department of Animal Science Date

Dean, Graduate School Date

ACKNOWLEDGEMENTS

The completion of this thesis is the product of much assistance from my advisor, classmates, and my family.

First, I would like to thank my advisor, Dr. Michael Gonda, for all the guidance and direction that he has provided. His patience has given me a drive to succeed even when times were tough. He has helped me gain a fundamental understanding of my field and has held my hands during lab work and thesis writing.

I want to extend my gratitude to all committee members, Dr. Benoit St Pierre, Dr. Xijin Ge, and Dr. Wanlong Li, for their support and time.

I want to thank my classmate, Stephanie Perkins, for helping me with RNA extraction. I could not have done these extractions without her.

I would like to thank my husband, Miran, for always being there for me. He was my copilot, from babysitting our son to comforting me when my work wasn't advancing at times.

My family and siblings have always been there for me. They have encouraged me to keep going and to see the finish line. They have helped me financially and emotionally.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

USING RNA SEQ TO DEFINE GENES AND PATHWAYS AFFECTING SALMONELLA VACCINE RESPONSE IN BEEF CATTLE

SARA AWRAHMAN

2017

A *Salmonella* siderophore receptor and porin (SRP) proteins vaccine is an intervention strategy to control *Salmonella* burden in cattle. The vaccine works by stimulating immunity to produce antibodies against bacterial SRPs as siderophore molecules play a major role in transporting iron to bacteria. Blocking iron uptake system by antibodies causes death of bacterial cells. The vaccine can also be useful to protect humans against salmonellosis, which causes high rates of illness and death annually, by reducing shedding of *Salmonella* in the feces of cattle. Though other researchers have evaluated efficacy and immune response of this *Salmonella* SRP vaccine, genes and pathways affected by vaccination have not been investigated. Therefore, the aim of this study was to identify differentially expressed genes and pathways in bovine whole blood after vaccination with *Salmonella* SRPs by RNA sequencing. Five Angus cattle were vaccinated with the *Salmonella* SRP vaccine, and blood samples were collected at first day of vaccination, day 21 post-vaccination (time of booster vaccination) and finally day 48 post-vaccination. To perform RNA seq, total RNA was extracted from blood samples using the PAXgene^{TM} Blood RNA kit. Library samples were prepared using the Illumina TruSeq Stranded Total RNA Library PrepKit and sequenced on an Illumina HiSeq sequencer. After processing raw sequencing reads, filtered data was aligned to the bovine

reference genome (Bta_4.6.1). Cuffdiff was used for pairwise differential expression analysis. The transcript abundance was normalized in Fragments Per Kilobase of transcript per Million mapped reads (FPKM). FDR adjusted p-values (q-value<0.05) were used to identify differentially expressed genes between 1) day 21 and day 0, and 2) day 48 and day 0 of vaccination. Of the 848 differentially expressed (DE) genes at day 21 post-vaccination, one gene was increased in abundance and 26 genes were decreased in abundance with a fold-change >2. Of the 1155 DE genes at day 48 post-vaccination, 20 genes were increased in abundance and 39 genes were decreased in abundance with a fold-change >2. DAVID bioinformatics was used to annotate a list of differentially expressed genes to their correlated GO terms and KEGG pathways using Bos Taurus as a background. Most decreased in abundance genes were annotated to biosynthetic processes of heme and protoporphyrinogen IX. FoxO signaling pathway, AMPK signaling pathway, and porphyrin and chlorophyll metabolism were found significantly decreased in abundance at day 21 post-vaccination. Real-time RT-PCR was used to validate RNA sequencing results for the day 48 to day 0 comparison. All three genes tested (HMB, ALAS2, and ATPIF1) were decreased in abundance at day 48 (P<0.05), confirming our RNA sequencing observations. Our analysis suggests that *Salmonella* SRP vaccination decreased in abundance pathways involving heme (iron) metabolism in bovine blood.

Chapter 1

LITERATURE REVIEW

Introduction

Vaccination is delivery of a foreign antigen into a body in order to stimulate its immunity by producing antibodies and memory cells to protect the body from actual disease. Livestock species are considered a potential source for foodborne *Salmonella* that can be transmitted from animals to humans through direct or indirect contact or consumption of contaminated meat. *Salmonella* are estimated to cause about 35% of hospitalizations and 28% of deaths by foodborne pathogens in the United States each year (Scallan et al., 2011). Therefore, attempts have been made to develop pre-harvest interventions that aid in the control of *Salmonella* in farm animals. For example, a siderophore based (SRP) vaccine is an intervention designed to induce animal immune response against bacterial siderophore proteins with the aim of reducing *Salmonella* populations. The efficacy of this vaccine has been evaluated. Some studies found the vaccine to be ineffective at reducing fecal shedding of *Salmonella* (Heider et al., 2008: Dodd et al., 2011; Cernicchiaro et al., 2016) but at least one study observed a reduction in fecal shedding (Loneragan et al., 2012). The reason for these conflicting results is unknown.

Sequencing of the transcriptome by next generation RNA sequencing ("RNA seq") enables scientists to identify differentially expressed transcripts and pathways correlated with different conditions without prior knowledge of gene sequences. In general, any NGS platform can be used to perform RNA seq though template preparation and data analysis steps may vary. Therefore, it is important for researchers to take into account the advantages and disadvantages of each system in terms of reagent cost, run time, read length, and other factors affecting cost, efficiency, and quantity and quality of

sequencing reads. RNA seq offers a wide range of applications such as discovery of novel transcripts and identification of genetic variants. For example, transcriptome profiling of bovine peripheral blood mononuclear cells by RNA seq has resulted in the discovery of putative biomarkers associated with vaccine response and disease resistance (Bhuju et al., 2012).

Studying links between vaccination and differentially expressed genes in beef cattle will provide greater insights into the molecular and biological mechanisms of immune response. This information may contribute to the design of new vaccines. Also, mRNA transcriptome studies could allow for identification of biomarkers and pathways associated with immune response to vaccines in beef cattle. This literature review will include an overview of *Salmonella* subspecies in cattle and the interventions that can be taken to reduce its prevalence as well as a general background on RNA seq, next generation sequencing platforms, and applications of RNA seq in livestock species.

Salmonella spp

Salmonella (nontyphoidal) species are common worldwide foodborne bacteria with over 2500 distinctive serotypes (WHO, 2013) based upon the structures of surface antigens (CDC, 2015). Nevertheless, some of these serotypes are specific to one host (*S. Dublin* in cattle and *S. Choleraesuis* in swine) while other serotypes can be present in a wide variety of hosts (*S. Typhimurium* and *S. Enteritidis*). All serotypes can be pathogenic to humans (WHO, 2013). *Salmonella* can survive outside the host and also colonize the gastrointestinal tract of domestic animals like cattle and poultry (Toldrá, 2009) without obvious signs of disease (Sanchez et al., 2002). Transmission of *Salmonella* to humans usually occurs by the ingestion of raw foods and animal products contaminated with manure, or by contact with infected animals and their environment (Nester et al., 2004). Common manifestations of salmonellosis are abdominal pain, fever, diarrhea, and occasionally vomiting. These illnesses can be severe in infants, elderly people and immunocompromised individuals (WHO, 2013).

The disease and economic burden of *Salmonella* are more substantial than other foodborne bacteria due to its high morbidity and mortality rate (Scallan et al., 2011). An average of 93.8 million cases of non-typhoidal *Salmonella* infections are estimated to occur each year, including 155,000 deaths (Majowicz et al., 2010). Non-typhoidal *Salmonella* is a leading cause of foodborne illness in the United States, resulting in approximately one million illnesses annually (Scallan et al., 2011). Among 15 foodborne pathogens, *Salmonella* was the costliest disease examined with an estimated loss of 3.7 billion US dollars each year (Hoffmann et al., 2015). Since cattle are primary reservoirs

for *Salmonella enterica* subspecies, people can become ill after consuming meat contaminated with *Salmonella.* To reduce and control the prevalence of *Salmonella* in cattle, pre-harvest interventions such as antibiotics and vaccines have been used. However, the effectiveness of each intervention can vary.

Though antibiotics can be beneficial to treat and control salmonellosis in animals, their use has been restricted due to increasing resistance of serotypes to antibiotics (Barrow and Methner, 2013). These multidrug-resistant serotypes could be a major problem to public health because they can be transmitted from animals to humans through consumption of contaminated foods (DiMarzio et al., 2013). Furthermore, some antibiotics used in livestock are similar or identical to antibiotics used by humans (McEwen and Fedorka-Cray, 2002) although many antibiotics used in human medicine are too expensive to use in livestock and are not approved for use in animals. *Salmonella typhimurium* (Definitive Type 104), in particular, is resistant to a number of antibiotics, which includes ampicillin, chloramphenicol, streptomycin, sulphonamides, and tetracycline (Mather et al., 2013). However, this article also presents evidence that transmission of resistant DT 104 between animals and humans was limited, suggesting that livestock may not be important reservoirs of antibiotic-resistant bacteria in humans. More research into transmission of antibiotic-resistant bacteria to humans and alternatives to the use of antibiotics for control of *Salmonella* are necessary.

Vaccination is an alternate route of intervention to prevent and control *Salmonella*. Two groups of vaccines are available against *Salmonella*: inactivated and live-attenuated vaccines. Inactivated vaccines often stimulate weaker immune responses against *Salmonella* relative to live-attenuated vaccines because inactivated vaccines use

killed pathogens as the antigen. Thus, inactivated vaccines are incapable of mimicking real infections (nonimmunogenic) and elicit only humoral immunity (Elgert, 2009). Therefore, these vaccines can be ineffective at protecting animals against *Salmonella* serotypes. Subunit vaccines are a type of inactivated vaccine which is composed of antigenic peptide fragments of a pathogen. These peptides evoke an immune response. A subunit vaccine has been developed that is labeled to reduce the shedding of *Salmonella* in cattle; this vaccine may also provide cross-protection against more than one *Salmonella* strain (Heider et al., 2008). In contrast to inactivated vaccines, live attenuated vaccines are capable of triggering stronger immune responses against *Salmonella* and stimulate both cell-mediated and antibody-mediated (humoral) immunity (Barrow and Methner, 2013) because the live-vaccines include a weakened form of disease-causing agents (Elgert, 2009). For instance, in White Leghorn chickens, a live-attenuated vaccine was more effective than an inactivated vaccine in the induction of cell-mediated immunity because of the increased proliferation of lymphocytes against *Salmonella enteritidis* antigen among chickens (Babu et al., 2003).

The *Salmonella* Newport Bacterial Extract vaccine manufactured by Epitopix targets Siderophore Receptors and Porin (SRP) proteins derived from the outer membrane of gram-negative *Salmonella* bacteria (Hermesch et al., 2008). These cell membrane proteins transport iron, which is a vital nutrient for bacterial survival and growth, into *Salmonella* bacteria. The vaccine is capable of activating the immune system to create antibodies that block the SRP proteins (iron transport system). Ultimately, the bacterial cell dies from lack of iron (Thomson et al., 2009).

Previous studies on *Salmonella* SRP vaccination have shown conflicting results on the efficacy of this vaccine as defined by reduction in shedding of *Salmonella* in the feces of beef and dairy cattle. Numerous experimental studies have indicated no significant difference in fecal shedding or prevalence of *Salmonella* between vaccine and control groups in dairy and feedlot cattle (Hermesch et al., 2008; Heider et al., 2008; Dodd et al., 2011; Cernicchiaro et al., 2016), whereas the result of an observational study on dairy cull cows indicated a difference in fecal prevalence of *Salmonella* between dairy herds that used *Salmonella* SRP vaccine and dairy herds that did not (Loneragan et al., 2012).

Dodd and colleagues (2011) carried out a study to ascertain how *Salmonella* SRP vaccination affects beef cattle performance, health, and fecal shedding of *Salmonella* bacteria. A total of 1,591 feedlot cattle with an average weight of 256 kg were enrolled in the experiment and were allocated to two treatment groups using 20 pens (10 pens per treatment group). One group (n=795) was inoculated with 2mL of *Salmonella* Newport SRP vaccine while a second group (n=796) was administrated a 2mL placebo. Cattle were revaccinated 21 days later with a booster shot. Feces were collected from each pen floor on day 0, 60, 120, and the day of harvest (day is relative to time of initial vaccination). The fecal samples were cultured for *Salmonella* detection. Weight gain, feed intake, and data on cattle health including morbidity and mortality were recorded. Data analysis of pen fecal prevalence were performed using linear mixed models. No statistically significant effect of the SRP vaccine on the performance and health of treatment (vaccine) vs. control groups, and also on the fecal shedding of *Salmonella* in vaccinated (10.2%) and control (10.9%) cattle were found. The authors concluded that

immunization of feedlot cattle with *Salmonella* SRP vaccine did not have an effect on the fecal shedding of *Salmonella* or on health and performance.

Another study by Hermesch and colleagues (2008) examined the effects of *Salmonella* SRP vaccination on milk yield, somatic cell count (SCC), and fecal shedding of *Salmonella* bacteria in female dairy cattle. A total of 180 non-lactating Holstein cows and pregnant Holstein heifers were enrolled without apparent signs of salmonellosis. Cattle were randomly assigned to vaccinated or control groups. The initial dose of vaccine or control solution was administered to cattle approximately 45 to 60 days prior the expected parturition. The booster dose was inoculated 14 to 21 days prior to parturition. Blood and fecal samples for measurement of antibody response to *Salmonella* SRP vaccine and for detection of *Salmonella,* respectively, were collected on the day of initial inoculation, at 7 to 14 days of lactation, and at 28 to 35 days of lactation. Milk samples were collected to estimate somatic cell count (SCC). No statistically significant difference was found between treatment and control groups for milk production. The SCC in milk samples of vaccinated cattle was significantly lower than control cattle at 30 to 60 days of lactation. However, no significant difference in SCC was detected between groups at other time points. Antibody titers of vaccinated cattle were higher than that of control cattle on days 7 to 14 of lactation and days 28 to 35 of lactation. No difference in prevalence of fecal shedding of *Salmonella* was detected between the two groups. Nonetheless, a significant decrease in the prevalence of *Salmonella* was found in both vaccinated and control groups at days 7-14 of lactation. The authors concluded immunization of healthy dairy cattle with the *Salmonella* Newport SRP protein vaccine

had no effect on reducing shedding of *Salmonella* bacteria even though the *Salmonella* SRP vaccine can decrease SCC and elicit an antibody response.

Unlike the two previous findings, Loneragan et al. (2012) suggested that the use of *Salmonella* SRP vaccine at herd-level can help reduce the prevalence of *Salmonella* in dairy farms. Though the main objective of their study was to characterize the epidemiology of *Salmonella* harbored by dairy cattle culled from farms in the Texas High Plains, they also estimated the prevalence and concentration of *Salmonella* in feces. In total nine dairies were enrolled, of which three dairies vaccinated their entire herd with the *Salmonella* SRP vaccine while the remainder did not. The *Salmonella* SRP vaccine was correlated (P=0.05) with reduced prevalence of *Salmonella*. The prevalence was 8% among farms that used this vaccine while 36.8% among farms that didn't use this vaccine. The authors concluded that the SRP vaccine might be useful for decreasing *Salmonella* prevalence. However, other factors besides vaccination might play a role in reducing *Salmonella* prevalence like biosecurity measures. *Salmonella* is still a serious problem for both animals and humans even with available pre-harvest interventions. Therefore, further research on different interventions can enable scientists to understand how to better control the spread of this bacteria. Genetic selection has potential to reduce salmonella prevalence by selecting cattle that elicit a stronger, more protective immune response after vaccination. It is important to mention that no study to our knowledge has investigated the genetic response of cattle to *Salmonella* SRP vaccination.

Overview of RNA Sequencing

RNA-seq is an application of next-generation sequencing (NGS) methods to study ribonucleic acids (RNA molecules). This high-throughput sequencing technique is used for quantification of the transcriptome in a biological sample and for analysis of transcriptional changes between different experimental conditions in order to determine which genes or pathways may be affected by treatments. Besides differential gene expression (DGE), RNA-seq can be used for annotation and discovery of transcripts and also for identification of polyadenylation and alternative splicing. With this novel technique, quantitative and qualitative information can be obtained. Quantitative data is used for measuring differentially expressed genes, alternative transcription start sites (TSS), polyadenylation sites and alternative splicing under different conditions. Qualitative data allows for the identification of transcripts, transcription start sites, 5' and 3' untranslated regions (UTR), and exon-intron boundaries (Kaundal, 2016).

Transcriptomics aims to identify transcript function (small RNAs, non-coding RNAs, and messenger RNAs), to ascertain a gene's transcriptional structure (splicing mechanisms, start sites and post-transcriptional modifications) and to quantify changes in the level of each expressed transcript in a given sample (Wang et al., 2009). For example, studies have used RNA-seq to ascertain patterns of alternative splicing (Wang et al., 2008; Tremblay et al., 2016), levels of steady-state mRNA molecules (Bremer and Moyes, 2014), and discovery of novel non-coding RNAs and mRNAs (Guttman et al., 2010; Djebali et al., 2012). Some applications related to livestock include the discovery of genetic variants (SNPs) (Cánovas et al., 2010), identification of small RNAs (Jin et al., 2009), long noncoding RNAs (Billerey et al., 2014), and alternative splicing events

(Wang et al., 2016). Since RNA-seq has become a very common research tool, it is important to understand how this powerful technique works before the initiation of a RNA-seq study.

One of the most common applications of RNA-Seq is to investigate differential gene expression (DGE) (Soneson et al., 2016). With DGE, up-regulated and down-regulated genes can be identified after application of a specific experimental treatment (Wit et al., 2012). For DGE, polyadenylated mRNA is isolated from total RNAs via oligo deoxythymidine (dT) beads (Wang et al., 2009; Wilhelm and Landry, 2009; Corney and Basturea, 2013). Ribosomal RNA (rRNA) constitutes about 90% of cellular RNAs (Wilhelm and Landry, 2009; Zeng and Mortazavi, 2012). The rRNA is removed to allow increased sequencing coverage of the lower abundance mRNAs. The mRNA is converted to a library of double-stranded complementary DNA (cDNA) fragments with adapters ligated to both ends of each cDNA molecule (Wang et al., 2009; Chu and Corey, 2012; Wolf, 2013). After amplification of the cDNA library, cDNA fragments are deep sequenced to produce short sequences (Wang et al., 2009; Marguerat and Bähler, 2010; Corney and Basturea, 2013) from either one direction (single end) or both directions (pair end). These short sequences are called reads. Sequencing platforms generate digital data that include millions of short reads. The length of the reads is usually 30 to 400 base pairs and differs among sequencing platforms (Wang et al., 2009). These reads are directly mapped to a reference genome (Wang et al., 2009; Marguerat and Bähler, 2010; Wolf, 2013; Finotello and Di Camillo, 2015) for transcript identification and DGE (Finotello and Di Camillo, 2015). The estimation of expression levels of a transcript is transformed to number of Reads Per Kilobase per Million mapped reads (RPKM) (Mortazavi et al.,

2008) or to number of Fragments Per Kilobase of transcript per Million fragments mapped reads (FPKM) (Trapnell et al., 2010; Griffith et al., 2015). The RPKM and FPKM are used for DGE to prevent biased read distributions. A biased distribution may result from differences in sequencing depth among samples or differences in the length of RNA molecules in a sample. In principle, any NGS platform can be utilized to perform RNA seq (Finotello and Di Camillo, 2015; Chu and Corey, 2012). At present, the major NGS platforms are Illumina, Applied Biosystems (SOLiD), Roche 454, Ion Torrent, Pacific Biosciences, and Oxford Nanopore (Korpelainen et al., 2015; Levy and Myers, 2016). Each platform utilizes distinct sequencing chemistry and detection techniques (Meldrum et al., 2011).

Next Generation Sequencing Platforms

The NGS technologies above involve numerous steps that can be generally categorized as library construction, sequencing, imaging, and lastly data analysis. Library (or template) preparation for the sequencing reaction is classified as either amplification dependent or independent. In the amplification dependent platforms, after fragmentation and adapter ligation of cDNA, the DNA fragments are either attached to a flow cell, beads or ion surfaces. The bound DNA fragments are then subjected to clonal PCR amplification (Metzker, 2010), which is performed by emulsion PCR (Dressman et al., 2003) or solid-phase bridge PCR (Fedurco et al., 2006). On the other hand, in the amplification independent platforms, the DNA is prepared by either the Single Molecule Real Time SMRTbell template technique or leader-hairpin template technique (Goodwin et al., 2016). With SMRTbell template, DNA is fragmented, and then two hairpin adapters are added to each end of the DNA. The hairpin adapters allow for continuous circular sequencing of the DNA. With leader-hairpin template technology used by Nanopore, after fragmentation, two distinct adapters (hairpin and leader) are added to each end of a DNA molecule. The hairpin adapter allows bidirectional sequencing while the leader interacts with a pore and motor protein to direct cDNA during sequencing. A single DNA molecule is then immobilized to either a Zero-Mode Waveguides (ZMW) well (Ambardar et al., 2016) or MinION flowcell (Lu et al., 2016) where the sequencing occurs.

Additionally, NGS platforms use different strategies for sequencing and signal detection: cyclic reversible termination (Ju et al., 2006), single nucleotide addition (Metzker, 2010), sequencing by ligation (Garrido-Cardenas et al., 2017), single molecule

real time sequencing (Rhoads and Au, 2015) and single molecule nanopore sequencing (Levy and Myers, 2016). One major difference among these sequencing approaches aside from nanopore sequencing (no enzyme) is that either a DNA polymerase or ligase enzyme is required for nucleotide incorporation (Ambardar et al., 2016). Another difference is that signal detection is directly or indirectly recorded. Real time PCR (Pacific Biosciences) or electric current (Oxford Nanopore) is used for direct detection of signals. With indirect detection, signals are recorded at the end of each cycle by either a Charge Coupled Device (CCD) camera (Illumina, SOLiD, and Roche 454) or field-effect transistor sensors (Ion Torrent) (Levy and Myers, 2016). Sequencing platforms differ in data format and output, read length, error rate, and run time. As a result of these differences, a diversity of bioinformatics approaches is needed to handle data for appropriate quality control, alignment, assembly, and processing (Zhang et al., 2011; Egan et al., 2012).

Three widespread deep-sequencing platforms that have been used by research laboratories, sequencing centers, and core facilities are Illumina, Roche 454 and SOLiD (Lister et al., 2009; Bai et al., 2012). The Illumina is the most commonly used technology and utilizes solid-phase amplification for cluster generation (Buermans and Den Dunnen, 2014) and a cyclic reversible terminator (CRT) for sequencing (Kircher and Kelso, 2010). Clonally amplified clusters are produced by multiple amplification cycles of a single molecule DNA template (Voelkerding et al., 2009). The DNA fragments with ligated adapters on both ends are loaded onto a flow cell coated with immobilized oligonucleotides. These oligonucleotides then hybridize to the adapters attached to the DNA fragments (Hodkinson and Grice, 2015). Bridge amplification of each DNA

fragment results in the creation of a cluster that includes thousands of copies of the DNA fragment (Voelkerding et al., 2009; Liu et al., 2012). After cluster generation, the CRTs sequence the cluster of DNA fragments. The CRT sequencing uses reversible terminator (RT) nucleotides to terminate DNA synthesis temporarily after single nucleotide incorporation. These RT nucleotides are fluorescently labeled and contain a blocking group at the 3' hydroxyl (OH) molecule of the ribose sugar that allows one single base to be incorporated each cycle (Mardis, 2013). All four fluorescently labeled dNTPs (nucleotides) and polymerase are added to the flow cell and then nucleotides are added sequentially and detected by the presence of a unique fluorescent dye on each nucleotide (Corney and Basturea, 2013). With RT nucleotides, a single nucleotide is incorporated per sequencing cycle (Corney and Basturea, 2013; Ambardar et al., 2016). Nucleotide incorporation results in the production of a fluorescent signal, which is visualized by a camera. Hence, DNA base composition is identified via fluorescent dyes (Egan et al., 2012). The Illumina error rate per base across all systems is ≤ 0.1 percent (Glenn, 2011). However, each Illumina system results in different read lengths, a yield of data per run, cost, and run time (www.illumina.com/systems/sequencing-platforms.html). For instance, the HiSeq 2000 system generates 2×150 base pair reads with 600 Gb of data per run in 11 days (Quail et al., 2012), although one main concern of this system is the high computation needs (Glenn, 2011) to handle large amounts of data. The newest version of HiSeq systems 3000/4000 is improved to provide more rapid and higher performance sequencing of over 200 Gb per a day (Illumina, 2014).

The Roche 454 technology relies on emulsion PCR and pyrosequencing chemistry (Mardis, 2008; Voelkerding et al., 2009; Siqueira et al., 2012). A DNA library is

constructed by fragmentation of genomic DNA (for genome sequencing) or cDNA (for RNA-seq). These DNA fragments undergo adapter ligation where adapters are added to both ends of the DNA fragments and then separated into single-stranded DNA (ssDNA) (Siqueira et al., 2012). The emulsion amplification process is initiated by immobilization of the ssDNA library onto DNA capture beads (one fragment per bead) (Hodkinson and Grice, 2015). Beads with the attached ssDNA and PCR reagents are emulsified in oil to form a water-in-oil microreactor (Lin et al., 2008; Radford et al., 2012). Ideally, each microreactor contains one bead along with one DNA fragment. The emulsion is then subjected to amplification by which the DNA fragment is amplified to generate monoclonal DNA beads (Radford et al., 2012). Clonally amplified beads are added into wells of a picotiter plate with a set of sequencing enzymes including DNA polymerase, ATP sulfurylase, and luciferase. The pyrosequencing reaction is initiated by sequential addition of four nucleotides (dNTPs) into the picotiter plate. Nucleotide incorporation results in pyrophosphate release and then ATP sulfurylase enzyme converts pyrophosphate to adenosine triphosphate (ATP). This ATP is used by the enzyme luciferase to convert luciferin to oxyluciferin that generates light signals that are detected by a camera to identify the base incorporated into a DNA fragment (Lin et al., 2008; Voelkerding et al., 2009; Hodkinson and Grice, 2015). The major strengths of the Roche 454 technology are long read length and high-speed (Meldrum et al., 2011; Liu et al., 2012). The latest 454 GS FLX system generates approximately one million reads per run (Hodkinson and Grice, 2015) in seven hours (Jazayeri et al., 2015). The read length of this system can reach 700 bp with 99.9% accuracy (Lin et al., 2008). The main limitations of the system are the high cost of reagents per MB ($7) and run ($6200)

(Glenn, 2011) and homopolymer errors (Meldrum et al., 2011), which are a number of miscalled bases that occur because of consecutive repetitions of a single nucleotide in a sequence (Fakruddin et al., 2012). Unfortunately, as of this writing the company has ceased operations but one may still find applications of this sequencing technology in the literature.

The Applied Biosystems SOLiD (Sequencing by Oligonucleotide Ligation Detection) technology uses emulsion PCR that is similar to the Roche 454 technology (Dhiman et al., 2009) and di-base probes to perform sequencing by ligation (Roy et al., 2016). The library preparation involves fragmentation of genomic DNA or cDNA by sonication, followed by ligation of adapters to both ends of these fragments (Yegnasubramanian, 2013). A water in oil emulsion PCR is performed to amplify the DNA fragments attached to polystyrene beads (Simon et al., 2009). The beads containing clonally amplified DNA fragments (template) are deposited onto a glass slide for sequencing (Dhiman et al., 2009). The ligase-mediated sequencing starts by annealing a complementary sequencing primer to the adapters on the amplified fragments (Mardis, 2008; Daniels et al., 2012), and then a pool of four di-base probes that are fluorescently labeled are added into the flow cell, where they compete for ligation to the sequencing primers (Vogl et al., 2012). The probe is 8 nucleotides in length (Zhang et al., 2011), and consists of two known bases and six degenerate bases (Ambardar et al., 2016). Each probe is linked with one of four fluorescent labels at its 5' end and contains ligation and cleavage sites (Mardis, 2008; Voelkerding et al., 2009). The ligation site is at the 3' end of the di-base probe that hybridizes to the DNA template-primer and DNA ligase immobilizes the complex. The cleavage site is between the 5th and 6th base, which is the

position where a cleavage agent (e.g. silver) cleaves the last three bases and fluorescent dye from the probe. The 5' end of unexpended fragments is capped with phosphatase treatment. Phosphatase prevents any further ligation to maintain sequencing cycles (Goodwin et al., 2016). Following each ligation cycle, the fluorescent reporter is removed from the ligated probes, and the fluorescent signal is detected to determine the bases incorporated into the target sequence. A benefit of the SOLiD technology is high accuracy due to the two base sequencing approach (Liu et al., 2012). However, the limitation of the SOLiD technology is short read assembly (Hodkinson and Grice, 2015). The SOLiD 5500xl system generates 30 Gb per run with 99.99% accuracy and 85 bp read length in seven days (Liu et al., 2012).

Application of RNA-seq in livestock

In recent years livestock including cattle, swine, poultry, sheep, and goats have undergone considerable genome annotation by RNA-seq. Analysis of the livestock transcriptome is important for understanding gene function and improving livestock health and performance. Nowadays, researchers can sequence the whole transcriptome at a relatively low price in a short time. A major focus of the discipline of animal breeding and genetics is the selection of genetically superior animals for economically relevant traits. A genetic tool like RNA-seq may make identifying genetically superior animals more effective and less challenging (Bai et al., 2012). Below are some common applications that are used to investigate the transcriptome of livestock with RNA-seq. What follows is not a comprehensive assessment of livestock studies using RNA-seq. Instead, the focus is on how RNA-seq can be applied to answer different research questions in the animal sciences.

Quantifying gene expression has been the most common RNA seq application. Many studies have compared the abundance of transcripts between experimental treatments and different breeds, production levels, and time courses. For example, the first study of marbling formation in snow dragon beef cattle (Chinese cattle breed) was performed using RNA seq; transcripts from adult beef *longissimus dorsi* (LD) muscle were sequenced in these cattle (Chen et al., 2015). The muscle samples were collected from animals (same age, genetic background, and under similar conditions) immediately after slaughter. These samples were scored for marbling and cattle with high and low marbling scores were identified. Approximately 13 Gb raw paired-end reads were generated. Among 16,020 analyzed genes, 749 genes showed differential expression

between the two groups (383 genes up-regulated in the low marbling group and 366 up-regulated in the high marbling group) and were identified as putative candidates linked to marbling formation in LD muscle. This experiment is a good example of how RNA-seq can be used to identify transcriptomic differences between animals with differing levels of production.

In another study, transcript differences were compared between four different tissues in the *longissimus dorsi* (LD) of adult cattle (Lee et al., 2014). The four tissues were muscle, intramuscular adipose (IMA), subcutaneous adipose (SUA), and omental adipose (OMA). The tissue samples were collected from the LD section of nine Hanwoo cattle (3 bulls, 3 steers, and 3 cows) under same condition and diet, immediately after slaughter. The average number of raw reads from each tissue was 34.2 Mb for muscle, 35.8 Mb for IMA, 35.1 Mb for SUA, and 38.1 Mb for OMA. After data analysis, 7,282 differentially expressed genes were identified among the four tissues. They found an inverse correlation between muscle and adipose tissues in gene expression patterns, such that genes that were up-regulated in muscle were down-regulated in other tissues, and vice versa. The IMA tissue exhibited district pattern of gene expression from SUA and OMA tissues. However, gene expression pattern of SUA and OMA tissues were similar. They concluded that intramuscular adipose tissue can be essential for regulation of structure and development of the LD tissues and for communication between muscle and adipose. This manuscript is a good example of the use of RNA-seq to better understand biological pathways expressed in different tissue types.

The study of gene expression is not limited to bovine, but also includes other livestock species including chickens, pigs, horses, goats, and sheep. For instance, Wang

et al. (2014) used RNA seq to compare transcripts and signaling pathways associated with resistance to Avian Influenza Virus (AIV) infection between two chicken lines (Fayoumi and Leghorn). The two lines were genetically distinct in susceptibility to AIV infection with the Fayoumi line being more resistant than the Leghorn line. One group of chickens (n=2) from each line was inoculated with AIV (infected group) at 3 weeks of age and compared with controls (non-infected birds) from each line. Lungs and trachea samples were collected from 8 chickens (2 birds per treatment-by-line combination) at 4 day post-inoculation. Total RNA from the samples was isolated to examine the AIV replication using real-time RT-PCR, and then to construct cDNA library for transcriptome sequencing. Virus replication was detected in all infected birds while virus replication was negative in control birds. Virus titers were also significantly higher in the Leghorn line than the Fayoumi line. Sequencing generated approximately 29.6 M and 28.1 M reads for infected and non-infected Leghorn, respectively, and 28.7 M and 28.8 M reads for infected and non-infected Fayoumi, respectively. Under AIV inoculation, 117 genes were differentially expressed between the two lines. Only 6 of these genes were highly expressed in Fayoumi birds. Five of these genes were associated with the hemoglobin family, indicating a putative role of hemoglobin family genes in resistance to AIV infection. The authors concluded that up-regulation of hemoglobin family genes and down-regulation of cell adhesion molecules (CAMs) were highly correlated with disease resistance to AIV. Their results may be useful for understanding the molecular process of AIV infection and for development of new vaccines in poultry.

Another RNA seq application is the identification of long noncoding RNAs (lncRNAs) that play a regulatory role in different biological processes. These lncRNAs

are non-coding transcripts that are more than 200 nucleotides in length (Perkel, 2013). This class of transcripts was known before next-generation sequencing approaches became available. Nevertheless, the extent of their existence was not completely understood until the RNA seq technique was able to discover the diversity of lncRNA species inside living organisms (Ilott and Ponting, 2013). In a recent study, differentially expressed lncRNAs from endometrium tissues of 3 pregnant and 3 non-pregnant Duroc gilts were analyzed using RNA seq (Wang et al., 2017). Samples were collected from both groups on day 12 of pregnancy. After total RNA isolation, libraries for lncRNA sequencing were prepared. A total of 1490 lncRNAs were identified. Thirty-four differentially expressed lncRNAs were identified between pregnant and non-pregnant pigs, of which 29 lncRNAs were up-regulated in pregnant sows while the remainder was down-regulated in pregnant sows on day 12 of pregnancy. The study identified two up-regulated lncRNA loci that may be vital for embryo implantation. This experiment is a good example of the use of RNA-seq in identifying transcriptomic changes in non-coding RNA species.

RNA seq has also been used to identify small noncoding RNAs. Small RNAs are 20 to 30 nucleotides in length and include different types such as microRNAs (miRNAs) and piwi-interacting RNAs (piRNAs) (Großhans and Filipowicz, 2008). These small RNAs have an essential regulatory role in the maintenance and transcriptional gene silencing during spermatogenesis (De Mateo and Sassone-Corsi, 2014). In a recent study, Capra et al. (2017) compared miRNA and piRNA expression from semen of four Holstein bulls. For each bull, 12 semen samples were obtained and then grouped into high and low motile sperm fractions. The high and low fractions were evaluated for

sperm characteristics such as motility, viability, and acrosomal status; three replicates per fraction per bull were used to isolate total RNA. Small RNA libraries were prepared for sequencing. A total of 110,394,322 reads were obtained after trimming and 79 putative piRNAs and 83 miRNA clusters were differentially expressed between the high and low motile sperm. Among the DE miRNAs, 40 had been previously identified while the remainder had not. Among the known miRNAs, 26 were highly expressed in high motile fractions and the remainder (14) were highly expressed in low motile sperm. Some DE miRNAs in the high and low fraction targeted gene pathways related to apoptosis and alteration in spermatogenesis, and three pathways were found to be highly affected by dysregulation of these miRNAs. The authors concluded that based on specific DE miRNAs in the high versus low motile sperm, low motile sperm exhibited evidence of increased cell apoptosis and changes in cell function. It is clear that RNA-seq can be a useful application for many types of RNAs.

Single nucleotide polymorphisms (SNPs) can be detected by RNA-seq. A SNP is a single nucleotide change at a locus in the genome. SNPs can act as a molecular marker that predicts genetic merit because of their linkage disequilibrium with causative polymorphisms. In a recent study, RNA seq was applied to Milk Somatic Cells (MSCs) from two sheep breeds (Churra and Assaf used in Spain) to detect genetic variants within coding regions associated with milk production traits such as milk yield, and fat and protein percentage (Suárez-Vega et al., 2017). Milk samples were collected from eight ewes (both breeds) at four lactation time points: 10, 50, 120, and 150 days post-lambing. Milk samples were collected for the recording of milk production traits (milk somatic cells for DNA isolation only need to be collected once). After RNA isolation and library

preparation, samples were sequenced. A total of 216,637 variants from > 1,116 M paired-end reads were detected. Among these variants, 57,795 variants were identified within Quantitative Trait Loci (QTL) regions related to milk traits, 78.56% of which were known variants and 21.44% of which were novel variants. These SNPs might be used as potential markers for genetic prediction in the dairy sheep industry. This experiment is a good example of how RNA-seq can be utilized for discovery of genetic variants.

Conclusion

The above studies illustrate the power of RNA seq to identify genes and pathways associated with biological differences in livestock. Our study investigated changes in the whole blood transcriptome of beef cattle after vaccination with a commercially available *Salmonella* SRP vaccine. Our focus will be on changes in mRNA molecules in whole blood after vaccination. To date, we could not find published research that investigated how the transcriptome changes in response to this vaccine. The analysis of gene expression after vaccination could allow for discovery of newly discovered transcripts, novel functions of known transcripts, and putative gene-gene interactions.

Literature Cited

Ambardar, S., R. Gupta, D. Trakroo, R. Lal, and J. Vakhlu. 2016. High throughput sequencing: an overview of sequencing chemistry. Indian Journal of Microbiology 56(4): 394-404.

Babu, U., M. Scott, M.J. Myers, M. Okamura, D. Gaines, H.F. Yancy, H. Lillehoj, R.A. Heckert, and R.B. Raybourne. 2003. Effects of live attenuated and killed Salmonella vaccine on T-lymphocyte mediated immunity in laying hens. Veterinary Immunology and Immunopathology 91(1): 39-44.

Bai, Y., M. Sartor, and J. Cavalcoli. 2012. Current status and future perspectives for sequencing livestock genomes. Journal of Animal Science and Biotechnology 3(1): 8.

Barrow, P.A., and U. Methner. 2013. Salmonella in domestic animals. Centre for Agriculture and Biosciences International (CABI). Croydon, United Kingdom. pp. 249-250.

Bhuju, S., E. Aranday-Cortes, B. Villarreal-Ramos, Z. Xing, M. Singh, and H.M. Vordermeier. 2012. Global gene transcriptome analysis in vaccinated cattle revealed a dominant role of IL-22 for protection against bovine tuberculosis. PLoS Pathogens 8(12): e1003077.

Billerey, C., M. Boussaha, D. Esquerré, E. Rebours, A. Djari, C. Meersseman, C. Klopp, D. Gautheret, and D. Rocha. 2014. Identification of large intergenic non-coding RNAs in bovine muscle using next-generation transcriptomic sequencing. BMC Genomics 15(1): 499.

Bremer, K., and C.D. Moyes. 2014. mRNA degradation: an underestimated factor in steady-state transcript levels of cytochrome c oxidase subunits. Journal of Experimental Biology 217: 2212-2220.

Buermans, H.P.J., and J.T. Den Dunnen. 2014. Next generation sequencing technology: advances and applications. Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease 1842(10): 1932-1941.

Cánovas, A., G. Rincon, A. Islas-Trejo, S. Wickramasinghe, and J.F. Medrano. 2010. SNP discovery in the bovine milk transcriptome using RNA-Seq technology. Mammalian Genome 21(11-12): 592-598.

Capra, E., F. Turri, B. Lazzari, P. Cremonesi, T.M. Gliozzi, I. Fojadelli, A. Stella, and F. Pizzi. 2017. Small RNA sequencing of cryopreserved semen from single bull revealed altered miRNAs and piRNAs expression between high-and low-motile sperm populations. BMC Genomics 18(1): 14.

Centers for Disease Control and Prevention (CDC). 2015. Serotypes and the importance of serotyping Salmonella. https://www.cdc.gov/salmonella/reportspubs/salmonella-atlas/serotyping-importance.html (Accessed March 9, 2015).

Cernicchiaro, N., S.E. Ives, T.S. Edrington, T.G. Nagaraja, and D.G. Renter. 2016. Efficacy of a Salmonella siderophore receptor protein vaccine on fecal shedding and lymph node carriage of Salmonella in commercial feedlot cattle. Foodborne Pathogens and Disease 13(9): 517-525.

Chen, D., W. Li, M. Du, M. Wu, and B. Cao. 2015. Sequencing and characterization of divergent marbling levels in the beef cattle (longissimus dorsi muscle) transcriptome. Asian-Australasian Journal of Animal Sciences 28(2): 158-165.

Chu, Y., and D.R. Corey. 2012. RNA sequencing: platform selection, experimental design, and data interpretation. Nucleic Acid Therapeutics 22(4): 271-274.

Corney, D.C., and G.N. Basturea. 2013. RNA-seq using next generation sequencing: a comprehensive review of RNA-seq methodologies. Materials and Methods 3: 203.

Daniels, M., F. Goh, C.M. Wright, K.B. Sriram, V. Relan, B.E. Clarke, E.E. Duhig, R.V. Bowman, I.A. Yang and K.M. Fong. 2012. Whole genome sequencing for lung cancer. Journal of Thoracic Disease 4(2): 155-163.

De Mateo, S., and P. Sassone-Corsi. 2014. Regulation of spermatogenesis by small non-coding RNAs: role of the germ granule. Seminars in Cell and Developmental Biology 29: 84-92.

Dhiman, N., D.I. Smith, and G.A. Poland. 2009. Next-generation sequencing: a transformative tool for vaccinology. Expert Review of Vaccines 8(8): 963-967.

DiMarzio, M., N. Shariat, S. Kariyawasam, R. Barrangou, and E.G. Dudley. 2013. Antibiotic resistance in Salmonella enterica serovar Typhimurium associates with CRISPR sequence type. Antimicrobial Agents and Chemotherapy 57(9): 4282-4289.

Djebali, S., C.A. Davis, A. Merkel, A. Dobin, T. Lassmann, A. Mortazavi, A. Tanzer, J. Lagarde, W. Lin, F. Schlesinger, C. Xue, G.K. Marinov, J. Khatun, B.A. Williams, C. Zaleski, J. Rozowsky, M. Röder, F. Kokocinski, R. F. Abdelhamid, T. Alioto, I. Antoshechkin, M.T. Baer, N.S. Bar, P. Batut, K. Bell, I. Bell, S. Chakrabortty, X. Chen, J. Chrast, J. Curado, T. Derrien, J. Drenkow, E. Dumais, J. Dumais, R. Duttagupta, E. Falconnet, M. Fastuca, K. Fejes-Toth, P. Ferreira, S. Foissac, M.J. Fullwood, H. Gao, D. Gonzalez, A. Gordon, H. Gunawardena, C. Howald, S. Jha, R. Johnson, P. Kapranov, B. King, C. Kingswood, O.J. Luo, E. Park, K. Persaud, J.B. Preall, P. Ribeca, B. Risk, D. Robyr, M. Sammeth, L. Schaffer, L. See, A. Shahab, J. Skancke, A.M. Suzuki, H. Takahashi, H. Tilgner, D. Trout, N. Walters, H. Wang, J. Wrobel, Y. Yu, X. Ruan, Y. Hayashizaki, J. Harrow, M. Gerstein, T. Hubbard, A. Reymond, S.E. Antonarakis, G. Hannon, M.C. Giddings, Y. Ruan, B. Wold, P. Carninci, R. Guigó, and T.R. Gingeras. 2012. Landscape of transcription in human cells. Nature 489(7414): 101-108.

Dodd, C.C., D.G. Renter, D.U. Thomson, and T.G. Nagaraja. 2011. Evaluation of the effects of a commercially available Salmonella Newport siderophore receptor and porin protein vaccine on fecal shedding of bacteria and health and performance of feedlot cattle. American Journal of Veterinary Research 72(2): 239-247.

Dressman, D., H. Yan, G. Traverso, K.W. Kinzler, and B. Vogelstein. 2003. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. Proceedings of the National Academy of Sciences 100(15): 8817-8822.

Egan, A.N., J. Schlueter, and D.M. Spooner. 2012. Applications of next-generation sequencing in plant biology. American Journal of Botany 99(2): 175-185.

Elgert K.D. 2009. Immunology: understanding the immune system. John Wiley and Sons, New Jersey, USA. Chapter 17. pp. 603-605.

Fakruddin, M., A. Chowdhury, M.N. Hossain, K.S. Mannan, and R.M. Mazumda. 2012. Pyrosequencing-principles and applications. International Journal of Life Science and Pharma Research Review 2(2): 65-76.

Fedurco, M., A. Romieu, S. Williams, I. Lawrence, and G. Turcatti. 2006. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. Nucleic Acids Research 34(3): e22.

Finotello, F., and B. Di Camillo. 2015. Measuring differential gene expression with RNA-seq: challenges and strategies for data analysis. Briefings in Functional Genomics 14(2): 130-142.

Garrido-Cardenas, J.A., F. Garcia-Maroto, J.A. Alvarez-Bermejo, and F. Manzano-Agugliaro. 2017. DNA sequencing sensors: an overview. Sensors 17(3): 588.

Glenn, T.C. 2011. Field guide to next-generation DNA sequencers. Molecular Ecology Resources 11(5): 759-769.

Goodwin, S., J.D. McPherson, and W.R. McCombie. 2016. Coming of age: ten years of next-generation sequencing technologies. Nature Reviews Genetics 17(6): 333-351.

Griffith, M., J.R. Walker, N.C. Spies, B.J. Ainscough, and O.L. Griffith. 2015. Informatics for RNA sequencing: a web resource for analysis on the cloud. PLOS Computational Biology 11(8): e1004393.

Großhans, H., and W. Filipowicz. 2008. Molecular biology: the expanding world of small RNAs. Nature 451(7177): 414-416.

Guttman, M., M. Garber, J.Z. Levin, J. Donaghey, J. Robinson, X. Adiconis, L. Fan, M.J. Koziol, A. Gnirke, C. Nusbaum, J.L. Rinn, E.S. Lander, and A. Regev. 2010. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nature Biotechnology 28(5): 503-510.

Heider, L.C., R.W. Meiring, A.E. Hoet, W.A. Gebreyes, J.A. Funk, and T.E. Wittum. 2008. Evaluation of vaccination with a commercial subunit vaccine on shedding of Salmonella enterica in subclinically infected dairy cows. Journal of the American Veterinary Medical Association 233(3): 466-469.

Hermesch, D.R., D.U. Thomson, G.H. Loneragan, D.R. Renter, and B.J. White. 2008. Effects of a commercially available vaccine against Salmonella enterica serotype Newport on milk production, somatic cell count, and shedding of Salmonella organisms in female dairy cattle with no clinical signs of salmonellosis. American Journal of Veterinary Research 69(9): 1229-1234.

Hodkinson, B.P., and E.A. Grice. 2015. Next-generation sequencing: a review of technologies and tools for wound microbiome research. Advances in Wound Care 4(1): 50-58.

Hoffmann, S., B. Maculloch, and M. Batz. 2015. Economic burden of major foodborne illnesses acquired in the United States. Current Politics and Economics of the United States, Canada, and Mexico 17(4): 543-616.

Illumina. 2014. Specification sheet: sequencing. https://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/hiseq-3000-4000-specification-sheet-770-2014-057.pdf.

Ilott, N.E., and C.P. Ponting. 2013. Predicting long non-coding RNAs using RNA sequencing. Methods 63(1): 50-59.

Jazayeri, S.M., L.M. Melgarejo Muñoz, and H.M. Romero. 2015. RNA-Seq: a glance at technologies and methodologies. Acta Biológica Colombiana 20(2): 23-35.

Jin, W., J.R. Grant, P. Stothard, and S.S. Moore. 2009. Characterization of bovine miRNAs by sequencing and bioinformatics analysis. BMC Molecular Biology 10(1): 90.

Ju, J., D.H. Kim, L. Bi, Q. Meng, X. Bai, Z. Li, X. Li, M.S. Marma, S. Shi, J. Wu, J.R. Edwards, A. Romu, and N.J. Turro. 2006. Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. Proceedings of the National Academy of Sciences 103(52): 19635-19640.

Kaundal, R. 2016. NGS sequence analysis with R/Bioconductor: RNA-Seq workflow. http://biocluster.ucr.edu/~rkaundal/workshops/R_mar2016/RNAseq.html.

Kircher, M., and J. Kelso. 2010. High-throughput DNA sequencing–concepts and limitations. Bioessays 32(6): 524-536.

Korpelainen, E., J. Tuimala, P. Somervuo, M. Huss, and G. Wong. 2015. RNA-seq data analysis: a practical approach. CRC Press, Taylor & Francis Group, Boca Raton. Chapter 1. pp. 9-13.

Lee, H.J., H.S. Park, W. Kim, D. Yoon, and S. Seo. 2014. Comparison of metabolic network between muscle and intramuscular adipose tissues in Hanwoo beef cattle using a systems biology approach. International Journal of Genomics 2014: 679437.

Levy, S.E., and R.M. Myers. 2016. Advancements in Next-Generation Sequencing. Annual Review of Genomics and Human Genetics 17: 95-115.

Lin, B., J. Wang, and Y. Cheng. 2008. Recent patents and advances in the next-generation sequencing technologies. Recent Patents on Biomedical Engineering 2008(1): 60-67.

Lister, R., B.D. Gregory, and J.R. Ecker. 2009. Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. Current Opinion in Plant Biology 12(2): 107-118.

Liu, L., Y. Li, S. Li, N. Hu, Y. He, R. Pong, D. Lin, L. Lu, and M. Law. 2012. Comparison of next-generation sequencing systems. Journal of Biomedicine and Biotechnology 2012: 251364.

Loneragan, G.H., D.U. Thomson, R.M. McCarthy, H.E. Webb, A.E. Daniels, T.S. Edrington, D.J. Nisbet, S.J. Trojan, S.C. Rankin, and M.M. Brashears. 2012. Salmonella diversity and burden in cows on and culled from dairy farms in the Texas high plains. Foodborne Pathogens and Disease 9(6): 549-555.

Lu, H., F. Giordano, and Z. Ning. 2016. Oxford nanopore minion sequencing and genome assembly. Genomics, Proteomics and Bioinformatics 14(5): 265-279.

Majowicz, S.E., J. Musto, E. Scallan, F.J. Angulo, M. Kirk, S.J. O'Brien, T.F. Jones, A. Fazil, and R.M. Hoekstra. 2010. The global burden of nontyphoidal Salmonella gastroenteritis. Clinical Infectious Diseases 50(6): 882-889.

Mardis, E.R. 2013. Next-generation sequencing platforms. Annual Review of Analytical Chemistry 6: 287-303.

Mardis, E.R. 2008. Next-generation DNA sequencing methods. Annual Review of Genomics and Human Genetics 9: 387-402.

Marguerat, S., and J. Bähler. 2010. RNA-seq: from technology to biology. Cellular and Molecular Life Sciences 67(4): 569-579.

Mather, A.E., S.W.J. Reid, D.J. Maskell, J. Parkhill, M.C. Fookes, S.R. Harris, D.J. Brown, J.E. Coia, M.R. Mulvey, M.W. Gilmour, L. Petrovska, E. de Pinna, M. Kuroda, M. Akiba, H. Izumiya, T.R. Connor, M.A. Suchard, P. Lemey, D.J. Mellor, D.T. Haydon, and N.R. Thomson. 2013. Distinguishable epidemics within different hosts of the multidrug resistant zoonotic pathogen Salmonella Typhimurium DT104. Science 341(6153): 1514-1517.

McEwen, S.A., and P.J. Fedorka-Cray. 2002. Antimicrobial use and resistance in animals. Clinical Infectious Diseases 34: S93-S106.

Meldrum, C., M.A. Doyle, and R.W. Tothill. 2011. Next-generation sequencing for cancer diagnostics: a practical perspective. The Clinical Biochemist Reviews 32(4): 177-195.

Metzker, M.L. 2010. Sequencing technologies—the next generation. Nature Reviews Genetics 11(1): 31-46.

Mortazavi, A., B.A. Williams, K. Mccue, L. Schaeffer, and B. Wold. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Methods 5(7): 621-628.

Nester, E.W., D.G. Anderson, C.E. Roberts, Jr., N.N. Pearsall, M.T. Nester, and D. Hurley. 2004. Microbiology: a human perspective. McGraw-Hill, New York, NY, USA. Chapter 24. pp. 616-617.

Perkel, J.M. 2013. Visiting "noncodarnia". Biotechiques 54(6): 301-304.

Quail, M.A., M. Smith, P. Coupland, T.D. Otto, S.R. Harris, T.R. Connor, A. Bertoni, H.P. Swerdlow, and Y. Gu. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 13(1): 341.

Radford, A.D., D. Chapman, L. Dixon, J. Chantrey, A.C. Darby, and N. Hall. 2012. Application of next-generation sequencing technologies in virology. The Journal of General Virology 93(9): 1853-1868.

Rhoads, A., and K.F. Au. 2015. PacBio sequencing and its applications. Genomics, Proteomics and Bioinformatics 13(5): 278-289.

Roy, S., W.A. LaFramboise, Y.E. Nikiforov, M.N. Nikiforova, M.J. Routbort, J. Pfeifer, R. Nagarajan, A.B. Carter, and L. Pantanowitz. 2016. Next-generation sequencing informatics: challenges and strategies for implementation in a clinical environment. Archives of Pathology and Laboratory Medicine 140(9): 958-975.

Sanchez, S., C.L. Hofacre, M.D. Lee, J.J. Maurer, and M.P. Doyle. 2002. Animal sources of salmonellosis in humans. Journal of the American Veterinary Medical Association 221:492-497.

Scallan, E., R.M. Hoekstra, F.J. Angulo, R.V. Tauxe, M.A. Widdowson, S.L. Roy, J.L. Jones, and P.M. Griffin. 2011. Foodborne illness acquired in the United States--major pathogens. Emerging Infectious Diseases 17(1): 7-15.

Simon, S.A., J. Zhai, R.S. Nandety, K.P. McCormick, J. Zeng, D. Mejia, and B.C. Meyers. 2009. Short-read sequencing technologies for transcriptional analyses. Annual Review of Plant Biology 60:305-333.

Siqueira, J.F., A.F. Fouad, and I.N. Rôças. 2012. Pyrosequencing as a tool for better understanding of human microbiomes. Journal of Oral Microbiology 4: 10743.

Soneson, C., M.I. Love, and M.D. Robinson. 2016. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000 Research 4:1521.

Suárez-Vega, A., B. Gutiérrez-Gil, C. Klopp, G. Tosser-Klopp, and J.J. Arranz. 2017. Variant discovery in the sheep milk transcriptome using RNA sequencing. BMC Genomics 18(1): 170.

Thomson, D.U., G.H. Loneragan, A.B. Thornton, K.F. Lechtenberg, D.A. Emery, D.T. Burkhardt, and T.G. Nagaraja. 2009. Use of a siderophore receptor and porin proteins-based vaccine to control the burden of Escherichia coli O157: H7 in feedlot cattle. Foodborne Pathogens and Disease 6(7): 871-877.

Toldrá, F. 2009. Safety of meat and processed meat. Springer, New York, NY, USA. Chapter 1. pp. 3-5.

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

Tremblay, M.-P., V.E.S. Armero, A. Allaire, S. Boudreault, C. Martenon-Brodeur, M. Durand, E. Lapointe, P. Thibault, M. Tremblay-Létourneau, J.-P. Perreault, M.S. Scott, and M. Bisaillon. 2016. Global profiling of alternative RNA splicing events provides insights into molecular differences between various types of hepatocellular carcinoma. BMC Genomics 17(1): 683.

Voelkerding, K.V., S.A. Dames, and J.D. Durtschi. 2009. Next-generation sequencing: from basic research to diagnostics. Clinical Chemistry 55(4): 641-658.

Vogl, I., S.H. Eck, A. Benet-Pages, P.A. Greif, K. Hirv, S. Kotschote, M. Kuhn, A. Gehring, C. Bergmann, H.J. Bolz, M. Stuhrmann, S. Biskup, K.H. Metzeler, and

H.-G. Klein. 2012. Diagnostic applications of next generation sequencing: working towards quality standards. Laboratoriumsmedizin 36(4): 227-239.

Wang, E.T., R. Sandberg, S. Luo, I. Khrebtukova, L. Zhang, C. Mayr, S.F. Kingsmore, G.P. Schroth, and C.B. Burge. 2008. Alternative isoform regulation in human tissue transcriptomes. Nature 456(7221): 470-476.

Wang, X.G., Z.H. Ju, M.H. Hou, Q. Jiang, C.H. Yang, Y. Zhang, Y. Sun, R.L. Li, C.F. Wang, J.F. Zhong, and J.M. Huang. 2016. Deciphering transcriptome and complex alternative splicing transcripts in mammary gland tissues from cows naturally infected with Staphylococcus aureus mastitis. PloS One 11(7): e0159719.

Wang, Y., B. Lupiani, S.M. Reddy, S.J. Lamont, and H. Zhou. 2014. RNA-seq analysis revealed novel genes and signaling pathway associated with disease resistance to avian influenza virus infection in chickens. Poultry Science 93(2): 485-493.

Wang, Y., T. Hu, L. Wu, X. Liu, S. Xue, and M. Lei. 2017. Identification of non-coding and coding RNAs in porcine endometrium. Genomics 109(1): 43-50.

Wang, Z., M. Gerstein, and M. Snyder. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews. Genetics 10(1): 57-63.

Wilhelm, B.T., and J.-R. Landry. 2009. RNA-Seq—quantitative measurement of expression through massively parallel RNA-sequencing. Methods 48(3): 249-257.

Wit, P.D., M.H. Pespeni, J.T. Ladner, D.J. Barshis, F. Seneca, H. Jaris, N.O. Therkildsen, M. Morikawa, and S.R. Palumbi. 2012. The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. Molecular Ecology Resources 12(6): 1058-1067.

Wolf, J.W. 2013. Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. Molecular Ecology Resources 13(4): 559-572.

World Health Organization (WHO). 2013. Salmonella (non-typhoidal). http://www.who.int/mediacentre/factsheets/fs139/en/ (Accessed December 2016).

Yegnasubramanian, S. 2013. Preparation of fragment libraries for next-generation sequencing on the applied biosystems SOLiD platform. Methods in Enzymology 529: 185-200.

Zeng, W., and A. Mortazavi. 2012. Technical considerations for functional sequencing assays. Nature Immunology 13(9): 802-807.

Zhang, J., R. Chiodini, A. Badr, and G. Zhang. 2011. The impact of next-generation sequencing on genomics. Journal of Genetics and Genomics 38(3): 95-109.

Chapter 2

Using RNA Seq to Define Genes and Pathways Affecting Salmonella Vaccine Response in Beef Cattle

Introduction

Vaccination prevents a wide range of bacterial and viral diseases that result in economic losses to livestock producers and decreased yield, fertility, morbidity, and mortality in beef and dairy cattle. For example, pre-calving vaccination of dams with a killed *S. Typhimurium* vaccine deceased calf mortality by 75% in calves challenged with *Salmonella typhimurium* (Jones et al., 1988). Vaccination of livestock may also be useful for protecting humans against zoonotic diseases like salmonellosis. It is estimated that each year *Salmonella* (nontyphoidal) causes an average of one million illnesses and 380 deaths in the United States (Scallan et al., 2011). Therefore, advances in livestock vaccine programs against *Salmonella* species can aid in the decrease of disease incidence in humans.

Salmonella Newport Bacterial Extract vaccine with siderophore receptor and porin protein (SRP) technology is a vaccine that cattle producers can employ to control salmonellosis and reduce shedding of *Salmonella* in feces. This vaccine works by stimulating immunity to create antibodies that target bacterial siderophore and porin proteins. These proteins are found on the bacterial cell membrane and scavenge iron present in the body of infected animals. The vaccine impairs the iron uptake system by producing antibodies that block entry of iron into *Salmonella* through SRPs. Because iron is essential for survival of *Salmonella*, the bacterial cells begin to die (Thomson et al., 2009).

To the extent of our knowledge, no study of the host genetic response to *Salmonella* SRP vaccination has been reported in beef cattle. Understanding how the host responds to vaccination is important for developing more effective vaccines that protect a

larger fraction of a population. One approach for studying the host genetic response involves transcriptome profiling by RNA-seq. Unlike microarrays, this approach does not require us to define the genes we want to examine before initiation of the experiment. With RNA-seq, we can investigate biological pathways affected by vaccination independent of whether we know the sequence or biological function of the gene. Our hypothesis was that vaccination of beef cattle with *Salmonella* SRPs would change gene expression in whole blood at 21 days and 48 days post-vaccination. Therefore, the objectives of the study were to identify differentially expressed genes and pathways in whole blood after vaccination with *Salmonella* SRPs and generate a list of candidate genes important for mounting an immune response to this vaccine in beef cattle.

Material and Methods

Animal and Sample Collection

A total of five cows were sampled. All animals were Black Angus, born in 2011, housed at the SDSU Cow/Calf Research and Teaching Unit, and had never been immunized with *Salmonella* SRP vaccine. The cows were administered with 2 mL of *Salmonella* Newport SRP vaccine subcutaneously according to the manufacturer's instructions (http://www.epitopix.com/prod-cattle-salmonella). Blood samples were collected at three time points: time of the initial vaccination (day 0), time of booster vaccination (day 21), and 27 days after booster (day 48). Blood was collected in PAXgene tubes because the tubes provide an immediate stabilization of RNA after the blood is collected. The same five cows were used at all three time points, resulting in a total of fifteen blood samples. These samples were stored at -80 °C prior to RNA extraction.

RNA Extraction

Total RNA was isolated from all blood samples using the PAXgene^{TM} Blood RNA kit (PreAnalytiX, Qiagen, BD, Germany) according to the manufacturer's instructions (Appendix A). Concentration and purity of RNA samples were assessed using a Nanodrop ND-1000 spectrophotometer (Thermo Scientific, Waltham, MA). Before constructing RNA seq libraries, the integrity of each RNA sample was evaluated with an Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA).

Library Preparation and Sequencing

The RNA samples with dry ice were submitted to LC Sciences (Houston, Texas USA) for library construction and sequencing. Samples were prepared using the Illumina® TruSeq Stranded Total RNA Library PrepKit (Illumina, San Diego, CA) following the manufacturer's instructions. In the initial step of library preparation, polyadenylated RNA was isolated from the total RNA with poly (T) oligonucleotides attached to magnetic beads and then fragmented using divalent metal cations under high temperatures. First strand cDNA synthesis was performed using random primers and reverse transcriptase followed by synthesis of second strand cDNA to generate double-stranded cDNA. To form blunt ends, ds cDNA was subjected to end repair followed by addition of an adenine (A) nucleotide to the 3' ends. Adapters were ligated to the ends of the cDNA fragments followed by solid-phase PCR amplification to produce the sequencing library. Each sequencing library was indexed with a distinct adapter and was assessed for quality using an Agilent 2100 Bioanalyzer. The indexed library samples were normalized and pooled in equimolar concentrations prior to sequencing. Single-read sequencing (150 cycles) was performed on a single-lane flow cell of an Illumina HiSeq 2000 system using the HiSeq SR (Single Read-150 bases) Cluster Kit v4 at LC Sciences (Houston, Texas).

Processing of raw data

The raw sequencing reads were stored as fastq files, including the read summary of each sample. With the FastQC software (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), the quality of the reads was assessed to determine sequencing errors, contamination, and sequencing or PCR artifacts. Reads with low quality score and length less than 10 bp were removed. Also, 3'

adapter sequence and repetitive elements were trimmed using Cutadapt (https://cutadapt.readthedocs.io/en/stable/). After filtering, reads with length greater than 10 bp were used for mapping to the bovine genome.

Bioinformatics

Bioinformatics analyses (except DAVID analysis; see below) were completed by LC Sciences (Houston, Texas, USA). The trimmed reads were aligned to the Bovine reference genome (Bta_4.6.1) using the Bowtie/Tophat_v2.0.12 (Trapnell et al., 2012; Kim et al., 2013) to identify transcripts. Tophat utilized Bowtie for alignment of cDNA sequencing reads against the virtual transcriptome build and then mapped the reads to the reference genome to extract transcript sequences. The resulting alignment was stored in a BAM file format (Li et al., 2009) and filtered for mismatched reads.

A filtered BAM file of the read alignments was submitted to Cufflinks v2.2.1 (Trapnell et al., 2012) to assemble transcripts and estimate their abundance for each sample. Cufflinks first assigned genomic regions that were overlapped by sequencing reads as being potential exons and then utilized mapped junctions to assemble transcripts. The Cuffdiff module of Cufflinks was used for pairwise differential expression analysis at the gene level. In this process, the transcript abundance was normalized and evaluated in Fragments Per Kilobase of transcript per Million mapped reads (FPKM) (Mortazavi et al., 2008) using Cuffdiff. The FPKM values for all genes were used to estimate differential expression between samples at different time points. Tests for statistical significance (*P*-values) accounted for multiple testing by computing the False Discovery Rate p-value (FDR *P*-value) using the Benjamini-Hochberg correction (Benjamini and Hochberg, 1995). The FDR adjusts *P*-values for the number of hypothesis tests

performed to control type I error rate (proportion of false positives). Differences in expression were statistically significant when q-value (FDR adjusted *P*-value) was less than 0.05.

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of the differentially expressed genes were completed by an online software called Database for Annotation, Visualization, and Integrated Discovery (DAVID) (Huang et al., 2009). DAVID is a bioinformatics resource containing an integrated biological knowledgebase and analytic tools that aim to systematically extract enriched biological and functional meaning behind a given list of genes or proteins with q-value (Benjamini) < 0.05. Initially, a list of gene IDs was imported into the Functional Annotation Tool of DAVID where the most relevant GO terms and KEGG pathways were identified. The GO terms are a set of structured and controlled vocabularies to describe attributes of genes and gene products based on three biological domains. These domains are shared by all organisms and consist of biological processes, cellular components, and molecular functions (The Gene Ontology Consortium, 2008). The KEGG pathways are a compilation of manually drawn diagrams displaying networks of molecular interaction and biochemical reactions in various cellular processes (http://www.genome.jp/kegg/pathway.html). Thus, the GO term category of DAVID matches the input gene lists to their corresponding biological terms while the KEGG pathway category represents the summary of enriched pathways that are over- and underrepresented among differentially expressed genes.

Real Time RT-PCR

To validate differentially expressed genes from sequencing results by the real-time reverse transcription polymerase chain reaction (RT-PCR) method, candidate genes were selected based on the following criteria: 1) presence of detectable expression across all three time points and 2) fold change > 2 between day 0 and day 48 samples. Performing the real-time RT-PCR reaction was only possible with RNA samples from two time points: day 0 and day 48. Day 21 samples were excluded because sufficient RNA was not available from samples at this time point. For both selected genes and a reference gene, a pair of primers (forward and reverse) was designed using the PrimerQuest Tool from Integrated DNA Technologies (https://www.idtdna.com). Initially, the mRNA sequence of each gene was retrieved as a FASTA file from the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/). Then, the FASTA file for each gene was submitted to the UCSC Genome Browser (https://genome.ucsc.edu) where the BLAST-like alignment tool (BLAT) was used to align the mRNA sequence to the genome. Mapping the mRNA sequences to the genome with BLAT allows identification of exon-intron boundaries. Primers designed in PrimerQuest were mapped to the bovine genome to ensure primers overlapped at least one exon-intron boundary. All primer sets were ordered from Integrated DNA Technologies. Before resuspending dry oligonucleotides in water, the tubes were briefly centrifuged to collect the oligo pellets at the bottom of tubes. A resuspension calculator (https://www.idtdna.com/Calc/resuspension/) was used for calculating the volume needed to resuspend the oligonucleotides to the desired concentration that was based on the quantity and molecular weight of the oligo and final concentration desired (400μM).

Prior to real-time RT-PCR, an equal quantity of RNA from each sample was pooled to create a standard curve. The reason for performing the standard curve (1:2 dilution series) was to estimate the efficiency of the primer pair during PCR. Real-time RT-PCR was performed using the iTaq™ Universal SYBR® Green One-Step Kit (Bio-Rad, Hercules, CA) including iTaq universal SYBR® Green reaction mix (2x) and iScript reverse transcriptase. The 10 µl reaction setup was prepared by adding all required components according to the manufacturer's recommendations (Appendix B). The PCR plate was loaded into a MxPro3005P real-time thermocycler (Agilent Technologies, Santa Clara, CA) and was programmed as follows: reverse transcription reaction at 50 °C for 10 minutes; polymerase activation and DNA denaturation at 95°C for 1 minute; and 40 cycles of amplification including denaturation at 95°C for 15 seconds and the annealing/extension step at 60°C for 30 seconds. A melt-curve was generated by increasing the temperature by 0.5°C every 5 seconds from 65°C to 95°C. Forward and reverse primers for each gene are provided in Table 2.1. Glyceraldehyde-3-Phosphate Dehydrogenase (GAPDH) was used to normalize expression levels among samples.

Results

In the present study, day 0 (time of initial vaccination) was our control group. We chose day 0 as our control because vaccination would not have affected gene expression at this time point. Day 0 was compared with day 21 and day 48 after immunization to identify DE genes and to determine changes in the expression patterns of each gene at all time points.

The raw number of reads obtained for each sample ranged from 2 198 017 to 12 890 843 million single-end fragments (one read per fragment), amounting to a total of 151 709 746 million reads for the entire experiment. Additionally, the overall rate of read alignment to the bovine genome attained by TopHat ranged between 93.7 to 98.9% (Table 2.2). This data indicated a high consistency between samples for the number of and proportion of mapped reads.

Cufflinks analysis revealed a total of 848 genes that differed significantly in expression in whole blood (q-value <0.05) at 21 days after vaccination. Five hundred and eleven of the DE genes were increased in abundance and the remaining 337 genes were decreased in abundance relative to day 0. At day 48 post-vaccination, 1155 genes were differentially expressed (q-value <0.05), of which 922 were increased in abundance while the remaining 234 genes were decreased in abundance. Further analysis of the increased- and decreased in abundance genes was carried out for genes that were both differentially expressed and had a fold-change ≥ 2 and ≤ 2, respectively. The median FPKM of the five samples at each time point was used to compute the fold change. This further analysis identified one increased in abundance gene at day 21 post-vaccination (Table 2.3) and 26 decreased in abundance genes at this same time point (Table 2.5). At day 48, 20 genes

were increased in abundance (Table 2.4) and 39 genes were decreased in abundance (Table 2.6) with fold-change ≥ 2 and ≤ 2, respectively, relative to day 0. In total, 22 decreased in abundance genes were differentially expressed at both day 21 and day 48 post-vaccination.

Gene ontology analysis completed by DAVID for genes found in the comparison of 1) day 21 samples vs. day 0 samples and 2) day 48 samples vs. day 0 samples identified a list of significant decreased and increased in abundance GO terms (adjust P-value< 0.05) and revealed enriched functionally-related gene sets. The genes included in biological process clusters for both days 21 and 48 were associated with functions including heme biosynthetic process and protoporphyrinogen IX biosynthetic process. The list of genes involved in the heme biosynthetic process was ALAD, HMBS, SLC25A39, ATPIF1 and ALAS2 (Table 2.7). Interestingly, hemoglobin binds to iron in the bloodstream and antibodies generated by the vaccine block iron entry into *Salmonella* (Thomson et al., 2009). All five DEG genes that are components of the heme biosynthetic pathway were decreased in abundance at day 21 and day 48 (Figure 2.1). The cellular component cluster at day 21 included genes involved in cytoplasm and membrane. While none of the cellular components was found at day 48. Finally, the genes in a molecular function cluster for day 21 were related to DNA helicase activity, while no significant molecular function cluster for day 48 was identified. The increased in abundance GO terms for day 21 and day 48 were involved in a wide range of general biological processes and molecular activity which were less likely to be directly related to a vaccine response. Therefore, we focused on the decreased in abundance genes present in these clusters.

Three decreased in abundance pathways were identified by KEGG pathway analysis at day 21 post-vaccination including the FoxO signaling pathway (Figure 2.2), AMPK signaling pathway, and porphyrin and chlorophyll metabolism (Table 2.9). Significant pathways found decreased in abundance at day 48 post-vaccination were porphyrin and chlorophyll metabolism, *Staphylococcus aureus* infection, prion diseases, and chagas disease (American trypanosomiasis). Increased in abundance pathways for day 21 relative to day 0 included pathways affecting the ribosome, oxidative phosphorylation, and spliceosome, while for day 48 relative to day 0 were the FoxO signaling pathway and NOD-like receptor signaling pathways. Other significantly increased in abundance pathways at day 48 were involved in endocytosis, chronic myeloid leukemia, hepatitis B and hepatitis C.

Real-time RT-PCR results were analyzed with the Relative Expression Software Tool (REST). REST is a software package designed to analyze gene expression from real-time RT-PCR amplification data, and then determines statistical significance between samples and controls. It also takes into account differences in PCR efficiency and normalizes results to reference gene(s) (Pfaffl et al., 2002). The GAPDH reference gene was used to normalize samples, and the comparison was between day 0 and day 48 samples. The genes that were tested are ALAS2, ATPIF1, and HMB. All these genes were decreased in abundance (same direction as for RNA-seq) and statistically significant between day 0 and day 48 (Table 2.8).

Discussion

The ability of RNA-seq analysis to uncover patterns of expression across the whole genome allows an investigation of a wide range of genes in an experiment. As a large number of genes are explored, the number of differentially expressed genes associated with the same biological function can be determined. Nevertheless, interpretation of the data can be challenging. Our study uncovered differentially expressed genes that were increased and decreased in abundance in whole blood at 21 days and 48 days post-vaccination with *Salmonella* SRP vaccine with a fold change ≥ 2. Although we expected to observe more increased in abundance genes and pathways after vaccination, a decrease in abundance of genes was a predominant trend in this study and occurred in all five animals. One gene was increased in abundance and 26 genes were decreased in abundance (fold-change ≥ |2|) at day 21 relative to day 0 in all samples. The following sections are a discussion of some of the decreased in abundance genes and pathways identified in this experiment as well as the potential role of SRP vaccination against *Salmonella*.

A decrease in abundance of genes and pathways in our study could be due to several reasons. The first reason is lack of vaccine efficacy to stimulate a strong immune response after vaccination. Previous studies indicated that antibody concentration in blood serum of vaccinated dairy cattle with *Salmonella* SRPs were significantly higher (*P*-value = 0.01) than non-vaccinated cattle (Hermesch et al, 2008; Smith et al., 2014). These studies provide evidence of the vaccine's effectiveness in eliciting an immune response against *Salmonella* bacteria, but the vaccine might not produce a strong enough immune response to cause an increase in abundance of immune response genes in whole

blood at the times examined in this study. This vaccine is an inactivated vaccine that can stimulate only the humoral response and requires adjuvants and multiple doses to provide long-lasting immunity to animals. Another possible reason for the lack of an increase in abundance of genes related to the immune response is that cattle are reservoirs for *Salmonella* species, indicating *Salmonella* Newport SRPs. Therefore, these SRPs might not be strongly recognized as a foreign antigen by the host.

However, with regard to a genetic response to *Salmonella* SRP vaccine as measured by deep sequencing techniques, equivalent studies have not been conducted. Similar to our experiment, a previous study demonstrated a potential for whole blood transcriptome profiling by RNA seq to identify immune-related biomarkers. This study found that most DE genes were up-regulated in dairy cattle following vaccination (Demasius et al., 2013). However, this group used a viral antigen to stimulate immunity while we used a bacterial antigen. Response to an antigen can vary depending on pathogen types and how an antigen is prepared. Another study of whole blood transcriptome profiling with a microarray explored bovine immune response after a bacterial challenge (Rodrigues et al., 2015). Interestingly, consistent with our results, they found that the majority (84.9%) of DE genes and pathways were down-regulated (decreased in abundance), including pathways involved in the lymphatic system and B-lymphocytes.

Five decreased in abundance genes (ALAD, HMBS, SLC25A39, ATPIF1, and ALAS2) in our study were highly associated with heme and porphyrin-containing compound biosynthetic process. A decrease in abundance of these genes might be due to

the host immune response to *Salmonella* SRP antigen. Gram-negative bacteria like *Salmonella enterica* utilize numerous iron storage and transport proteins (e.g. host heme, transferrin, siderophores) to fulfill its nutritional need for iron (Krewulak and Vogel, 2008). Most of the iron inside a host is found within heme proteins, specifically hemoglobin (Burman, 1974). Siderophore receptors enable *Salmonella* to extract iron from their environment as iron is required for bacterial metabolism (Neilands, 1995). The SRP vaccine produces sufficient antibodies that target *Salmonella* siderophore receptors, preventing transport of iron molecules into the bacteria. One possibility is that the vaccine has pleiotropic effects on genes related to iron uptake in the host.

It is possible that the SRP vaccine was not responsible for a decrease in abundance of these genes, and instead other factors such as the environment, different pathogens, or diets could be causing these changes in gene expression. However, the only known change in the animals' environment during the study period was the administration of the *Salmonella* SRP vaccine. Animals' diet was not changed and no disease was reported during the entire experiment. Still, we cannot rule out the possibility that an unknown environmental effect may have caused some of the changes observed in gene expression. It is important to mention that more immune-related genes and pathways could have been detected in the present study if peripheral blood mononuclear cells have been used instead of whole blood. Transcriptome analysis of whole blood captures changes in gene expression among all red and white blood cells (including neutrophils), while PBMCs focus on only changes in lymphocyte and monocyte gene expression.

All five DE genes related to heme biosynthesis process in our study were previously found to affect heme biosynthesis using microarrays (Nilsson et al., 2009). Zebrafish was used as an *in vivo* vertebrate to identify genes that are required for functional heme biosynthesis in red blood cells. They found that the SLC25A39 enzyme is required for proper hemoglobin synthesis. Their results support a major role of SLC25A39 in the regulation of heme levels and maintenance of mitochondrial iron homeostasis. The ALAD enzyme catalyzes the second step in the heme biosynthetic and porphyrin pathway (Kelada et al., 2001) and plays a major role in lead toxicity in blood (Behera et al., 2016). The HMBS enzyme catalyzes the third step of the heme biosynthetic pathway (Ulbrichova et al., 2009), and is highly associated with an inherited disorder called porphyria, which is caused by abnormal heme production (Bustad et al., 2013). The ALAS2 gene has an essential role in the formation of hemoglobin by erythroid cells (Whatley et al., 2008). ATPIF1 modulates heme synthesis during differentiation of erythroid cells (Shah et al., 2012).

Siderophores are small molecules used by microbes to scavenge iron from a host and are considered a major virulence factor during infection with bacteria. Sassone-Corsi et al. (2016) demonstrated that immunization of mice with siderophore protein conjugates (derived from *cholerae* bacteria) induced an antibody response against siderophores and resulted in reduction of the intestinal and systemic colonization of *Salmonella enterica*. The authors concluded that the immunogenic carrier protein could be a potential tool to limit microbial burden and reduce the outcomes of infection. Similar to the previous study, Cox et al. (2017) used *Salmonella enterica* Enteritidis (SE) which is a siderophore receptor-porin vaccine in chickens. They found that the SE SRP vaccine was effective in

reduction of *Salmonella* shedding and prevention of reproductive tract colonization in vaccinated chickens. However, other studies found that the SRP antigen (derived from *Salmonella* Newport) was not associated with *Salmonella* reduction in feces (Dodd et al., 2011; Heider et al., 2008). One reason for conflicting results is *Salmonella* Newport SRPs might not be effective against all *Salmonella* serotypes in cattle (Cernicchiaro et al., 2016). Another reason might be due to differences in digestive systems between mice/poultry and cattle. Cattle contain diverse microbial populations in their digestive tract compared with mice/poultry, which lack a rumen.

Conclusion

In the present study, we investigated differentially expressed genes and pathways in whole blood of beef cattle after vaccination with *Salmonella* SRPs. We found that most DE genes and pathways were down-regulated rather than up-regulated at both time points (d21 and d48 post-vaccination). Genes in the heme biosynthetic process were down-regulated after vaccination. These genes play an important role in regulation of heme biosynthesis and metabolism. Several KEGG pathways were identified as well including AMPK and FoxO signaling pathways and the porphyrin metabolism pathway. This study could be a first step toward better understanding of the molecular mechanisms responsible for a vaccine response.

Literature Cited

Behera, R., M.D. Kothekar, D.S. Kale, K. Krishnamurthi, A.R. Sirothia, D.R. Kalorey, and M.S. Patil. 2016. Study of mutations in aminolevulinic acid dehydratase (ALAD) gene in cattle from fly ash zone in Maharashtra, India. Indian Journal of Animal Research 50(1): 19-22.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological) 57(1): 289-300.

Burman, D. 1974. Iron metabolism in infancy and childhood. In: Iron in biochemistry and medicine. (Eds. Jacobs, A and M. Worwood). Academic Press, New York, NY. pp. 544-563.

Bustad, H.J., M. Vorland, E. Rønneseth, S. Sandberg, A. Martinez, and K. Toska. 2013. Conformational stability and activity analysis of two hydroxymethylbilane synthase mutants, K132N and V215E, with different phenotypic association with acute intermittent porphyria. Bioscience Reports 33(4): e00056.

Cernicchiaro, N., S.E. Ives, T.S. Edrington, T.G. Nagaraja, and D.G. Renter. 2016. Efficacy of a Salmonella siderophore receptor protein vaccine on fecal shedding and lymph node carriage of Salmonella in commercial feedlot cattle. Foodborne Pathogens and Disease 13(9): 517-525.

Cox, G.J.M., B. Griffith, M. Reed, J.D. Sandstrom, M.P. Peterson, and D.E. Straub. 2017. Vaccination against Salmonella enterica Enteritidis using Siderophore

Receptor and Porin Proteins. 66th Western Poultry Disease Conference, Davis, CA. p. 40.

Demasius, W., R. Weikard, F. Hadlich, K.E. Müller, and C. Kühn. 2013. Monitoring the immune response to vaccination with an inactivated vaccine associated to bovine neonatal pancytopenia by deep sequencing transcriptome analysis in cattle. Veterinary Research 44(1): 93.

Dodd, C.C., D.G. Renter, D.U. Thomson, and T.G. Nagaraja. 2011. Evaluation of the effects of a commercially available Salmonella Newport siderophore receptor and porin protein vaccine on fecal shedding of bacteria and health and performance of feedlot cattle. American Journal of Veterinary Research 72(2): 239-247.

Heider, L.C., R.W. Meiring, A.E. Hoet, W.A. Gebreyes, J.A. Funk, and T.E. Wittum. 2008. Evaluation of vaccination with a commercial subunit vaccine on shedding of Salmonella enterica in subclinically infected dairy cows. Journal of the American Veterinary Medical Association 233(3): 466-469.

Hermesch, D.R., D.U. Thomson, G.H. Loneragan, D.R. Renter, and B.J. White. 2008. Effects of a commercially available vaccine against Salmonella enterica serotype Newport on milk production, somatic cell count, and shedding of Salmonella organisms in female dairy cattle with no clinical signs of salmonellosis. American Journal of Veterinary Research 69(9): 1229-1234.

Huang, D.W., B.T. Sherman, and R.V. Lempicki. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols 4(1): 44-57.

Jones P.W., P. Collins, and M.M. Aitken. 1988. Passive protection of calves against experimental infection with *Salmonella* typhimurium. Veterinary Record 123: 536-541.

Kelada, S.N., E. Shelton, R.B. Kaufmann, and M.J. Khoury. 2001. δ-Aminolevulinic acid dehydratase genotype and lead toxicity: a huge review. American Journal of Epidemiology 154(1): 1-13.

Kim, D., G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, and S.L. Salzberg. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology 14(4): R36.

Krewulak, K.D., and H.J. Vogel. 2008. Structural biology of bacterial iron uptake. Biochimica et Biophysica Acta (BBA)-Biomembranes 1778(9): 1781-1804.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25(16): 2078-2079.

Mortazavi, A., B.A. Williams, K. Mccue, L. Schaeffer, and B. Wold. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Methods 5(7): 621-628.

Neilands, J.B. 1995. Siderophores: structure and function of microbial iron transport compounds. Journal of Biological Chemistry 270(45): 26723-26726.

Nilsson, R., I.J. Schultz, E.L. Pierce, K.A. Soltis, A. Naranuntarat, D.M. Ward, J.M. Baughman, P.N. Paradkar, P.D. Kingsley, V.C. Culotta, J. Kaplan, J. Palis, B.H. Paw, and V.K. Mootha. 2009. Discovery of genes essential for heme biosynthesis through large-scale gene expression analysis. Cell Metabolism 10(2): 119-130.

Pfaffl M.W., G.W. Horgan, and L. Dempfle. 2002. Relative Expression Software Tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Research 30(9): E36.

Rodrigues, V., P. Holzmuller, C. Puech, H. Wesonga, F. Thiaucourt, and L. Manso-Silván. 2015. Whole blood transcriptome analysis of Mycoplasma mycoides subsp. mycoides-infected cattle confirms immunosuppression but does not reflect local inflammation. PloS One 10(10): e0139678.

Sassone-Corsi, M., P. Chairatana, T. Zheng, A. Perez-Lopez, R.A. Edwards, M.D. George, E.M. Nolan, and M. Raffatellu. 2016. Siderophore-based immunization strategy to inhibit growth of enteric pathogens. Proceedings of the National Academy of Sciences of the United States of America 113(47): 13462-13467.

Scallan, E., R.M. Hoekstra, F.J. Angulo, R.V. Tauxe, M.A. Widdowson, S.L. Roy, J.L. Jones, and P.M. Griffin. 2011. Foodborne illness acquired in the United States--major pathogens. Emerging Infectious Diseases 17(1): 7-15.

Shah, D.I., N. Takahashi-Makise, J.D. Cooney, L. Li, I.J. Schultz, E.L. Pierce, A. Narla, A. Seguin, S.M. Hattangadi, A.E. Medlock, N.B. Langer, T.A. Dailey, S.N. Hurst, D. Faccenda, J.M. Wiwczar, S.K. Heggers, G. Vogin, W. Chen, C. Chen, D.R. Campagna, C. Brugnara, Y. Zhou, B.L. Ebert, N.N. Danial, M.D. Fleming, D.M. Ward, M. Campanella, H.A. Dailey, J. Kaplan, and B.H. Paw. 2012. Mitochondrial Atpif1 regulates heme synthesis in developing erythroblasts. Nature 491(7425): 608-612

Smith, G.W., M.L. Alley, D.M. Foster, F. Smith, and B.W. Wileman. 2014. Passive immunity stimulated by vaccination of dry cows with a Salmonella bacterial extract. Journal of Veterinary Internal Medicine 28(5): 1602-1605.

The Gene Ontology Consortium. 2008. The gene ontology project in 2008. Nucleic Acids Research 36(Database issue): D440-D444.

Thomson, D.U., G.H. Loneragan, A.B. Thornton, K.F. Lechtenberg, D.A. Emery, D.T. Burkhardt, and T.G. Nagaraja. 2009. Use of a siderophore receptor and porin proteins-based vaccine to control the burden of Escherichia coli O157: H7 in feedlot cattle. Foodborne Pathogens and Disease 6(7): 871-877.

Trapnell, C., A. Roberts, L. Goff, G. Pertea, D. Kim, D. Kelley, H. Pimentel, S. Salzberg, J. Rinn, and L. Pachter. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols 7(3): 562-578.

Ulbrichova, D., M. Hrdinka, V. Saudek, and P. Martasek. 2009. Acute intermittent porphyria–impact of mutations found in the hydroxymethylbilane synthase gene on biochemical and enzymatic protein properties. The FEBS Journal 276(7): 2106-2115.

Whatley, S.D., S. Ducamp, L. Gouya, B. Grandchamp, C. Beaumont, M.N. Badminton, G.H. Elder, S.A. Holme, A.V. Anstey, M. Parker, A.V. Corrigall, P.N. Meissner, R.J. Hift, J.T. Marsden, Y. Ma, G. Mieli-Vergani, J.-C. Deybach, and H. Puy. 2008. C-Terminal deletions in the ALAS2 gene lead to gain of function and cause x-linked dominant protoporphyria without anemia or iron overload. American Journal of Human Genetics 83(3): 408-414.

Table 2.1 Forward and Reverse Primers Used for Real-Time RT-PCR*

Gene Symbol	Gene Name	Direction	Primer Sequence	PCR Efficiency	Amplicon Length
GAPDH	Glyceraldehyde-3-Phosphate Dehydrogenase	FWD REV	5'- GAT GCT GGT GCT GAG TAT GT -3' 5'- GCA GAA GGT GCA GAG ATG AT -3'	94.7%	113bp
ALAS2	5'-Aminolevulinate synthase 2	FWD REV	5'- CTT GCC AGG GTG TGA GAT TTA -3' 5'- CAG GAT CAT TGT GCC TGA AGA -3'	85.5%	107bp
ATPIF1	ATPase inhibitory factor 1	FWD REV	5'- TAC TTC CGA GCT CGT GCT AA -3' 5'- CTG CTT ATG CCG CTC AAT TTC -3'	70.0%	120bp
HBM	Hemoglobin Subunit Mu	FWD REV	5'- TCT TCA CTG TGT ACC CTA GCA -3' 5'- GGA TCA CCA GTG GGA AGT TG -3'	95.9%	213bp

* Real-time reverse transcription polymerase chain reaction

Table 2.2 Summary of Next Generation Sequencing Reads for each Sample

Sample & Time Point	**Number of reads**	**Number (%) reads mapped to bovine genome**
116—d 0	10 515 662	10 363 117 (98.55%)
116—d21	9 617 720	9 417 720 (97.92%)
116—d48	12 269 345	11 505 992 (93.78%)
133—d0	11 037 438	10 879 003 (98.56%)
133—d21	9 154 869	8 901 246 (97.23%)
133—d48	10 773 153	10 446 866 (96.97%)
140—d0	10 122 006	9 780 734 (96.63%)
140—d21	7 303 897	7 217 419 (98.82%)
140—d48	10 705 198	10 562 128 (98.66%)
143—d0	9 423 631	9 303 833 (98.73%)
143—d21	2 198 017	2 059 846 (93.71%)
143—d48	12 470 610	12 095 478 (96.99%)
148—d0	12 598 709	12 467 396 (98.96%)
148—d21	12 890 843	12 716 807 (98.65%)
148—d48	10 628 648	10 388 605 (97.74%)

Table 2.3 Differentially Expressed Genes Increased in Abundance at 21 Days Post-Vaccination with a Fold-Change ≥ 2

ID[1]	Gene	FC[2]	FDR[3]-Adj p-value	FPKM[4]_D0 Numbers		FPKM_D21 Numbers	
				Avg[5]	SD[6]	Avg[5]	SD[6]
XLOC_000432	BT.89425	2.98	4.57E-02	0.25	0.22	1.39	1.2

[1] Locus Identifier

[2] Fold Change

[3] False Discovery Rate

[4] Fragments Per Kilobase of transcript per Million mapped reads

[5] Average of five samples for day 0 and day 21, respectively.

[6] Standard deviation

Table 2.4 Differentially Expressed Genes Increased in Abundance at 48 Days Post-Vaccination with a Fold-Change ≥ 2

				FPKM[4]_D0 Numbers		FPKM_D48 Numbers	
ID[1]	**Gene**	**FC[2]**	**FDR[3]-adj p-value**	**Avg[5]**	**SD[6]**	**Avg[5]**	**SD[6]**
XLOC_019465	DMBT1	3.86	1.60E-03	0.23	0.15	4.24	1.85
XLOC_020745	ROBO3	3.54	1.60E-03	9.84	20.89	751.66	1231.57
XLOC_024014	TNPO3	3.20	1.60E-03	34.34	39.98	140.93	138.47
XLOC_003247	TPC3	3.19	1.60E-03	0.93	0.4	7.85	1.41
XLOC_017235	C2	2.99	1.60E-03	7.26	3.22	55.25	14.53
XLOC_012518	KRT10	2.78	4.14E-03	0.15	0.11	3.09	3.88
XLOC_029167	BT.87687	2.68	1.60E-03	0.63	0.27	4.62	0.66
XLOC_024940	C1R	2.65	1.60E-03	10.41	3.38	72.83	20.36
XLOC_026254	BT.64790	2.61	1.60E-03	0.36	0.18	2.31	0.79
XLOC_014106	ALPL	2.57	1.60E-03	6.37	6.37	23.03	11.99
XLOC_021226	BT.88775	2.54	1.60E-03	3.33	1.28	21.68	8.85
XLOC_022268	SLAMF8	2.43	5.24E-03	1.19	0.84	6.26	2.55
XLOC_011764	bta-mir-10a	2.37	1.60E-03	301.34	507.32	1460.55	2379.98
XLOC_022150	SPP2	2.34	1.60E-03	1.71	1.26	8.54	5.99
XLOC_028332	GDF9	2.27	2.96E-02	0.08	0.04	5.03	10.77
XLOC_018836	NUPR1	2.21	1.60E-03	24.74	7.23	144.55	59.76
XLOC_009447	MN1	2.13	2.26E-02	0.30	0.09	1.33	0.49
XLOC_018123	BT.66080,SERPINB4	2.09	8.95E-03	1.86	1.83	4.46	3.02

XLOC_020995	BT.40283	2.03	1.60E-03	0.59	0.42	2.59	1.11
XLOC_006657	TEX12	2.01	1.07E-02	0.74	0.43	2.01	0.91

[1] Locus Identifier

[2] Fold Change

[3] False Discovery Rate

[4] Fragments Per Kilobase of transcript per Million mapped reads

[5] Average of five samples for day 0 and day 48, respectively.

[6] Standard deviation

Table 2.5 Differentially Expressed Genes Decreased in Abundance at 21 Days Post-Vaccination with a Fold-Change ≤ 2

				FPKM[4]_D0 Numbers		FPKM_D21 Numbers	
ID[1]	**Gene**	**FC[2]**	**FDR[3]-Adj p-value**	**Avg[5]**	**SD[6]**	**Avg[5]**	**SD[6]**
XLOC_018239	HBM	8.21	1.15E-02	188.37	149.47	1.21	1.49
XLOC_030110	ARG1	6.45	1.60E-03	11.84	8.39	0.21	0.11
XLOC_029129	EPB49	6.07	1.60E-03	10.42	7.82	0.16	0.12
XLOC_031117	BT.43709	5.91	1.60E-03	35.35	26.87	0.58	0.54
XLOC_023921	AQP1	5.73	1.60E-03	3.31	3.07	0.65	1.09
XLOC_027423	SEMA6B	5.65	2.50E-02	4.60	4.89	0.13	0.15
XLOC_031733	ALAS2	5.32	1.60E-03	1551.37	1233.4	27.48	9.97
XLOC_002105	EPB42	5.22	1.60E-03	10.22	7.062	0.40	0.31
XLOC_002793	BT.57494	4.51	2.93E-03	2.75	2.048	0.12	0.09
XLOC_029496	BT.67346	4.42	1.80E-02	3.59	1.87	0.11	0.07
XLOC_006403	BT.49731	3.91	1.60E-03	76.23	39.94	4.88	1.94
XLOC_027297	KLF1	3.74	1.60E-03	42.96	32.55	3.14	1.24
XLOC_015638	BT.88457	3.47	1.60E-03	26.00	18.12	1.86	0.90
XLOC_000110	ENSBTAG00000023666	3.44	1.60E-03	26.57	19.04	1.43	0.41
XLOC_007905	TMCC2	3.23	1.60E-03	62.05	53.45	5.19	1.43
XLOC_027545	TRIM58	3.05	1.60E-03	10.92	6.75	1.09	0.30
XLOC_020086	DUSP26	3.01	2.93E-03	4.57	2.37	0.56	0.35
XLOC_018657	TMEM8	2.80	1.60E-03	72.34	58.42	6.64	1.70

XLOC_014041	ATPIF1	2.79	1.60E-03	4734.68	4266.27	395.44	34.78
XLOC_002859	CHAC2	2.51	1.60E-03	7.53	4.32	1.52	0.27
XLOC_027205	PLVAP	2.44	1.60E-03	3.31	1.89	0.50	0.13
XLOC_016056	IFRD2	2.32	1.60E-03	145.30	93.01	24.29	4.13
XLOC_010349	TMEM86B	2.28	1.60E-03	34.34	27.32	4.87	1.60
XLOC_027303	TNPO2	2.52	2.96E-02	126.70	121.08	65.12	105.27
XLOC_005534	C20ORF123	2.17	3.18E-02	3.56	2.15	0.92	0.58
XLOC_018530	WBSCR27	2.14	4.71E-02	4.94	5.23	0.68	0.47

[1] Locus Identifier

[2] Fold Change

[3] False Discovery Rate

[4] Fragments Per Kilobase of transcript per Million mapped reads

[5] Average of five samples for day 0 and day 21, respectively.

[6] Standard deviation

Table 2.6 Differentially Expressed Genes Decreased in Abundance at 48 Days Post-Vaccination with a Fold-Change ≤ 2

				FPKM[4]_D0 Numbers		FPKM_D48 Numbers	
ID[1]	**Gene**	**FC[2]**	**FDR[3]-Adj p-value**	**Avg[5]**	**SD[6]**	**Avg[5]**	**SD[6]**
XLOC_019406	BT.35938	12.20	1.60E-03	37215.74	37573.37	414.24	913.26
XLOC_018239	HBM	6.58	5.24E-03	188.37	149.47	1.13	0.64
XLOC_029129	EPB49	6.13	1.60E-03	10.42	7.82	0.16	0.09
XLOC_031117	BT.43709	6.07	1.60E-03	35.35	26.87	0.74	0.59
XLOC_031733	ALAS2	6.05	1.60E-03	1,551.37	1233.41	14.60	4.58
XLOC_002911	REEP1	5.75	1.60E-03	8.05	5.39	0.12	0.04
XLOC_030110	ARG1	5.66	1.60E-03	11.84	8.39	0.35	0.17
XLOC_009241	GYPB	5.23	1.60E-03	8.06	6.74	0.36	0.34
XLOC_002793	BT.57494	4.87	7.27E-03	2.75	2.05	0.07	0.03
XLOC_023921	AQP1	4.69	1.60E-03	3.31	3.07	0.12	0.05
XLOC_027743	bta-mir-378-1	4.32	1.60E-03	36766.80	57965.01	1544.26	2206.33
XLOC_007613	ENSBTAG00000040320	4.17	2.73E-02	27.12	44.06	1.61	2.12
XLOC_002105	EPB42	3.87	1.60E-03	10.22	7.06	0.97	0.56
XLOC_006403	BT.49731	3.85	1.60E-03	76.23	39.94	5.57	1.38
XLOC_015417	ENSBTAG00000048096	3.84	3.24E-02	4.31	2.92	0.55	0.66
XLOC_027205	PLVAP	3.77	4.14E-03	3.31	1.89	0.20	0.14
XLOC_027595	ELA2	3.43	1.60E-03	8.87	3.94	0.74	0.37
XLOC_027297	KLF1	3.39	1.60E-03	42.96	32.55	3.59	0.28

XLOC_000110	ENSBTAG00000023666	3.23	1.60E-03	26.57	19.04	1.95	0.93
XLOC_020086	DUSP26	3.12	1.60E-03	4.57	2.37	0.45	0.12
XLOC_029153	ADAMDEC1	2.97	1.60E-03	13.52	10.34	1.00	0.67
XLOC_010807	PGLYRP1	2.91	1.23E-02	6.67	4.23	1.01	0.48
XLOC_002859	CHAC2	2.90	1.60E-03	7.53	4.32	1.47	0.57
XLOC_015638	BT.88457	2.81	1.60E-03	26.00	18.12	3.94	2.26
XLOC_026522	ZFYVE28	2.78	1.60E-03	56.77	15.41	11.48	8.73
XLOC_010349	TMEM86B	2.72	1.60E-03	34.34	27.32	4.10	0.98
XLOC_007905	TMCC2	2.52	1.60E-03	62.05	53.45	9.36	2.97
XLOC_012151	ENSBTAG00000039504	2.43	2.93E-03	11.72	6.06	1.75	1.08
XLOC_007611	PRG3	2.38	1.60E-03	19.42	25.65	1.85	1.66
XLOC_018657	TMEM8	2.35	1.60E-03	72.34	58.42	7.26	0.83
XLOC_014041	ATPIF1	2.33	1.60E-03	4734.68	4266.27	559.58	107.09
XLOC_016056	IFRD2	2.25	1.60E-03	145.30	93.01	29.73	10.16
XLOC_016871	BT.26590	2.17	1.60E-03	11.33	7.73	2.07	0.81
XLOC_000819	MLF1	2.13	4.97E-02	1.52	0.97	0.36	0.12
XLOC_027545	TRIM58	2.12	1.60E-03	10.92	6.75	2.12	0.53
XLOC_018242	HBQ1	2.11	2.93E-03	19.34	18.55	1.81	0.69
XLOC_016080	bta-mir-191	2.07	1.60E-03	4576.13	6988.25	878.88	1139.58
XLOC_009397	SDSL	2.05	1.60E-03	11.46	3.99	2.47	1.02
XLOC_013341	STRADB	2.00	1.60E-03	15.86	7.56	3.89	1.24

[1] Locus Identifier [2] Fold Change [3] False Discovery Rate [4] Fragments Per Kilobase of transcript per Million mapped reads
[5] Average of five samples for day 0 and day 48, respectively. [6] Standard deviation

Table 2.7 Heme Biosynthetic Process Genes Decreased in Abundance at 21 Days and 48 Days Post-Vaccination in All Samples

Gene Symbol	Gene name	Log2 (FC)[1] day21	Log2 (FC)[1] day48	Adj P-value[2]
ALAD	Aminolevulinate dehydratase	-1.24	-1.47	1.60E-03
HMBS	Hydroxymethylbilane synthase	-1.46	-1.63	1.60E-03
SLC25A39	Solute carrier family 25, member 39	-1.93	-1.75	1.60E-03
ATPIF1	ATPase inhibitory factor 1	-3.58	-3.08	1.60E-03
ALAS2	5'-Aminolevulinate synthase 2	-5.82	-6.73	1.60E-03

[1] Fold Change

[2] False Discovery Rate Adjusted P-value

Table 2.8 Validation of RNA-seq by Real-Time RT-PCR (d 48 vs. d 0 post-vaccination)

Gene	Expression	Std. Error	95%Confidence Interval	P-value
ALAS2	0.016	0.005 - 0.052	0.004 - 0.068	0.002
ATPIF1	0.244	0.092 - 0.475	0.072 - 0.553	0.014
HMB	0.019	0.008 - 0.065	0.007 - 0.071	0.001

ALAS2 5'-Aminolevulinate Synthase 2

ATPIF1 ATPase inhibitory factor 1

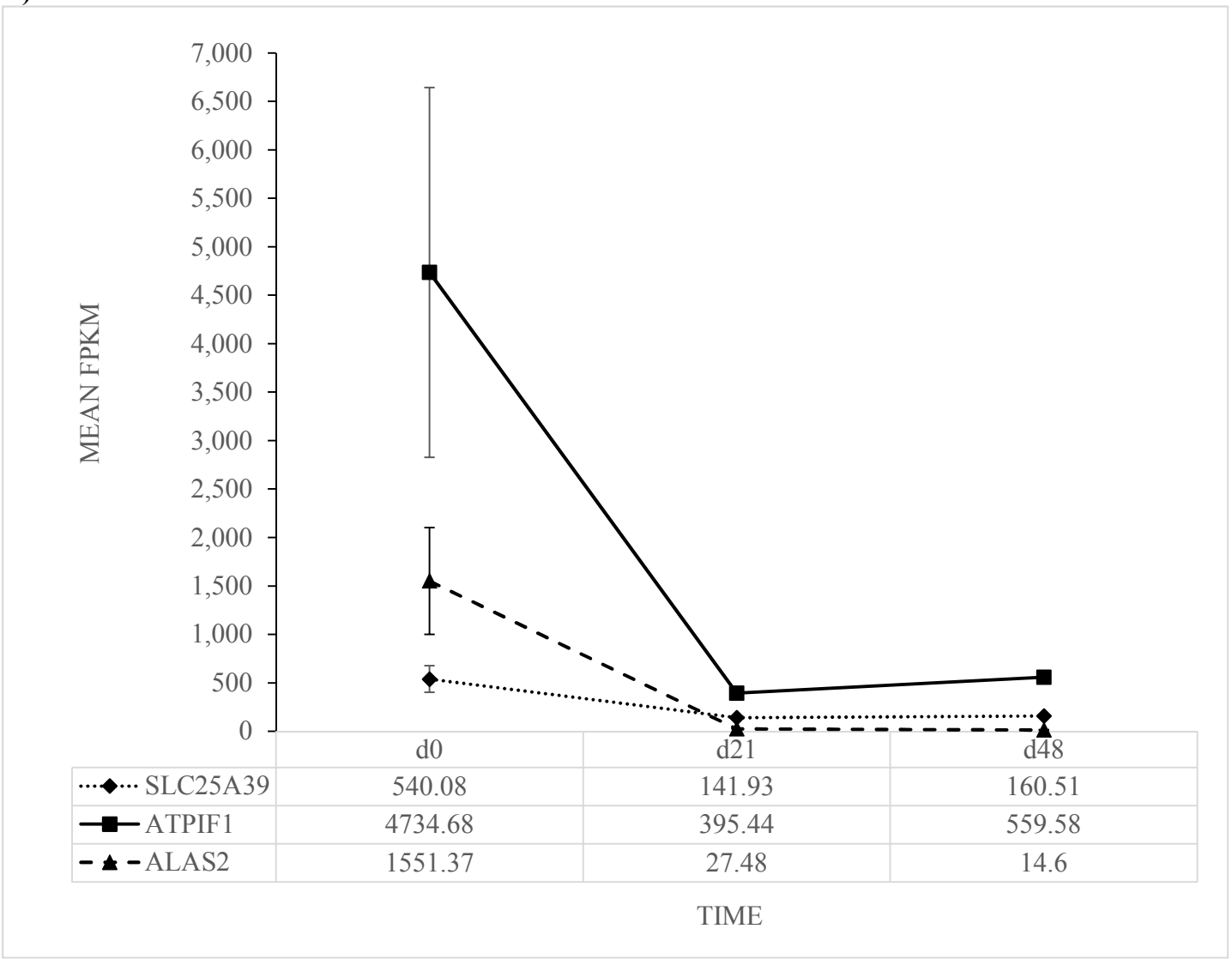
HMB Hemoglobin, mu

Table 2.9 Decreased in Abundance Pathways at Day 21 Post-Vaccination

KEGG Pathway	Adjusted p-values	Gene Set-Size	Genes
FoxO signaling pathway	8.90E-03	11	BCL6, CREBBP, GABARAPL1, FOXO1, FOXO3, IGF1R, IRS2, PIK3CD, SGK1, SOD2, SOD2
AMPK signaling pathway	1.20E-02	10	STRADB, ADIPOR1, CPT1A, FOXO1, FOXO3, GYS1, IGF1R, IRS2, PIK3CD, PFKL
Porphyrin and chlorophyll metabolism	1.50E-02	6	ALAS2, ALAD, BLVRB, EPRS, HMBS, UROD

Figure 2.1 Heme Biosynthetic Gene Expression Levels from Time of Initial Vaccination to 48 Days Post-Vaccination

A) Genes with FPKM > 500 at d0.



	d0	d21	d48
SLC25A39	540.08	141.93	160.51
ATPIF1	4734.68	395.44	559.58
ALAS2	1551.37	27.48	14.6

X-axis represents days of sample collection

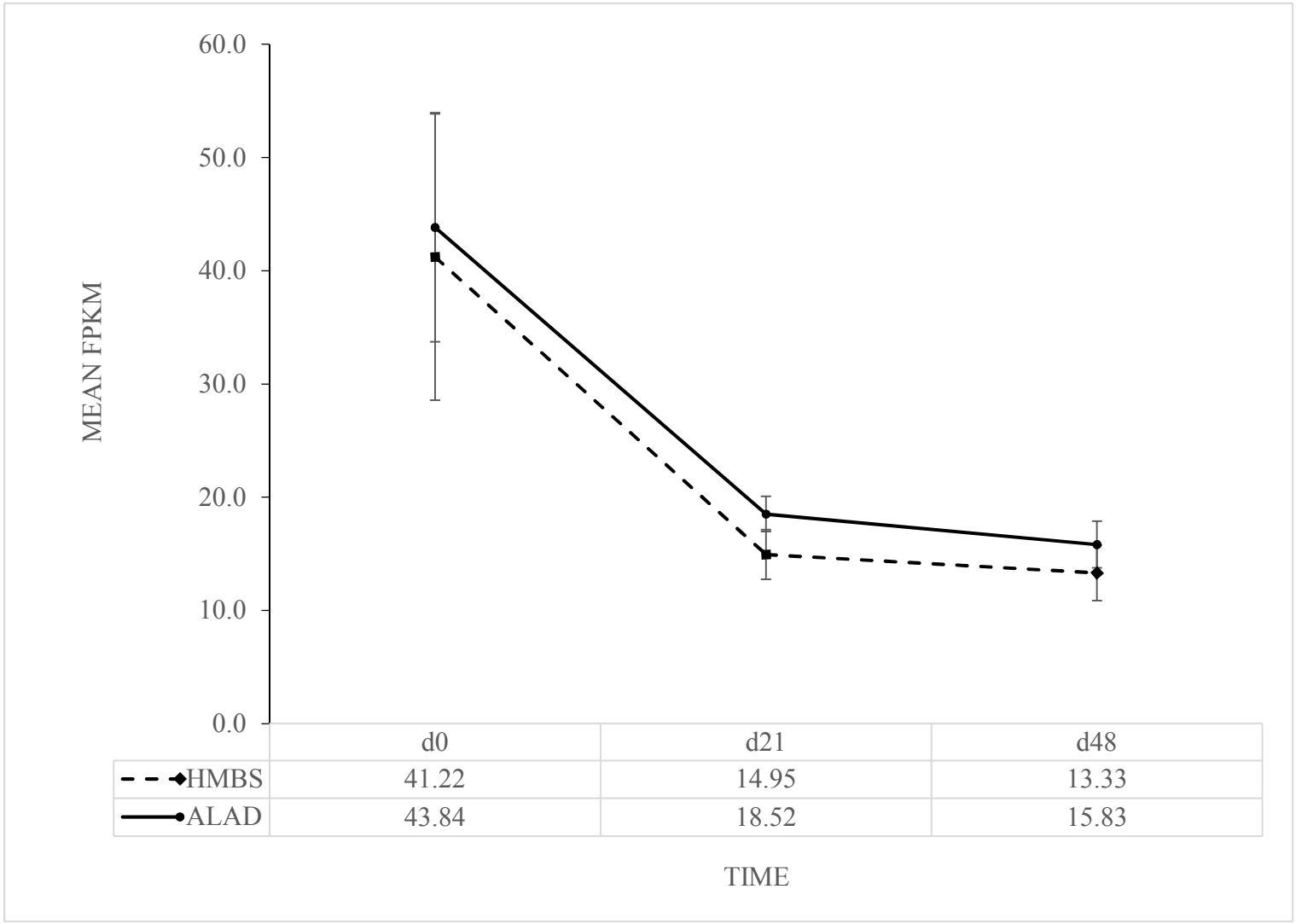
Y-axis represents average of reads in Fragments Per Kilobase Million

SLC25A39 Solute carrier family 25 member 39

ATPIF1 ATPase inhibitory factor 1

ALAS2 5'-Aminolevulinate Synthase 2

B) Genes with FPKM < 500 at d0



	d0	d21	d48
HMBS	41.22	14.95	13.33
ALAD	43.84	18.52	15.83

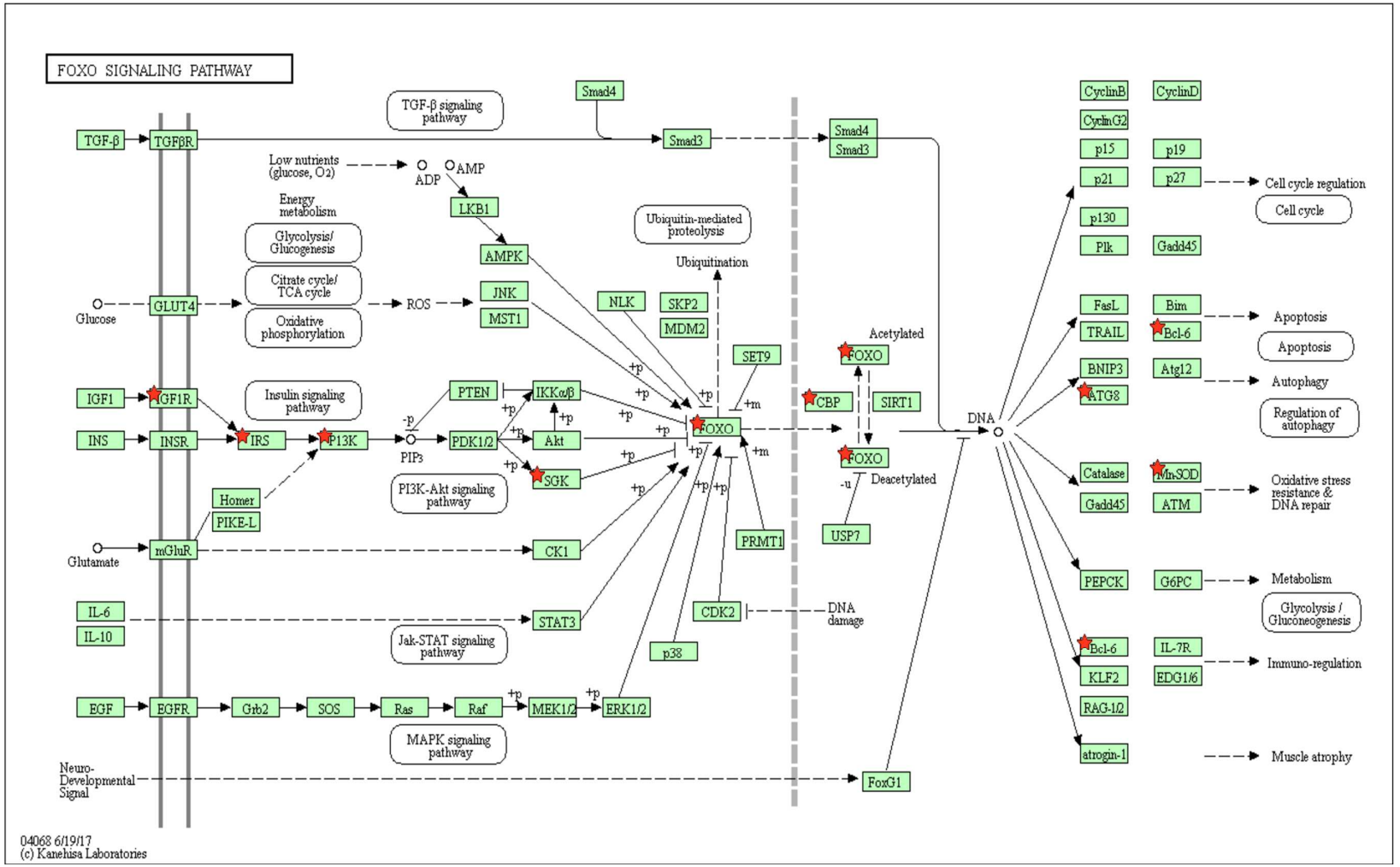
X-axis represents days of sample collection

Y-axis represents average of reads in Fragments Per Kilobase Million`

HMBS Hydroxymethylbilane Synthase

ALAD Delta-aminolevulinic acid dehydratase

Figure 2.2 FOXO Signaling Pathway for Decreased in Abundance Genes at Day 21 Post-Vaccination



Red stars represent significant differential gene expression in our study.

Appendix A: PAXgene Blood RNA kit Extraction Protocol

Blood samples were thawed at room temperature for at least two hours before starting the extraction and centrifuged for 10 minutes at 4000xg. After removing the supernatant, 6 mL of RBC lysis buffer was added to each sample and vortexed until pellet was dissolved for approximately 5 minutes. Tubes were centrifuged for 10 minutes at 3000xg and the supernatant was discarded. Each tube was loaded with 4 mL of RNase free water. The tubes were vortexed until pellet was dissolved and centrifuged again for 10 minutes at 3000xg. After discarding all of the supernatant, 350 µl of BR1 buffer was added to each tube and vortexed until pellet was dissolved. Samples were transferred into a 1.5 mL microcentrifuge tube. BR2 buffer (300 µl) and 40 µl of proteinase K were added to each sample, samples were vortexed for 5 seconds, and then incubated for 10 minutes at 55°C in a shaker-incubator (~140 rpm on the settings). Approximately 700 µl of lysate was pipetted into a LILAC spin column and spun for 3 minutes at 14,000xg. Without bumping pellet, flow-through was transferred into a fresh 1.5 mL microcentrifuge tube. 100% ethanol (350 µl) was added to each tube, vortexed, and briefly centrifuged. Sample (700 µl) was pipetted into the RED spin column, and spun for 1 minute at 14,000xg. The spin column was placed into a new collection tube, and the old tube with flow-through was discarded. BR3 (350 µl) buffer was added onto the spin column and centrifuged for 1 minute at 14,000xg. The spin column was placed into a new collection tube and the old tube with flow-through was discarded a second time. DNase I (10 µl) was added to 70 µl of RDD buffer in a new 1.5mL tube and then gently flicked to mix. DNase I and RDD buffer mixture (80 µl) was added onto the spin column and incubated at room temperature for 15 minutes. BR3 buffer (350 µl) was added onto the

spin column and spun for 1 minute at 14,000xg. The spin column was placed into a new collection tube and the old tube and flow-through was discarded. BR4 buffer (500 µl) was added onto the spin column and spun for 1 minute at 14,000xg. The spin column was placed into a new collection tube and the old tube with flow-through was discarded again. Another 500 µl of BR4 buffer was added onto the spin column and spun for 3 minutes at 14,000xg. Again, the spin column placed into new collection tube and the old tube with flow-through was discarded. To remove the remaining buffer, the spin column was spun again for 1 minute at 14,000xg and flow-through tube was discarded. Finally, the spin column was placed into a new 1.5 mL microcentrifuge tube. BR5 buffer (40 µl) was directly pipetted onto the spin column membrane and spun for 1 minute at 14,000xg. The previous step was repeated with the addition of 40 µl of BR5 buffer using the same collection tube. After the spin column was discarded, the 1.5 mL tube was then incubated for exactly 5 minutes at 65°C using a hot-plate (no-shaking), and immediately was moved to ice. The RNA samples were quantified with a Nanodrop ND-1000 spectrophotometer and stored in -80°C freezer.

Appendix B: SYBR Green One-Step Kit Reaction Mix Preparation Protocol

Frozen iTaq Universal SYBR Green reaction mix (2x) and iScript reverse transcriptase were thawed to 4°C or room temperature and then mixed thoroughly and centrifuged briefly for 3 to 4 seconds. First, the reaction setup was prepared for all PCR reactions by adding each required component except RNA including 5 µl of iTaq Universal SYBR Green, 0.125 µl of iScript reverse transcriptase and 0.3 µl of each forward and reverse primer. The reaction setup was spun briefly to ensure homogeneity and then 5.725 µl of reaction setup was pipetted into each reaction well of a PCR plate. Finally, 4 µl of RNA samples (40 ng) were added to the wells and 0.275 µl of nuclease-free water was added to each well for a total volume of 10 µl. The PCR reaction for each RNA sample was performed in duplicate. Also, one well of the PCR plate was a negative control with no template and another well had no reverse transcriptase (no RT enzyme).