A COMPARISON OF CONTROL SPLEEN GENE EXPRESSION USING MICROARRAY AND HIGH THROUGH-PUT SEQUENCING

by

Anne S. Mirabito

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Bachelor of Science in Animal Science with Distinction.

Spring 2010

A COMPARISON OF CONTROL SPLEEN GENE EXPRESSION USING MICROARRAY AND HIGH THROUGH-PUT SEQUENCING

by

Anne S. Mirabito

Approved: _____

Calvin Keeler, Ph.D.
Professor in charge of thesis on behalf of the Advisory Committee

Approved: _____

Conrad Pope, D.V.M.
Committee member from the Department of Animal and Food Sciences

Approved: _____

Nicole Donofrio, Member, Ph.D.
Committee member from the Board of Senior Thesis Readers

Approved: _____

Ismat Shah, Ph.D.
Chair of the University Committee on Student and Faculty Honors

ACKNOWLEDGMENTS

I would like to thank all the members of the Keeler Lab for their help, support, and advice throughout the process of the project. In addition, I would like to thank Dr. Schmidt for his initial analysis the Illumina Data, and Dr. Don Ritter for providing the birds for the project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Gene expression is the process of transcribing the DNA that resides on a gene into mRNA and translating the mRNA into protein. Studying allows researchers to decipher the functions of genes and their gene products, i.e. proteins. The objective of this project was to compare two methods of monitoring gene expression, microarray technology and Illumina sequencing technology, using the same avian spleen sample. A microarray consists of a glass slide that contains samples of many genes distributed in an ordered pattern. In this case, a 4,949-element avian innate immunity cDNA microarray (AIIM) was used. Illumina sequencing technology is based on massively parallel sequencing of millions of fragments using reversible terminator based sequencing chemistry in conjunction with the Illumina Genome Analyzer. The results obtained using these two methods were compared on a global level, in terms of housekeeping genes, and JAK pathway gene expression. In addition to these comparisons, some practical considerations about the two processes were made. Many of the most highly expressed genes in the Illumina data also showed expression levels with the array, through there was not a clear trend between them. Both of these technologies can be implemented to study gene expression. Though Illumina technology may be more quantitative and possibly more sensitive, the data generated from Illumina sequencing does not yet have a standard quantification method, as does microarray data.

Chapter 1

INTRODUCTION

1.1 Gene Expression: Levels of Regulation

There are thousands of genes in a eukaryotic cell and all genetic information is made up of sequences of bases in these genes (23). This information determines the structure, composition, and function of all cellular proteins (23). Large and complicated cellular machinery is responsible for the synthesis of proteins, the first step of which is transcription of DNA into mRNA, and secondly the translation of mRNA into proteins (23). The regulation of these steps increases or decreases gene expression, and allows for the differentiation of cells in multicellular organisms as well as the adjustment of an organism to changing environmental conditions (23).

Transcription is the first and a major level of regulation for gene expression (23). Transcription is similar to the process of replication except only a single strand of DNA is used as a template. Transcription generates not only the mRNAs, or the blueprints for the assembly of proteins, but all the different types of RNA that are needed for RNA processing and protein synthesis. RNA polymerase catalyzes the development of phosphodiester links between nucleotides from the 5' end of the RNA (18). In eukaryotic cells, there are at least four different kinds of RNA polymerases that locate and bind to promoter sequences. The promoter identifies the start of gene transcription. The mRNA produced is referred to as an mRNA transcript. A single gene may be transcribed many times. After transcription has occurred, the DNA strands re-anneal.

During translation, the triplicate base code carried on the mRNA is converted into the sequence of amino acids that code for a protein. There are three steps in this process: initiation, elongation, and termination.

Initiation begins when the cap binding protein binds to the 5' end of the mRNA (18). One of the subunits of this protein then unwinds any RNA secondary structure that may have formed. This complex then associates with the 40S and 60S ribosomal units (18). Lastly, the mRNA migrates so that it is in the proper position for interaction with charged tRNA that will bind to the peptide site (18).

The second step of translation is elongation. During elongation, the next available codon on the mRNA pairs with a newly charged tRNA. This guides its amino acyl residue to the A site on the ribosome, which is then transferred onto the amino group of the incoming amino acyl residue. The uncharged tRNA separates from the ribosome and the charged tRNA and the mRNA moves in relation to ribosome. An additional elongation factor is used and a GTP molecule is hydrolyzed. This moves the peptidyl-tRNA to the P- site so that another charged tRNA can enter the A site and base pair with the next codon on the mRNA (18). This process is repeated many times in order to build the polypeptide chain.

Termination is the final step of translation. It occurs when the termination codon is reached. At this point, there is no corresponding tRNA and the polypeptide chain is liberated from the last tRNA by hydrolysis (18). As the last, now uncharged, tRNA separates from the complex, the ribosomal subunits also separate from the mRNA and from each other, and are added to the pool of subunits that are available for use in a new initiation complex (18).

1.2 Importance of Studying Gene Expression

The study of gene expression enables scientists to decipher the function of genes and their protein products. The transcriptional state of a cell under given conditions provides a profile that can be compared to that of other cells under different conditions (23). These profiles can be used to group genes on the basis of their response to a given stimulus (23). Genes that are in common pathways are usually regulated in a similar way (23). Thus, similarities in gene expression patterns may be used to determine the function of unknown genes. A better understanding of the function of genes and the proteins they encode will lead to a more complete model of the complex regulatory networks that control all biological processes (1). Sequencing the genome of a variety of organisms has provided a wealth of information about many different genes and the roles that they play in the health of the organism. However, the DNA sequence alone is not enough to determine how cells function and what is going wrong when a disease occurs. To gain more knowledge about these processes, researchers are studying the conditions under which genes in the are expressed.

1.3 Relevance of Studying Spleen

The avian spleen has a similar role to that of its mammalian counterpart, though it has been studied less extensively. The spleen is composed of the red pulp and the white pulp. Its functions include phagocytosis of erythrocytes in the red pulp, lymphocyte production in the white pulp, and antibody production in both areas. The spleen removes the aged and abnormal red blood cells from the circulating blood. The T and B lymphocytes of the spleen produce antibodies In addition to these functions, the spleen

plays in an important role in the avian immune system, is transcriptionally active, easy to collect, and has given high yields of RNA when previously in the lab. used

The spleen is a principal organ of the immune system and the avian spleen is considered the largest secondary lymphoid organ. In most birds, lymph nodes are absent or rudimentary. This lack of lymph nodes increases the importance of the spleen in disease resistance. It is primarily a defensive organ and is often involved in systemic disease. Birds are comparable to mammals in their ability to generate antibodies and express cell-mediated immunity (41).

The spleen is small, soft, round, soft organ similar to the liver in color. In relation to body weight, the spleen is significantly smaller in birds than in mammals. In the chicken, the spleen is located near the gizzard and the proventriculus. A normal chicken spleen is about one quarter the length of the proventriculus, or about 2 cm (17). The avian spleen only has a thin connective tissue capsule with a minimal amount of smooth muscle tissue. However, it is large enough that is easy to identify and gather.

Spleen tissue typically yields a large amount of RNA. In previous experiments that have involved RNA isolation and amplification, spleen tissue has given good results. This means that a second round of RNA amplification is not usually needed which saves time in the overall experimental process.

1.4 Methods of Studying Gene Expression

Typically, the study of gene expression involves the comparison of mRNA populations in two different samples, control (untreated) versus treated, diseased versus healthy, or one species versus another species. There are many different methods to achieve this purpose. Some techniques for monitoring gene expression include Northern

Blot, RT-PCR, microarrays, and next generation sequencing technologies, such as Illumina.

1.4.1 Northern Blot

The Northern Blot technique allows for the detection of specific sequences of in RNA. This technique is performed in three steps. First, the RNA that is to be analyzed is electrophoresed under denaturing conditions in an agarose gel (23) Next the fractionated RNA is transferred from the denaturing agarose gel to a nitrocellulose or nylon membrane by capillary or vacuum transfer. Lastly, the RNA sequences of interest are hybridized to a labeled DNA or RNA probe (23). During hybridization, a probe/target complex, or hybrid, is formed between the probe molecule and the target molecule. In the case of Northern Blot, hybrids form between the single stranded RNA target on the membrane and a labeled complementary DNA probe or between the RNA target strand and a secondary complementary RNA molecule (23).

1.4.2 RT-PCR

Basic PCR is a fast, inexpensive way to amplify small segments of DNA. The process includes heating the sample to denature the DNA and using the enzyme Taq polymerase to construct two new strands of DNA, using the original strands as templates. The cycles of denaturing and synthesizing occur many times and result in many copies of the target of DNA. Analysis would typically occur after this amplification. RT-PCR (real-time fluorescence based reverse transcription polymerase chain reaction) combines the amplification and analysis steps of the PCR reaction, making processing after PCR unnecessary. This is achieved by monitoring the amount of DNA produced during each

PCR cycle. RT-PCR offers streamlined assay development, reproducible results, and a large dynamic range. There are currently five techniques that use florescent dyes and combine the processes of RNA amplification and detection to allow for the monitoring of PCR reactions in real time (23). The simplest methods use fluorescent dyes that bind specifically to double stranded DNA. The other methods rely on the hybridization of fluorescence-labeled oligonucleotides to the correct amplicon (23).

1.4.3 Microarrays

Microarrays are tools that can be used to monitor gene expression. A microarray consists of a glass slide that contains samples of many genes distributed in an ordered pattern. Thousands of different gene sequences can be fixed to the glass slide (4). These sequences can consist of DNA, cDNA, or oligonucleotides (4). Once a sample of fluorescently labeled nucleic acid molecules is added to the slide and binds to complementary nucleotide sequences, fluorescence can be measured using a laser scanner. The ratio of fluorescence of the spot to the background is used to ascertain the expression of individual genes. This method enables researchers to detect the expression of thousands of genes at one time.

The microarray has proven to be an important tool in the study of immune function. Several low-density and high-density cDNA based microarrays have been developed for use with avian species (8). Arrays such as these have been used to study gene expression in avian systems and tissues such as the immune system, metabolic/somatic system, neuroendocrine/reproductive system, and the liver (8). An example of this type of research was the evaluation the transcriptional response of

chicken spleen to H5N1 (13). The array has also been successfully used to determine gene expression in avian species in numerous tissues.

Microarray experiments require the spotted slide itself and the device for producing the spotted slide, a system for hybridization, a scanner to read the microarrays after hybridization, and software to quantify and interpret the results. Versions of these components are available commercially. It should be stressed that value of the data that can be obtained through use of the microarray depends on the quality of arraying (28).

1.4.3.1 Avian Innate Immunity Microarray

A 4,949-element avian innate immunity cDNA microarray (AIIM) has been created and used to examine the avian innate immune response (20). The array was created from EST libraries of stimulated chicken macrophages in addition to genes of interest from several specific innate immune pathways. Three EST libraries were created from commercial broiler chicken macrophages. The macrophages were either control (unstimulated), stimulated with INF-γ, or stimulated with LPS derived from *Salmonella* (13). The last library was created from HD11 avian macrophages that were treated with *Salmonella*-derived LPS (8). Most of these elements (4,772) come from clones derived from these EST libraries. Additionally, 134 of the AIIM elements acquired from the University of Delaware's chickEST library project (http://www.chickest.udel.edu/). More genetic elements to complete specific immunological pathways were obtained from 13 full-length cDNA clones that were provided by Dr. Peter Kaiser (Institute of Animal Health, Compton, UK). The remaining 40 elements consist of genes that were amplified directly from avian tissues using RT-PCR (20).

Each element is spotted in triplicate on the slide using an Omnigrid Accent Spotter (Gene Machine; San Carlos, CA). Several blank spots in each subarray serve as negative controls. Spotting in triplicate increases the statistical power of the microarray. The final array contains 28 elements that are part of the TLR pathway and viral interferon response pathway (13). In addition, the AIIM contains transcription factors, cytokines, chemokines, and lymphokines (20).

Previous studies have shown the effectiveness of the AIIM using different tissues types, including air sac, lung, liver, spleen, embryonic spleen, thymus and duodenum (20). All of these tissues were capable of hybridizing to the AIIM resulting in positive hybridization signals in 72%-92% of the genetic elements (20), thus validating the ability of the AIIM to monitor gene expression from different tissues involved in avian immunity. In a separate experiment, RNA was isolated from chicken, duck, and turkey spleens. This RNA was amplified and hybridized to the array to determine the utility of the array in analyzing other samples from other avian species besides chicken. 27.5% of the genetic elements hybridized to turkey spleen aRNA and 35.6% hybridized to duck spleen aRNA on the AIIM (20), confirming that the AIIM can be used to analyze duck and turkey samples, as well as chicken.

1.4.4 Illumina Sequencing Technology

This method is based on massively parallel sequencing of millions of small DNA fragments using reversible terminator based sequencing chemistry in conjunction with the Illumina Genome Analyzer (2). The technology relies on the attachment of randomly fragmented genomic DNA to a planar, transparent surface (2). The attached fragments are extended and “bridge amplified” to create a sequencing flow cell with hundreds of

millions of DNA “clusters”, each containing about 1,000 copies of the same template (2). The Illumina sequencing technology uses four proprietary, fluorescently labeled reversibly terminated nucleotides and a special polymerase to quickly sequence the tens of millions clusters base by base in parallel (3). Laser excitation and total internal reflection optics allow fluorescence detection to be highly sensitive (2). Sequence reads are then aligned against a reference genome. Illumina does have a specifically designed analysis software but other commercially available pipelines do exist. In addition, different sample preparation methods allow the same system to be used for a wide range of applications including gene expression, small RNA discovery, and protein-nucleic acid interactions.

Illumina Sequencing Technology can be used for measuring mRNA levels. This process involves creating a library. The first step in library creation is the isolation of poly-A RNA. This RNA is fragmented and randomly primed for reverse transcription in order to generate the cDNA fragments to are to be sequenced (3). Sequencing adaptors are ligated to the fragments, which are then size selected by gel electrophoresis and excision (3). A PCR step reduces contamination of the RNA and amplifies the target. These mRNA derived template libraries are sequenced the same way DNA samples are sequenced. They are loaded onto a fully automated Cluster Station. At this point they bind to complementary adapter oligos that have been attached to the flow cell surface (3). The cluster station isothermally amplifies these cDNA constructs to create colonal clusters of about 1,000 copies each (3). The resulting array of template clusters on the flow cell is directly sequenced by the fully automated Genome Analyzer.

1.5 Objectives

The detection of mRNA levels of specific genes is one of the cornerstones of molecular biology (27). There are many different methods through which mRNA can be detected, each with its own benefits and detriments. The AIIM is a useful tool for monitoring transcriptional response of immune cells in turkeys, ducks, and chickens (20). Though this technology has been used extensively, it involves lengthy protocols, considerable expense, and is very sensitive to environmental conditions. Next generation sequencing technology has been proposed as an alternative method of monitoring gene expression. It is thought that Illumina technology will prove to be more quantitative and possibly more sensitive than microarray technology. The objective of this project was to compare Illumina sequencing with microarray technology, as a method of monitoring gene expression. This comparison was made on global level, in terms of housekeeping genes, and the Janus kinase/signal transducers and activators of transcription (JAK/STAT) Pathway. The JAK/STAT pathway transmits information from signals outside the cell, through the cell membrane, to gene promoters on the DNA in the cell nucleus, which then causes DNA transcription and activity in the cell. This pathway was chosen for comparison because it contains a reasonable amount of genes and is will represented on the array. A schematic of the experimental design can be seen in Figure 1.1.

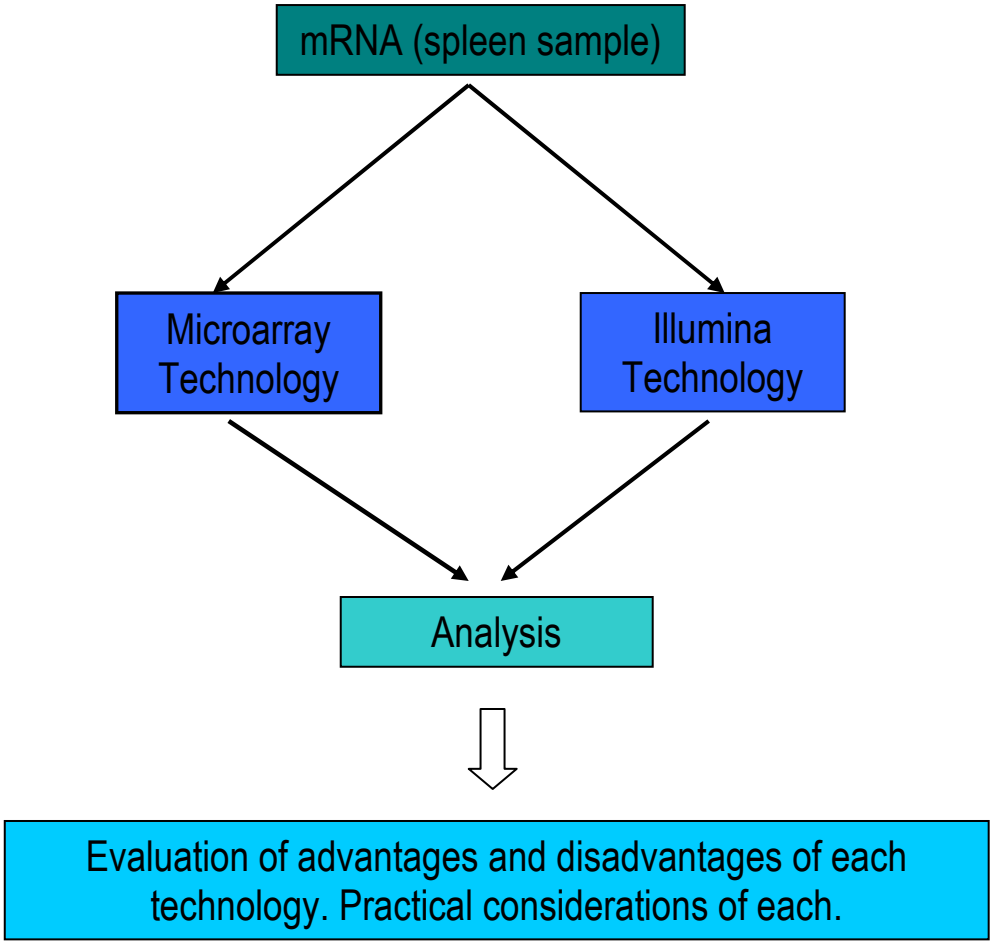

Figure 1.1 Schematic of Overall Experimental Design.

Chapter 2

METHODS AND MATERIALS

2.1 Collection of Tissues

Four, 4-week old male commercial broiler chickens were obtained from Dr. Don Ritter at Mount Aire Farms. Necropsies were performed on the four birds and the following tissues were collected: brain, liver, spleen, and testes. The tissue from each of the four birds was flash frozen in liquid nitrogen and then crushed. The pulverized brain, liver, spleen, and testes tissues from each bird were stored separately at -80 C. Spleen tissue was used in this experiment.

2.2 AIIM Slide Preparation

AIIM slides are stored under vacuum after they are spotted. Slides are blocked by heating them at 100^{o}C for 3 minutes, soaking in blocking solution (5xSSC, 0.1%SDS, 1%BSA) and heating at 50 ^{o}C for 30 minutes before use. Slides were rinsed 3 times in water and centrifuged for 1 minute at 1200 RPM to dry.

2.3 RNA Isolation from chicken spleen samples.

The RNeasy Midi RNA Purification Kit (Qiagen Inc., Valencia, CA) was used to obtain total cellular RNA from the spleen samples according to the manufacturer's protocol. Spleen tissue (25 mg) from each bird (4 total) was pooled for a total of 100 mg. Powdered spleen tissue was homogenized in 4 ml of Buffer RLT using a rotor

homogenizer (Get manufacturer and Location). The resulting lysate was then centrifuged for 10 minutes at 3000xg to remove cellular debris. Following centrifugation, one volume of 70% ethanol was added to the homogenized lysate, which was then applied to on RNeasy midi column. The column was centrifuged for 5 minutes at 3000xg.

An On-Column DNase Digestion was performed as the RNA was to be used in microarray experiments that were sensitive to DNA contamination. Following the On-Column DNase Digestion, 4 ml of buffer RW1 was added to the column and centrifuged for 5 minutes at 3000xg. Next, 2.5 ml of Buffer RPE was added to the column, which was centrifuged for 2 minutes at 3,000xg to wash the column. Another 2.5 ml of Buffer RPE was added, and then the column was centrifuged for 5 minutes to dry it. The RNA was eluted with 150uL of RNase-free water. The column stood for 1 minute and was then centrifuged for 3 minutes at 3,000xg. The elution step was then repeated. The purity and quality of the RNA were tested using a spectrophotometer and the Agilent RNA 6000 Nano Assay Protocol (Agilent Technologies) in the Agilent 2100 Bioanalzyer. RINs, or RNA Integrity Numbers, were determined using the bioanalyzer for each sample to confirm the quality of the sample. RINS range from 1-10, where 1 is the poorest quality RNA and 10 is the highest quality (14).

2.4 RNA Amplification

The Ambion Amino Allyl MessageAmp II aRNA amplification Kit (Ambion Inc., Austin, TX) was used amplify the mRNA found in 1 ug of total cellular RNA into amino allyl modified RNA or aRNA. Total spleen RNA underwent reverse transcription with a modified oligo (dT) primer and reverse transcriptase to synthesize essentially full-length single strands of cDNA. This cDNA underwent second strand synthesis to create

double stranded cDNA. cDNA was purified using 250ul of cDNA binding buffer. This buffer binds synthesized cDNA. The solution was passed through a cDNA filter, which was washed using 500uL of wash buffer, eliminating unincorporated nucleotides. The cDNA was eluted two times with 9ul of nuclease free water that had previously been heated to 55 ^{o}C. The purified, double stranded cDNA was then used as a template for in vitro transcription (IVT) with T7 RNA polymerase. The IVT reaction was incubated for 14 hours at 37 ^{o}C. The resulting aRNA was purified by adding 350 uL of aRNA binding buffer, passing the solution through an aRNA filter cartridge, washing the column with 650 uL of wash buffer, and eluting the aRNA with 100 uL of nuclease-free water that was preheated to 50 ^{o}C.

2.5 Labeling and Hybridization

10 ug of aRNA in 9 uL of coupling buffer was fluorescently labeled using a mixture of Alexa Fluor 555 that had been resuspended in 11 uL DMSO and incubated at room temperature for 3 hours in the dark. Labeled aRNA was purified to remove any excess dye and the concentration and the labeling efficiencies of the aRNA were determined spectrophotometrically. The labeling efficiency refers to the number of dye molecules per 100 bases. The dye: base ratio calculator (http://probes.invitrogen.com) was used to perform this calculation. Labeled aRNA was mixed with hybridization solution for a final volume of 50 uL for each sample. This mixture was incubated at 100 ^{o}C for 1 minute to denature the RNA. After this incubation, the mixture was pipetted onto blocked AIIM slides (Slide 399 and Slide 400) that were covered with NuncM Series Lifter Slips (Nunc Brand; Rochester, NY). The samples were hybridized overnight in a 2-slide hybridization chamber in a 42 ^{o}C water bath.

2.6 Washing and Scanning the AIIM Slides

After the overnight hybridization, the slides were washed to remove any nucleotides that were not bound. The washing process involved several steps. First slides were rinsed in 0.5xSSC and 0.01% SDS solution at room temperature. The next wash was in 0.2xSSC, 0.2%SDS at 42 ^{o}C for 15 minutes. During this wash, the slides are placed in a nutator to provide continuous gentle mixing. The slides were then washed 3 times at room temperature in 0.2xSSC for 1 minute. The last step in the washing process was to rinse the slides 3 times in room temperature Millipore water. The slides were centrifuged for 5 minutes at 1200RPM to dry them. After centrifugation, they were placed in 50 mL conical tubes that had been covered with foil to prevent photobleaching. Nitrogen gas was added to the tubes to help preserve the fluorescent dye. An Axon GenePix4000B scanner (Molecular Devices, Sunnyvale, CA) and a one-dye scan were used to scan the microarray slides.

2.7 Slide Analysis

Spot and background intensities were obtained using GenePix Pro 4.1 (Molecular Devices, Sunnyvale, CA). Abnormal spots, such as dust and air bubbles, were flagged as absent and removed from the analysis. Spot intensity was determined using a local background subtraction method (15) and a GenePix Results file was generated in Microsoft Excel 2007 (Microsoft; Seattle, WA). The percentage of spots that were not flagged was calculated to determine if the results were satisfactory for further evaluation. Data from the analyzed slides were imported to GeneSpring v7.0 (Silicon Genetics, Redwood City, CA).

2.8 Creation of the Illumina Library

An RNA spleen library to be used in Illumina sequencing was created using the mRNA-Seq Sample Prep Kit (Illumina, Inc., San Diego, CA). This kit is used to build libraries for both single-read and paired-end sequencing on the Genome Analyzer.

2.8.1 Purifying the mRNA

This process purifies mRNA molecules containing poly-A using poly-T oligo attached beads. Ten ug total RNA was diluted with nuclease free water to 50 uL in a 1.5 mL RNase- free non-sticky tube. The sample was heated in a pre-heated 65°C heat block for 5 minutes in order to disrupt secondary structures and then placed on ice. 15 uL of Sera-Mag oligo (dT) beads were aliqouted into a 1.5 mL RNase-free non-sticky tube. The beads were washed twice with 100 uL of Bead Binding Buffer and the supernatant was removed. The beads are then resuspended in 50 uL of Bead Biding Buffer and added to total RNA. The tube was rotated at room temperature for 5 minutes and the supernatant was removed.

While the tube was incubating, 50 uL of Binding Buffer was aliqouted to a new 1.5mL RNase-free non-sticky tube. After the 5 minute incubation, the beads were washed twice with 200 uL of Washing Buffer and the supernatant was removed. After 50 uL of 10mM Tris-HCl was added to the beads, they were heated in a preheated heat block at 80°C for 2 minutes to elute the mRNA from the beads. The tube was immediately placed on a magnetic stand and the supernatant (mRNA) was transferred to the fresh tube containing 50 uL Binding Buffer. The beads were then washed twice 200 uL of Washing

Buffer. The samples were heated a second time on the preheated 65°C for minutes to disrupt the secondary structures and then placed on ice. The iced samples were added to the washed beads and rotated at room temperature for 5 minutes.

Following this process the supernatant was removed. The beads were washed twice more with 200 uL of Washing Buffer and the supernatant was again removed. 17 uL of 10mM Tris-HCl to the beads and they were heated on the preheated heat block at 80 °C for 2 minutes to elute the mRNA from the beads. Following this step, the tube was immediately placed on the magnet stand and the supernatant (mRNA) was transferred to a fresh 200 uL thin wall PCR tube.

2.8.2 mRNA Fragmentation

mRNA is fragmented into small pieces using divalent cations at raised temperatures. A reaction mix of 5X fragmentation buffer (4uL) and mRNA (16 uL) was prepared in a 200 uL thin wall PCR tube. The tube was incubated in the thermal cycler for exactly 5 minutes at 94°C . After this incubation 2 uL of Fragmentation Stop Solution was added to the tube and the tube was placed on ice. The solution was then transferred to a 1.5 mL RNase-free non-sticky tube. 3 M NaOAC (2 uL), glycogen (2 uL), and 100% EtOH (60 uL) was added to the tube and it was incubated for 30 minutes. Following this incubation, the tube was centrifuged at 14,000 rpm for 25 minutes at 4°C. The EtOH was removed. The pellet was washed with 300 uL of 70% of EtOH and centrifuged. The pellet was air-dried for 10 minutes at room temperature and then the RNA was resuspended in 11.1 uL of RNase-free water.

2.8.3 First Strand cDNA Synthesis

RNA fragments were transcribed into cDNA using reverse transcriptase and random primers. Random primers (1 uL) and mRNA (11.1 uL) were added to a 200 uL thin walled PCR tube. The sample was incubated at 65°C for 5 minutes and then placed on ice. The following reagents were then added in the order listed: 5X First Strand Buffer (4 uL), 100mM DTT (2 uL), 25mM dNTP Mix (0.4), and RNase Inhibitor (0.5 uL). 6.9 uL of this mixture was added to the PCR tube and mixed well. The sample was then 25°C for 2 minutes. One uL Superscript II was added to the sample and it was incubated in the thermal cycler with the following program: 25°C for 10 minutes, 42°C for 50 minutes, 70°C for 15 minutes, hold at 4°C. After this incubation, the tube was placed on ice.

2.8.4 Second Strand cDNA Synthesis

Second strand cDNA synthesis is achieved by removing the RNA template and synthesizing a replacement strand, generating double-stranded cDNA. Ultra pure water (62.8 uL) was added to the first strand cDNA synthesis mix. GEX Second Strand Buffer (10 uL) and 25 mM dNTP mix (1.2 uL) were added to the mix and then it was incubated on ice for 5 minutes. After the incubation on ice, RNase-H (1 uL) and DNA Pol I (5 uL) were added and the mix was incubated at 16°C in the thermal cycler for 2.5 hours. The sample was purified using the QIAquick PCR Purification Kit (QIAGEN Inc., Valencia, CA) and eluted in 50 uL of QIAGEN EB Buffer.

2.8.5 Performing End Repair

During this step, the frayed or overhanging ends are converted into blunt ends using T4 DNA polymerase and Klenow DNA polymerase. The 3' to 5' exonuclease activity of these enzymes eliminates the 3' overhangs and the polymerase activity fills in the 5' overhangs. The following reaction mix was prepared in a 1.5 mL RNase-free non-sticky tube: eluted DNA (50 uL), water (27.4 uL), 10X End Repair Buffer (10 uL), 25 mM dNTP Mix (1.6 uL), T4 DNA Polymerase (5 uL), Klenow DNA Polymerase (1 uL), and T4 PNK (5 uL). This mix was incubated in a heat block at 20°C for 30 minutes. The QIAquick PCR Purification Kit (QIAGEN Inc., Valencia, CA) was used to purify the sample and then it was eluted in 32 uL of QIAGEN EB Buffer.

2.8.6 Adenylating 3' Ends

This step adds an 'A' base to the 3' end of the blunt phosphorylated DNA fragments, taking advantage of the polymerase activity of Klenow fragment. This readies the DNA fragments for ligation to the adaptors, which have one 'T' base overhang on the 3' end. The following reaction mix was prepared in a 1.5 mL RNase-free non-sticky tube: eluted DNA (32 uL), A-Tailing Buffer (5 uL), 1mM dATP (10 uL), and Klenow exo (3' to 5' exo minus) (3 uL). The sample was incubated in a heat block at 37°C for 30 minutes. The MinElute PCR Purification Kit (QIAGEN Inc., Valencia, CA) was used to purify the sample and it was eluted in 23 uL of QIAGEN EB Buffer.

2.8.7 Ligating the Adapters

This step ligates adapters to the end of the DNA fragments, preparing them to be hybridized to a single read flow cell. The following mix was prepared in a 1.5 mL

RNase-free non-sticky tube: eluted DNA (23 uL), 2X Rapid T4 DNA Ligase Buffer (25 uL), PE Adaptor Oligo Mix (1 uL), and T4 DNA Ligase (1 uL). The mix was incubated at room temperature for 15 minutes, purified using the MinElute PCR Purification Kit (QIAGEN Inc., Valencia, CA), and eluted in 10 uL of QIAGEN EB Buffer.

2.8.8 Purifying the cDNA Templates

This step purifies the products of the ligation reaction on agarose gel to select a size range for enrichment. A 50 mL, 2% agarose gel was run with distilled water and TAE. The samples were loaded into a gel and run at 120 V for 1 hr. A region of the gel containing the material in the 200 bp range was cut with a clean gel excision tip. The QIAquick Gel Extraction Kit (QIAGEN Inc., Valencia, CA) was used to purify the DNA sample, which was eluted in 30 uL of QIAGEN EB Buffer.

2.8.9 Enriching the Purified cDNA

This process uses PCR to amplify the cDNA library. The PCR was performed with two primers that anneal to the ends of the ligated adapters. The following PCR reaction was mixed in a 200 uL thin wall PCR tube: 5X Phusion Buffer (10 uL), PCR Primer PE 1.0 (1 uL), PCR 2.0 (1 uL), 25 mM dNTP Mix (0.5 uL), Phusion DNA Polymerase (0.5 uL), and water (7 uL). Thirty uL of purified ligation mix was added to the 200 uL PCR tube. The cDNA was amplified using the following PCR process: 30 seconds at 98°C, and 15 cycles of10 seconds at 98°C, 30 seconds at 65°C, 30 seconds at 72°C, and 5 minutes at 72°C. The QIAquick PCR Purification K (QIAGEN Inc., Valencia, CA) it was used to purify the sample and it was eluted into 30 uL of QIAGEN EB Buffer.

2.9 Illumina Sequencing

Illumina Sequencing involves four main steps. These include sample preparation, cluster generation, sequencing by synthesis, and data analysis. Sample preparation methods have been discussed previously. Cluster generation takes approximately 5 hours. The Cluster Station (Illumina Inc., San Diego, CA) amplifies the DNA from the samples on the flow cell surface to create clusters containing colonal copies. The resulting array of templates are sequenced using the Genome Analyzer (Illumina Inc., San Diego, CA). Sequencing by synthesis is performed in parallel with novel reversible terminator nucleotides that have been fluorescently labeled. Illumina has a standard pipeline to analyze the data generated but it was not used in this case.

2.10 Illumina Analysis

The initial data analysis took place in Dr. Carl Schmidt's lab using the TopHat and Bowtie softwares. Bowtie is a fast, memory effective short read aligner. It aligns and remaps the output to the avian genome. TopHat is a fast splice junction mapper, which maps the splice junctions. They are also able to recognize tags from one axon to the next.

Output from Bowtie and TopHat is used with Pearl scripts created by Dr. Schmidt. The scripts map to known or unknown sequence tags. The scripts are run on UDCANR Big Bird. This is a multi terabyte storage system. There are 8 processors with 64 GB RAM. This was output to GBrowse as a log plot. GBrowse allows for online access to information about the chicken, *Gallus gallus,* including genomic information. The GBrowse server is a valuable resource for researchers seeking genomic information about avian species (32).

Three spreadsheets were generated. The CalSpleen file includes the gene ID, the symbol, the raw counts, the counts per million, counts per million per kilobase, and the gene description. The GODavid spreadsheet is a simple GO analysis of the detected gene products. The Reactome Analysis has two sheets. One of these lists all the genes that have Reactome pathways. The other sheet shows the pathways and the probability that the pathway is "significantly represented" above the background list of pathways.

2.11 Comparison of Illumina and AIIM Results

Comparison of these two methods was carried out on a global level, in terms of housekeeping genes, and the JAK-STAT Pathway. The top 100 most highly expressed genes from the Illumina Sequencing were compared the genes that were on the AIIM. Gene IDs were used to for this comparison. The AIIM database was searched for the gene ID corresponding to the top 100 most highly expressed. If the genes was present, the normalized intensity value of the spot on the slide was compared to the number of counts per million per kilobase detected in the Illumina analysis. The second method of comparison involved 7 housekeeping genes that are found on the AIIM. The spreadsheets generated in the initial Illumina data analysis were searched for gene IDs corresponding to these housekeeping genes. If the gene was present, expression levels were again compared. Lastly, the two methods were compared based on the presence and expression levels of genes involved in the JAK-STAT Pathway.

Chapter 3

RESULTS

3.1 Microarray Results

Duplicate AIIM slides were hybridized to aRNA from the pooled spleen sample of four, four week old male chickens as described in the materials and methods. Slides were scanned using an ArrayWoRk scanner (Applied Precision, Issaquah, WA). Slide 399 was selected to be aligned using the GenePixPro Software Program (Molecular Devices, Silicon Valley, CA). 2,219 of 4,959 elements, or approximately 30%, of genes on Slide 399 passed with at least one of three replicate spots. Typically, it is expected that at least 75% of slide pass with one of three replicates. Therefore, Slide 157, was referenced for comparative analysis. Slide 157 was made by Lorna Dougherty using chicken spleen. 3,471 of 4959 elements, or approximately 70%, of genes on Slide 157 passed with at least two of three replicate spots (14). A comparison of the ten most highly expressed elements from Slide 157 and Slide 399 is shown in Table 3.1. As seen in the table, Slide 157 had higher levels of gene expression overall based on normalized intensity levels. Based on this observation, Slide 157 was used in place of Slide 399 in all future comparisons.

Table 3.1 Comparison of the 10 Most Highly Expressed Genes on 2 AIIM Slides. The ten genes with the highest fluorescent intensities from Slide 157 and Slide 399 were compared based on normalized intensity levels from GeneSpring. GeneSpring defines normalization as a method of standardization to enable differentiation between biological variations in gene expression levels from variations due to measurement process. Slide 157 was used in place of Slide 399 in all future comparisons because of its higher intensity levels.

Slide 399		Normalized Intensity Level	Slide 157		Normalized Intensity Level
Clone ID	Symbol		Clone ID	Symbol	
LPSm_F24	LSM5	19	LPSd_M05	N/A	27.48
Controle_N07	LTA4H	19	LPSk_L16	CLASP2	27.13
LPSf_I16	N/A	17	pmp1c.pk002.c05	N/A	17.69
Controlc_P04	LOC100087232 9	13	LPSh_F24	DPP4	16.92
Controlc_L08	N/A	11	LPSk_N21	STK4 RCJMB04_17i 1	16.33
IFNm_C07	LOC100221195	6	Control_C14	RHOA	12.39
IFNd_J22	LYSMD3	4	IFNi_J13	N/A	11.83
IFNm_I06	LDHA	0.01	LPSb_D07	N/A	11.77
LPS_H02	PKN2	0.01	LPSm_B19	N/A	11.44
IFNm_F23	C15	0.01	LPSl_A12	MBNL3	11.33

*N/A refers to a clone id that does not have a corresponding Symbol in the Array Database

GeneSpring Analysis of the slide allowed for the identification of most highly expressed genes on the array in this experiment. These genes were compared to a list of the most highly expressed genes from the Illumina sequencing data. The genes with the highest florescent intensities were compared with the genes that had the highest number of reads in Illumina. Following this, microarray results were compared to Illumina sequencing results based housekeeping genes (ABHD12, RPL35A, RPL9, MAF1, AFF4, LOC413586, PMM2) and the JAK-STAT pathway based on the same principles.

3.2 Illumina Results

The Illumina spleen library was created from the pooled spleen of four, 4 wk old male chickens using the mRNA-Seq Sample Prep Kit as described in Materials and Methods. Analysis of data generated by the Genome Analyzer, found that 503 million raw counts out of 511 million counts matched to the chicken genome. There are approximately 20,000-23,000 genes in the entire chicken genome The top 20 most highly expressed genes (counts per kilobase per million), along with their functions, can be seen in Table 3.1. Counts per kilobase per million is normalization process that takes the raw counts, how many times a particular sequence was seen on the machine, A majority of the top 20 most highly expressed genes had immunological or ribosomal functions.

Table 3.2 Top 20 Most Highly Expressed Genes from Illumina's Genome Analyzer

Gene ID	Symbol	Gene Function	Counts per KB per Million
416928	IGLL1	Immunoglobulin lambda like polypeptide 1	22022.94
395970	FTH1	Ferritin, heavy polypeptide	7238.64
414741	CD74	major histocompatibility complex, class II invariant chain	5871.69
769283	LOC769283	immunoglobulin, light chain, lambda, psi-V8 pseudogene	5411.27
373963	EEF1A1	eukaryotic translation elongation factor 1 alpha 1	5036.46
416929	LOC416929	immunoglobulin, light chain, lambda, psi-V7 pseudogene	4690.95
408047	TMSB4X	thymosin beta 4, X-linked	4269.03
724083	BLB1	MHC class II antigen B-F minor heavy chain	4145.27
693256	BLB2	MHC class II beta chain	4060.51
769981	RPL37A	ribosomal protein L37a	4012.67
769305	LOC769305	similar to immunoglobulin lambda chain	3921.81
423726	RPS24	ribosomal protein S24	3499.4
768348	LOC768348	similar to MHC class I antigen	3449.86
396148	RPS6	ribosomal protein S6	3443.69
374117	IGJ	immunoglobulin J polypeptide, linker protein for immunoglobulin alpha and mu polypeptides	3420.95
420001	RPL23	ribosomal protein L23	2822.74
430990	RPS20	ribosomal protein S20	2727.75
396425	UBC	ubiquitin C	2507.3
373957	RPL6	ribosomal protein L6	2452.48
396363	TPT1	tumor protein, translationally-controlled 1	2436.9

3.3 Comparison of Illumina and Microarray Results

This comparison was facilitated by gene ids. The array database contains gene id numbers as well as other information about the elements on the AIIM. 3,764, or 76%, of genes on the array have gene ids. The reason that the some of the genes on the array do not have ids is because the array contains unique elements are not detected when BLAST is performed. In order to simplify things, the top 100 most highly expressed genes in Illumina were compared the array.

Of these 100 genes, 77 were present on the array. Slide 157 was used in this comparison. In order to make a more fair comparison of microarray expression data, expression data from a better quality chicken spleen slide (Slide 157) has been included. This data can be seen in Table 3.2. Of 77 genes that were present, 69 showed some expression level. However, only 13 genes showed expression levels greater than 1.0. There does not seem to be clear trend between expression levels on the array and in the Illumina sequencing data. The most highly expressed gene in the Illumina data was IGLL1 (immunoglobulin lambda-like polypeptide 1), whereas the most highly expressed gene from Slide 157 was LPSd_M05 (Symbol Unavailable). Another issue is that the microarray slide that is being used for the majority of the comparison was not generated from the same sample as the Illumina data as the original slide was. Though these

samples are both from chicken spleen, there could be some differences, such age of the birds or their lineage, which could cause differences in gene expression.

Table 3.3 Comparison of Illumina and AIIM. The Top 100 most highly expressed genes from the Illumina data analysis from to normalized intensity levels from the AIIM. An X indicates that the gene was present on the AIIM. Expression levels generated by both technologies are also provided. ~ Indicates that the gene was not present on the array and thus was not able to show expression on it.

Gene ID	Symbol	Expression Level in Illumina (Cts Per KB per Million)	Present on Array	Normalized Intensity Level (Slide 157)
416928	IGLL1	22022.94	~	~
395970	FTH1	7238.64	X	0
414741	CD74	5871.69	X	0
769283	LOC769283	5411.27	~	~
373963	EEF1A1	5036.46	X	0.51
416929	LOC416929	4690.95	~	~
408047	TMSB4X	4269.03	X	0
724083	BLB1	4145.27	~	~
693256	BLB2	4060.51	~	~
769981	RPL37A	4012.67	~	~
769305	LOC769305	3921.81	~	~
423726	RPS24	3499.4	X	0
768348	LOC768348	3449.86	~	~
396148	RPS6	3443.69	X	0.80
374117	IGJ	3420.95	~	~
420001	RPL23	2822.74	X	0
430990	RPS20	2727.75	X	0.23
396425	UBC	2507.3	X	0.78
373957	RPL6	2452.48	X	0.97
396363	TPT1	2436.9	X	0.38
427703	LOC427703	2430.34	~	~
425389	BF2	2288.68	X	0.38
424931	PTMA	2210.15	X	1.06
424584	RPS8	2183.66	X	0.29
693260	HLA-G	2110.02	X	0
373937	RPL22	2096.25	X	0.37
419049	RPS11	2077.03	X	1.14
396280	RPL27	2072.21	X	0.03
419904	RPS10	1931.41	X	0.09
374133	RPL35	1889.09	X	0.05
418568	RPL8	1850.92	X	0.63
395835	RPLP0	1843.12	X	0.21
420182	RPL7	1807.11	X	1.18

396448	RPS15	1792.56	X	0.23
422509	LOC422509	1769.91	~	~
421698	RPS12	1761.35	X	0.71
420003	RPL19	1733.8	X	0.61
395269	RPL5	1706.5	X	0.11
Gene ID	**Symbol**	**Expression Level in Illumina (Cts Per KB per Million)**	**Present on Array**	**Normalized Intensity Level (Slide 157)**
414830	B2M	1692.62	X	0.058
374132	RPL39	1665.56	X	0.07
395849	RPL13	1632.69	X	0.71
374134	RPL14	1628.22	~	~
769505	LOC769505	1624.22	~	~
415551	RPL4	1604.57	~	~
427323	RPS23	1591.66	X	0.48
430997	RCJMB04_7l20	1563.96	X	0.36
416930	LOC416930	1513.66	~	~
417574	RPL23A	1508.51	X	0.12
396001	RPS4X	1498.51	X	0.39
396536	APOA1	1459.74	~	~
418016	RPL3	1447.41	X	0.08
772115	LOC772115	1445.26	~	~
427675	RPS15A	1423	~	~
419069	RPS3	1416.52	X	0.45
426845	RPL17	1413.03	X	0.27
419682	RPL11	1403.12	~	~
425416	RPL30	1367.44	X	1.66
768852	LOC768852	1356.45	~	~
416544	RPS2	1339.04	X	0.52
395723	RCJMB04_32c11	1328.81	X	4.32
417158	RPL7A	1313.91	X	2.5
395796	RPS27A	1287.42	X	1.77
428442	RPL15	1276.5	X	0
395916	NME2	1266.24	X	1.08
417264	RPL12	1258.32	X	2.29
419895	RPL10A	1245.81	X	0.09
418401	RPL24	1238.5	X	0.06
769327	LOC769327	1234.93	~	~
770777	LOC770777	1191.67	~	~
396203	NPM1	1190.1	X	0.50
373936	RPL36	1127.41	X	0.19
417044	GNB2L1	1082.84	X	2.46
374163	HSP90B1	1066.36	X	0.28
417871	RPS16	1062.81	X	0.02
395912	MGP	1027.6	X	0.82
396262	RPLP1	1024.64	X	0.17
395602	PSAP	1006.67	X	0.27

396447	PPIB	1001.91	X	0.43
769550	LOC769550	999.11	X	0.17
427186	RPL37	987.95	~	~
415448	RCJMB04_7b15	967.99	X	0.01
771528	RPL38	921.77	X	0.73
771147	LOC771147	905.3	X	0.04

Gene ID	Symbol	Expression Level in Illumina (Cts Per KB per Million)	Present on Array	Normalized Intensity Level (Slide 157)
770018	RPL27A	879.14	X	0.11
396325	EEF2	812.57	X	1.08
770722	RPS25	810.66	X	0.60
418710	RPL31	803.97	X	0.13
396190	UBB	797.49	X	0.10
425657	CTSS	792.85	X	0.18
422477	RCJMB04_13j13	784.69	X	0.77
396201	NCL	769.3	X	0.63
374072	SLC25A6	768.23	X	0.54
423668	RSFR	764.8	X	1.26
374053	RPS17	751.95	X	3.48
769368	EEF1D	748.5	X	1.22
395232	EIF4A2	743.77	X	1.59
770612	LOC770612	742.65	X	0.08
396017	RCJMB04_24e12	723.75	X	0.08
415549	MAP2K1	720.39	X	0
420627	RCJMB04_2g17	719.28	X	0.11

3.3.1 Comparison via Housekeeping Genes

The second way the two methods were compared was in terms of housekeeping genes. Housekeeping genes, or control genes, are typically used to normalize mRNA levels between different samples. This is because housekeeping genes are expressed at a constant level to maintain cellular function. The genes that were used were ABHD12, RPL35A, RPL9, MAF1, AFF4, LOC413586, and PMM2. Table 3.3 shows a comparison of seven housekeeping on AIIM and with Illumina. There was not a clear trend that could be seen between the Illumina data and the AIIM data. For example, the most highly expressed housekeeping gene on the array was LOC4134586, whereas the most highly expressed housekeeping gene was RPL35A.

3.3.2 Comparison Based on JAK-STAT Pathway

The final method of comparison made use of the Janus Kinase/signal transducers and activators of transcription JAK-STAT Pathway. This pathway is believed to integrate functions of several separate signaling cascades (29). It was useful for comparison since had a manageable number of gene. In addition, many of the genes contained in this pathway are present on the array. An overview of the pathway can be seen in Figure 3.1. A comparison of the genes from this pathway that were present on

array and expression levels can be seen in Table 4.1. 10 of 24 genes, or 42%, of those in the JAK/STAT pathway showed expression on the AIIM. 23 of 24 genes, or 96%, of genes in the JAK/STAT pathway showed expression in the Illumina data.

Table 3.4 Comparison of Illumina Sequencing Technology and AIIM based on Housekeeping Genes.

Control Gene	Gene ID	Present on Array	Normalized Intensity Levels (Slide 157)	Present in Illumina Data	Expression Level (Cts per KB per Million)
ABHD12	421249	X	0.21	X	31.84
RPL35A	424924	X	1.23	X	303.89
RPL9	425468	X	1.46	X	286.63
MAF1	610210	X	0.25	X	286.63
AFF4	416337	X	2.09	X	0.89
LOC413586	413586	X	4.06	X	0.21
PMM2	427679	X	0.94	X	16.63

Table 3.5 Comparison Illumina Sequencing Technology and AIIM Based on JAK-STAT Pathway.

Gene Name	Gene ID	Present on Array	Expressional Level (Slide 157)	Present in Illumina Data	Expression Level (Cts per KB per Million)
AKT1	395928	~	~	X	11.6
BCL2L1	373954	X	0.12	X	21.58
CBL	374038	X	0.24	X	7.97
CCND2	374047	X	0.57	X	3.56
CISH	395335	~	~	X	2.91
CREBBP	416667	X	2.53	X	14.86
GH	378781	~	~	X	3.97
GRB2	386572	~	~	X	88.28
IFNG	396054	X	0.08	X	0.03
IL15	395258	X	0.76	X	0.02
IL6	395337	X	0.03	X	0.02
JAK2	374199	X	0.02	X	1.65
LEPR	374223	~	~	X	0.11
MYC	420332	~	~	X	34.05
PIAS2	416383	X	0.28	X	3.72
PIK3R5	417319	X	0.46	X	21.5
PTPN11	395815	~	~	X	24.5
PTPN6	427947	~	~	X	198.71

SOCS2	395219	~	~	X	9.78
SOS2	423572	~	~	X	3.4
SPRED2	421271	~	~	X	2.75
SPRY1	395583	~	~	~	~
STAM	420517	~	~	X	0.14
STAT5B	395556	~	~	X	34.95

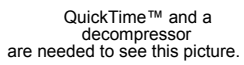

Figure 3.1 Overview of the JAK-STAT Pathway

Chapter 4

Discussion

4.1 Using Microarray Technology to Monitor Gene Expression

Arrays have led to many important advances in solving biological problems across the board, including the identification of gene expression differences between diseased and healthy tissues, responses as a result of drugs and the evolutionary development of gene regulation (28.) In addition, it is relatively accessible, cost effective technology. No highly specialized equipment is needed for hybridization. Data capture also can be carried out with equipment that is common in many labs. Another major advantage of microarray technology is that a standards confirming and publishing data have been established (11).

Microarray data is typically quantified or confirmed using qRT-PCR. Confirmational studies allow for discrepancies between gene expression between the microarray data and the independent PCR data to be identified (24). The Minimum Information About a Microarray Experiment (MIAME) guidelines outline the information must be provided to allow for the unambiguous interpretation of a microarray-based experiment (11). Acceptance of these guidelines has ensured that data used to support published conclusions is made available to researchers in a way that makes these data

meaningful (11). In this case, qRT-PCR confirmation was not performed since the microarray data was intended for journal publication.

As mentioned previously, microarray technology has definite advantages. The age of the technology is in its self somewhat of an advantage as this has allowed for the standardization of data produced in microarray publications. However, there are several limitations are inherent with the microarray process. Some of these limitations include sample purity, length of the protocol, variability, and the number of genes on the array. An example of this variability has been seen in this project. Five slides were made. Of these slides, one was completely blank, and the others were of differing quality. Some of this variability arises from background hybridization, hybridization that occurs (irrespective of the corresponding transcript's expression level (25). This can lead to accuracy problems with expression levels, especially for those transcripts that are low in abundance (25). As a result, though hybridization results across arrays can identify gene expression differences among samples (5), hybridization results from a single sample may not be able to provide an accurate measure of the relative expression of different transcripts (25).

Lastly, arrays are limited by the number of genes they contain as well as types of genes they contain. Thus, the largest limitation of arrays in comparison with Illumina is the number of genes that can be monitored. For example, the AIIM used in this research only contains 4,949 elements, whereas 20,000-23,000 genes in the chicken genome can be monitored with Illumina technology.

4.2 Using Illumina Technology to Monitor Gene Expression.

Recent improvements in the cost, efficiency, and quality of genome-wide sequencing have caused ultra-high through put sequencing (second generation or next generation sequencing) methods, such as Illumina's Genome Analyzer, to garner favor (7). Sequencing based approaches for monitoring gene expression have the potential to over come the limitations of the microarray (24).

A brief overview of an mRNA-Seq experiment is as follows. Experimental samples are used to prepare a library. Creation of the library involves several steps to convert the experimental RNA into small fragments of DNA that can be sequenced by the Illumina machine. These steps include poly-A RNA isolation, RNA fragmentation, reverse transcription to cDNA using random primers, adapter ligation, size selection from a gel, and PCR enrichment. The cDNA library is then put on one of the eight lanes of a flow-cell (7). Each individual cDNA fragment is attached to the surface of the lane and undergoes an amplification step. After this step, the fragments are converted to clusters of double stranded DNA (7). The flow-cell is then put in the sequencing machine. Once in the machine, each cluster is sequenced in parallel (7). In addition, at each cycle, the four fluorescently labeled nucleotides are added and the signals that are produced are recorded. For each flow-cell, this process is repeated for a specific number of cycles (7). The number of cycles established the length of the reads whereas the number of clusters determines the number of reads (7.)

In this experiment, Illumina sequencing technology was used to monitor gene expression in control chicken spleen samples. Illumina sequencing technology provided expression information on a broader level, as it was able to monitor expression of all the genes in the chicken genome. However, in order to confirm that the expression levels

provided by Illumina technology were correct, the experiment should be repeated and qRT-PCR should be performed.

A protocol has been suggested for monitoring gene expression with Illumina Sequencing Technologies. In this protocol, PCR was used to quantify both microarray and Illumina data. Though standardized guidelines for publishing Illumina sequencing have not yet been established, this is a step in right direction.

4.3 Summary Comparison of Microarray Technology and Illumina Sequencing Technology.

Other studies have shown that the information in a single lane of Illumina sequencing data appears to comparable to that in a single array for allowing identification of differentially expressed genes (24). In addition, Illumina sequencing technology allows for analyzes of low-expressed genes, alternative splice variants, and novel transcripts. The results of this experiment showed that Illumina sequencing technology could be used to monitor expression levels in control chicken spleen. Though Illumina sequencing technology gave good results, the fact that is has not yet been quantified or confirmed must be kept in minds. In order to prove that the data generated was "real", it would be beneficial to perform PCR.

Microarray have already been established as a standard method of monitoring gene expression. The microarray results in this particular experiment were not used in the comparison of the two technologies because of the low number of spots that passed on Slide 399. Also, there was a large amount of variability in the slides that were produced for the experiment. Reproducibility of microarray experiments is something that should be taken into account. Hybridization is also sometimes an issue with microarrays. There

can either be too little hybridization or cross hybridization. Both of these issues affect the accuracy of the gene expression levels.

The technical variance associated with Illumina sequence has shown to be slight. That is results of the experiments have not been seen to vary much. In an experiment conducted by Marioni et al, it was found that the results from one lane of Illumina sequencing data were comparable to the results from a single microarray for the purpose of identifying differently expressed genes. This study also found that Illumina sequencing is highly replicable. However, there has yet to be a published study that includes repetition of the processing step of the Illumina sequencing library, i.e. creating multiple libraries using the same sample. This kind of study would be costly and time consuming but may be beneficial in establishing the overall reproducibility of the Illumina sequencing technology protocol.

In summary, Illumina sequencing technology looks to be a promising method for monitoring gene expression as well as recognizing differentially expressed genes. Though further studies that prove the overall reproducibility and accuracy of Illumina sequencing technology still need to be carried out, this technology seems comparable, and perhaps superior in some ways to the existing array technology.

RESOURCES

1. Gene Expression and proteomics. Fundamental Genomics Research. http://ec.europa.eu/research/health/genomics/tabs/gene_exp_prot_en.htm. 2004.

2. Illumina Sequencing Technology. Illumina, Inc. http://www.illumina.com/technology/sequencing_technology.ilmn. 2009

3. Illumina. mRNA Expression Analysis. Illumina. Inc. www.illumina.com/documents/products/.../datasheet_mrna_expression.pdf. 2009.

4. Microarrays: Chipping away at the mysteries of science and medicine. NCBI. http://www.ncbi.nlm.nih.gov/About/primer/microarrays.html. 2007.

5. Allison, D., Cui, X., Page, G., Sabripour, M. Microarray data analysis: From disarray to consolidation and consensus. Nat. Rev. Genet. 7:55–65. 2006

6. Beebee T and Burke J. Gene Structure and Transcription. 2nd edition. Oxford University 1992. Pg 80

7. Baumbusch LO, Aarøe J, Johansen FE, Hicks J, Sun H, Bruhn L, Gunderson K, Naume B, Kristensen VN, Liestøl K, Børresen-Dale AL, Lingjaerde OC. Comparison of the Agilent, ROMA/NimbleGen and Illumina platforms for classification of copy number alterations in human breast tumors. BMC Genomics. 9:379. 2008 Aug 8

8. Bliss T. W., Dohms J. E., Emara M.G., and Keeler C. L. Gene expression profiling of avian macrophage activation. Vet. Immunol. Immunopathol. 105:289-299. 2005.

9. Borish L. C., and Steinke, J. W. Cytokines and chemokines. J Allergy Clin Immunol 11:S460-S475. 2003

10. Bullard James H, Purdom Elizabeth, Hansen Kasper D, and Dudoit Sandrine. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics. 11: 94. 2010

11. Brazma A. Minimum Information About a Microarray Experiment (MIAME)--successes, failures, challenges. ScientificWorldJournal. 9:420-3. 2009

12**.** Burgess JK and Hazelton RH. New developments in the analysis of gene expression. Redox Rep 5: 63–73, 2000.

13 Chung, I. Evaluation of the transcriptional response of chicken spleen to highly pathogenic avian influenza (H5N1). M.S. diss., University of Delaware, 2009.

14. Copois, V., Bibeau, F., Bascoul-Mollevi, C., Salvetat, N., Chalbos, P., Bareil, C., Candeil, L., Fraslon, C., Conseiller, E., Granci, V., Maziere, P., Kramar, A., Ychou, M., Pau, B., Martineau, P., Molina, F., and Del Rio, M. Impact of RNA degradation on gene expression profiles: Assessment of different methods to reliably determine RNA quality. J. Biotechnol. 127:549-559. 2007.

15. Dalloul RA, Bliss TW, Hong Y-H, Ben-Chouikha I, Park DW, Keeler CL, Lillehoj HS: Unique responses of the avian macrophages to different species of *Eimeria*. Molec. Immunol. 44: 558-566. 2007.

14. Dougherty L. Multi-Species Use of an Avian Microarray. Senior Thesis, University of Delaware. 2009.

15. Green E.D. Strategies for the systematic sequencing of complex genomes. Nature Reviews Genetics 2. 573-583. 2001.

16. Gershon D. Microarray technology: an array of opportunities. Nature 416: 885–891, 2002.

17. Jeremy L. John. The Avian Spleen: A Neglected Organ. The Quarterly Review of Biology. Vol. 69, No. 3 pp. 327-351 (Sep., 1994)

18 Hawkins, John D. Gene Structure and Expression. 1996. Third Edition.

19. Hughes, T.R., Marton M.J., Jones, A.R, Roberts, C.J., Stoughton, R., Armour, C.D, Bennett, H.A., Coffey, E., Dai, H., He, Y.D., Kidd, M.J., King, A.M., Meyer, M.R., Slade, d., Lum, P.Y, Stepaniants, S.B., Shoemaker, D.D, Gachotte, D., Chakraburtty, K., Simon J., Bard, M., and Friend, S.H. Functional Discovery via a compendium of expression profiles. 2000.

20. Keeler, C. L., Bliss, T. W., Lavric, M., and Maughan, M. N. A functional genomics approach to the study of avian innate immunity. 2007.

21. Kim CH, Lillehoj HS, Bliss TW, Keeler CL Jr, Hong YH, Park DW, Yamage M, Min W, Lillehoj. “Construction and Application of an avian intestinal intraepithelial

lymphocyte cDNA microarray (AVIELA) for gene expression profiling during Eimeria maxima infection." EP.Vet Immunol Immunopathol. 124(3-4): 341-54. 2008.

22. Kogut MH, Swaggerty C, He HQ, Pevzner I, Kaiser P: Toll-like receptor agonists stimulate differential functional activation and cytokine and chemokine gene expression in heterophils isolated from chickens with differential innate responses. Microbes and Infect. 8:1866-1874. 2006.

23. Lorkowski, Stefan and Cullen, Paul. Analyzing Gene Expression, A Handbook of Methods: Possibilities and Pitfalls. WILEY-VCH. 2003. Pg 3, 5, 167, 345, 411.

24. Marioni John C., Mason Christopher E., Mane Shrikant M., Stephens Matthew, and Gilad Yoav. RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. Genome Res.18(9): 1509–1517. 2008

25. Murphy, David. Gene Expression Studies Using Microarrays: Principles, Problems, and Prospects. Adv Physiol Educ. 2002

26. Passador-Gurgel, G., Hsieh, W., Hunt, P., Deighton, N., Gibson, G. Quantitative trait transcripts for nicotine resistance in *Drosophila melanogaster*. Nat. Genet. 2007

27. Perdew Gary H., Heuvel John P. Vanden, Peters Jeffery M. Regulation of Gene Expression. Molecular Mechanisms. Humana Press, Totowa, NJ. 11. 2006. Pg 11

28. Primrose S.B, Twyman R.M. Principles and Genome Analysis and Genomics. (3rd Edition). Blackwell Science, Oxford. 2003. Pg 9, 92

29. Rane, Sushil G., Reddy Premkumar E. Janus kinases: components of multiple signaling pathways.

30. Rifkin, S., Kim, J., White, K. Evolution of gene expression in the *Drosophila melanogaster* subgroup. Nat. Genet. 33:138–144. 2003

31. Scherf, U., Ross, D.T., Waltham, M., Smith, L.H., Lee, J.K., Tanabe, L., Kohn, K.W., Reinhold, W.C., Myers, T.G., Andrews, D.T. A gene expression database for the molecular pharmacology of cancer. Nat. Genet. 2000; 24:236–244.

32. Schmidt, C. J., Romanov, M., Ryder, O., Magrini, V., Hickenbotham, M., Glasscock, J., McGrath, S., Mardis, E., Stein, L.D. Gallus GBrowse a unified genomic database for the chicken Nucleic Acids Research. 36, D719-D723., in Nucl. Acids Res. 2008

33. Shi, L., Reid, L., Jones, W., Shippy, R., Warrington, J., Baker, S., Collines, P., de Longueville, F., Kawasaki, E., Lee, K., et al. The MicroArry Quality Control (MAQC)

project shows inter- and intraplatform reproducibility of gene expression measurements. Nat. Biotechnol. 24:1151–1161. 2006

34. Silver Nicholas**,** Best Steve**,** Jiang Jie, and Thein Swee Lay. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR

35. Steemers FJ, Gunderson KL. Illumina, Inc. Pharmacogenomics. 6(7): 777-82. 2005

36. Sugarbaker, D., Richards, W., Gordon, G., Dong, L., De Rienzo, A., Maulik, G., Glickman, J., Chirieac, L., Hartman, M., Taillon, B., et al. Transcriptome sequencing of malignant pleural mesothelioma tumors. Proc. Natl. Acad. Sci. 105:3521–3526. 2008

37. Swaggerty, CL, Kaiser P, Rothwell L, Pevzber IY, Kogut MH: Heterophil cytokine mRNA profiles from genetically distinct lines of chickens with differential heterophil-mediated innate immune responses. Avian Pathol. 35:102-109. 2006

38. t Hoen PA, Ariyurek Y, Thygesen HH, Vreugdenhil E, Vossen RH, de Menezes RX, Boer JM, van Ommen GJ, den Dunnen JT. Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. Nucleic Acids Res. 2008.

39. Torres, T., Metta, M., Ottenwälder, B., Schoötterer, C. Gene expression profiling by massively parallel sequencing. Genome Res. 1:172–177. 2008

40. Trapnell C., Pachter L, and Salzberg S.L. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 25:1105-11. 2009.

41. Vainio, O., and A. Toivanen. 1987. Cellular co-operation in immunity. In A. Toivanen and P. Toivanen (eds.), Avian Immunology: Basis and Practice, Vol. 2, pp. 1-12. 1987.

42. Weber, A., Weber, K., Carr, K., Wilkerson, C., Ohlrogge, J. Sampling the *Arabidopsis* transcriptome with massively parallel pyrosequencing. Plant Physiol. 144:32–42. 2007

43. White, K. Functional genomics and the study of development, variation and evolution. Nat. Rev. Genet. 2:528–537. 2001