8-2018

Genomic Approaches to Identify Important Traits in Avian Species

Bhuwan Khatri
University of Arkansas, Fayetteville

Genomic Approaches to Identify Important Traits in Avian Species

A dissertation submitted in partial fulfillment
of the requirements for degree of
Doctor of Philosophy in Poultry Science

by

Bhuwan Khatri
Tribhuvan University
Bachelor in Science in Microbiology, 2005
Tribhuvan University
Master in Science in Medical Microbiology, 2008

August 2018
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

_____
Byungwhi Caleb Kong, Ph.D.
Dissertation Director

_____
Walter G. Bottje, Ph.D.
Committee Member

_____
Young M. Kown, Ph.D.
Committee Member

_____
Charles F. Rosenkrans Jr., Ph.D.
Committee Member

ABSTRACT

ABSTRACT

This dissertation focusses on identifying different molecular markers that have impact on overall poultry production. Chapter one reviews microRNA (miRNA), copy number variation (CNV) and single nucleotide polymorphism (SNP) as markers suggested in different avian species by various studies. It reviews modern genomic approaches that are employed for next generation sequencing data analysis and verification.

Chapter two seeks to identify and validate the muscle specific miRNAs in the breast muscle of modern broilers and its foundational chicken line. Small RNA sequencing was performed to identify differentially expressed mature miRNAs in the breast muscles of these two chicken lines. Results showed that nine different mature miRNAs were differentially expressed (DE) in the breast muscle of modern broilers compared to foundational chicken lines. Target genes of DE miRNAs were involved in MAP/ERK1/2, calcium signaling, axonal guidance signaling and NRF2-mediated oxidative response pathways suggesting their roles in muscle growth and development.

Chapter three is focused on identifying and validating copy number variation in the whole genome of two divergently selected high and low stress quail lines. Whole genome sequencing was performed, and data were analyzed for copy number variation detection in genome of the quail lines. Results showed the unique sets of copy number variable regions and genes in the genomes of high stress and low stress birds. Importantly, these genes were involved in development of nervous/endocrine systems, and humoral/cell-mediated immune responses suggesting that they could be potential biomarkers for understanding effects of stress in the well-being and growth performance of avian species and other animals.

Chapter four focuses on identifying SNPs in whole genome of Arkansas Progressor (AP) and Regressor (AR) chicken lines selected for tumor progression and tumor regression upon *v-src*

oncogene induction. Whole genome sequencing was performed, SNPs were analyzed and validated using allele-specific PCR. Results showed the unique sets of SNPs in AP and AR lines. Based on the functional studies, the candidate SNPs were associated with ubiquitylation, and PI3K and NF-kB signaling pathways, suggesting their role in tumor regression in AR chickens.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor Dr. Byungwhi Caleb Kong for not only providing me golden opportunity to work in his lab but also for his continuous guidance and persistent support throughout my PhD. His tremendous mentorship helped me grow as an independent researcher. His continuous support has helped me to overcome many hurdles during my researches and finally come up with this dissertation successfully. I am very thankful to grow under his research expertise and knowledge which will be invaluable resources to me in future as I continue chasing my research goals.

I am equally thankful to my committee members: Dr. Walter G. Bottje, Dr. Young Min Kwon and Dr. Charles F. Rosenkrans Jr. for their directions, support and guidance in my research. I will always remember and appreciate their invaluable inputs.

Further, I would like to thank Stephanie S. Shouse for helping me in the lab for the research. I am also indebted to all my friends and staffs of Department of Poultry Science who directly or indirectly helped me during my study period.

Finally, and the most importantly, I would like to thank my wife Lakhi Maya Devi Khatri for her support and encouragement for everything I do. We would like to welcome our PhD daughter in our life. I thank to my sister, brothers, nephews and niece for the support of any kind that helped me to achieve this degree.

DEDICATION

I would like to dedicate this dissertation to my mom and dad. I will always remember them for their unwavering support and encouragement that has helped me to travel this far in my research career chasing my goals.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

List of Submitted Papers

Chapter 2: MicroRNA profiling associated with muscle growth in modern broiler compared to unselected chicken breed
B. Khatri[1], D. Seo[1], S. Shouse[1], N. Hudson[2], J. Kim[2], W. Bottje[1] and B. Kong[1]*
(Submitted to BMC Genomics)

Chapter 4: Additional analyses of genome resequencing of Arkansas Progressor and Regressor line chickens to identify SNPs associated with tumor regression
B. Khatri, A. Hayden, N. Anthony, and B. Kong*
(Submitted to Genome)

CHAPTER 1
Review of Literature

1.1 INTRODUCTION

The modern poultry industry has been successful to meet increasing global protein needs of consumers by providing good quality, affordable and safe meat products. Yet, there are many issues and obstacles in the industry to meet ever increasing consumer demands world-wide. Disease susceptibility, poor feed conversion rate and effect of stress on production have been major confronts among the list at present. Moreover, more challenging factors that affect poultry production are likely to evolve in years to come.

Poultry production has been affected by various diseases (Ojok 1993) and more recently by muscle abnormalities or myopathies (Kuttappan et al. 2009; Sihvo et al. 2014). Incidence of diseases like avian influenza, Marek's disease, inclusion body hepatitis, salmonellosis, campylobacteriosis, fowl cholera, coccidiosis and others have increased morbidity and mortality and thereby decreased public confidence for poultry products (Ojok 1993). Various factors are responsible for disease outbreaks in birds e.g. rearing of birds in a dense population, free-range rearing, low genetic diversity and restriction of prophylactics to control diseases. Currently, these problems have been alleviated to some extent by employing modern animal husbandry, vaccination programs and development of disease resistant chickens (Cheng et al. 2013). However, better control measures are always needed to impede upcoming threats to help the poultry industry flourish and promote sustainability. Additionally, in recent years muscle abnormalities or myopathies are commonly seen in modern commercial broilers due to intensive genetic selection for the improvement of feed conversion, growth rate, body size and breast yield (Petracci and Cavani 2011). The most common myopathies which include wooden breast and white striping,

though subclinical, have adversely affected poultry industries due to reduced aesthetic acceptability of the meat by consumers (Mudalal et al. 2015).

The field of genomics in poultry has been established after the initial release of the chicken's reference genome in 2004 (International Chicken Genome Sequencing Consortium 2004). Since then genomics has provided exciting avenues to ameliorate and solve issues that exist in the industry today. Genomics has helped in identifying heritable traits of agricultural benefits and selecting birds with various important traits such as disease resistance, improved feed efficiency, and stress resistance through marker-assisted selection. The ultimate goal of modern genomics has been to address the long-standing question of how genetic variation of an organism can be connected to its observed phenotypic variation.

This review provides information on how modern genomic approaches are being used to identify important traits in avian species and gives specific examples of the use of the next generation transcriptomics and whole genome re-sequencing data to identify candidate molecular markers associated with a particular trait of an organism. This review describes the recent gain of insights on three different genetic markers: microRNA (miRNA), single nucleotide polymorphisms (SNPs) and copy number variations (CNVs) focusing on their roles in growth and development of animals. A better understanding of their molecular functions can provide new knowledge and the markers that will contribute to the improvement of poultry production.

1.2 MICRO RNA, TRANSCRIPTOMICS AND REGULATORY FUNCTIONS

The transcriptome is a collection of all transcripts present in a cell and includes both coding and non-coding RNAs. These RNA species are essential to interpret functional elements of the genome and understand physiological conditions of an organism. Transcriptomics is performed to: a) profile all species of transcripts; b) know gene structure, its splicing pattern and other post-

transcriptional modifications; as well as d) study expression levels of transcripts under different conditions (Wang et al. 2009). Among the RNA molecules, miRNAs are a relatively new discovery consisting of small non-coding regulatory RNAs of ~22 nucleotides long. They are involved in various biological processes and regulate expression of specific genes at the post-transcriptional levels by targeting mRNAs for cleavage, deadenylation or translation inhibition (Ha and Kim 2014; Sharma et al. 2014). Initially, miRNAs are transcribed as long primary transcripts called pri-miRNA from host genes by RNA polymerase II. In animals, the majority of miRNAs are encoded by introns of non-coding and coding transcripts and some by exonic regions (Denli, et al. 2004; Han et al. 2004). The mature and active miRNAs are then subsequently generated from pri-miRNAs after two endonuclease processing steps of Drosha and Dicer (Lee et al. 2006; Chendrimada et al. 2005). The mature miRNAs are then ultimately incorporated into RNA-induced silencing complex (RISC) (Hammond 2015). The miRNA-RISC complex then moves to the target mRNAs that have a complementary sequence to the miRNA. In particular, a heptametrical sequence at position 2-7 from 5' end of mature miRNA, called "seed" region is used for binding to 3' untranslated region (UTR) of the target mRNA. The target mRNA is then exonucleolytically cleaved or inhibited for translation (Lai et al. 2005; Lewis et al. 2003). A miRNA can have a large number of target mRNA; and a specific mRNA can be targeted by more than one miRNAs (Brodersen and Voinnet 2009; Lewis et al. 2005).

The latest lists of miRNAs are deposited in an online database (miRBase.org): 740 precursors and 994 mature miRNAs have been identified in chicken. Many miRNAs are ubiquitously distributed in various tissues of animals; however, some are expressed in specific sites with particular functions (Lagos-Quintana, Mariana 2002; Horak, Martin 2016). For instance

miR-1, miR-133 and miR-206 are expressed in the skeletal muscles (Chen et al. 2006; Kim et al. 2006; Van et al. 2008).

1.3 MIRNA TARGETS PREDICTION

Identification of targets of a specific miRNA is essential to understand its biological functions. A validated miRNA could be the best biomarker for determining a particular trait of an organism. Generally, three different approaches are used for miRNA targets prediction: 1) bioinformatic approach, 2) biochemical approach and 3) omics (transcriptomic/proteomic) approach (Hammond 2015). Online methods like TargetScan, miRanda, miRDB, and PicTar are the most popular miRNA target prediction tools that work within the context of a bioinformatics based approach. All of these methods use the seed sequences of miRNAs as important determinants for target binding. However, because they also predict a large number of mRNAs as targets for a particular miRNA, there may be false positives. Therefore, to improve the precision, the target prediction algorithms used in the tools limit target sites to 3' untranslated regions of mRNAs because the best characterized targets are located in these regions (Peterson et al. 2014). The second general approach used for miRNA target prediction is a biochemical based method. This method uses physically associated complexes of miRNA/RISC and target mRNAs to identify miRNA targets. Anti-Argonaute antibodies are used to immunoprecipitate miRNA/RISC plus target mRNA complex and profiling of bound target mRNAs is performed using NGS or microarray methods (Karginov et al. 2007; Hafner et al. 2010). Alternatively, biotinylated miRNAs are also used to identify their corresponding target mRNAs (Lal et al. 2011). In the 'omics' based method, proteome or transcriptome study of a cell or tissue in the presence and absence of a miRNA is conducted to identify its corresponding mRNA targets. The proteomic approach quantitatively measures effects of miRNA on it corresponding target proteins, and selects

targets set accurately (Eichhorn et al. 2014). However, the method is technically challenging. The validations of miRNA targets identified by all the methods outlined above are essential.

1.4 MIRNA REGULATING MUSCLE GROWTH IN CHICKEN

Since 2006 when the first miRNA discovered in chicken (Xu et al. 2006), many miRNAs have been identified and associated with embryo development (Hicks et al. 2008), lipogenesis and cell proliferation (Hicks et al. 2010), skeletal and breast muscle developments (Li et al. 2012; Ouyang et al. 2015), avian influenza virus infection in trachea and lungs (Wang et al. 2009) and Marek's disease infection in embryo fibroblast (Burnside et al. 2008). This review will focus only on miRNAs associated with muscle growth and development in chickens.

In a systemic analysis of breed specific miRNAs from skeletal muscle of chickens, gga-miR-101, gga-miR-10a, miR-10b, gga-miR-1677, gga-let-7f and gga-miR-31were significantly up-regulated in layers while gga-let-7c, gga-miR-200b, gga-miR-16c, gga-miR-15b, gga-miR-15c, gga-miR-460, gga-miR-429 and gga-miR-2188 in broilers. Besides that list, muscle specific miRNAs: gga-miR-206, gga-miR-133 and miR-1 were found in skeletal muscles of both broilers and layers (Li et al. 2011). In miRNA expression analysis conducted by Ouyang et al. (2015), 22 highly differentially expressed miRNAs were identified in breast muscles of fast-growing compared to slow-growing broilers. The gaa-miR-146b-3p was reported to regulate the expression of growth hormone receptor (GHR) by dual-luciferase assay (Ouyang et al. 2015). Similarly, gga-let-7b was found to be associated with regulation of GHR in skeletal muscle of normal and dwarf chickens by genome-wide association study of growth trait and miRNAs (Xu et al. 2013). In study by Wang et al. (2013) reported that skeletal muscle specific miRNAs gga-miR-133a and gga-miR-1a were demonstrated to be important for muscle development in old age (14 to 49 weeks) compared to in early age (0 to 14 day) in chickens (Xu et al. 2013).

The miR-1, miR-133 and miR-206 are involved in muscle growth and development across different animals (Fatima and Morris 2013). Their synthesis, regulation or stimulation and optimal functioning require expression of specific transcription factors, such as MyoD, SRF and MEF2 (Liu et al. 2007). Transcription of miR-1, miR-133 and miR-206 is known to be regulated by the mammalian target of rapamycin (mTOR) signaling in *Myo*-dependent manner in animals. The expression of miR-1 is up-regulated with the help of MyoD that lies downstream of mTOR. MiR-1 then degrades a follistatin suppressor, *HDAC4*, which affects fusion of myocyte (Sun et al. 2010). Similarly, miR-133 and miR-206 are down-regulated by inhibition of mTOR by rapamycin in animals (Ge and Chen 2011). Alternatively, miR-1 and miR-133 have been found to regulate muscle growth and development through phosphatidylinositol 3-kinase/AKT (PI3K/Akt) signaling pathway. The PI3K/Akt pathway is active through PI3K-dependent phosphorylation by binding of insulin growth factor (IGF-1) and its receptor (IGF-1R) during skeletal muscle cell differentiation in animals (Wang 2013). The activated Akt then phosphorylates and inhibits expression of forkhead box O3 (Foxo3a). Foxo3a is a negative regulator of muscle growth and down-regulates promoter activity of miR-1. Thus, down-regulation of miR-1 helps in expression of IGF-1 and IGF-1R. Up-regulation of miR-133 occurs in the presence of myogenin, a myogenic transcription factor, and exogenous IGF-1. The up-regulated miR-133 blocks the expression of IGF-1R, inhibits PI3K/Akt signaling pathway and thereby reduce regulation of Akt phosphorylation through modulation of IGF-1R signaling pathway during myogenesis (Huang et al. 2011).

1.5 WHOLE GENOME RE-SEQUENCING AND GENETIC VARIATION IN CHICKENS

Whole genome re-sequencing of an organism provides the most comprehensive genetic assessment of that organism. It helps in gaining all the important information about genes, genetic variation and gene functions (Bentley 2006). Before the development of next generation

sequencing (NGS) method in 2005, the Sanger sequencing method was used to sequence whole genome of an organism (Sanger et al. 1977). The automated Sanger's method required a longer time and more cost to sequence whole genome of an organism. For example, this method required 13 years and 2.7 billion US dollars to sequence the complete human genome (Gyles 2008). However, a NGS method can sequence 1 tera base (Tb) of DNA in 6 days with an error rate of as low as < 1% (Reuter et al. 2015). Therefore, since 2000s NGS method has become the most popular, fast, accurate, high-throughput and yet cost-effective method for sequencing of genomes of organisms (Shendure and Ji 2008).

The NGS techniques are methods for massively parallel sequencing of short DNA fragments. Their basic workflow involves three steps (Shendure and Ji 2008; Metzker 2010). In the first step, libraries are prepared in which the input DNA molecules are randomly fragmented and ligated with common adapter sequences. The second step involves amplification of the libraries in which an individual library constructed in the first step is amplified such that original DNA molecules and their respective copies are clustered at the same position. The third step comprises the actual sequencing and imaging where sequencing is carried out in cycles with addition of fluorescent labeled nucleotides. The data presented in this thesis were generated by NGS method using chicken and Japanese quail whole genomes, and small RNA sequencing from breast muscles of broiler chickens.

1.6 APPLICATIONS OF NEXT GENERATION SEQUENCING

In animal genetics, NGS is mainly used to identify putative structural variants that can be linked with phenotypic traits of organisms. Whole genome, amplicon, and exome sequencings approaches are used for the purpose of detecting variants. The whole genome resequencing method is used for the global detection of genetic variations in a complete genome of an animal (DePristo

et al. 2011). The amplicon sequencing is used to study variants within individual genes or in small regions of a genome and often is conducted in multiple samples (Zhou et al. 2011). The exome sequencing is performed for the identification of novel genes associated with a phenotypic variation or if the list of putative variant genes is too long (Majewski et al. 2011). Several exome enrichment platforms under different sequencing companies including Agilent, Roche/Nimblegen and Illumina are available (Clark et al. 2011).

The NGS method is also used for messenger RNAs (mRNA) sequencing (RNAseq) of an organism. The mRNA is captured by polyA enrichment and reverse transcribed into cDNA that is processed similar to genomic DNA, as described above. The RNAseq data obtained are used for differential expression analyses between groups of samples, similar to expression data from microarray. However, unlike microarray data, RNAseq data can also can also identify novel transcripts, alternative splicing and allele specific expression (Wang et al. 2009).

The NGS method is also used to study epigenetics in model organisms. Chromatin immunoprecipitation (ChIP) followed by sequencing (ChIP-Seq) is used to identify genome-wide DNA binding positions for proteins of interest, e.g. transcription factors (Park 2009). Bisulfite treatment of DNA followed by bisulfite sequencing (BS-seq) has been a standard method for detection of methylation in genomic DNA (Laird 2010). Similarly, miRNA sequencing is another application of NGS approach as explained above.

1.7 VARIANTS DETECTION IN WHOLE GENOME RESEQUENICNG DATA

The detection of single nucleotide polymorphisms (SNPs) from raw whole genome resequencing data includes steps involving alignment, variants calling and annotation (Nielsen et al. 2011). The short reads obtained from NGS in the form of FASTQ files can be aligned in two different ways: *de novo* assembly and reference based assembly. In *de novo* assembly, the reads

are used to construct the genome without use of any information from the reference genome of the organism. Assemblers such as *ABySS* (Birol et al. 2009), *Velvet* (Zerbino and Birney 2008) or *SOAPdenovo* (Simpson and Durbin 2012) can be used for de novo assembly of sequenced reads. De novo assembly is useful when a full annotated reference genome is not available. In the reference based assembly, sequenced reads are directly aligned to the reference genome. Modern tools such as Bowtie (Langmead and Salzberg 2012) and Burrows-Wheeler Alignment tool (BWA) (Li and Durbin 2009) are commonly used in reference based assembly. The output files obtained from this alignment are then processed for variant calling. Variant calling is performed to identify positions in the genome that are different from the reference genome of the organism. In this process, a list of all sequenced bases aligned to the specific position is pileup, and the proportion of bases that are different to the reference genome is calculated. Various tools such as SAMtools (Li et al. 2009), GATK (McKenna et al. 2010) and Picard are used to call variants from the aligned files. After the variants calling step, variants are filtered based on different parameters such as minimum read depth or minimum average mapping quality of reads. SAMtools varFilter (Li and Durbin 2009) and GATK VarinatFiltration (McKenna et al. 2010) are used for variants filter. The filtered variants are then subsequently annotated based on additional information, such as variant with in a gene, accession number of variants or a conservation score of the affected position. ANNOVAR (Wang et al. 2010) and SnpEff (Cingolani et al. 2012) tools were used for this propose.

1.8 GENETIC VARIATION IN CHICKENS

Chickens are good animal models for genetic studies of their phenotypic traits because of the extensive diversity that have been selected for proposes of egg laying and meat production (International Chicken Polymorphism Map Consortium 2004; Rubin et al. 2010). Several

monogenic traits are well studied, but a large number of valuable traits are complex yet interesting and determined by many genes. The quantitative trait loci (QTLs) have been studied for a number of phenotypic traits, including for growth, meat quality, disease resistance, immune response, egg production, body composition and behavior (Abasht et al. 2006). However, linking of the putative causative genes for a trait is difficult because each locus usually controls only a small fraction of phenotypic variation (Atzmon et al. 2008). This review will focus on two important genetic variations: SNPs and CNVs associated with growth and development in chickens.

1.9 SINGLE NUCLEOTIDE POLYMORPHISMS

A number of SNPs located in different genes associated with growth and development traits in chickens have been identified. The SNP c.739 + 726T>C identified in growth hormone secretagogue receptor (GHSR) or ghrelin receptor was associated with chicken growth and carcass traits (Fang et al. 2010). Insulin-like growth factor-1 (IGF1) plays vital role in muscle growth and development in chickens. The SNP (g.570C>A) present in the promoter region of IGF1 is a putative marker for improved breast muscle yield in chickens (Sato et al. 2012). The chicken growth hormone (cGH) plays vital roles in regulating growth and metabolism. The SNP (G+1705A) found in intron of cGH gene is significantly associated with growth and carcass related traits (Nie et al. 2005). The SNP c.782G>A located in exon region of adipose triglyceride lipase (ATGL) gene was significantly associated with chicken growth and fat traits (Nie et al. 2010). A single nucleotide transversion adenine (A) to thymidine (T), producing a change from amino acid asparagine (Asn) to isoleucine (Ile), was found at position 980 of the open reading frame of pituitary-specific transcription factor gene (PIT1) in chickens. The PIT1 SNP was positively related with early growth rate (Jiang, R 2004). Similarly, three SNPs (rs13849241, rs15231472

and rs13849381) present in intron of high mobility group AT-hook 2 (HMGA2) gene of chicken have been found for their significant associations with body weight gain (Song et al. 2011).

Several insertion-deletion (indel) events have been identified in genes that are related with growth in chickens. A 62-bp indel was found in 5' untranslated region of transforming growth factor-beta 2 (TGFB2) gene that was associated with body weight in chickens (Tang et al. 2011). An 8-bp indel located in exon 1 of chicken Ghrelin (cGHRL) gene was significantly associated with body weight and composition trait (Fang et al. 2007). A 9-bp indel and SNPs (A197835978G and G197836086A) in thyroid hormone responsive spot 14α (THRSPα) gene in chickens were significantly associated with both growth and fat traits (D'Andre Hirwa et al. 2010).

Various genome-wide association studies (GWAS) on chicken growth traits have found significant SNPs in several chromosomes. Five different SNPs were identified in 1.5 Mb karyopherin alpha 3 (KPNA3)-forkhead box O1 (FOXO1A) region of chromosome 1 that had highest significant effect for growth traits. The SNPs were located at 8.9 Kb upstream of KPNA3-FOXO1A region, 1.9 Kb downstream of FOXO1A, 20.9 Kb downstream of ENSGALG0000002273, and in integrator complex subunit 6 (INTS6) and KPNA3 genes (Xie et al. 2012). A region of approximately 8.6 Mb in length in chromosome 4 had significant SNP effects on late growth in chicken F2 resource population. The SNP (GGaluGA266058) in the LIM domain-binding factor 2 (LDB2) gene had the strongest effect on daily body weight gain (Gu et al. 2011). The LDB2 is known to bind various transcription factors, and play role in brain development (Ostendorff, Heather P 2006) and blood vessel formation (Javerzat et al. 2009).

1.10 COPY NUMBER VARIATION

Copy number variation (CNV) basically involves deletion or duplication of nucleotide sequences, or it can be combination of both. It is defined as a loss or gain of genomic segments

that varies in size from 1 kb or larger and is present at variable copy numbers in comparison to the reference genome. The deletion of a DNA segment leads to loss while its duplication results in gain of copy number of that region of a genome (Redon, Richard 2006). The formation of CNVs occurs by four different mechanisms i.e. non-allelic homologous recombination (NAHR), non-homologous end joining (NHEJ), Fork Stalling and Template Switching (FoSTeS) and Retrotransposition (Hastings et al. 2009). The CNVs potentially exert phenotypic diversity in animals through changes in gene structure, gene dosage, and gene expression by exposing recessive alleles (Bickhart and Liu 2014) or indirectly through the perturbation of regulatory region of genes (Zhang et al. 2009).

Four different approaches are used to detect copy number variations (CNVs) from whole genome NGS data sets. The **read-pair** approach uses the distance between the mapped first and the second reads of a pair to detect CNVs. If the read pairs at the position of insert are significantly larger than the average insert size, then it is assumed that a deletion event rather than an insertion has occurred at that position. Tools such as BreakDancer, PEMer, and Hydra use read-pair method to detect CNVs. These tools can detect medium sized insertions or deletions and fail to detect small sized insertions or deletions (Medvedev et al. 2009; Pirooznia et al. 2015).

The **split-read** method uses NGS pair end reads to find CNVs. In this approach only one read of the pair reliably maps while another read partially or completely fails to map to the reference genome. The unmapped reads provide potential start and end of breakpoints at one base pair level. The split read based approach including Pindel, SVseq2, and Prism fail to detect large CNVs (Zhang et al. 2011; Pirooznia et al. 2015). The **read-depth** approach uses depth of coverage of a region of genome to call deletion or duplication. In this method sliding window approach is used to detect CNVs from aligned files of NGS reads. In brief, the mapped reads to the reference

genome are divided into small regions called windows and average read depths are calculated for each window. Duplication or deletion is called if a window or its adjacent part has a significantly higher or lower read depth. Using this approach, large size CNVs can be detected which are otherwise difficult to identify using read-pair and split-pair methods. CNVnator, CNV-seq, ERDS, and Cn.MOPS tools use read-depth approach to detect CNVs from NGS data (Abyzov et al. 2011; Pirooznia et al. 2015). In this thesis, we have used CNVnator to detect CNVs in whole genome resequencing data of Japanese quail. The ***de novo* assembly** approach can theoretically detect all forms of CNVs if the reads are long and accurate. This method first generates contigs which are then compared with the reference genome to detect structural variants. Unfortunately, this method is rarely used in CNV detection because of high computational demands. In addition, due to presence of significant amounts of repeats and segmental duplications in eukaryotic genomes, de novo assembly approach can be less accurate and more complex to perform (Nijkamp et al. 2012; Xi et al. 2012).

The first genomic CNV map in chicken was generated after developing an array comparative genomic hybridization (aCGH) assay. Wang et al. 2010, identified 96 CNVs that encompassed approximately 1.3% of the chicken genome, 27 were high confidence CNVs occurring in more than one individual (Wang et al. 2010). Since then several studies have reported CNVs in genomes of different chickens. A total of 3,154 CNVs grouped into 1556 CNV regions (CNVRs) were reported from 64 commercial and experimental lines of chickens. In their analysis, Crooijmans et al. 2013 reported CNVs covered 60 Mb that represented approximately 5.4% of the chicken genome (Crooijmans et al. 2013). The CNVs that are widely distributed in both macro and microchromosomes in chickens have been reported. In *Gallus gallus* chromosomes (GGA) from 1 to 15, CNVRs covering 5-14% of DNA sequences have been reported. The coverages of DNA

sequence by CNVRs in microchromosomes were also found to be similar to macrochromosome. However, GGA16, GGA25, E64 and W are the exceptions (Wang et al. 2014).

The association of CNV with the agricultural traits such as residual feed intake (Hou, Yali 2012), growth rate and disease resistance (Hou, Yali 2012) have been reported in cattle. A very few efforts have been made to establish relationship between CNV and economic traits in chickens. In their study, Luo et al. 2013 reported a 50 kb deleted region on GGA19 as a high confident CNV responsible for Marek's disease-resistance (Luo et al. 2013). They reported 83.5 kb CNVR associated with loss on uncharacterized chromosomal region encompassing a general transcription factor IIi (GTF2I) as a high confidence candidate responsible for Marek's disease susceptibility.

1.11 PHENOTYPIC VARIATION DUE TO CNV

A CNV can affect the phenotypic properties of an individual by various mechanisms as explained previously. A CNV associated with protein coding genes can not only affect protein translation or function of encoded protein but can also influence the expression of genes in its vicinity and its effect can extend up to half a megabase distance (Henrichsen et al. 2009). However, it is also possible that CNV can have no or very little effect (Wang et al. 2014). Phenotypes including pea-comb (Wright et al. 2009), late-feathering (Elferink et al. 2008), dermal hyperpigmentation (Dorshorst et al. 2010), and dark brown plumage color (Gunnarsson et al. 2011) have been characterized to be associated with CNV in chickens.

The pea-comb phenotype in chickens that exhibit showing reduced sizes of comb and wattles has been associated with duplication of intron 1 of sex determining region Y (SRY)-Box 5 (SOX5) gene on GGA1. The massive amplification of the intron 1 interferes with the expression of SOX5 and regulates gene expression during cell differentiation required for comb and wattles development. SOX5 is a transcription factor that plays role in regulation of embryonic

development and cell fate determination. The protein encoded by SOX5 plays a role in chondrocyte formation. Pea-comb phenotype is found in chickens living in cold climate which is advantageous to reduce heat loss and susceptibility to frost lesions (Wright et al. 2009).

The late feathering phenotype in chickens is associated with duplication of the prolactin receptor gene (PRLR) and the sperm flagellar protein 2 gene (SPEF2). This trait is linked with reduced fertility and late development of flight feathers. Beside increased PRLR gene expression, altered mRNA expression level was observed in many keratin-related genes during early development. A partially dominant sex-determined K locus on GGAZ is responsible for this trait (Elferink et al. 2008).

The dermal hyperpigmentation or fibromelanosis (FM) trait in Chinese Silkie chicken is characterized by distinct phenotypes such as large feathers on head, fluffy plumage, black skin, and bones, blue ear lobes, feathered legs and feet, and five toes on each foot. The dark color in connective tissues is due to excessive pigmentation or unusual melanogenesis. The FM is due to inverted duplication and junction of the two genomic regions separated by more than 400 kb in wild birds. The duplicated region contains endothelin 3 (EDN3) gene that plays vital role in promoting the proliferation of melanoblasts (Dorshorst et al. 2010).

Dark brown plumage color phenotype in chicken is determined by deletion of 8.3 kb region present upstream of SRY-Box 10 (SOX10) gene that plays role in melanocyte migration and survival. The deletion of this upstream region leads to decreased expression of SOX10 gene. This phenomenon in turn reduces the expression of key enzymes such as tyrosinase needed for pigment synthesis. Therefore, the reduced tyrosinase activity leads to development of pheomelanistic (reddish) plumage color, a characteristics of dark brown plumage trait (Gunnarsson et al. 2011).

In this dissertation, various genomic approaches were utilized to analyze genome wide copy number variations and single nucleotide polymorphisms in Japanese quail and chicken lines respectively to characterize economically traits in these avian species. We also performed small RNA transcriptome analysis to characterize muscle specific miRNAs in modern broilers and their foundational breeds to understand their regulatory roles that could play for rapid muscle growth and development in modern broilers.

1.12 OBJECTIVES

Various genomic approaches have been used to decipher the entire genome of an organism and relate its economically important phenotypic trait to its genotypic character. Such approaches can provide insight into the modern methods used for animal production, improvement of their well-being and growth performances. In avian species, many traits have been defined at the molecular levels, yet very little information is available on how the defined biomarkers would be appropriate for animal selection in commercial poultry industry. The main hypothesis for this project is that there are many biomarkers yet to be discovered that could have an impact the betterment of poultry. Thus, the general objective of this dissertation is to characterize various muscle specific miRNAs, immobilization stress related copy number variable genes and regions, and tumor regression associated SNPs in different lines of birds.

Specific objectives for this dissertation are as follows:

1. To characterize breast muscle specific miRNAs in foundational Barred Plymouth Rock chickens, and modern broilers selected for high feed efficiency.
2. To identify copy number variable regions and genes in Japanese Quail selected for susceptibility and resistance to immobilization stress.
3. To characterize the SNPs in a) Arkansas Progressor and, b) Arkansas Regressor chickens selected for tumor regression property.

REFERENCES

Abasht, B., Dekkers, J., & Lamont, S. (2006). Review of quantitative trait loci identified in the chicken. *Poultry Science, 85*(12), 2079-2096.

Abyzov, A., Urban, A. E., Snyder, M., & Gerstein, M. (2011). CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Research, 21(6), 974-984.

Atzmon, G., Blum, S., Feldman, M., Cahaner, A., Lavi, U., & Hillel, J. (2008). QTLs detected in a multigenerational resource chicken population. Journal of Heredity, 99(5), 528-538.

Bentley, D. R. (2006). Whole-genome re-sequencing. Current Opinion in Genetics & Development, 16(6), 545-552.

Bickhart, D. M., & Liu, G. E. (2014). The challenges and importance of structural variation detection in livestock. Frontiers in Genetics, 5, 37.

Birol, I., Jackman, S. D., Nielsen, C. B., Qian, J. Q., Varhol, R., Stazyk, G., et al. (2009). De novo transcriptome assembly with ABySS. Bioinformatics, 25(21), 2872-2877.

Brodersen, P., & Voinnet, O. (2009). Revisiting the principles of microRNA target recognition and mode of action. Nature Reviews Molecular Cell Biology, 10(2), 141.

Burnside, J., Ouyang, M., Anderson, A., Bernberg, E., Lu, C., Meyers, B. C., et al. (2008). Deep sequencing of chicken microRNAs. BMC Genomics, 9(1), 185.

Chen, J., Mandel, E. M., Thomson, J. M., Wu, Q., Callis, T. E., Hammond, S. M., et al. (2006). The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. Nature Genetics, 38(2), 228.

Chendrimada, T. P., Gregory, R. I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K., et al. (2005). TRBP recruits the dicer complex to Ago2 for microRNA processing and gene silencing. Nature, 436(7051), 740.

Cheng, H. H., Kaiser, P., & Lamont, S. J. (2013). Integrated genomic approaches to enhance genetic resistance in chickens. Annu.Rev.Anim.Biosci., 1(1), 239-260.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of drosophila melanogaster strain w1118; iso-2; iso-3. Fly, 6(2), 80-92.

Clark, M. J., Chen, R., Lam, H. Y., Karczewski, K. J., Chen, R., Euskirchen, G., et al. (2011). Performance comparison of exome DNA sequencing technologies. Nature Biotechnology, 29(10), 908.

Crooijmans, R. P., Fife, M. S., Fitzgerald, T. W., Strickland, S., Cheng, H. H., Kaiser, P., et al. (2013). Large scale variation in DNA copy number in chicken breeds. BMC Genomics, 14(1), 398.

D'Andre Hirwa, C., Yan, W., Wallace, P., Nie, Q., Luo, C., Li, H., et al. (2010). Effects of the thyroid hormone responsive spot 14α gene on chicken growth and fat traits. Poultry Science, 89(9), 1981-1991.

Denli, A. M., Tops, B. B., Plasterk, R. H., Ketting, R. F., & Hannon, G. J. (2004). Processing of primary microRNAs by the microprocessor complex. Nature, 432(7014), 231.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics, 43(5), 491.

Dorshorst, B., Okimoto, R., & Ashwell, C. (2010). Genomic regions associated with dermal hyperpigmentation, polydactyly and other morphological traits in the silkie chicken. Journal of Heredity, 101(3), 339-350.

Eichhorn, S. W., Guo, H., McGeary, S. E., Rodriguez-Mias, R. A., Shin, C., Baek, D., et al. (2014). mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. Molecular Cell, 56(1), 104-115.

Elferink, M. G., Vallée, A. A., Jungerius, A. P., Crooijmans, R. P., & Groenen, M. A. (2008). Partial duplication of the PRLR and SPEF2 genes at the late feathering locus in chicken. BMC Genomics, 9(1), 391.

Fang, M., Nie, Q., Luo, C., Zhang, D., & Zhang, X. (2007). An 8 bp indel in exon 1 of ghrelin gene associated with chicken growth. Domestic Animal Endocrinology, 32(3), 216-225.

Fang, M., Nie, Q., Luo, C., Zhang, D., & Zhang, X. (2010). Associations of GHSR gene polymorphisms with chicken growth and carcass traits. Molecular Biology Reports, 37(1), 423-428.

Fatima, A., & Morris, D. G. (2013). MicroRNAs in domestic livestock. Physiological Genomics, 45(16), 685-696.

Ge, Y., & Chen, J. (2011). MicroRNAs in skeletal myogenesis. Cell Cycle, 10(3), 441-448.

Gu, X., Feng, C., Ma, L., Song, C., Wang, Y., Da, Y., et al. (2011). Genome-wide association study of body weight in chicken F2 resource population. PloS One, 6(7), e21872.

Gunnarsson, U., Kerje, S., Bed'hom, B., Sahlqvist, A., Ekwall, O., Tixier-Boichard, M., et al. (2011). The dark brown plumage color in chickens is caused by an 8.3-kb deletion upstream of SOX10. Pigment Cell & Melanoma Research, 24(2), 268-274.

Gyles, C. (2008). The DNA revolution. The Canadian Veterinary Journal = La Revue Veterinaire Canadienne, 49(8), 745-746.

Ha, M., & Kim, V. N. (2014). Regulation of microRNA biogenesis. Nature Reviews Molecular Cell Biology, 15(8), 509.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., et al. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell, 141(1), 129-141.

Hammond, S. M. (2015). An overview of microRNAs. Advanced Drug Delivery Reviews, 87, 3-14.

Han, J., Lee, Y., Yeom, K. H., Kim, Y. K., Jin, H., & Kim, V. N. (2004). The drosha-DGCR8 complex in primary microRNA processing. Genes & Development, 18(24), 3016-3027.

Hastings, P. J., Lupski, J. R., Rosenberg, S. M., & Ira, G. (2009). Mechanisms of change in gene copy number. Nature Reviews Genetics, 10(8), 551.

Henrichsen, C. N., Vinckenbosch, N., Zöllner, S., Chaignat, E., Pradervand, S., Schütz, F., et al. (2009). Segmental copy number variation shapes tissue transcriptomes. Nature Genetics, 41(4), 424.

Hicks, J., Tembhurne, P., & Liu, H. (2008). MicroRNA expression in chicken embryos. Poultry Science, 87(11), 2335-2343.

Hicks, J. A., Trakooljul, N., & Liu, H. (2010). Discovery of chicken microRNAs associated with lipogenesis and cell proliferation. Physiological Genomics, 41(2), 185-193.

Horak, M., Novak, J., & Bienertova-Vasku, J. (2016). Muscle-specific microRNAs in skeletal muscle development. Developmental Biology, 410(1), 1-13.

Hou, Y., Bickhart, D. M., Chung, H., Hutchison, J. L., Norman, H. D., Connor, E. E., et al. (2012). Analysis of copy number variations in holstein cows identify potential mechanisms contributing to differences in residual feed intake. Functional & Integrative Genomics, 12(4), 717-723.

Hou, Y., Liu, G. E., Bickhart, D. M., Matukumalli, L. K., Li, C., Song, J., et al. (2012). Genomic regions showing copy number variations associate with resistance or susceptibility to gastrointestinal nematodes in angus cattle. Functional & Integrative Genomics, 12(1), 81-92.

Huang, M., Xu, H., Xie, S., Zhou, H., & Qu, L. (2011). Insulin-like growth factor-1 receptor is regulated by microRNA-133 during skeletal myogenesis. PLoS One, 6(12), e29173.

International Chicken Genome Sequencing Consortium. (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature, 432(7018), 695.

International Chicken Polymorphism Map Consortium. (2004). A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. Nature, 432(7018), 717.

Javerzat, S., Franco, M., Herbert, J., Platonova, N., Peille, A., Pantesco, V., De Vos, J., Assou, S., Bicknell, R., Bikfalvi, A., (2009). Correlating global gene regulation to angiogenesis in the developing chick extra-embryonic vascular system. PloS one 4, e7856.

Jiang, R., Li, J., Qu, L., Li, H., & Yang, N. (2004). A new single nucleotide polymorphism in the chicken pituitary-specific transcription factor (POU1F1) gene associated with growth rate. Animal Genetics, 35(4), 344-346.

Karginov, F. V., Conaco, C., Xuan, Z., Schmidt, B. H., Parker, J. S., Mandel, G., et al. (2007). A biochemical approach to identifying microRNA targets. Proceedings of the National Academy of Sciences of the United States of America, 104(49), 19291-19296.

Kim, H. K., Lee, Y. S., Sivaprasad, U., Malhotra, A., & Dutta, A. (2006). Muscle-specific microRNA miR-206 promotes muscle differentiation. The Journal of Cell Biology, 174(5), 677-687.

Kuttappan, V., Brewer, V., Clark, F., McKee, S., Meullenet, J., Emmert, J., et al. (2009). Effect of white striping on the histological and meat quality characteristics of broiler fillets. Poult.Sci, 88, 136-137.

Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., & Tuschl, T. (2002). Identification of tissue-specific microRNAs from mouse. Current Biology, 12(9), 735-739.

Lai, E. C., Tam, B., & Rubin, G. M. (2005). Pervasive regulation of drosophila notch target genes by GY-box-, brd-box-, and K-box-class microRNAs. Genes & Development, 19(9), 1067-1080.

Laird, P. W. (2010). Principles and challenges of genome-wide DNA methylation analysis. Nature Reviews Genetics, 11(3), 191.

Lal, A., Thomas, M. P., Altschuler, G., Navarro, F., O'Day, E., Li, X. L., et al. (2011). Capture of microRNA–bound mRNAs identifies the tumor suppressor miR-34a as a regulator of growth factor signaling. PLoS Genetics, 7(11), e1002363.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. Nature Methods, 9(4), 357.

Lee, Y., Hur, I., Park, S. Y., Kim, Y. K., Suh, M. R., & Kim, V. N. (2006). The role of PACT in the RNA silencing pathway. The EMBO Journal, 25(3), 522-532.

Lewis, B. P., Burge, C. B., & Bartel, D. P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell, 120(1), 15-20.

Lewis, B. P., Shih, I., Jones-Rhoades, M. W., Bartel, D. P., & Burge, C. B. (2003). Prediction of mammalian microRNA targets. Cell, 115(7), 787-798.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics, 25(14), 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. Bioinformatics, 25(16), 2078-2079.

Li, T., Wang, S., Wu, R., Zhou, X., Zhu, D., & Zhang, Y. (2012). Identification of long non-protein coding RNAs in chicken skeletal muscle using next generation sequencing. Genomics, 99(5), 292-298.

Li, T., Wu, R., Zhang, Y., & Zhu, D. (2011). A systematic analysis of the skeletal muscle miRNA transcriptome of chicken varieties with divergent skeletal muscle growth identifies novel miRNAs and differentially expressed miRNAs. BMC Genomics, 12(1), 186.

Liu, N., Williams, A. H., Kim, Y., McAnally, J., Bezprozvannaya, S., Sutherland, L. B., et al. (2007). An intragenic MEF2-dependent enhancer directs muscle-specific expression of microRNAs 1 and 133. Proceedings of the National Academy of Sciences of the United States of America, 104(52), 20844-20849.

Luo, J., Yu, Y., Mitra, A., Chang, S., Zhang, H., Liu, G., et al. (2013). Genome-wide copy number variant analysis in inbred chickens lines with different susceptibility to marek's disease. G3 (Bethesda, Md.), 3(2), 217-223.

Majewski, J., Schwartzentruber, J., Lalonde, E., Montpetit, A., & Jabado, N. (2011). What can exome sequencing do for you? Journal of Medical Genetics, 48(9), 580-589.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 20(9), 1297-1303.

Medvedev, P., Stanciu, M., & Brudno, M. (2009). Computational methods for discovering structural variation with next-generation sequencing. Nature Methods, 6(11s), S13.

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

Mudalal, S., Lorenzi, M., Soglia, F., Cavani, C., & Petracci, M. (2015). Implications of white striping and wooden breast abnormalities on quality traits of raw and marinated chicken meat. Animal, 9(4), 728-734.

Nie, Q., Sun, B., Zhang, D., Luo, C., Ishag, N., Lei, M., et al. (2005). High diversity of the chicken growth hormone gene and effects on growth and carcass traits. Journal of Heredity, 96(6), 698-703.

Nie, Q., Fang, M., Xie, L., Shen, X., Liu, J., Luo, Z., et al. (2010). Associations of ATGL gene polymorphisms with chicken growth and fat traits. Journal of Applied Genetics, 51(2), 185-191.

Nielsen, R., Paul, J. S., Albrechtsen, A., & Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. Nature Reviews Genetics, 12(6), 443.

Nijkamp, J. F., van den Broek, Marcel A, Geertman, J. A., Reinders, M. J., Daran, J. G., & de Ridder, D. (2012). De novo detection of copy number variation by co-assembly. Bioinformatics, 28(24), 3195-3202.

Ojok, L. (1993). Diseases as important factor affecting increased poultry production in uganda. Der Tropenlandwirt-Journal of Agriculture in the Tropics and Subtropics, 94(1), 37-44.

Ostendorff, H. P., Tursun, B., Cornils, K., Schlüter, A., Drung, A., Güngör, C., et al. (2006). Dynamic expression of LIM cofactors in the developing mouse neural tube. Developmental Dynamics, 235(3), 786-791.

Ouyang, H., He, X., Li, G., Xu, H., Jia, X., Nie, Q., et al. (2015). Deep sequencing analysis of miRNA expression in breast muscle of fast-growing and slow-growing broilers. International Journal of Molecular Sciences, 16(7), 16242-16262.

Park, P. J. (2009). ChIP–seq: Advantages and challenges of a maturing technology. Nature Reviews Genetics, 10(10), 669.

Peterson, S. M., Thompson, J. A., Ufkin, M. L., Sathyanarayana, P., Liaw, L., & Congdon, C. B. (2014). Common features of microRNA target prediction tools. Frontiers in Genetics, 5, 23.

Petracci, M., & Cavani, C. (2011). Muscle growth and poultry meat quality issues. Nutrients, 4(1), 1-12.

Pirooznia, M., Goes, F. S., & Zandi, P. P. (2015). Whole-genome CNV analysis: Advances in computational approaches. Frontiers in Genetics, 6, 138.

Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., et al. (2006). Global variation in copy number in the human genome. Nature, 444(7118), 444.

Reuter, J. A., Spacek, D. V., & Snyder, M. P. (2015). High-throughput sequencing technologies. Molecular Cell, 58(4), 586-597.

Rubin, C., Zody, M. C., Eriksson, J., Meadows, J. R., Sherwood, E., Webster, M. T., et al. (2010). Whole-genome resequencing reveals loci under selection during chicken domestication. Nature, 464(7288), 587.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America, 74(12), 5463-5467.

Sato, S., Ohtake, T., Uemoto, Y., Okumura, Y., & Kobayashi, E. (2012). Polymorphism of insulin-like growth factor 1 gene is associated with breast muscle yields in chickens. Animal Science Journal, 83(1), 1-6.

Sharma, M., Juvvuna, P. K., Kukreti, H., & McFarlane, C. (2014). Mega roles of microRNAs in regulation of skeletal muscle health and disease. Frontiers in Physiology, 5, 239.

Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. Nature Biotechnology, 26(10), 1135.

Sihvo, H., Immonen, K., & Puolanne, E. (2014). Myodegeneration with fibrosis and regeneration in the pectoralis major muscle of broilers. Veterinary Pathology, 51(3), 619-623.

Simpson, J. T., & Durbin, R. (2012). Efficient de novo assembly of large genomes using compressed data structures. Genome Research, 22(3), 549-556.

Song, C., Gu, X., Feng, C., Wang, Y., Gao, Y., Hu, X., et al. (2011). Evaluation of SNPs in the chicken HMGA2 gene as markers for body weight gain. Animal Genetics, 42(3), 333-336.

Sun, Y., Ge, Y., Drnevich, J., Zhao, Y., Band, M., & Chen, J. (2010). Mammalian target of rapamycin regulates miRNA-1 and follistatin in skeletal myogenesis. The Journal of Cell Biology, 189(7), 1157-1169.

Tang, S., Ou, J., Sun, D., Zhang, Y., & Xu, G. (2011). A novel 62-bp indel mutation in the promoter region of transforming growth factor-beta 2 (TGFB2) gene is associated with body weight in chickens. Animal Genetics, 42(1), 108-112.

Van Rooij, E., Liu, N., & Olson, E. N. (2008). MicroRNAs flex their muscles. Trends in Genetics, 24(4), 159-166.

Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Research, 38(16), e164-e164.

Wang, X., Nahashon, S., Feaster, T. K., Bohannon-Stewart, A., & Adefope, N. (2010). An initial map of chromosomal segmental copy number variations in the chicken. BMC Genomics, 11(1), 351.

Wang, Y., Brahmakshatriya, V., Zhu, H., Lupiani, B., Reddy, S. M., Yoon, B., et al. (2009). Identification of differentially expressed miRNAs in chicken lung and trachea with avian influenza virus infection by a deep sequencing approach. BMC Genomics, 10(1), 512.

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-seq: A revolutionary tool for transcriptomics. Nature Reviews Genetics, 10(1), 57.

Wang, X. H. (2013). MicroRNA in myogenesis and muscle atrophy. Current Opinion in Clinical Nutrition and Metabolic Care, 16(3), 258-266.

Wright, D., Boije, H., Meadows, J. R., Bed'Hom, B., Gourichon, D., Vieaud, A., et al. (2009). Copy number variation in intron 1 of SOX5 causes the pea-comb phenotype in chickens. PLoS Genetics, 5(6), e1000512.

Xi, R., Lee, S., & Park, P. J. (2012). A survey of Copy-Number variation detection tools based on High-Throughput sequencing data. Current Protocols in Human Genetics, , 7.19. 1-7.19. 15.

Xie, L., Luo, C., Zhang, C., Zhang, R., Tang, J., Nie, Q., et al. (2012). Genome-wide association study identified a narrow chromosome 1 region associated with chicken growth traits. PLoS One, 7(2), e30910.

Xu, H., Wang, X., Du, Z., & Li, N. (2006). Identification of microRNAs from different tissues of chicken embryo and adult chicken. FEBS Letters, 580(15), 3610-3616.

Xu, Z., Nie, Q., & Zhang, X. (2013). Overview of genomic insights into chicken growth traits based on genome-wide association study and microRNA regulation. Current Genomics, 14(2), 137-146.

Zerbino, D. R., & Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de bruijn graphs. Genome Research, 18(5), 821-829.

Zhang, F., Gu, W., Hurles, M. E., & Lupski, J. R. (2009). Copy number variation in human health, disease, and evolution. Annual Review of Genomics and Human Genetics, 10, 451-481.

Zhang, Z. D., Du, J., Lam, H., Abyzov, A., Urban, A. E., Snyder, M., et al. (2011). Identification of genomic indels and structural variations using split reads. BMC Genomics, 12(1), 375.

Zhou, J., Wu, L., Deng, Y., Zhi, X., Jiang, Y., Tu, Q., et al. (2011). Reproducibility and quantitation of amplicon sequencing-based detection. The ISME Journal, 5(8), 1303.

CHAPTER 2

MicroRNA profiling associated with muscle growth in modern broiler compared to unselected chicken breed

B. Khatri[1], D. Seo[1], S. Shouse[1], N. Hudson[2], J. Kim[2], W. Bottje[1] and B. Kong[1]*

[1]Department of Poultry Science, Center of Excellence for Poultry Science

[2]Department of Human Environmental Science

University of Arkansas, Fayetteville, Arkansas

[3]The University of Queensland, Brisbane, St Lucia, QLD, Australia

*Correspondence to: Dr. Byungwhi Caleb Kong

Department of Poultry Science,

University of Arkansas, Fayetteville, AR 72701

Tel: (479) 575-5494 FAX: (479) 575-7139

E-mail: bkong@uark.edu

2.1 ABSTRACT

Genetically selected modern broiler chickens have acquired outstanding production efficiency through rapid growth and improved feed efficiency compared to unselected chicken breeds. Recently, we analyzed transcriptome of breast muscle tissues obtained from modern pedigree male (PeM) broilers (rapid growth and higher efficiency) and foundational Barred Plymouth Rock (BPR) chickens (slow growth and poorer efficiency). In this study, in addition to mRNA expression, differential abundance of microRNA (miRNA) was analyzed from the same tissues and the results were integrated with differentially expressed (DE) mRNA of breast muscles of PeM and BPR. To investigate the expression level of miRNA, small RNA sequencing was performed with 6 muscle samples per group using 1×50 bp single end read method of Illumina sequencing. After quality assessment and adapter trimming of raw reads, the clean reads were aligned to chicken reference mature miRNA sequences and read counts were normalized by reads per millions. The miRNA showing average read counts >5 in each group, p-value <0.05 by t-test, and fold change >1.2 were considered as DE miRNAs between PeM and BPR. Differentially expressed miRNA were validated by quantitative PCR and were subjected to target-predictions with miRDB online software. A total of 994 miRNA were identified in PeM and BPR chicken lines. After initial filtering and statistical analyses, miR-2131-5p, miR-221-5p, miR-126-3p, miR-146b-5p, miR-10a-5p, let-7b, miR-125b-5p, and miR-146c-5p were up-regulated whereas miR-206 was down-regulated in PeM compared to BPR breast muscle. Based on inhibitory regulations of miRNAs on the mRNA abundance, our computational analysis predicated that 118 down-regulated mRNAs may be targeted by the up-regulated miRNAs, while 35 up-regulated mRNAs appear to be due to a down-regulated miRNA (i.e., miR-206). This study may provide foundation data for elucidating molecular mechanisms that govern muscle growth in chickens.

Key words: miRNA, pedigree male broiler, Barred Plymouth Rock, production efficiency, differential expression

2.2 INTRODUCTION

Genetically selected modern broiler chickens are characterized with rapid growth and improved feed efficiency compared to unselected counterparts. These traits are beneficial to meet the global protein needs for an ever increasing human population (Kong et al. 2016; Kong et al. 2017). Understanding the mechanism behind rapid muscle growth and high feed efficiency in chickens will help in maintaining a sustainable protein source through improved animal production system.

MicroRNAs are short 18-24 nucleotides long non-coding regulatory RNAs that target mRNAs for cleavage, deadenylation or translational inhibition of gene expression at post-transcriptional level (Bushati and Cohen 2007; Rana 2007). In recent years, many studies have shown the vital roles of miRNAs in various aspects of biological phenomenon associated with growth and development (Anderson et al. 2006; Gangaraju and Lin 2009; Li et al. 2011; Wienholds et al. 2005). Recently, a total of 921 miRNAs were identified from breast muscles of fast and slow growing broilers (Ouyang et al. 2015). Of note, let-7b was experimentally validated through genetic analyses in chickens to affect signaling pathways regulating skeletal muscle growth (Lin et al. 2012). The miR-1 was shown to promote myogenesis by targeting histone deacetylase (HDAC) 4, a transcriptional repressor protein of muscle gene expression. The miR-133 was proven to enhance myoblast proliferation by repressing serum response factor (SRF) (Chen et al. 2006). The miR-26a was reported to accelerate the process of myogenesis through induction of creatine kinase and up-regulation of *myoD* and *myogenin* (Wong and Tellam 2008). All these reports suggest that miRNAs are important regulators of muscle growth and development in vertebrate animals.

Extensive genetic selection has led to rapid growth rate and large muscle mass in modern broilers compared to unselected chicken breeds (Lopez et al. 2007). Previously, we have identified differentially expressed genes associated with breast muscle myogenesis in pedigree male (PeM) broilers (rapidly growing, higher efficiency, and large muscle mass) compared with Barred Plymouth Rock (BPR) chickens (slowly growing, poorer efficiency, and small muscle mass) (Kong et al. 2017). This transcriptomic analysis indicated that rapid growth and large muscle mass shown in modern broilers may be due to altered mitochondrial functions, growth signaling pathways, oxidative stress pathway, and/or hormone receptor pathways. To elucidate regulatory roles of miRNAs on muscle growth and production efficiency, we profiled differentially expressed (DE) miRNAs using small RNA sequencing followed by prediction of potential target mRNAs; and eventually, miRNA profiling results were integrated with trasncriptomic data of DE genes obtained by previous mRNA sequencing study (Kong et al. 2017).

2.3 METHODS

2.3.1 ETHICS STATEMENT

The present study was conducted in accordance with the recommendations in the guide for the care and use of laboratory animals of the National Institutes of Health. All procedures for animal care complied with the University of Arkansas Institutional Animal Care and Use Committee (IACUC): Protocol #14012.

2.3.2 SAMPLES

Breast muscle tissues were obtained from pedigree male broilers (PeM), highly selected for growth and feed efficiency (Bottje et al. 2002; Kong et al. 2011), and Barred Plymouth Rock (BPR) as described elsewhere (Kong et al. 2017). Briefly, immature PeM and BPR chickens (≤8 weeks old, n=6 per breed) were killed by an overdose of sodium pentobarbital (i.v.) and breast muscle tissue was obtained and flash frozen in liquid nitrogen. Total RNAs were extracted from the muscle tissue using TRIzol reagent (Thermo-Fisher Scientific, Carlsbad, CA) following manufacturer's protocol. Extracted RNA samples were treated with DNase I (New England Biolabs Inc., Ipswich, MA) and purified again using TRIzol reagent. RNA quality was then assessed using Agilent 2200 TapeStation instrument (Santa Clara, CA). All RNA samples showed high enough quality and quantity (data not shown) and were subjected to miRNA sequencing.

2.3.3 MICRORNA SEQUENCING AND DATA ANALYSIS

Library preparation for individual samples and sequencing were performed by Research Technology Support Facility at Michigan State University (East Lansing, MI). Illumina TruSeq system 1×50 bp single end read method was used for miRNA sequencing. Quality of raw reads were determined using FastQC tool kit (Andrews 2010) and adapters were trimmed using bbduk.sh command line of BBMap toolkit (http://sourceforge.net/projects/bbmap/). The clean reads were

then aligned to reference mature miRNA sequences of *Gallus gallus* obtained from miRBase (http://mirbase.org/) using Arraystar program in Lasergene software package (DNAStar, Madison, WI) and read counts were normalized by reads per millions (RPM) to stabilize the variance. Differential expression with normalized read counts was further analyzed using JMP Genomics 9 (SAS Institute Inc., Cary, NC). MicroRNAs with less than 5 average read counts in both comparison groups were not considered for further analysis. The t-statistics was used to compare abundances between PeM and BPR, and miRNAs showing fold change >1.2 and p-value <0.05 were considered as DE.

2.3.4 HIERARCHICAL CLUSTERING

DE miRNAs of PeM and BPR were subjected to hierarchical cluster analysis using JMP Genomics Program. A matrix with as many columns as birds (12) and as many rows as DE miRNA (9) were imported in which each cell contained log2 transformed fold change value for that miRNA and bird into JMP Genomics Program, normalizing on rows. After, hierarchical clustering on both rows and columns were applied followed by dendrogram image production.

2.3.5 TARGET PREDICTION OF DE MIRNA

Online miRNA target prediction tool, miRDB (http://www.mirdb.org/) was used to predict the targets of DE miRNAs. The predicted targets of DE miRNA were then integrated with DE mRNA list obtained from same breast muscle tissue in earlier study (Kong et al. 2017). The mRNA showing opposite direction of expression to the corresponding miRNA were chosen as targets of DE miRNAs and used for subsequent bioinformatics analyses.

2.3.6 INGENUITY PATHWAY ANALYSIS

Ingenuity Pathway Analysis (IPA; Qiagen, Valencia, CA; http://www.ingenuity.com) software was used for construction of interaction network between DE miRNA and their candidate

targets. Since IPA is based on bioinformatics in humans, functionalities for DE miRNAs in the chicken datasets are principally based upon mammalian biological mechanisms. As investment in biomedical research biases the functional annotations towards human disease, we have attempted to draw plausible conclusions based on avian literature (Kong et al. 2017). All target genes of DE miRNAs were subjected to IPA analysis for functional annotation and canonical pathways mapping among which all target genes were identified by IPA.

2.3.7 SMALL RNA PURIFICATION, CDNA SYNTHESIS AND QUANTITATIVE REAL TIME PCR (QPCR)

Sixty micrograms of total RNA samples from 6 muscle samples each for PeM and BPR were used for small RNA enrichment and subsequent validation of miRNA sequencing results by qPCR. Small RNAs were enriched using mirVana miRNA isolation kit (Ambion, Carlsbad, CA) following manufacturer's instructions. Enriched small RNAs were polyadenylated using Poly(A) Polymerase (Ambion) and re-purified using QIAquick Nucleotide Removal Kit (Qiagen). The polyadenylated small RNAs were then ligated with RNA oligonucleotide adapter (Table 1), treated with RNaseOUT and reverse transcribed to cDNA using adapter primer and SuperScript III reverse transcriptase (Thermo-Fisher Scientific). The cDNA samples were diluted to 1:10 ratio and a portion (2 µl) of cDNA was used for qPCR reaction using ABI prism 7500HT system (Thermo-Fisher Scientific) with PowerUp SYBR Green Master Mix (Thermo-Fisher Scientific). Primers were synthesized by Integrated DNA Technologies (Coralville, IA), and are listed in Table 2.1. The qPCR condition was as follows: 1 cycle at 95 °C for 2 min, 40 cycles at 95 °C for 30 s, 60 °C for 30 s. The chicken 5S ribosomal RNA was used as internal control. Dissociation curves were generated at the end of amplification process for validating data quality. All qPCR reactions were conducted three times and values of average cycle threshold (Ct) were determined for each sample,

and $2^{-\Delta\Delta Ct}$ values for the comparison of PeM and BPR were used for relative quantification by fold-change and statistical significance.

2.4 RESULTS

2.4.1 MICRORNA PROFILING IN PEM AND BPR CHICKENS BY MIRNA SEQUENCING

MicroRNA Sequencing of twelve samples yielded 33,727,148 and 51,310,328 raw sequence reads from PeM and BPR samples, respectively. After adapter trimming, 17,751,585 and 22,922,027 clean reads remained in PeM and BPR, respectively (data not shown). After alignment of clean reads to chicken reference mature miRNA collections, a total of 994 mature miRNAs were identified in both PeM and BPR. Rarely expressed mature miRNAs (i.e., raw read count < 5) were filtered out, resulting in 38 miRNAs remained as meaningfully expressed and were used for subsequent analyses (Supplementary 2.1).

2.4.2 DIFFERENTIALLY EXPRESSED MIRNAS IN PEM COMPARED TO BPR

Nine DE miRNAs showing p-value <0.05 and fold change >1.2 were identified in PeM compared with BPR (Table 2.2). Among 9 DE miRNAs, 8 miRNAs including miR-2131-5p, miR-221-5p, miR-126-3p, miR-146b-5p, miR-10a-5p, let-7b, miR-125b-5p, and miR-146c-5p were up-regulated while miR-206 was down-regulated in PeM compared to BPR breast muscle (Table 2.2). All DE miRNAs were validated using qPCR (Table 2.2). Our qPCR results indicated that expression patterns of 8 out of 9 miRNAs were in good agreement with miRNA sequencing data in terms of their direction and magnitude of change. One (miR-126-3p) out of 9 miRNAs did not match with miRNA sequencing which may be due to different approaches for data normalization. Additionally, hierarchical clustering showed clear discrimination of 12 birds into correct group of origin (Figure 2.1).

2.4.3 TARGET PREDICTION AND NETWORK ANALYSIS

In our recent study, DE mRNAs were identified by mRNA sequencing analysis in the same breast muscle tissues of PeM and BPR (Kong et al. 2017). To investigate potential interactions between miRNA and mRNA expression, miRDB (http://mirdb.org), an online tool for miRNA target prediction and functional annotations was used to predict target genes of validated, DE miRNAs. A total of 2,194 genes (mRNAs) were predicted as potential targets for 8 qPCR validated DE miRNAs (except miR-126-3p which showed inconsistent fold change values between miRNAseq and qPCR) (data not shown). Target mRNAs for DE miRNA were integrated with our DE mRNA dataset (retrieved from Kong et al., 2017). Expressions showing opposite direction to corresponding miRNA (e.g., down-regulated transcripts that are targets of up-regulated miRNA in PeM) were chosen for further pathway analysis using the IPA. According to the miRNA-mRNA interaction criteria, 153 candidate target genes for 8 miRNAs were identified (Supplementary Table 2.2) including 118 down-regulated transcripts potentially targeted by 7 up-regulated miRNAs (miR-2131-5p, miR-221-5p, miR-146b-5p, miR-10a-5p, let-7b, miR-125b-5p, and miR-146c-5p). Similarly, it was predicted that 35 up-regulated transcripts might be modulated by down-regulated miRNA (miR-206) in PeM muscle samples.

As results of pathway analyses with DE miRNA and their target genes, the top biological functions of target genes were identified by the IPA using its features, “Top Canonical Pathways” and “Physiological System Development and Function” (Table 2.3). The most relevant biological functions of DE miRNA and their target DE mRNA in skeletal muscle included axonal guidance signaling, glycine degradation, calcium signaling, serine biosynthesis, zymosterol biosynthesis, endocrine system development and function, embryonic development, organismal development, skeletal and muscular system development and function, and tissue organismal development.

2.5 DISCUSSION

In this study, an extensive set of miRNAs was identified by small RNA sequencing and their potential roles in muscle growth and feed efficiency were determined in PeM and BPR chickens. Among a total of 994 mature miRNAs identified, the 20 most abundant mature miRNAs were identified in this study. Of those, miR-21-5p was the most abundant with average read counts of 9,763 and 15,132 in PeM and BPR chickens, respectively (Supplementary Table 2.1). Previous studies have reported upregulated expression of miR-21 in breast muscle of low body weight chickens (Ouyang et al. 2015). In separate studies conducted in chickens and rats, miR-21 inhibited cell proliferation (Wang et al. 2016; Lin et al. 2014), suggesting that lowered expression of miR-21 in PeM may be a factor responsible for rapid muscle growth compared to BPR. The other miRNA, miR-133c-3p, that was abundantly expressed in chicken breast muscle (Sweetman et al. 2008), enhances skeletal muscle proliferation and differentiation by repressing serum response factor (Chen et al. 2006). In addition, miR-22-3p, miR-30a-5p, miR-30d, miR-10b, miR-148a, miR-146c-5p and miR-199 were also known as abundantly expressed in breast muscles (Ouyang et al. 2015). Altogether, evidences in literature suggest that abundant miRNAs identified in our samples may play role in enhanced growth and development of breast muscle of modern broilers.

Of eight DE miRNAs validated with qPCR, expression of miR-146b-5p, miR-125b-5p, miR-2131-5p, let-7b, miR-221-5p miR-146c-5p, miR-126-3p, and miR-10a-5p were higher in PeM muscle while the miR-206 showed lowered expression in PeM. All DE miRNAs have been shown to be involved in muscle development in various animal species (Anderson at al. 2006; Hu et al. 2014; Huang et al. 2010; Khanna et al. 2014; Li et al. 2011; Ouyang et al. 2015). MiR-206 is specifically expressed in skeletal muscle and functions in muscle differentiation and cell proliferation in chickens and human (Jia et al. 2016; Koutsoulidou et al. 2011; Kim at al. 2006).

Let-7b has been reported to be abundantly expressed in breast and skeletal muscle in chickens and to play roles in growth regulation via targeting growth hormone receptor (Lin et al. 2012). MiR-10a is a well characterized miRNA and is known to implicate with muscle development and myogenesis regulation in various animals (Ørom et al. 2008; Huang et al. 2010; Hu et al. 2014; Li et al. 2011). MiR-146b is a known regulator of skeletal myoblast differentiation in vitro and muscle regeneration in mice (Khanna et al. 2014). MiR-126 mediates vascular integrity and angiogenesis. It also elicits direct effects on regulation of skeletal muscle growth and activation of insulin like growth factor 1(IGF-1) (Wang et al. 2008; Rivas et al. 2014). MiR-125b is known to regulate calcification of vascular smooth muscle cells. It targets IGF in both regenerating muscles and myoblasts (Goettsch et al. 2011). There is evidence of expression of miR-221 controlled by the Ras-MAPK pathway. It is involved in vascular smooth muscle proliferation and plays a role in progression from myoblasts to myocytes developing into fully differentiated phenotype (Cardinali et al. 2009; Liu et al. 2009). Therefore, all DE miRNAs identified in this study seem to be closely related with muscle growth and development in chickens to their mammalian counterparts.

To understand cellular and physiological mechanisms in chickens muscle development, DE miRNA and their target DE mRNA were subjected to in silico pathway analysis using the IPA program. The network analysis of DE miRNA-DE mRNA pairs (showing opposite expression patterns between miRNA and DE mRNA) were shown to be interlinked with P38 MAPK, ERK1/2, PI3K, and insulin-signaling pathways (Figure 2.2). Further, functions associated with these genes include embryonic development, organ development, organismal development, and molecular transport. In the network, the up-regulation of annexin A2 (ANXA2), tropomycin 3 (TPM3) and eukaryotic translation initiation factor 2 alpha kinase 3 (EIF2AK3) genes seemed to be directly regulated by down-regulation of miR-206 and their close association with P38 MAPK, ERK1/2,

PI3K and insulin signaling cascades. ANXA2, lipocortin II or p36, which is a 36-kDa Ca^{2+}-dependent protein of the annexin superfamily play regulatory functions in proliferation, migration and cytoskeletal formation in muscle cells (Babiychuk et al. 1999; Chen et al. 2014). TPM3 binds to actin filaments in muscle cells. In association with troponin complex, TPM3 has central role in controlling contraction of striated muscle in vertebrates (Lawlor et al. 2010). EIF2AK3, a metabolic stress sensing protein kinase, phosphorylates eukaryotic translation initiation factor 2 and is also involved in controlling mitochondrial morphology and function (Shi et al. 1999; De et al. 2017). Therefore, it is possible that the interaction between miR-206 and its target genes may reflect the rapid myogenesis shown in breast muscle of PeM chickens. Interestingly, the down-regulated target genes of miR-146, miR-10a-5p, miR-125b-5p, miR-2131-5p and let-7b are also assigned to these signaling pathways.

From the IPA canonical pathway analysis, we found that the targeted DE mRNAs were associated with calcium signaling (p-value 3.03E-02; Figure 2.3), axonal guidance signaling (p-value 9.91E-03; Figure 2.4), and NRF2-mediated oxidative stress response (p-value 3.61E-02; Figure 2.5) pathways. The target genes involved in calcium signaling pathway include: HDAC 11 gene (target of miR-10a-5p); a member of RAS oncogene family (RAP1B) and nuclear factor of activated T-cells 1 (NFATC1) genes (targets of miR-2131-5p); and TPM3 and bone morphogenetic protein 6 genes (targets of miR-206). It is well established that the homeostasis of intracellular calcium level is important for muscle growth and development. The increased intracellular level of Ca^{2+} can occur due to both poor Ca^{2+}-ATPase activity and disturbance of sarcolemma integrity which results in hypercontraction of myofibers and degeneration of muscle mass (Mutryn et al. 2015; Oberc and Engel 1977; Byrd 1992; Emery and Burt 1980); the IPA suggests that this pathway might have been modulated by miR-10a-5p and miR-2131-5p thereby

leading to larger muscle mass observed in PeM compared to BPR chickens. We also found the involvement of target genes of miR-2131-5p in axonal guidance signaling pathway; the predicted target mRNAs of miR-2131-5p included RAP1B, leucyl and cystinyl aminopeptidase, CRK like proto-oncogene, adapter protein, Rho associated coiled-coil containing protein kinase 2, BMP7, and NFATC1. Interestingly, Mutryn et al. (2015) recently suggested possible roles of axonal guidance signaling pathway in the breast muscle myopathy in chickens (Mutryn et al. 2015). Hence, it is postulated that DE miR-2131-5p can play roles in regulation of axonal guidance signaling pathway in breast muscle of chickens. Further investigations and validation works might be warranted with regards to the implications of calcium metabolism and axonal guidance signaling in PeM samples and responsible physiological roles of DE miRNAs therein.

Last, we identified EIF2AK3 and glutathione S-transferase theta 1 (GSTT1) genes (targets of miR-206) involved in NRF2-mediated oxidative stress response pathway. NRF2 is one of the main factors responding to both oxidative and xenobiotic stresses. It plays a critical role in neutralizing oxidative stress by activating the expression of antioxidants and detoxifying enzymes (Choi et al. 2017). The activation of EIF2AK3 and GSTT1 genes is associated with decreased reactive oxygen species levels (Ramnarayanan et al. 2016; Simic et al. 2009). Previously, the augmentation of the NRF2-mediated oxidative stress response pathway is observed in breast muscle of higher feed efficient and rapidly growing chickens (Zhou et al. 2015). Therefore, increased expression of EIF2AK3 and GSTT1 genes, as a result of down-regulation of miR-206, implied the activation of canonical NRF2-mediated response pathway for scavenging reactive oxygen species from breast muscle of rapidly growing PeM chickens.

2.6 CONCLUSION

Using miRNA sequencing, integrated analyses of miRNA-mRNA data and IPA, we were able to identify breast muscle specific miRNAs and their target genes whose concerted actions can contribute to rapid growth and higher feed efficiency in modern broiler chickens. We believe our comprehensive analysis enables us to better understand miRNA and their physiological roles in breast muscle growth in chickens. Future validation studies are warranted in regard to interactions between miRNA and target genes (e.g., in vitro transfection studies) to characterize functions of miRNAs and their specific targets in the context of rapid muscle growth and development.

REFERENCES

Kong B, Lassiter K, Piekarski-Welsher A, Dridi S, Reverter-Gomez A, Hudson NJ, et al. Proteomics of Breast Muscle Tissue Associated with the Phenotypic Expression of Feed Efficiency within a Pedigree Male Broiler Line: I. Highlight on Mitochondria. PloS one. 2016; 11(5):e0155679.

Kong BW, Hudson N, Seo D, Lee S, Khatri B, Lassiter K, et al. RNA sequencing for global gene expression associated with muscle growth in a single male modern broiler line compared to a foundational Barred Plymouth Rock chicken line. BMC Genomics. 2017 Jan 13; 18(1): 82,016-3471-y.

Bushati N, Cohen SM. microRNA functions. Annu Rev Cell Dev Biol. 2007; 23:175-205.

Rana TM. Illuminating the silence: understanding the structure and function of small RNAs. Nature reviews Molecular cell biology. 2007; 8(1):23-36.

Anderson C, Catoe H, Werner R. MIR-206 regulates connexin43 expression during skeletal muscle development. Nucleic Acids Res. 2006; 34(20):5863-71.

Gangaraju VK, Lin H. MicroRNAs: key regulators of stem cells. Nat Rev Mol Cell Biol. 2009 Feb; 10(2):116-25.

Li T, Wu R, Zhang Y, Zhu D. A systematic analysis of the skeletal muscle miRNA transcriptome of chicken varieties with divergent skeletal muscle growth identifies novel miRNAs and differentially expressed miRNAs. BMC Genomics. 2011; 12(1):186.

Wienholds E, Kloosterman WP, Miska E, Alvarez-Saavedra E, Berezikov E, de Bruijn E, et al. MicroRNA expression in zebrafish embryonic development. Science. 2005 Jul 8; 309(5732):310-1.

Ouyang H, He X, Li G, Xu H, Jia X, Nie Q, et al. Deep Sequencing Analysis of miRNA Expression in Breast Muscle of Fast-Growing and Slow-Growing Broilers. International journal of molecular sciences. 2015; 16(7):16242-62.

Liu X, Cheng Y, Zhang S, Lin Y, Yang J, Zhang C. A necessary role of miR-221 and miR-222 in vascular smooth muscle cell proliferation and neointimal hyperplasia. Circ Res. 2009 Feb 27; 104(4):476-87.

Chen J, Mandel EM, Thomson JM, Wu Q, Callis TE, Hammond SM, et al. The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. Nat Genet. 2006; 38(2):228-33.

Wong CF, Tellam RL. MicroRNA-26a targets the histone methyltransferase Enhancer of Zeste homolog 2 during myogenesis. J Biol Chem. 2008 Apr 11; 283(15):9836-43.

Lopez G, De Lange K, Leeson S. Partitioning of retained energy in broilers and birds with intermediate growth rate. Poult Sci. 2007; 86(10):2162-71.

Bottje W, Iqbal M, Tang ZX, Cawthon D, Okimoto R, Wing T, et al. Association of mitochondrial function with feed efficiency within a single genetic line of male broilers. Poult Sci. 2002 Apr; 81(4):546-55.

Kong BW, Song JJ, Lee JY, Hargis BM, Wing T, Lassiter K, et al. Gene expression in breast muscle associated with feed efficiency in a single male broiler line using a chicken 44K oligo microarray. I. Top differentially expressed genes. Poult Sci. 2011 Nov; 90(11):2535-47.

Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010.

Wang W, Cheng M, Qiao S, Wang Y, Li H, Wang N. Gga-miR-21 inhibits chicken pre-adipocyte proliferation in part by down-regulating Kruppel-like factor 5. Poult Sci. 2016; 96(1):200-10.

Lin L, Gan H, Zhang H, Tang W, Sun Y, Tang X, et al. MicroRNA-21 inhibits SMAD7 expression through a target sequence in the 3'untranslated region and inhibits proliferation of renal tubular epithelial cells. Molecular medicine reports. 2014; 10(2):707-12.

Sweetman D, Goljanek K, Rathjen T, Oustanina S, Braun T, Dalmay T, et al. Specific requirements of MRFs for the expression of muscle specific microRNAs, miR-1, miR-206 and miR-133. Dev Biol. 2008; 321(2):491-9.

Hu R, Pan W, Fedulov AV, Jester W, Jones MR, Weiss ST, et al. MicroRNA-10a controls airway smooth muscle cell proliferation via direct targeting of the PI3 kinase pathway. FASEB J. 2014 May; 28(5):2347-57.

Huang H, Xie C, Sun X, Ritchie RP, Zhang J, Chen YE. MiR-10a contributes to retinoid acid-induced smooth muscle cell differentiation. J Biol Chem. 2010 Mar 26; 285(13):9383-9.

Khanna N, Ge Y, Chen J. MicroRNA-146b promotes myogenic differentiation and modulates multiple gene targets in muscle cells. PLoS One. 2014; 9(6):e100657.

Jia X, Lin H, Abdalla BA, Nie Q. Characterization of miR-206 Promoter and Its Association with Birthweight in Chicken. Int J Mol Sci. 2016 Apr 14; 17(4):559.

Koutsoulidou A, Mastroyiannopoulos NP, Furling D, Uney JB, Phylactou LA. Expression of miR-1, miR-133a, miR-133b and miR-206 increases during development of human skeletal muscle. BMC developmental biology. 2011; 11(1):34.

Kim HK, Lee YS, Sivaprasad U, Malhotra A, Dutta A. Muscle-specific microRNA miR-206 promotes muscle differentiation. J Cell Biol. 2006 Aug 28; 174(5):677-87.

Lin S, Li H, Mu H, Luo W, Li Y, Jia X, et al. Let-7b regulates the expression of the growth hormone receptor gene in deletion-type dwarf chickens. BMC Genomics. 2012; 13(1):306.

Ørom UA, Nielsen FC, Lund AH. MicroRNA-10a binds the 5′ UTR of ribosomal protein mRNAs and enhances their translation. Mol Cell. 2008; 30(4):460-71.

Wang S, Aurora AB, Johnson BA, Qi X, McAnally J, Hill JA, et al. The endothelial-specific microRNA miR-126 governs vascular integrity and angiogenesis. Developmental cell. 2008; 15(2):261-71.

Rivas DA, Lessard SJ, Rice NP, Lustgarten MS, So K, Goodyear LJ, et al. Diminished skeletal muscle microRNA expression with aging is associated with attenuated muscle plasticity and inhibition of IGF-1 signaling. FASEB J. 2014 Sep; 28(9):4133-47.

Goettsch C, Rauner M, Pacyna N, Hempel U, Bornstein SR, Hofbauer LC. MiR-125b regulates calcification of vascular smooth muscle cells. The American journal of pathology. 2011; 179(4):1594-600.

Cardinali B, Castellani L, Fasanaro P, Basso A, Alema S, Martelli F, et al. Microrna-221 and microrna-222 modulate differentiation and maturation of skeletal muscle cells. PloS one. 2009; 4(10): e7607.

Liu X, Cheng Y, Zhang S, Lin Y, Yang J, Zhang C. A necessary role of miR-221 and miR-222 in vascular smooth muscle cell proliferation and neointimal hyperplasia. Circ Res. 2009 Feb 27; 104(4):476-87.

Babiychuk EB, Palstra RJ, Schaller J, Kampfer U, Draeger A. Annexin VI participates in the formation of a reversible, membrane-cytoskeleton complex in smooth muscle cells. J Biol Chem. 1999 Dec 3; 274(49):35191-5.

Chen L, Li YS, Cui J, Ning JN, Wang GS, Qian GS, et al. MiR-206 controls the phenotypic modulation of pulmonary arterial smooth muscle cells induced by serum from rats with hepatopulmonary syndrome by regulating the target gene, annexin A2. Cell Physiol Biochem. 2014; 34(5):1768-79.

Lawlor MW, DeChene ET, Roumm E, Geggel AS, Moghadaszadeh B, Beggs AH. Mutations of tropomyosin 3 (TPM3) are common and associated with type 1 myofiber hypotrophy in congenital fiber type disproportion. Hum Mutat. 2010; 31(2):176-83.

Shi Y, An J, Liang J, Hayes SE, Sandusky GE, Stramm LE, et al. Characterization of a mutant pancreatic eIF-2alpha kinase, PEK, and co-localization with somatostatin in islet delta cells. J Biol Chem. 1999 Feb 26; 274(9):5723-30.

De Mario A, Quintana-Cabrera R, Martinvalet D, Giacomello M. (Neuro) degenerated Mitochondria-ER contacts. Biochem Biophys Res Commun. 2017; 483(4):1096-109.

Mutryn MF, Brannick EM, Fu W, Lee WR, Abasht B. Characterization of a novel chicken muscle disorder through differential gene expression and pathway analysis using RNA-sequencing. BMC Genomics. 2015; 16(1):399.

Oberc MA, Engel WK. Ultrastructural localization of calcium in normal and abnormal skeletal muscle. Lab Invest. 1977 Jun; 36(6):566-77.

Byrd SK. Alterations in the sarcoplasmic reticulum: a possible link to exercise-induced muscle damage. Med Sci Sports Exerc. 1992 May; 24(5):531-6.

Emery AE, Burt D. Intracellular calcium and pathogenesis and antenatal diagnosis of Duchenne muscular dystrophy. Br Med J. 1980 Feb 9; 280(6211):355-7.

Choi B, Da Young Ryu, In-Geun Ryoo, Kwak M. NFE2L2/NRF2 silencing-inducible miR-206 targets c-MET/EGFR and suppresses BCRP/ABCG2 in cancer cells. Oncotarget. 2017; 8(63):107188.

Ramnarayanan S, Kyathanahalli C, Ingles J, Park-York M, Jeyasuria P, Condon JC. The Unfolded Protein Response Regulates Uterine Myocyte Antioxidant Responsiveness During Pregnancy. Biol Reprod. 2016; 95(6):120, 1-9.

Simic T, Savic-Radojevic A, Pljesa-Ercegovac M, Matic M, Mimic-Oka J. Glutathione S-transferases in kidney and urinary bladder tumors. Nature Reviews Urology. 2009; 6(5):281.

Zhou N, Lee WR, Abasht B. Messenger RNA sequencing and pathway analysis provide novel insights into the biological basis of chickens' feed efficiency. BMC Genomics. 2015; 16(1):195.

APPENDIX

Table 2.1 Primers used for qPCR. The first column indicates primer names and the second column shows their sequences.

Name	Sequence
RTQ_primer	CGAATTCTAGAGCTCGAGGCAGGCGACATGGCTGGCTAGTTAAGCTTG GTACCGAGCTCGGATCCACTAGTCCTTTTTTTTTTTTTTTTTTTTTTTVN
RTQ-UNIr	CGAATTCTAGAGCTCGAGGCAGG
miR-146c-5p	TGAGAACTGAATTCCATGGACTG
miR-146b-5p	TGAGAACTGAATTCCATAGGCG
miR-10a-5p	TACCCTGTAGATCCGAATTTGT
miR-2131-5p	CTGTTACTGTTCTTCTGATGG
miR-221-5p	AACCTGGCATACAATGTAGATTTCTGT
miR-10b-5p	TACCCTGTAGAACCGAATTTGT
let-7b	TGAGGTAGTAGGTTGTGTGGTT
miR-125b-5p	TCCCTGAGACCCTAACTTGTGA
miR-206	TGGAATGTAAGGAAGTGTGTGG
5S_rRNA-F1	AAGCCTACAGCACCCGGTAT

Table 2.2 Comparison of fold change between miRNAseq and qPCR in breast muscle tissue of PeM compared to BPR broilers.

miRNA	Sequence	miRNAseq *	qPCR*
miR-2131-5p	AUGCAGAAGUGCACGGAAACAGCU	2.62	1.45
miR-221-5p	AACCUGGCAUACAAUGUAGAUUUCUGU	2.42	1.43
miR-126-3p**	UCGUACCGUGAGUAAUAAUGCGC	1.49	-0.87**
miR-146b-5p	UGAGAACUGAAUUCCAUAGGCG	2.44	1.88
miR-10a-5p	UACCCUGUAGAUCCGAAUUUGU	1.96	1.17
miR-206	UGGAAUGUAAGGAAGUGUGUGG	-0.72	-0.62
let-7b	UGAGGUAGUAGGUUGUGUGGUU	1.20	1.20
miR-125b-5p	UCCCUGAGACCCUAACUUGUGA	1.51	1.17
miR-146c-5p	UGAGAACUGAAUUCCAUGGACUG	1.42	1.31

* Values denote linear fold changes

** Indicate inconsistent fold change between RNAseq and qPCR

Table 2.3 Top biological functions of target genes between PeM and BPR as presented by IPA.

Top Canonical Pathways	**p-value**
Axonal Guidance Signaling	0.00991
Glycine Degradation (Creatine Biosynthesis)	0.0131
Calcium Signaling	0.0303
Serine Biosynthesis	0.0325
NRF2-mediated oxidative stress response	0.0361
Zymosterol Biosynthesis	0.0389
Physiological System Development and Function	**p-value range**
Endocrine System Development and Function	1.96E-02 - 2.56E-04
Embryonic Development	2.86E-02 - 4.33E-04
Organismal Development	2.92E-02 - 4.33E-04
Skeletal and Muscular System Developmental and Function	2.61E-02 - 6.36E-04
Tissue Development	2.86E-02 - 8.86E-04

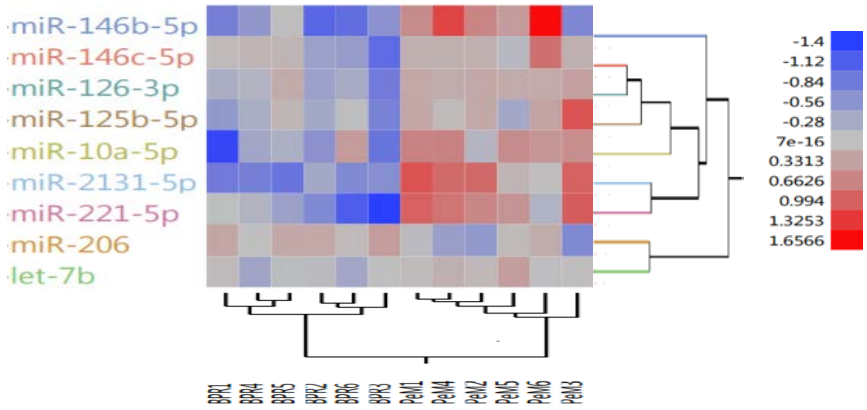

Figure 2.1 Hierarchically clustered heat map of 9 DE miRNA. Red and blue represent up and down-regulated expression in PeM respectively. Color density indicated level of fold change.

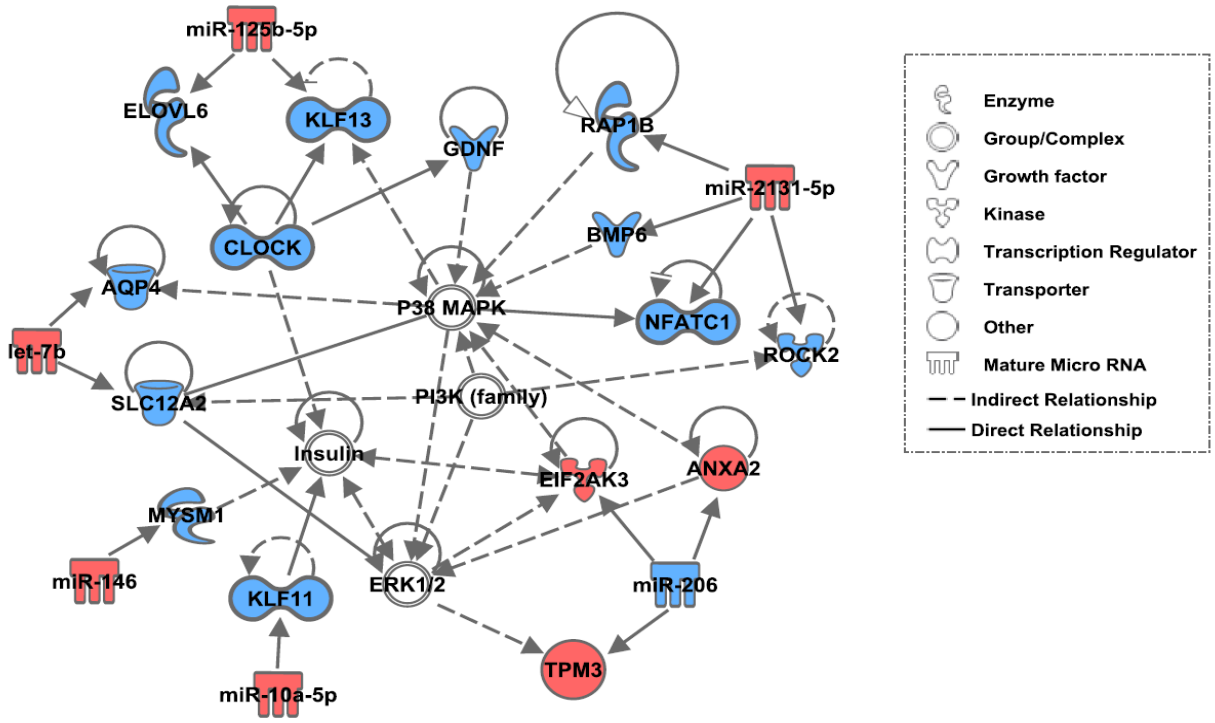

Figure 2.2 Network associated with P38 MAPK, ERK1/2, PI3K and insulin signaling pathways. Genes and miRNA filled with red are up-regulated while symbols in blue color are down-regulated in PeM in comparison to BPR.

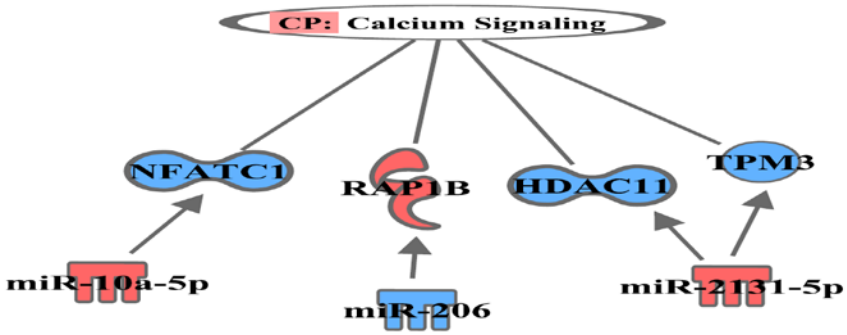

Figure 2.3 Pathway of calcium signaling predicted by IPA. Color symbols were indicated in the legend of Figure 2.2.

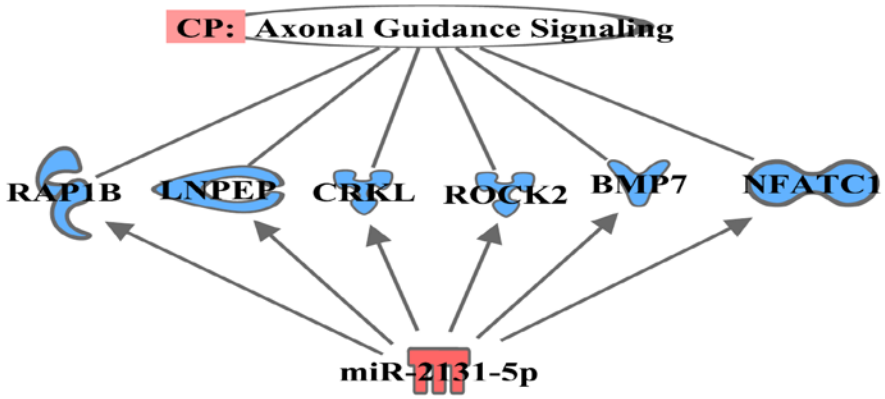

Figure 2.4 Pathway of axonal guidance. Color symbols were indicated in the legend of Figure 2.2.

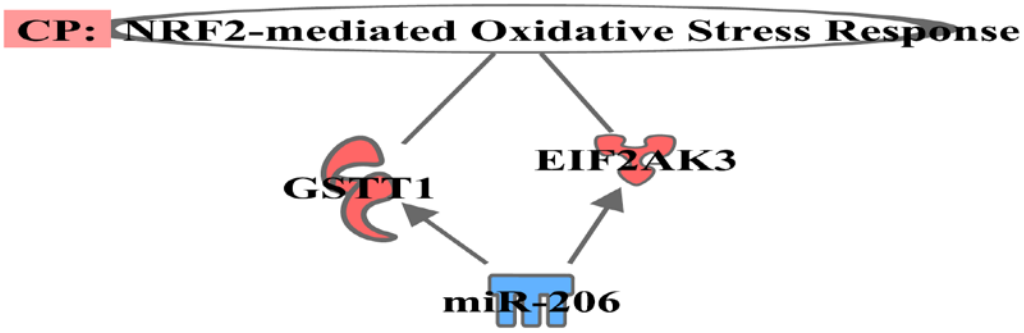

Figure 2.5 Pathway of NRF2-mediated oxidative stress response. Color symbols were indicated in the legend of Figure 2.2.

CHAPTER 3

Copy number variation study in Japanese quail associated with stress related traits using whole genome re-sequencing data

B. Khatri, S. Kang, S. Shouse, N. Anthony, W. Kuenzel and B. Kong*

Department of Poultry Science, Center of Excellence for Poultry Science

University of Arkansas, Fayetteville, Arkansas

*Correspondence to: Dr. Byungwhi Caleb Kong

Department of Poultry Science,

University of Arkansas, Fayetteville, AR 72701

Tel: (479) 575-5494 FAX: (479) 575-7139

E-mail: bkong@uark.edu

3.1 ABSTRACT

Background: Copy number variation (CNV) is a major driving factor for genetic variation and phenotypic diversity in animals. Japanese quail is an important model for understanding effects of stress in well-being and growth performances of animals. To detect CNVs and understand genetic components underlying stress related traits, we performed whole genome re-sequencing analysis in High Stress (HS) and Low Stress (LS) birds.

Results: We assembled Illumina HiSeq data using reference based assembly on the quail genome sequences and performed bioinformatics analyses. Our bioinformatics analyses were then complemented with experimental validation using real time quantitative PCR (qPCR). The depth of coverage for re-sequenced data using pooled DNA of each 12 HS and LS birds reached to 41.4x and 42.5x respectively. Using read-depth approach and CNVnator tool to detect CNVs in the aligned data sets, we found 262 (235 deletions and 27 duplications) and 168 (148 deletions and 20 duplications) CNV regions (CNVRs) affecting 15.20 Mb (1.6%) and 18.17 Mb (1.9%) of reference genome in HS and LS birds respectively. Using BEDOPS tool and custom bash script, we extracted genes from CNVRs and found 454 unique genes in HS and 493 in LS birds associated with CNV. Results of Ingenuity Pathway Analysis (IPA) showed that the CNV genes were significantly enriched to phospholipase C signaling, neuregulin signaling, reelin signaling in neurons, CD27 signaling in lymphocytes and associated with endocrine and nervous development, humoral immune response, and carbohydrate and amino acid metabolisms in HS birds compared to different sets of signaling pathways, cell-mediated immune response, protein and lipid metabolisms in LS birds.

Conclusion: We identified different sets of genes affected by CNVs in HS and LS birds, most notably involved in nervous and endocrine systems development, humoral and cell-mediated

immune response and different metabolisms, which may be suggested as candidate marker genes responsible for stress related traits.

Keywords: whole genome re-sequencing, copy number variation, restraint stress, Japanese quail

3.2 BACKGROUND

Understanding the evolutionary process that leads to divergence in animals requires study of their genetic variation. Genomic variation is a principal factor responsible for phenotypic diversity in animals (Pezer et al. 2015). Basically, genomic variation can encompass a wide range of alterations from small indels to large deletion or duplication of the entire genome. The deletion or duplication of a certain region of DNA causes change in copy number variation (CNV) (Redon et al. 2006). CNV is arbitrarily defined as DNA segment that is 1 kb or larger and present at variable copy number in comparison with a reference genome (Feuk et al. 2006). It is estimated that DNA region that have CNV can account for 4.8-9.5% of human genome and surpass the diversity caused by single nucleotide polymorphisms (Stankiewicz and Lupski 2010; Zarrei et al. 2015).

Due to their larger sizes and abundances, CNVs may have impact on functions of many genes and consequently fitness in animals (Schrider and Hahn 2010). To date, four different mechanisms have been proposed for the formation of CNVs including non-allelic homologous recombination (NAHR), non-homologous end joining (NHEJ), Fork Stalling and Template Switching (FoSTeS) and Retrotransposition (Hastings et al. 2009). CNVs potentially influence to gene structure, gene dosage, and gene expression by exposing recessive alleles (Bickhart and Liu 2014) or indirectly through the perturbation of regulatory region of genes (Zhang et al. 2009).

Several studies have identified CNVs associated with phenotypic variations and complex disorders in human, such as schizophrenia, developmental delay, mental retardation, autism, systemic lupus erythematosus, diabetes, obesity, psoriasis, neuroblastoma and susceptibility to HIV infection (Girirajan et al. 2011; Henrichsen et al. 2009; Diskin et al. 2009). Phenotypic diversity associated with CNV has also been characterized in various domestic animals. The pea-

comb phenotype characterized by decrease in comb size in male and female chickens is due to duplication of the first intron of sex determining region Y (SRY)-box5 (*Sox5)* gene (Wright et al. 2009). A late feathering phenotype in chicken is due to partial duplication of prolactin receptor (*PRLR)* and sperm flagella 2 (*SPEF2)* genes (Elferink et al. 2008). Similarly excessive black pigmentation phenotype in chickens is due to duplication of 130 kb locus containing endothelin 3 (*EDN3)* gene (Shinomiya et al. 2012). White coat phenotypes in sheep and pigs are due to duplications of agouti signaling protein (*ASIP)* and KIT proto-oncogene receptor tyrosine kinase (*KIT)* genes, respectively (Fontanesi et al. 2009; Giuffra et al. 2002). The dorsal hair ridge in Rhodesian and Thai dogs and their susceptibility to dermoid sinus is caused by duplication of fibroblast growth factors (*FGF3*, *FGF4*, and *FGF19)* and oral cancer overexpressed 1 (*ORAOV1)* genes (Hillbertz et al. 2007). CNVs have also been reported to be associated with disease resistance and developmental disorders in animals. Loss of MHC class I antigen-presenting proteins are associated with Marek's disease resistance in chicken (Luo et al. 2013). Gain of class II major histocompatibility complex transactivator (*CIITA)* is associated with nematode resistance in cattle (Liu, George E 2011). Likewise, cone-rod dystrophy 3 (Goldstein et al. 2010), startle diseases in dogs (Gill et al. 2011), and osteopetrosis, abortion and stillbirths in cattle have been linked to CNVs (Meyers et al. 2010; Flisikowski et al. 2010). From these findings, we hypothesize that CNVs can be important biomarkers for phenotypic traits or disease resistance in animals.

Two genetically distinct line of Japanese quail named as high stress (HS) and low stress (LS) have been selected by divergent plasma corticosterone response to restraint stress during the 1980s (Satterlee, DG 1988). Since then these two lines have been used as stress responding animal models in poultry. In LS line, the mean corticosterone level is approximately one-third lower compared to HS line. Compared to HS line, LS line is less fearful and more social, higher in body

weight, higher in egg production, and lower in heterophil/lymphocyte ratio. LS line shows lower stress-induced osteoporosis, accelerated onset of puberty, and heightened male sexual activity and efficiency compared to HS line (Satterlee and Johnson 1988; Huff et al. 2013). In this study, we have performed CNV analysis with whole genome re-sequencing data of high and low stress lines of Japanese quail particularly focusing on identifying full length genes within CNVs. These genes could be relevant for divergence and adaptation of the two lines of quail.

3.3 MATERIALS AND METHODS

3.3.1 ETHICS AND STATEMENT

This study was conducted following the recommended guidelines for the care and use of laboratory animals for the National Institutes of Health. All procedures for animal care were performed according to the animal use protocols that were reviewed and approved by the University of Arkansas Institutional Animal Care and Use Committee.

3.3.2 BIRDS AND DNA SEQUENCING

The early process of development and selection of HS and LS lines of Japanese quail for their plasma corticosterone response to immobilization for up to 12 generations was explained by Satterlee and Johnson (1988). Since then, an independent random mating condition has been used for their maintenance (Satterlee et al. 2000; Satterlee et al. 2002; Suhr et al. 2010). These research lines were shipped to University of Arkansas at generation 44 from Louisiana State University and maintained at Arkansas Agricultural Experimentation Station, Fayetteville, AR (Huff et al. 2013).

We used adult male HS and LS birds for this study because of their stable physiology. We collected blood samples (3 mL) from 20 birds each from HS and LS lines. Genomic DNA was purified from each sample using QiaAmp DNA mini kit (Qiagen, Hilden, Germany) following manufacturer's method. DNA quality was assessed using NanoDrop 1000 (Thermo Scientific, Waltham, MA) and agarose gel electrophoresis. Twelve samples showing highest DNA quality per line were pooled to represent each line. Library preparation and Illumina sequencing for the pooled DNA samples were performed by the Research Technology Support Facility at Michigan State University (East Lansing, MI) using Illumina HiSeq 2×150 bp paired end read technology.

3.3.3 DATA QUALITY ASSESSMENT AND SEQUENCE ASSEMBLY

We used the FastQC tool (v0.11.6) to assess the quality of raw reads obtained after sequencing in form of FASTQ files (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). After quality assessment, the low-quality reads were trimmed out using Trimmomatic tool (v0.32) (Bolger et al. 2014). The clean reads were then mapped onto the Japanese quail reference genome obtained from NCBI (https://www.ncbi.nlm.nih.gov/genome/113) using Bowtie2 (v2.3.3.1) with the default settings for the parameters (Langmead, Ben 2012). We removed PCR duplicates using the rmdup command line of SAMtools (v0.1.19) and SAMtools was further used to convert SAM to BAM files and then to sorted BAM files to save run time in subsequent analysis (Li et al. 2009).

3.3.4 CNV DETECTION AND COPY NUMBER ESTIMATION

We used CNVnator software (v0.3.3) to predict CNV in sorted BAM files relative to reference quail genome (Abyzov et al. 2011). Optimal bin sizes of 1200 and 1500 were chosen for HS and LS respectively according to author's recommendations, in which the ratio of average read-depth signal to its standard deviation was between 4 and 5 (Abyzov et al. 2011; Pezer et al. 2015; Harr at al. 2016). All the CNV calls in both HS and LS samples were greater than 1 kb. CNV calls were filtered according to criteria recommended by (Abyzov et al. 2011). CNV showing P-value <0.01 (e-val1 calculated using t-test statistics), size >1 kb, and q0 < 0.5 (q0: fraction of mapped reads with zero quality) were filtered and used for downstream analysis.

We estimated gene copy number (CN) in HS and LS birds across genome length using the "-genotype" option of CNVnator. We wrote a custom bash script and retrieved CNV genes from CNV regions (CNVRs) of HS and LS lines using RefSeq genes from NCBI and BEDOPS tool (v2.4.30) (Neph et al. 2012).

3.3.5 REAL TIME QUANTITATIVE PCR FOR CNV VALIDATION

Real time quantitative PCR (qPCR) was used to validate CNVs detected by CNVnator with HS and LS lines. A total of 9 genes showing CNV were randomly chosen and primers were designed using Primer3 software and listed in Table 3.1. Primer specificities were checked using Primer-BLAST tool of NCBI. A segment of the β-actin gene, which is present in two copies per diploid and showed no CNV in either line of quail, was chosen as control in all reactions. Five nanogram of genomic DNA was subjected to qPCR (total volume of 25 µL) in triplicate reactions using ABI prism 7500HT system (ThermoFisher Scientific) with PowerUp^{TM} SYBER® Green Master Mix (ThermoFisher Scientific). The conditions of real-time qPCR amplification were as follows: 1 cycle at 95°C (10 min), 40 cycles at 95°C (15 s each), followed by 60°C for 1 min. We used ΔΔCt method for calculating relative copy number of each gene. First, the cycle threshold (Ct) value of each gene was normalized against the control gene, and then ΔCt value was determined between test gene and reference gene predicted as normal copy number by CNVnator. Finally values around 3 or above were considered as duplications or gain and around 1 or less as deletion or loss.

3.3.6 FUNCTIONAL ANNOTATION OF CNV GENES

We analyzed genes retrieved from CNVRs in terms of gene ontology and molecular networks using Ingenuity Pathway Analysis (IPA; http://www.ingenuity.com; Qiagen, Valencia, CA). We imported lists of unique genes overlapped with CNVRs of HS and LS lines of quail into IPA separately and subsequently mapped to their corresponding annotations in the Ingenuity Pathway Knowledge Base. IPA identifies networks accommodating these unique genes in comparison with comprehensive global networks. IPA illustrates each molecular network with an assigned relevance score, the number of focus molecules, and top functions of the network. During

analysis, we set each network to the limit of 35 molecules by default and only human was chosen for the species option. We used evidence of experimentally observed for the confidence level. Molecules in network are represented by nodes, distinguished by their shapes based on their functional category, and are connected by distinct edges based on interaction among molecules.

3.4 RESULTS AND DISCUSSION

3.4.1 GENOME RE-SEQUENCING AND DISTRIBUTION OF CNVS

We performed whole genome resequencing of pooled DNA samples from 12 birds each from HS and LS lines of the quail and produced ~250 and ~257 million reads of 150 bp respectively and their respective depth of coverage reached to ~41x and ~42x for HS and LS, respectively (Table 3.2). Of four basic strategies for CNV analyses including as read pair, read-depth, split-read and sequence assembly of next generation sequencing data, we used the software tool CNVnator (Abyzov et al. 2011) that works under read-depth approach as the most suitable method to detect CNVs in our data and address our hypothesis. CNVnator has several advantages over other methods with respect to accurate CNV detection, precise break point resolution, and detection of different sizes of CNVs, from a few hundred bases to several megabases in the whole genome. In addition, CNVnator has high sensitivity (86-96%), low false discovery rate (3-20%) and high genotyping accuracy (93-95%) (Abyzov et al. 2011; Medvedev et al. 2009; Zhao et al. 2013). With analyzing data to call CNVs from the mapped data, we considered calls (deletions or duplications) ≥1 kb length in our analysis which makes more reliable to detect CNVs (Abyzov et al. 2011).

A total of 262 and 168 CNVRs were identified in HS and LS lines, respectively. Among these, 235 and 27 were deletion and duplication CNVRs, respectively, in HS, while 148 and 20 were deletion and duplication CNVRs, respectively, in LS lines (Table 3.3). The distribution of

deletions and duplications over each chromosome is shown in Figure 3.1. Interestingly, no CNVRs were detected in chromosome 6 and 16 of LS line but CNVRs were present in HS line. The number of CNVRs in each chromosome was proportional to its length. Replication and recombination mechanisms have been suggested as possible events for the CNV formation across genome. Since the recombination rate is generally higher in longer DNA, more CNVRs may present in large chromosomes in our study (Jia et al. 2013; Bickhart and Liu 2014). The chromosome 16 in chickens has the major histocompatibility complex (MHC) genes that encode key proteins regulating immune responses (Delany et al. 2009). A study by Huff et al. (2013) reported HS birds were more susceptible to infection of *Salmonella* species as compared to LS birds (Huff et al. 2013). Therefore, the deletion event detected in chromosome 16 of HS birds might be the cause for more susceptibility of HS birds to diseases.

We found fewer copy numbers with zero state deletions (e.g. genes completely deleted or homozygous deletion) compared to one state deletion (e.g. one copy deleted or hemizygous deletion) in both HS and LS lines of quail, which was similar to those observed in chickens (Locke et al. 2015). The deletions outnumber duplications by a ratio of 8.70:1 in HS and 7.4:1 in LS (Table 3.4), which is consistent with previous studies where more deletion events were discovered than duplications (Crooijmans et al. 2013). The length of CNVRs ranged from 6.0-1341.6 kb in HS and 7.5-1101 kb in LS lines (Figure 3.2). The total length of deletion CNVRs accounted for 13.8 Mb in HS and 17.02 Mb in LS lines. Similarly, the total lengths of CNVRs associated with duplication were 1.32 Mb in HS and 1.15 Mb in LS lines. The average lengths of CNVRs were ~50 kb in HS and ~100 kb in LS lines (Table 3.3). The CNVRs covered 1.6 and 1.9 % of quail genome in HS and LS, respectively. We found the amount of quail genome affected by CNVs similar to those reported for chickens (1.42%, 2.61%) (Jia et al. 2013; Zhou et al. 2014), dogs (1.08%) (Berglund

et al. 2012), and Holstein cattle (1.61%) (Jiang, Li 2013) but lower than in swine (4.23%) (Wang et al. 2012), mice (6.87% or 8.15%) (Locke et al. 2015) and human (5.9% and 12%) (Shaikh et al. 2009; Redon et al. 2006). However, these values could be affected by sample size, diversity of samples, sequencing technology and CNV calling methods per each study (Locke et al. 2015).

3.4.2 CNV VALIDATION USING QPCR

In this study, we used pooled DNA samples from each line for whole genome re-sequencing. We validated genes associated with CNV in 16 individual birds each from HS and LS lines using qPCR. We randomly selected 9 different genes retaining each CNVRs for their validation. We used ΔΔCt method for determining relative CN of the genes. We found ~80% (7 out of 9) of our qPCR results agreed with the CN state predicted by CNVnator (Table 3.9). The result clearly showed that there is difference in CNV in 9 genes between HS and LS lines of quail. Thus, differential CNVs observed in genes between HS and LS lines can be causes for their phenotypic variations in stress responses.

3.4.3 GENE CONTENT OF CNVRS AND BIOINFORMATICS ANALYSIS

We used a custom bash script, BEDOPS tool and reference genome annotation file (in GFF format) of Japanese quail from NCBI to extract genes from CNVRs of both the lines. We retrieved a total of 948 genes completely overlapped within CNVRs in HS and 982 in LS lines. The total number of genes overlapped with deletion CNVRs was 895 in HS and 922 in LS lines while their respective number of genes overlapped with duplication CNVRs was 53 and 60. Among the genes overlapped with deletion CNVRs, 436 were unique in HS and 471 in LS lines (Table 3.4). Similarly, we found 18 unique genes overlapped with duplication CNVRs in HS and 22 in LS lines (Table 3.4). Structural genetic variations have been known to accumulate during inbreeding process in animals (Katju and Bergthorsson 2013). However, the effects of the inbreeding process

in accumulation of genetic variation in quail populations was not known to date. We have identified several hundred genes that were fully deleted in HS and LS lines of quail, which supports a phenomenon of perpetual gene turnover in the two quail populations and their genetic differences. Duplication of whole genes has been known to impact gene expression by altering gene dosage (Stingele et al. 2012; Veitia 2004). If a duplication of a gene is adaptive, it is usually favored and retained more frequently in a population (Pezer et al. 2015). We found 23 genes in HS and 32 in LS lines that, on average, had 10 or more copies and are considered as high copy number genes. These gene lists included both annotated and unannotated genes with 11 genes having compatible copy number between HS and LS lines. The high copy number annotated genes in HS included PCR11 cleavage and polyadenhylation factor (PCF11), ankyrin repeat domain 42 (ANKRD42), obscurin, cytoskeletal calmodulin and titin-interacting RhoGEF (OBSCN), chromosome 2 H6orf52 homolog (C2H6orf52), nucleoporin 153 (NUP153), core-binding factor alpha subunit 2 (CBFA2T2), syntrophin alpha 1 (SNTA1), and dynein axonemal intermediate chain 1 (DNAI1). The high copy number annotated genes in LS lines were PCF11, ANKRD42, and hydroxysteroid dehydrogenase like 2 (HSDL2). The high copy number genes (i.e. duplication genes) in HS line were associated with cellular assembly and organization, cellular morphology, nervous system development and function.

We used IPA to characterize the biological functions, describe molecular interaction networks and canonical pathways implicated by unique genes overlapped with deletion CNVRs in HS and LS lines. We identified five canonical pathways significantly enriched (p-value < 0.01) by the deletion genes in HS lines (Table 3.5). The pathways include Phospholipase C Signaling, Reelin Signaling in Neurons, ERK5 Signaling, CD27 Signaling in Lymphocytes and Neuregulin Signaling. Similarly, six canonical pathways were significantly enriched (p-value <0.01) by the

deletion genes in LS lines which include Type II Diabetes Mellitus Signaling, GP6 Signaling, nNOS Signaling in Skeletal Muscle Cells, Hepatic Fibrosis/Hepatic Stellate Cell Activation, and role of CHK Proteins in Cell Cycle Checkpoint Control (Table 3.5). We found the top diseases and biological functions of the deletion genes in HS line related to endocrine system disorders, organismal injury and abnormalities, neurological disease, and gastrointestinal disease. Similarly, the top diseases and bio functions of the deletion genes in LS line were related to endocrine system disorderorganismal injury and abnormalities, connective tissue disorders, and reproductive system disease (Table 3.6). The unique deletion genes in HS that are involved in endocrine system disorder are listed in Table 3.7. The deletion genes in LS line produced a network associated with lipid metabolism. Thus, in contrast to LS line, canonical signaling pathways in HS are related to regulation of immune response, stress and neurological diseases. Therefore, a higher level of mean corticosterone level seen in HS compared to LS lines may be associated with the genes with CNVs. These differences might implicate CNV as an adaptive change in response to restraint stress between HS and LS lines of Japanese quail. This type of adaptive variation at DNA level can improve the fitness of organisms to new and challenging environments (Hull et al. 2017).

We identified a total of 17 gene networks in HS and 18 in LS lines with score not less than 10 among which 4 different networks in HS and 5 in LS lines were significantly involved in nervous and endocrine systems development (Table 3.7). A score of 10 implies that there will be less than a 10^{-10} probability that the genes in the network are associated with each other by chance. The topmost network involving deletion genes in HS line was specifically associated with cell to cell signaling and interaction, cellular assembly and organization, nervous system development and function (Figure 3.3). The genes associated with loss in this network are involved in signaling pathways of ERK1/2 connected to CaMKII (Ca^{2+}/calmodulin-dependent protein kinase II),

PPP1R9B (Protein phosphatase 1 regulatory subunit 9B), APH1A (Aph-1 homolog A), proinsulin and growth hormone. CaMKII, PPP1R96, and APH1A are involved in nervous system development and functions. CaMKII functions in various cells by phosphorylating proteins involved in synaptic plasticity, electrical excitability and neurotransmitter synthesis (Tsui,J. 2005). PPP1R9B gene is expressed in dendritic spines and plays a role in receiving signals from the central nervous system (Carnero 2012). This gene is switched off during chronic stress condition (Aalling et al. 2018). APH1A gene encodes a component of gamma secretase complex that is involved in proteolysis of amyloid precursor protein (Fortna et al. 2004). The deletion of APH1A gene impairs the function of secretase gamma that can render neurons susceptible to stress and normal functioning secretase gamma is essential for neuronal integrity (Kallhoff-Munoz et al. 2008). The ERK1/2 signaling pathway activated by upstream mitogen-activated protein kinases regulates synaptogenesis, neuronal excitability and histone modification. It is found to be hypoactive in brain of stress-induced depressed individuals (Dwivedi and Zhang 2016). The connection in this network therefore suggests loss of CaMKII, PPP1R9B, APH1A and other genes in this pathway may impair functional interactions of ERK1/2 signaling pathway with growth hormones, proinsulin and secretase gamma. This may be a reason for high stress and therefore reduced growth rate and low basal weight observed in HS compared to LS birds.

Reduced heterophil/lymphocyte ratio is observed in LS compared to HS line of Japanese quail (Huff et al. 2013). Interestingly, we found humoral immune response in HS and cell-mediated immune response in LS lines associated with gene deletion (Figure 3.4). In this network, the deletion genes are associated with signaling pathway of P38 MAPK connected to CDKN1A (cyclin dependent kinase inhibitor 1A), PRKCE (protein kinase C epsilon) and CSF3 (colony stimulating factor 3). The protein encoded by CDKN1A inhibits cyclin-dependent kinase 2

(CDK2) and function in regulation of cell cycle progression at the G1 phase (Cloonan et al. 2008). PRKCE is involved in lipopolysaccharide (LPS)-mediated signaling in activating macrophages and also functions in controlling anxiety-like behavior (Castrillo et al. 2001). The protein product of CSF3 is cytokine that controls production, differentiation and functions of granulocytes (Kowanetz et al. 2010). Therefore, molecular interactions of P38 MAPK with T-cell receptor (TCR), B-Cell Receptor (BCR) complex, and interferon gamma may be impaired due to deletion of CDKN1A, PRKCE, CSF3 and other CNV related genes. It might indicate for suppression of cellular response leading to reduced heterophil counts in LS birds of quail.

3.5 CONCLUSIONS

We identified sets of genes affected by CNVs in HS and LS lines of quail, most importantly involved in nervous/endocrine systems development and humoral/cell-mediated immune responses. This result supports our hypothesis that CNVs have impact in increasing genotypic diversity and thereby phenotypic traits observed in quail. The quail will continue to evolve as an important research animal model for understanding well-being and production performances in avian species and other animals.

REFERENCES

Aalling N, Hageman I, Miskowiak K, Orlowski D, Wegener G, Wortwein G: Erythropoietin prevents the effect of chronic restraint stress on the number of hippocampal CA3c dendritic terminals-relation to expression of genes involved in synaptic plasticity, angiogenesis, inflammation, and oxidative stress in male rats. J Neurosci Res 2018, 96(1); 103-116.

Abyzov A, Urban AE, Snyder M, Gerstein M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. 2011 Jun; 21(6):974-84.

Berglund J, Nevalainen EM, Molin A, Perloski M, Andre C, Zody MC, et al. Novel origins of copy number variation in the dog genome. Genome Biol. 2012; 13(8): R73.

Bickhart DM, Liu GE. The challenges and importance of structural variation detection in livestock. Front Genet. 2014 Feb 18; 5:37.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30(15):2114-20.

Buchanan JA, Scherer SW. Contemplating effects of genomic structural variation. Genetics in Medicine. 2008; 10(9):639-47.

Carnero A. PPP1R9B (protein phosphatase 1, regulatory subunit 9B). 2012.

Castrillo A, Pennington DJ, Otto F, Parker PJ, Owen MJ, Boscá L. Protein kinase Cϵ is required for macrophage activation and defense against bacterial infection. J Exp Med. 2001; 194(9):1231-42.

Cloonan N, Brown MK, Steptoe AL, Wani S, Chan WL, Forrest AR, et al. The miR-17-5p microRNA is a key regulator of the G1/S phase cell cycle transition. Genome Biol. 2008; 9(8): R127.

Crooijmans RP, Fife MS, Fitzgerald TW, Strickland S, Cheng HH, Kaiser P, et al. Large-scale variation in DNA copy number in chicken breeds. BMC Genomics. 2013; 14(1):398.

Delany ME, Robinson CM, Goto RM, Miller MM. Architecture and organization of chicken microchromosome 16: order of the NOR, MHC-Y, and MHC-B subregions. J Hered. 2009; 100(5):507-14.

Diskin SJ, Hou C, Glessner JT, Attiyeh EF, Laudenslager M, Bosse K, et al. Copy number variation at 1q21. 1 associated with neuroblastoma. Nature. 2009; 459(7249):987-91.

Duan J, Zhang J, Deng H, Wang Y. Comparative studies of copy number variation detection methods for next-generation sequencing technologies. PloS one. 2013; 8(3): e59128.

Dwivedi Y, Zhang H: Altered ERK1/2 signaling in the brain of learned helpless rats: Relevance in vulnerability to developing stress-induced depression. Neural Plast 2016, 2016;.

Elferink MG, Vallée AA, Jungerius AP, Crooijmans RP, Groenen MA. Partial duplication of the PRLR and SPEF2 genes at the late feathering locus in chicken. BMC Genomics. 2008; 9(1):391.

Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. Nature Reviews Genetics. 2006; 7(2):85-97.

Flisikowski K, Venhoranta H, Nowacka-Woszuk J, McKay SD, Flyckt A, Taponen J, et al. A novel mutation in the maternally imprinted PEG3 domain results in a loss of MIMT1 expression and causes abortions and stillbirths in cattle (Bos taurus). PloS one. 2010; 5(11): e15116.

Fontanesi L, Beretti F, Riggio V, Gomez Gonzalez E, Dall'Olio S, Davoli R, et al. Copy number variation and missense mutations of the agouti signaling protein (ASIP) gene in goat breeds with different coat colors. Cytogenet Genome Res. 2009; 126(4):333-47.

Fortna RR, Crystal AS, Morais VA, Pijak DS, Lee VM, Doms RW. Membrane topology and nicastrin-enhanced endoproteolysis of APH-1, a component of the gamma-secretase complex. J Biol Chem. 2004 Jan 30; 279(5):3685-93.

Gill JL, Capper D, Vanbellinghen J, Chung S, Higgins RJ, Rees MI, et al. Startle disease in Irish wolfhounds associated with a microdeletion in the glycine transporter GlyT2 gene. Neurobiol Dis. 2011; 43(1):184-9.

Girirajan S, Campbell CD, Eichler EE. Human copy number variation and complex genetic disease. Annu Rev Genet. 2011; 45:203-26.

Giuffra E, Törnsten A, Marklund S, Bongcam-Rudloff E, Chardon P, Kijas JM, et al. A large duplication associated with dominant white color in pigs originated by homologous recombination between LINE elements flanking KIT. Mammalian Genome. 2002; 13(10):569-77.

Goldstein O, Mezey JG, Boyko AR, Gao C, Wang W, Bustamante CD, et al. An ADAM9 mutation in canine cone-rod dystrophy 3 establishes homology with human cone-rod dystrophy 9. Mol Vis. 2010 Aug 11; 16:1549-69.

Harr B, Karakoc E, Neme R, Teschke M, Pfeifle C, Pezer Ž, et al. Genomic resources for wild populations of the house mouse, Mus musculus and its close relative Mus spretus. Scientific data. 2016; 3:160075.

Hastings PJ, Lupski JR, Rosenberg SM, Ira G. Mechanisms of change in gene copy number. Nature Reviews Genetics. 2009; 10(8):551.

Henrichsen CN, Chaignat E, Reymond A. Copy number variants, diseases and gene expression. Hum Mol Genet. 2009; 18(R1): R1-8.

Hillbertz NHS, Isaksson M, Karlsson EK, Hellmen E, Pielberg GR, Savolainen P, et al. Duplication of FGF3, FGF4, FGF19 and ORAOV1 causes hair ridge and predisposition to dermoid sinus in Ridgeback dogs. Nat Genet. 2007; 39(11):1318-20.

Huff G, Huff W, Wesley I, Anthony N, Satterlee D. Response of restraint stress-selected lines of Japanese quail to heat stress and Escherichia coli challenge. Poult Sci. 2013; 92(3):603-11.

Hull RM, Cruz C, Jack CV, Houseley J. Environmental change drives accelerated adaptation through stimulated copy number variation. PLoS biology. 2017; 15(6): e2001333.

Jia X, Chen S, Zhou H, Li D, Liu W, Yang N. Copy number variations identified in the chicken using a 60K SNP BeadChip. Anim Genet. 2013; 44(3):276-84.

Jiang L, Jiang J, Yang J, Liu X, Wang J, Wang H, et al. Genome-wide detection of copy number variations using high-density SNP genotyping platforms in Holsteins. BMC Genomics. 2013; 14(1):131.

Kallhoff-Munoz V, Hu L, Chen X, Pautler RG, Zheng H: Genetic dissection of gamma-secretase-dependent and -independent functions of presenilin in regulating neuronal cell cycle and cell death. J Neurosci 2008, 28(44); 11421-11431.

Katju V, Bergthorsson U. Copy-number changes in evolution: rates, fitness effects and adaptive significance. Front Genet. 2013 Dec 10; 4:273.

Kowanetz M, Wu X, Lee J, Tan M, Hagenbeek T, Qu X, et al. Granulocyte-colony stimulating factor promotes lung metastasis through mobilization of Ly6G+Ly6C+ granulocytes. Proc Natl Acad Sci U S A. 2010 Dec 14; 107(50):21248-55.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods. 2012; 9(4):357-9.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009; 25(16):2078-9.

Liu GE, Brown T, Hebert DA, Cardone MF, Hou Y, Choudhary RK, et al. Initial analysis of copy number variations in cattle selected for resistance or susceptibility to intestinal nematodes. Mammalian Genome. 2011; 22(1-2):111-21.

Locke MEO, Milojevic M, Eitutis ST, Patel N, Wishart AE, Daley M, et al. Genomic copy number variation in Mus musculus. BMC Genomics. 2015; 16(1):497.

Luo J, Yu Y, Mitra A, Chang S, Zhang H, Liu G, et al. Genome-wide copy number variant analysis in inbred chicken lines with different susceptibility to Marek's disease. G3 (Bethesda). 2013 Feb; 3(2):217-23.

Malisch JL, Satterlee DG, Cockrem JF, Wada H, Breuner CW. How acute is the acute stress response? Baseline corticosterone and corticosteroid-binding globulin levels change 24h after an acute stressor in Japanese quail. Gen Comp Endocrinol. 2010; 165(2):345-50.

Medvedev P, Stanciu M, Brudno M. Computational methods for discovering structural variation with next-generation sequencing. Nature methods. 2009; 6: S13-20.

Meyers SN, McDaneld TG, Swist SL, Marron BM, Steffen DJ, O'Toole D, et al. A deletion mutation in bovine SLC4A2 is associated with osteopetrosis in Red Angus cattle. BMC Genomics. 2010; 11(1):337.

Neph S, Kuehn MS, Reynolds AP, Haugen E, Thurman RE, Johnson AK, et al. BEDOPS: high-performance genomic feature operations. Bioinformatics. 2012; 28(14):1919-20.

Pezer Z, Harr B, Teschke M, Babiker H, Tautz D. Divergence patterns of genic copy number variation in natural populations of the house mouse (Mus musculus domesticus) reveal three conserved genes with major population-specific expansions. Genome Res. 2015 Aug; 25(8):1114-24.

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. Nature. 2006; 444(7118):444-54.

Satterlee D, Cadd G, Jones R. Developmental instability in Japanese quail genetically selected for contrasting adrenocortical responsiveness. Poult Sci. 2000; 79(12):1710-4.

Satterlee D, Johnson W. Selection of Japanese quail for contrasting blood corticosterone response to immobilization. Poult Sci. 1988; 67(1):25-32.

Satterlee D, Marin R, Jonest R. Selection of Japanese quail for reduced adrenocortical responsiveness accelerates puberty in males. Poult Sci. 2002; 81(7):1071-6.

Schrider DR, Hahn MW. Gene copy-number polymorphism in nature. Proc Biol Sci. 2010 Nov 7; 277(1698):3213-21.

Shaikh TH, Gai X, Perin JC, Glessner JT, Xie H, Murphy K, et al. High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. Genome Res. 2009 Sep; 19(9):1682-90.

Shinomiya A, Kayashima Y, Kinoshita K, Mizutani M, Namikawa T, Matsuda Y, et al. Gene duplication of endothelin 3 is closely correlated with the hyperpigmentation of the internal organs (Fibromelanosis) in silky chickens. Genetics. 2012 Feb; 190(2):627-38.

Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. Annu Rev Med. 2010; 61:437-55.

Stingele S, Stoehr G, Peplowska K, Cox J, Mann M, Storchova Z. Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells. Mol Syst Biol. 2012; 8:608.

Suhr C, Schmidt J, Treese S, Satterlee D. Short-term spatial memory responses in aged Japanese quail selected for divergent adrenocortical stress responsiveness 1. Poult Sci. 2010; 89(4):633-42.

Tsui J, Inagaki M, Schulman H. Calcium/calmodulin-dependent protein kinase II (CaMKII) localization acts in concert with substrate targeting to create spatial restriction for phosphorylation. J Biol Chem. 2005 Mar 11; 280(10):9210-6.

Veitia RA. Gene dosage balance in cellular pathways: implications for dominance and gene duplicability. Genetics. 2004 Sep; 168(1):569-74.

Wang J, Jiang J, Fu W, Jiang L, Ding X, Liu J, et al. A genome-wide detection of copy number variations using SNP genotyping arrays in swine. BMC Genomics. 2012; 13(1):273.

Wright D, Boije H, Meadows JR, Bed'Hom B, Gourichon D, Vieaud A, et al. Copy number variation in intron 1 of SOX5 causes the Pea-comb phenotype in chickens. PLoS genetics. 2009; 5(6):e1000512.

Zarrei M, MacDonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. Nature Reviews Genetics. 2015; 16(3):172-83.

Zhang F, Gu W, Hurles ME, Lupski JR. Copy number variation in human health, disease, and evolution. Annual review of genomics and human genetics. 2009; 10:451-81.

Zhao M, Wang Q, Wang Q, Jia P, Zhao Z. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. BMC Bioinformatics. 2013; 14(11): S1.

Zhou W, Liu R, Zhang J, Zheng M, Li P, Chang G, et al. A genome-wide detection of copy number variation using SNP genotyping arrays in Beijing-You chickens. Genetica. 2014; 142(5):441-50.

APPENDIX

Table 3.1 Primers used for validation of CNV by qPCR. β-actin was used as internal control for qPCR.

Gene	Forward	Reverse	Size
NPTN	TGTCTGCACTGCCTATCAAG	ACGTTGTGTTTCCCATGGTA	158 bp
UBA7	TTGAACTCATCACGAGCCCA	TTTGGTGTCCCATCCCATCT	140 bp
RPHA	AACAGCAGGAAGCTGGGAAT	TCTGCAGGTGCAGCAATGCT	140 bp
CACNG2	TAGAGGAGGATCCACTCAGA	ACAGGATGTGCCAGACCTGA	140 bp
LRRC16B	TCTGCTTGGGATTCCACTGA	AGACTGGGCAACCATCTCTA	160 bp
PCF11	ACAGACCTCTTCCAGTCTAG	ATACATCCACCACTGCCCTT	124 bp
CBFA2T2	AGAGGATATCTGCTGGTAAC	GAGCACGTACTTCAGGTAGA	142 bp
PIH1D3	TGCTGCTGTGACGTGGAATT	GAGACTTGCCAACGTTCTGA	140 bp
FAM219A	ACAGCAGAGATACAGCAGAG	TTGTTGGAGCCCTGCTATTA	140 bp
β-Actin	CTCCTCCTCCCACCCATTTC	GCAGGGACTTCCTTTGTCCC	121 bp

Table 3.2 Sequencing and Mapping data of High and Low stress lines of Japanese quail.

Line	# of raw reads	# of mapped reads	Coverage
HS	250,617,546	85,577,152	41.45x
LS	257,535,422	88,195,797	42.59x

Table 3.3 Summary of CNV in High and Low stress lines of Japanese quail.

Line	CNVnator bin size	Average RD per bin ± StDev	Total # of CNVRs	# of Deletions CNVRs	# of Duplications CNVRs	Deletion CNVRs (Mb)	Duplication CNVRs (Mb)	Total CNVRs(Mb)	Average CNVRs size (Mb)
HS	1200	78.1745± 17.0184	262	235	27	13.80	1.32	15.20	0.05
LS	1500	55.9714 ± 13.2859	168	148	20	17.02	1.15	18.17	1.08

Table 3.4 Number of genes overlapped with CNVRs in High and Low stress lines of Japanese quail.

Quail Lines	Total # of Genes overlapped with CNVRs	# of Deletion Genes	# of Duplication Genes	# of Unique Deletion Genes	# of Unique Duplication Genes
HS	948	895	53	436	18
LS	982	922	60	471	22

Table 3.5 Unique genes overlapped with deletion CNVRs in HS and LS lines of Japanese quail associated with Canonical pathways.

Canonical Pathways	Molecules
HS line:	
Phospholipase C Signaling	ARHGEF11, ARKGEF12, BTK, HDAC5, ITGA3, ITPR1, MEF2B, MEF2D, MPRIP, PLA2G3, PLD6
Neuregulin Signaling	CDK5R1, ERBB2, GRB7, ITGA3, PIK3R2
Reelin Signaling in Neurons	ARHGEF11, ARKGEF12, CDK5R1, ITGA3, MAPT, PIK3R2
ERK Signaling	MAP2K5, MEF2B, MEF2D, NTRK1, SH2D2A
CD27 Signaling in Lymphocytes	CASP9, MAP2K5, MAP3K13, MAP3K14
LS line:	
Type II Diabetes Mellitus Signaling	ACSBG2, ADIPOR2, CACNA1G, CACNA2D4, CACNG3, PIK3C2B, PRKCB, SLC27A3
GP6 Signaling Pathway	COL16A1, COL18A1, COL5A1, COL6A1, COL6A2, COL9A2, PIK3C2B, PRKCB
nNOS Signaling in Skeletal Muslce cells	CACNA1G, CACNA2D4, CACNG3, NOS1
Hepatic Fibrosis / Hepatic Stellate Cell Activation	COL16A1, COL18A1, COL5A1, COL6A1, COL6A2, COL9A2, ECE1, SMAD7
Role in CHK Proteins in Cell Cycle Checkpoint Control	CDKN1A, RAD9A, RFC5, SLC19A1

Table 3.6 Unique genes overlapped with deletion CNVRs in HS and LS lines of Japanese quail associated with Top Disease and Bio functions.

Name	p-value	#Molecules
HS line:		
Neurological Disease	1.46E-02 - 2.11E-04	21
Endocrine System Disorder	1.46E-02 - 2.84E-05	7
Organismal injury and Abnormalities	1.46E-02 - 2.84E-05	76
Gastrointestinal Disease	1.46E-02 - 2.84E-05	12
LS line:		
Endocrine System Disorder	2.80E-02 - 5.63E-05	14
Organismal Injury and Abnormalities	2.80E-02 - 1.66E-05	130
Connective Tissue Disorder	2.80E-02 - 1.66E-05	14
Reproductive System Disease	2.09E-02 - 1.66E-05	32

Table 3.7 Unique genes overlapped with deletion CNVRs in HS and LS lines of Japanese quail associated with Endocrine System Disorder

HS line:			
Symbol	**Entrez Gene Name**	**Location**	**Type(s)**
CDK5R1	cyclin dependent kinase 5 regulatory subunit 1	Nucleus	kinase
CSF3R	colony stimulating factor 3 receptor	Plasma Membrane	transmembrane receptor
ERBB2	erb-b2 receptor tyrosine kinase 2	Plasma Membrane	kinase
HSD11B2	hydroxysteroid 11-beta dehydrogenase 2	Cytoplasm	enzyme
POLE	DNA polymerase epsilon, catalytic subunit	Nucleus	enzyme
POLE3	DNA polymerase epsilon 3, accessory subunit	Nucleus	enzyme
TRIM29	tripartite motif containing 29	Cytoplasm	transcription regulator
LS line:			
AMH	anti-Mullerian hormone	Extracellular Space	growth factor
CACNA2D4	calcium voltage-gated channel auxiliary subunit alpha2delta 4	Plasma Membrane	ion channel
CDKN1A	cyclin dependent kinase inhibitor 1A	Nucleus	kinase
COL16A1	collagen type XVI alpha 1 chain	Extracellular Space	other
COL18A1	collagen type XVIII alpha 1 chain	Extracellular Space	other
COL5A1	collagen type V alpha 1 chain	Extracellular Space	other
COL6A1	collagen type VI alpha 1 chain	Extracellular Space	other
COL6A2	collagen type VI alpha 2 chain	Extracellular Space	other
COL9A2	collagen type IX alpha 2 chain	Extracellular Space	other
NCOA1	nuclear receptor coactivator 1	Nucleus	transcription regulator
PRKCB	protein kinase C beta	Cytoplasm	kinase
REN	renin	Extracellular Space	peptidase
THRA	thyroid hormone receptor, alpha	Nucleus	ligand-dependent nuclear receptor
TP73	tumor protein p73	Nucleus	transcription regulator

Table 3.8 Significant interaction networks of unique genes overlapped with deletion CNVRs and involved in nervous system and endocrine development in HS and LS lines of quail

SN	Molecules in Network	Score	Focus Molecules	Top Diseases and Functions
HS line:				
1	14-3-3,APH1A,ATP6V0D1,ATP6V1A,ATP6V1G1,atypical protein kinase C,BSN,CAMK2N2,CaMKII,ERK1/2,Glycogen synthase, GPATCH8,Growth hormone,IBA57,IL1RAPL1,INSRR,LLGL1, LSG1,MIOX,MLXIPL,NECTIN1,PEBP4,PP1 protein complex group,PPP1R9B,Proinsulin,pyruvate kinase,RAB3A,RASD1,RPH3A, Secretase gamma,STX1A,STXBP1,TNFRSF13B,Vacuolar H+ ATPase,VWA5B2	37	24	Cell-To-Cell Signaling and Interaction, Cellular Assembly and Organization, Nervous System Development and Function
2	AGMAT,AMPK,BHLHE40,CDC25A,Cg,Ck2,Creb,DUSP23, FAM3D,FSH,GABPB2,Gsk3,HDL,Lh,NCAN,NCL,Nr1h,NUDT 15,NUP153,OSBPL2,p70 S6k,PDGF BB,PEPCK,phosphatase,PI3K (family),Pkc(s),POLE,POLE3,PRUNE1,RNA polymerase II,Rnr,SLC36A4,SREBF1,SUGP1,UBTF	24	18	Cancer, Endocrine System Disorders, Gastrointestinal Disease
3	1700030F18Rik,AKNA,APP,ARMC9,ATAT1,ATXN7L3,C16orf78, C4orf46,CARMIL3,CBFA2T2,CD40,CSAG1,DBF4B,DNAJB7, EPB41L4A,FAM212A,GLRA4,GPR6,GPR12,GPR15,GPR61,GPR78,GPR85,JTB,LMF2,MARCH10,MED9,NUP62CL,OCEL1, RXFP3,SLC13A3,SPEN,SRPK2,TMEM41A,VIPR2	17	14	Cell-To-Cell Signaling and Interaction, Inflammatory Response, Nervous System Development and Function
4	ADH7,B3GNT7,BAG6,CADM3,CTRC,Epsin,ESR2,FAM84B,FGD2,HEBP1,KAZALD1,KLHL12,LSM12,Macf1,MRPL55,NAA38,NBPF10 (includes others),OTP,PABPC5,PCMTD2,POU5F1,RALBP1,Rplp1 (includes others),RUNDC3A,SDK1,SLC5A7,SLC6A1, SMAD4,SNRNP25,TBRG1,TCTA,Ubb,UBC,UBL7,ZFHX3	11	10	Nervous System Development and Function, Neurological Disease, Organ Morphology

Table 3.8 (Contd.)

SN	Molecules in Network	Score	Focus Molecules	Top Diseases and Functions
LS line:				
1	ACSBG2,ADAMTS9,ADAMTS15,ARHGEF9,ASB18,ATRN,CACFD1,COPS5,CREB3,GCFC2,HSD11B1L,LSM12,MGST2,NUDT1,PGAM5,PHRF1,PPP1CA,PPP1R15B,PRPF6,PRPF39,RELL2,RHEB,RIMS3,RNPC3,RRP7A,SF3A2,SNRNP35,SNRPE,TFIP11,TMEM222,U4 snRNP,U5 snRNP,U6 snRNP,VCAN,ZMAT5	19	15	Developmental Disorder, Hereditary Disorder, Neurological Disease
2	20s proteasome,26s Proteasome,Alpha tubulin,AMPK,Calcineurin protein(s),CDT1,CPEB1,Cyclin A,Cyclin D,cytochrome C, cytochrome-c oxidase,DFFB,EIF4G3,ELP3,ERK,HISTONE, Histone H1,MEAF6,Mitochondrial complex 1,MRPL48,MTORC2, NFE2L1,Nos,NOS1,OAZ1,PARP,PCDH1,PDE3A,PP2A,PPME1,Ppp2c,PRKAA,Rb,SURF1,TIP60	17	14	Hereditary Disorder, Metabolic Disease, Neurological Disease
3	AKT1,AMIGO2,ARHGAP33,C1orf174,CCL5,CIART,CREB1,CSRNP1,ETNK2,GPR65,GPR83,IGSF9B,JPT1,MMP14,MMP23B,NGF,NR3C1,NRBP2,NTSR1,P2RX3,PCOLCE,RAP1GAP2,SLC17A6,SORCS3,SPATA20,SPOCK3,SRPK2,SRXN1,STON1, TIPARP,VPS26B,VSTM2L,YPEL4,ZDHHC5,ZNF395	11	10	Cell Morphology, Cellular Function and Maintenance, Nervous System Development and Function

Table 3.9 Experimental validation of 9 CN genes using qPCR in larger number of HS (16) and LS (16) birds.

CNV Type	Coordinates	Gene	Copy Number (CNVnator)	Copy Number (qPCR)
Deletion	chr10:1573201-2205600	NPTN	1.53	1.27
Deletion	chr12:1348801-1467600	UBA7	1.42	0.72
Deletion	chr15:5568001-5653200	RPHA	1.38	1.19
Deletion	chr1:47198401-47250000	CACNG2	1.35	1.39
Deletion	chr1:6001-28800	LRRC16B	1.00	1.43
Duplication	chr1:169119601-169160400	PCF11	33.60	1.33*
Duplication	chr20:1994401-2090400	CBFA2T2	25.90	20.57
Deletion	chr4:1858801-2012400	PIH1D3	1.45	1.46
Duplication	chrZ:7086001-7182000	FAM219A	7.80	1.03*

*Indicate inconsistency between CNVnator output and qPCR results

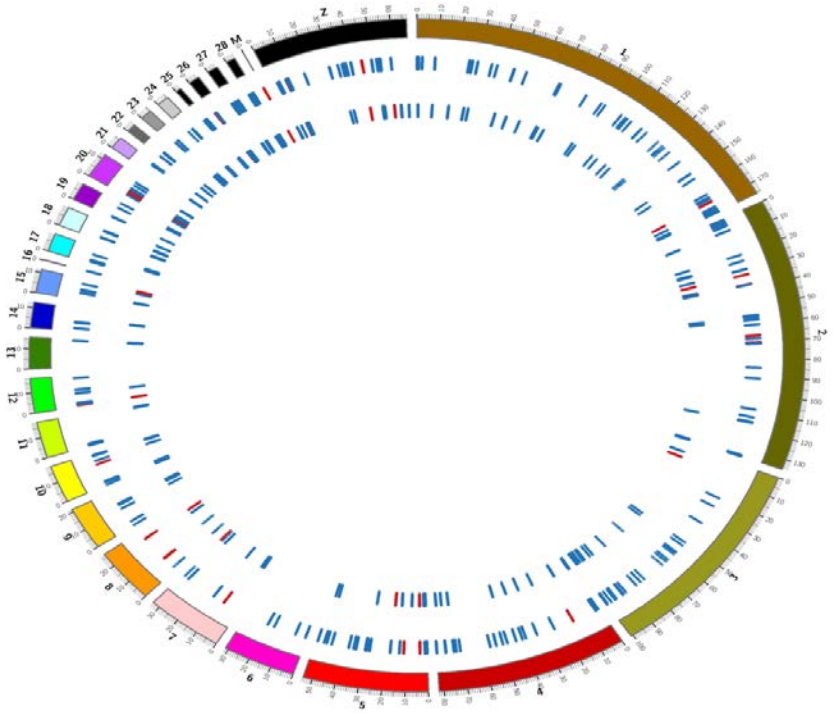

Figure 3.1 Genome-wide distribution of CNVRs in quail. CNVRs are represented in individual tracks as bars, where the outer track depicts CNVRs in HS and inner in LS line of quail. In the tracks, CNVRs indicated by blue bars are deletions and red bar are duplications with respect to the reference assembly.

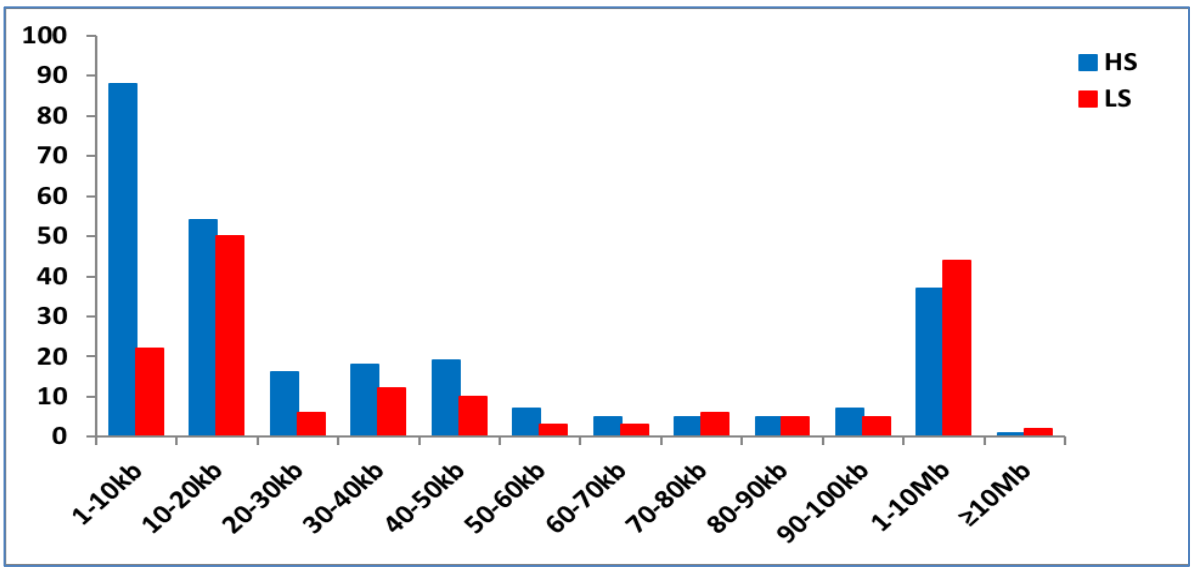

Figure 3.2 Size and Frequency distribution of CNVRs in HS and LS lines of quail.

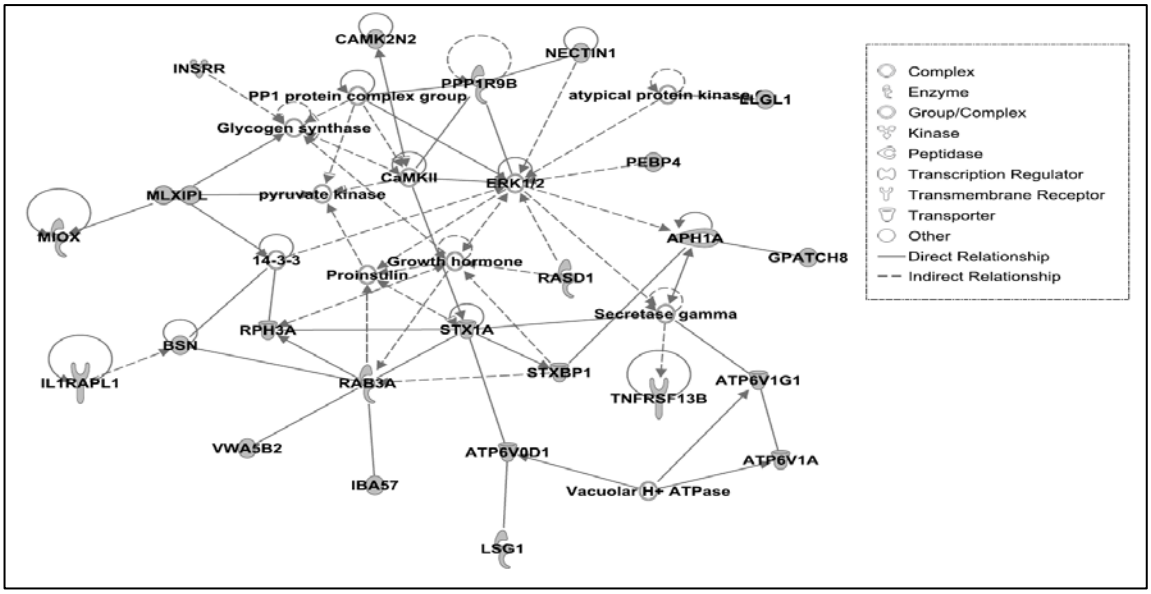

Figure 3.3 Top-scoring multi-gene networks associated with Cell-To-Cell Signaling and Interaction, Cellular Assembly and Organization, Nervous System Development and Function in HS line of quail. The genes overlapped with deletion CNVRs are molecules in gray.

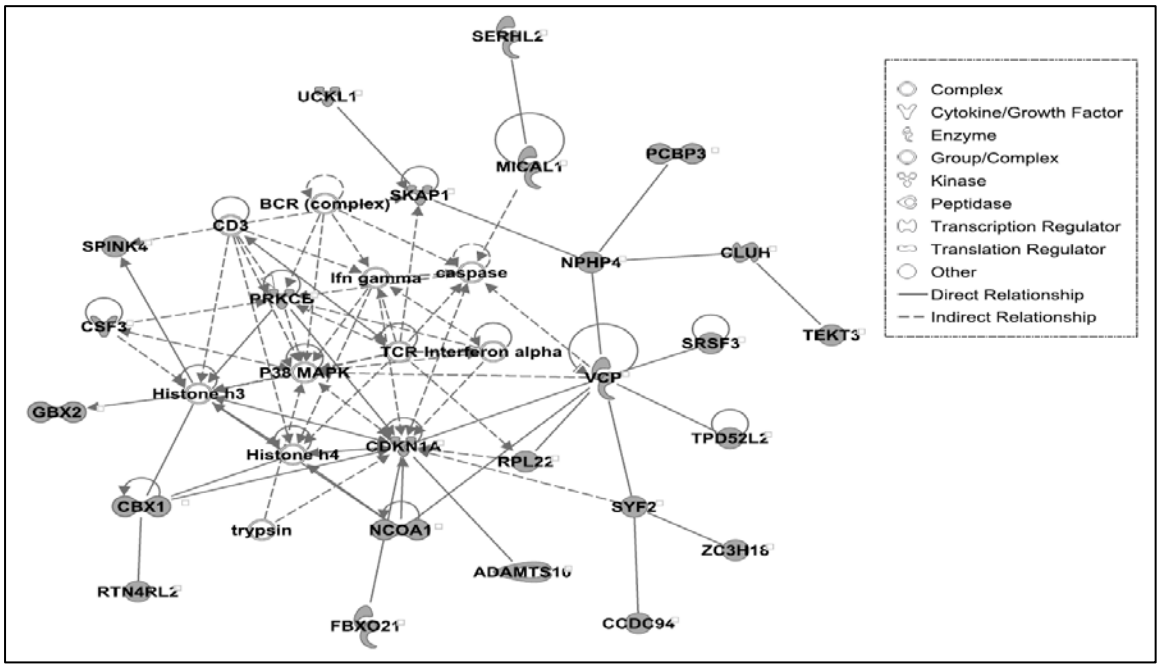

Figure 3.4 Top-scoring multi-gene networks associated with Cell-mediated Immune Response, Cellular Development, Cellular Function and Maintenance in LS line of quail. The genes overlapped with deletion CNVRs are molecules in gray.

CHAPTER 4

Additional analyses of genome resequencing of Arkansas Progressor and Regressor line chickens to identify SNPs associated with tumor regression

B. Khatri, A. Hayden, N. Anthony, and B. Kong*

Department of Poultry Science, Center of Excellence for Poultry Science

University of Arkansas, Fayetteville, Arkansas

*Correspondence to: Dr. Byungwhi Caleb Kong

Department of Poultry Science,

University of Arkansas, Fayetteville, AR 72701

Tel: (479) 575-5494 FAX: (479) 575-7139

E-mail: bkong@uark.edu

4.1 ABSTRACT

Background: Regression of *v-src* oncogene induced tumors is an important attribute in Arkansas Regressor (AR) chickens. For understanding the genetic factors that are responsible for the tumor regression properties, whole genome resequencing was carried out in AR chickens with confirmed tumor regression property and compared with Arkansas Progressor (AP) chickens, which develop tumor upon *v-src* oncogene induction.

Results: Pooled DNA samples from each 10 AR and 10 AP chickens were sequenced using Illumina Hi-Seq 2×100 paired end method and the sequences were aligned to the chicken reference genome for Red Jungle Fowl using NGen program. The sequences for AR and AP chickens reached the coverage of 11x and 14x respectively. A total of 7.1 and 7.3 million SNPs were present in the AR and AP genomes respectively. Through a series of filtration processes, a total of 12,242 SNPs were identified in AR chickens that were associated with induction of mutations such as non-synonymous, frameshift, nonsense, no-start and no-stop. Further filtering of the SNPs on the basis of read depth ≥ 10, SNP % ≥ 0.75 and non-synonymous SNPs, identified 63 reliable marker SNPs were chosen for gene network analysis by Ingenuity Pathway Analysis (IPA). The gene network analysis, which represents the intermolecular connections among interacting genes based on the functional knowledge inputs, revealed that the candidate genes identified in AR birds play roles in networks centered to UBC, PI3K and NF-kB complexes. Thereby, IPA suggested that the tumor regression property in AR chickens may be associated with ubiquitylation (UBC), PI3K and NF-kB pathways.

Conclusion: Several candidate non-synonymous SNP markers encoding varying amino acids and likely to be associated with tumor regression trait were identified in AR chickens. Based on functional studies, the reliable candidate SNPs containing genes were associated with deubiquitylation, and PI3K and NF-kB signaling pathways, suggesting their role in the tumor

regression in AR chickens. This study may provide an insight into genetic factors that may be responsible for tumor regression property.

Key words: tumor regression, v-src, SNPs, resequencing, chicken

4.2 BACKGROUND

Chickens are not only an important supply of proteins for human population but are also outstanding animal models in several fields of biology. They provide excellent opportunities for unraveling the genetic basis of phenotypic variations (Andersson 2001). A vast diversity of phenotypic breeds of chickens has been created by several years of extensive selective inbreeding for the development of beneficial agricultural traits. Their larger population size and increased longevity provide a greater chance of evolution of variants of chickens and then being selected for the important agronomic traits, providing exceptional opportunity to discover novel functions of the specific genes (Fan et al. 2013; Andersson 2004). There are a wide range of variants of chickens integrating various mutations that affect disease resistance and susceptibility, growth rate, body weight, muscle color, reproduction, behavior, feather color, structure and distribution, and comb shape (Li et al. 2013; Zhou et al. 2005; Kong et al. 2018; Ou et al. 2009; Dorshorst et al. 2011; Wright et al. 2009).

Arkansas Regressor (AR) and Arkansas Progressor (AP) chickens are important animal models for studying molecular basis of resistance and susceptibility for development of tumors. The AR and AP line chickens were developed in 1965 by inbreeding of White Leghorn (susceptible to Rous Sarcoma virus: RSV) and Giant Jungle Fowl strains (resistant to RSV) (Hayden 2016). The AR birds regress the tumor induced by *v-src* RSV oncogene unlike the AP birds which upon the *v-src* activation develop malignant tumor in connective tissue known as sarcoma. The tumor regression process in AR chickens may be due to suppression of cell division, induction of apoptosis, DNA damage repair or inhibition of metastasis by various tumor suppressor genes. In chicken models, tumor regression has been found to be strongly associated with both B complex haplotype that encodes major histocompatibility complex (MHC) molecules and non-MHC molecules which include T-lymphocytes and B-cell alloantigens (Devaney et al. 1982; Sun and Yang 2010; Taylor 2004). At the genetic level, this

difference may be due to the occurrence of two or more bases in the same position of the genomic DNA sequences. These single nucleotide polymorphisms (SNPs) are very common and occur at the rate of ~5 SNPs per kilobase (kb) in chicken (Rubin, Carl-Johan 2010) and ~1 SNP per 1-2 kb in human (Sachidanandam et al. 2001). A SNP in the coding region of DNA may change the encoded amino acid (nonsynonymous) thereby altering the structure and function of an encoded protein or it can be silent (synonymous) or simply occurs in the noncoding region of DNA. The SNPs may influence gene expression, mRNA stability and localization of mRNAs/proteins in subcellular compartments and therefore may develop phenotypic traits (Shastry 2009). The SNPs have become important biomarkers and are utilized for the study of population genetic and evolutionary changes (Syvanen 2001; Sachidanandam et al. 2001).

Several genetic variation analyses are performed to identify SNPs associated with disease resistance traits in chickens. Two SNPs, one on SPARC-related modular calcium-binding 1 (SMOC1) gene on chicken chromosome 5 (GGA5) and the other on protein tyrosine phosphate, non-receptor type 3 (PTPN3) gene on GGA2, associated with Marek's disease resistance trait were identified by genome-wide association study (GWAS) (Li et al. 2013). The genetic variations study in response to colonization of pathogenic *Salmonella enteritidis* identified SNPs in T-cell specific surface protein (CD28) and myeloid differentiation protein-2 (MD-2) genes significantly associated with bacterial load in different organs and vaccine antibody response in chickens, suggesting roles of the SNPs in disease resistance (Malek et al. 2004). Major quantitative loci associated with immune response to Newcastle disease (Zhang et al. 2015) and infectious disease (Luo et al. 2014) viruses have been identified in chicken by GWAS. However, there is limited information about genetic factors responsible for disease resistance mechanism against RSV in chickens. Therefore, in this study, whole genome resequencing of AR and AP chickens was performed with a rationale to genotype SNPs at

genome-wide level to detect genomic regions in which the frequencies of the SNP alleles differ between AR and AP, and with reference to Red Jungle Fowl. This study was aimed at identifying SNPs associated with resistance of tumor development by v-*src* oncogene in AR chicken, providing beneficial information for genetic disorder diagnosis and evolutionary study. Previously, genome wide SNPs in AP and AR lines were determined with the same genome sequencing data as used in this study and the thousands of SNPs were identified (Hayden 2016). In this study, we used the updated reference genome (galgal5) and another analytical method for more profound analysis, resulting in higher number of SNPs and more reliable variations were identified.

We identified millions of SNPs in these two lines of chickens and this study is focused only in reliable candidate SNPs that are associated with non-synonymous mutations and capable of inducing amino acid changes in encoded proteins. We analyzed the genes containing SNPs for their biological functions and molecular interactions. The SNPs resources presented here can be useful for breeding in chickens and comparative genetic study in related species.

4.3 METHODS

4.3.1 CHICKEN LINES AND DNA PREPARATION

Adult Arkansas progressor (AP) and Arkansas regressor (AR) chickens that are maintained by Dr. N. Anthony at the University of Arkansas (Fayetteville, AR) were used for this study. A blood specimen (5 mL) was collected from 12 birds from each line following an animal use protocol approved by the University of Arkansas Institutional Animal Care and Use Committee (IACUC; approval number: 14012). Genomic DNA was extracted from whole blood sample using QiaAmp DNA mini kit (Qiagen, Hilden, Germany) following manufacturer's instructions. Quality of DNA was checked using NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA) and agarose gel electrophoresis. The 10 samples having the best quality from each line were pooled to represent each chicken line.

4.3.2 ILLUMINA SEQUENCING AND SEQUENCE ASSEMBLY

Library construction and whole genome sequencing for the pooled DNA samples were done by the National Center for Genome Resources (NCGR; Santa Fe, NM). Illumina HiSeq 2×100 bp paired end read method was used for genome sequencing. The quality of raw sequencing data was determined by using FastQC toolkit (Andrews 2010) and low-quality reads were removed using reformat.sh in BBMap (Bushnell 2016). The clean reads were then aligned to the chicken reference genome sequence for Red Jungle Fowl (*galgal5*) retrieved from NCBI. For the reference based genome alignment, the NGen genome sequence assembly program of the Lasergene software package (DNAStar, Madison, WI) was used. Assembly parameters were as follows: file format, BAM; mer Size, 21; mer skip query, 2; minimum match percentage, 93; maximum gap size, 6; minimum aligned length, 35; match score, 10; mismatch penalty, 20; gap penalty, 30; SNP calculation method, diploid Bayesian; minimum SNP percentage, 5; SNP confidence threshold, 10; minimum SNP count, 2; minimum base

quality score, 5. After assembly, the SeqMan Pro program of the Lasergene package (DNAStar) was used for further analyses (Jang et al. 2014).

4.3.3 SNP DETECTION AND ANALYSIS

The JMP genomics (SAS Institute, Inc., Cary, NC) program was used for filtering unique SNPs for tumor regression in the AR chickens. SNPs occurring in both AR and AP lines were filtered out, leaving behind the unique SNPs for each line. To identify highly fixed and homozygous SNPs, the SNPs were filtered based on SNP percentages (SNP%). The SNPs with a SNP% of ≥ 0.75 (for example, number of SNP = 3 of read depth = 4) were chosen. The 75% cutoff for SNP selection was set by considering potential sequencing errors that can be generated by the massively parallel sequencing method. Potential causal tumor regressing SNPs that induce non-synonymous changes in CDS regions were chosen and unique SNPs in either AP or AR showing ≥10 read depths were selected as reliable SNPs. To reduce false positives, reliable SNPs chosen by criteria described above were confirmed by double-checking the initial assembly results with alignment view in SeqMan Pro program of Lasergene package (DNAStar).

4.3.4 VALIDATION OF SNPS

For validation purposes, nine different SNPs associated with induction of amino acids changes in CDS regions were randomly selected and subjected to allele-specific PCR using a greater number of birds. For this, 96 phenotypically verified birds each from AR and AP lines were used for blood sampling and then genomic DNA isolation. Genomic DNA was purified from whole blood using Wizard SV 96 Genomic DNA Purification System (Promega; Madison, WI) following manufacturer's instructions. Quality and quantity of isolated DNA was determined using Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific) and agarose gel electrophoresis. DNAs from all samples were then diluted to 1 ng/µL in 96 well PCR formats. Allele-specific primers were designed corresponding to nine different SNPs

based on the RJF genome sequence (*galgal5*). Two sets of primers were designed with a common reverse primer for each SNP. The forward primer designated as F1 is the reference type whose terminal nucleotide at the 3' end matches with a base in the reference genome. The forward primer designated as F3 is the SNP type and terminal nucleotide at the 3' end matches with a base in the SNP. The third nucleotides from 3' end of both the forward primers (F1 and F3) were intentionally changed so that they mismatch with the genome in that position (Liu et al. 2012). All primers were commercially synthesized by Integrated DNA Technology (Ames, IA) (Table 4.1). Allele-specific PCR were conducted using F1 and F3 forward primers and a common reverse primer separately in 25 µL reaction volume in 96 well plates with cycles condition as follows: 95 ^{0}C for 1 min, 35 cycles of amplification (95^{o}C for 30 s, 55 or 63^{o}C for 1 min, 72^{o}C for 1min), and final extension 72^{o}C for 10 min in Applied Biosystems 2720 Thermal Cycler (Life Technologies, Carlsbad, CA). Formation of allele-specific products was verified by running PCR products in 1% agarose gel electrophoresis.

4.3.5 INGENUITY PATHWAYS ANALYSIS

Candidate SNPs (n=63) associated with regression of tumor after the filtering process were analyzed using Ingenuity Pathways (IPA; Qiagen; www.ingenuity.com) for understanding the gene ontology and molecular networks. Since IPA is based on human, mouse and rat bioinformatics, functionalities for SNP containing genes in chicken were interpreted based primarily on mammalian biological mechanisms. The number of molecules in the network was set to the limit of 35, leaving only the most important ones based on the number of connections for each focus gene (a subset of uploaded significant genes having direct interactions with other genes in the database) to other significant genes (Kong et al. 2011)

4.4 RESULT AND DISCUSSION

4.4.1 GENOME SEQUENCING AND ASSEMBLY

The result of Illumina sequencing of pooled whole genome from 10 AR and AP chickens yielded approximately 55 and 69 million sequence reads of each 200 bp length, respectively. Among these reads approximately 80% were used for alignment while the remaining 20% were not aligned due to their lower sequence count scores. The genome coverage for AP and AR therefore reached to 14× and 11× respectively, of the Red Jungle Fowl (Table 4.2). A large number of SNPs, 7.1 and 7.3 million per examined AR and AP line genome respectively, were found which might be due to data based on 2 read coverage depths. Most of the SNPs were found to be localized in chromosomes 1 to 4 (data not shown). For identification of the signature genetic biomarkers that may be associated with the tumor regression trait in AR chicken lines, the unique SNPs that were present only in AR were selected removing the SNPs that were also present in AP birds. Further steps of SNP filtration were carried out as described elsewhere by Jang et al. (2014). In brief, the SNPs having SNP% $\geq$ 0.75, present in CDS (protein coding) region only, associated with non-synonymous mutations (such as frameshift, nonsense, no-start- and no-stop changes), and showing $\geq$ 10 read depths were considered as potential candidate SNPs thereby included in this study. However, the process used in this study did not involve a typical SNP calling and filtering method based on quality score. From the unique AR SNPs, filtration based on SNP % $\geq$ 0.75 resulted about 1.2 million SNPs identified throughout the AR chicken genome. Further grouping of SNPs based on feature type of chromosome regions showed that 24,868 SNPs were present in CDS region and about 50% were found in intergenic region (Figure 4.1). The higher percentage of SNPs found to be present in intergenic or regulatory region in AR chicken genome was matching with a GWAS study in chickens (Pértille et al. 2016). About 42% of AR SNPs in the protein coding region were associated with synonymous mutations that did not lead to alternation of amino

acid changes. Since the main aim of this study was to identify the causal SNPs that are responsible for tumor regression in AR chickens, we focused on SNPs that were linked to changes in amino acid sequences after mutation. A total of 12,242 SNPs were identified to be linked with induction of mutations such as non-synonymous, frameshift, nonsense, no-start and no-stop. This suggests that these SNPs may play role in protein coding leading to tumor regression property in AR chickens. In order to consider more reliable SNP markers, the candidate 12,242 SNPs were further filtered on the basis of ≥ 10 read depth which yielded only 63 SNPs (Table 4.3). These 63 candidate SNP markers were chosen for further study.

4.4.2 SNP VALIDATION

To verify the SNPs identified by genome resequencing, 9 SNPs were randomly chosen from the 63 reliable candidate SNPs and subjected to validation using allele-specific PCR in a greater number of birds (Figure 4.2); specifically, 96 AR chickens with confirmed tumor regression property whereas 96 AP chickens with confirmed tumor progression property were used. The results clearly showed the segregation of SNP genotypes that majority of AR and AP birds showed SNP type and reference type, respectively (Table 4.4). Thus, the 63 SNPs chosen this study may become potential genetic biomarkers for tumor regression in AR chickens.

4.4.3 INGENUITY PATHWAY ANALYSIS OF CANDIDATE CAUSAL GENES

The Ingenuity Pathway Analysis (IPA) program was used to generate the data sets which include functional groups and networks for gene containing amino acid changes in AR chicken. The 63 SNPs were found in 58 genes associated with chromosomal open reading frames, and hypothetical proteins (Table 4.5). The genes were further grouped into 89 functional groups that are directly or indirectly related to tumor development.

4.4.4 GENE NETWORKS

Using IPA, molecular networks were generated with the interacting genes associated with amino acid changes in AR chickens based on functional knowledge inputs. Summary of the associated network functions of candidate genes is presented in Table 4.6.

The major functions of molecules associated with Network #1 are related to developmental, heredity and metabolic disorders. Similarly, the top functions of molecules in Network #2 include cell death and survival, hematological system development and function, and humoral immune response (Table 4.6). The molecules in Network #1 (Figure 4.3) and #2 (Figure 4.4) are centrally linked to ubiquitin C (UBC). The UBC functions in protein degradation, DNA repair, cell cycle regulation, kinase modification, endocytosis, and regulation of other cell signaling pathways. Ub ligase, an important enzyme in ubiquitination process, which functions for ligating the substrate molecule to ubiquitin via lysine residue can have role for regulating the stability of oncogenes or tumor suppressors-proteins (Popovic et al. 2014). In AR chickens, the amino acid lysine was found to be changed to glutamic acid in FAM208B (family with sequence similarity 208, member B). Similarly, lysine residues in proteins LAMB4 (laminin, beta 4) and IFT140 (intraflagellar transport 140 homolog) were identified to be changed to arginine residues. This might suggest that various cellular processes involving protein degradation by altered ubiquitinylation properties of proteins may play a significant role in the regression of tumors in AR chickens.

The candidate genes in Network #3 (Figure 4.5) are associated with the signaling pathway of phosphoinositide 3-kinases (PI3K) and NF-kB (nuclear factor kappa-light-chain-enhancer of activated B cells) connected to Arrestin Beta 1 (ARRB1) with insulin signaling in the center. The top functions of the genes are related to developmental, gastrointestinal and heredity disorders. It has been reported that PI3K signaling pathway is crucial for several aspects of cell growth and survival. Recent human cancer genomic studies have shown that

many components of this pathway are frequently targeted for the design of anticancer agents in human by many aberrations including mutation, amplification and rearrangement (Courtney et al. 2010; Hennessy et al. 2005; Liu et al. 2009). A study has also shown that the inhibition of PI3K pathway leads to partial tumor growth inhibition (Wee et al. 2009). In this study, the SNPs identified in the genes PIK3R4 and PIK3C2G, which are the components of PI3k pathway may have a role in the down regulation of PI3K pathway and may be responsible for tumor regression trait in AR chickens. NF-kB signaling pathway also plays role in oncogenesis as it regulates the expression of genes involved in many processes that play role in the development and progression of cancer such as proliferation, migration and apoptosis (Ismail et al. 2004; Dolcet et al. 2005). It has been shown that blocking of PI3-kinase leads to a marked reduction of constitutive NF-kB activity and promotes p53-mediated transcription. The p53 is a crucial cellular protein that regulates the cell cycle and functions as a tumor suppressor, preventing oncogenesis (Levine et al. 2004; Grandage et al. 2005). Therefore, the dual down regulation of PI3K and NF-kB signaling pathways due to the SNPs present in the components of PI3K pathway might block the anti-apoptotic pathway and lead to apoptosis by p53 tumor suppressing properties in AR chickens. BMX (BMX Non-Receptor Tyrosine Kinase) which is present in the Network #3 is a TEC (Tyrosine-protein kinase Tec) family of kinase and is found to be expressed in endothelial lineages and some cancers such as breast and prostate. It has been shown to have anti-apoptotic properties in prostate cancer lines and regulate PI3K signaling pathway (Potter et al. 2014; Qiu and Kung 2000; Vogt and Hart 2011). Due to occurrence of a single nucleotide polymorphism in BMX gene, the oncogenic functions of the BMX gene may have been turned off and along with its regulatory action on PI3K signaling pathway. This may be another reason behind the tumor regression character of AR chickens.

4.5 CONCLUSION

In this study, several candidate molecular markers encoding varying amino acids and likely to be associated with tumor regression trait in AR chickens were detected through high throughput genome re-sequencing. Based on functional studies, the reliable candidate SNPs containing genes were associated with ubiquitylation, and PI3K and NF-kB signaling pathways, suggesting their role in tumor regression in AR chickens. In future, further study focusing on allele specific expression of the marker genes with candidate SNPs in target tissues will be performed to gain better insight into the mechanism of tumor regression in AR chickens.

REFERENCES

Andersson, L. 2001. Genetic dissection of phenotypic diversity in farm animals. Nature Reviews Genetics 2: 130. DOI: 10.1038/35052563. PMID: 11253052.

Andersson, L., and Georges, M., 2004. Domestic-animal genomics: deciphering the genetics of complex traits. Nature Reviews Genetics 5, 202. DOI: 10.1038/nrg1294. PMID: 14970822.

Andrews, S., 2010. FastQC: A quality control tool for high throughput sequence data. Reference Source.

Bushnell, B., 2016. BBMap short read aligner. URL http://sourceforge.net/projects/bbmap.

Courtney, K.D., Corcoran, R.B., and Engelman, J.A., 2010. The PI3K pathway as drug target in human cancer. J. Clin. Oncol. 28, 1075-1083. DOI: 10.1200/JCO.2009.25.3641. PMID: 20085938.

Devaney, J.A., Gyles, N.R., and Lancaster, J.L.,Jr, 1982. Evaluation of Arkansas Rous sarcoma regressor and progressor lines and giant jungle fowl for genetic resistance to the northern fowl mite. Poult. Sci. 61, 2327-2330. DOI: 10.3382/ps.0612327. PMID: 6298763.

Dolcet, X., Llobet, D., Pallares, J., and Matias-Guiu, X., 2005. NF-kB in development and progression of human cancer. Virchows Arch. 446, 475-482. DOI: 10.1007/s00428-005-1264-9. PMID: 15856292.

Dorshorst, B., Molin, A., Rubin, C., Johansson, A.M., Strömstedt, L., Pham, M., et al. 2011. A complex genomic rearrangement involving the endothelin 3 locus causes dermal hyperpigmentation in the chicken. PLoS genetics 7, e1002412. DOI: 10.1371/journal.pgen.1002412. PMID: 22216010.

Fan, W., Ng, C.S., Chen, C., Lu, M.J., Chen, Y., Liu, C., et al. 2013. Genome-wide patterns of genetic variation in two domestic chickens. Genome biology and evolution 5, 1376-1392. DOI: 10.1093/gbe/evt097. PMID: 23814129.

Grandage, V.L., Gale, R.E., Linch, D.C., and Khwaja, A., 2005. PI3-kinase/Akt is constitutively active in primary acute myeloid leukaemia cells and regulates survival and chemoresistance via NF-kappaB, Mapkinase and p53 pathways. Leukemia 19, 586-594. DOI: 10.1038/sj.leu.2403653. PMID: 15703783.

Hayden, A., 2016. Identification of Biomarkers Associated with Rous Sarcoma Virus induced Tumors in Two Divergently Selected Chicken Lines.

Hennessy, B.T., Smith, D.L., Ram, P.T., Lu, Y., and Mills, G.B., 2005. Exploiting the PI3K/AKT pathway for cancer drug discovery. Nat. Rev. Drug Discov. 4, 988-1004. DOI: 10.1038/nrd1902. PMID: 16341064.

Ismail, H.A., Lessard, L., Mes-Masson, A.M., and Saad, F., 2004. Expression of NF-kappaB in prostate cancer lymph node metastases. Prostate 58, 308-313. DOI: 10.1002/pros.10335. PMID: 14743471.

Jang, H.M., Erf, G.F., Rowland, K.C., and Kong, B.W., 2014. Genome resequencing and bioinformatic analysis of SNP containing candidate genes in the autoimmune vitiligo Smyth line chicken model. BMC Genomics 15, 707-2164-15-707. DOI: 10.1186/1471-2164-15-707. PMID: 25151476.

Kong, B.W., Lee, J.Y., Bottje, W.G., Lassiter, K., Lee, J., and Foster, D.N., 2011. Genome wide differential gene expression in immortalized DF-1 chicken embryo fibroblast cell line. BMC Genomics 12, 571-2164-12-571. DOI: 10.1186/1471-2164-12-571. PMID: 22111699.

Kong, H.R., Anthony, N.B., Rowland, K.C., Khatri, B., and Kong, B.C., 2018. Genome re-sequencing to identify single nucleotide polymorphism markers for muscle color traits in broiler chickens. Asian-Australas J. Anim. Sci. 31, 13-18. DOI: 10.5713/ajas.17.0479. PMID: 28830129.

Levine, A.J., Finlay, C.A., and Hinds, P.W., 2004. P53 is a tumor suppressor gene. Cell 116, S67-9, 1 p following S69. PMID: 15055586.

Li, D., Lian, L., Qu, L., Chen, Y., Liu, W., Chen, S., et al. 2013. A genome-wide SNP scan reveals two loci associated with the chicken resistance to Marek's disease. Anim. Genet. 44, 217-222. DOI: 10.1111/j.1365-2052.2012.02395.x. PMID: 22812605.

Liu, J., Huang, S., Sun, M., Liu, S., Liu, Y., Wang, W., et al. 2012. An improved allele specific PCR primer design method for SNP marker analysis and its application. Plant Methods 8, 1. DOI: 10.1186/1746-4811-8-34. PMID: 22920499.

Liu, P., Cheng, H., Roberts, T.M., and Zhao, J.J., 2009. Targeting the phosphoinositide 3-kinase pathway in cancer. Nat. Rev. Drug Discov. 8, 627-644. DOI: 10.1038/nrd2926. PMID: 19644473.

Luo, C., Qu, H., Ma, J., Wang, J., Hu, X., Li, N., et al. 2014. A genome-wide association study identifies major loci affecting the immune response against infectious bronchitis virus in chicken. Infection, Genetics and Evolution 21, 351-358. DOI: 10.1016/j.meegid.2013.12.004. PMID: 24333371.

Malek, M., Hasenstein, J., and Lamont, S., 2004. Analysis of chicken TLR4, CD28, MIF, MD-2, and LITAF genes in a Salmonella enteritidis resource population. Poult. Sci. 83, 544-549. DOI: 10.1093/ps/83.4.544. PMID: 15109052.

Ou, J., Tang, S., Sun, D., and Zhang, Y., 2009. Polymorphisms of three neuroendocrine correlated genes associated with growth and reproductive traits in the chicken. Poult. Sci. 88, 722-727. DOI: 10.3382/ps.2008-00497. PMID: 19276414.

Pértille, F., Guerrero-Bosagna, C., Da Silva, V.H., Boschiero, C., da Silva Nunes, José de Ribamar, Ledur, M.C., Jensen, P., et al. 2016. High-throughput and cost-effective chicken genotyping using next-generation sequencing. Scientific reports 6, 26929. DOI: 10.1038/srep26929. PMID: 27220827.

Popovic, D., Vucic, D., and Dikic, I., 2014. Ubiquitination in disease pathogenesis and treatment. Nat. Med. 20, 1242-1253. DOI: 10.1038/nm.3739. PMID: 25375928.

Potter, D.S., Kelly, P., Denneny, O., Juvin, V., Stephens, L.R., Dive, C., et al. 2014. BMX acts downstream of PI3K to promote colorectal cancer cell survival and pathway inhibition sensitizes to the BH3 mimetic ABT-737. Neoplasia 16, 147-157. doi: 10.1593/neo.131376.

Qiu, Y., and Kung, H.J., 2000. Signaling network of the Btk family kinases. Oncogene 19, 5651-5661. DOI: 10.1038/sj.onc.120395. PMID: 11114746.

Rubin, C., Zody, M.C., Eriksson, J., Meadows, J.R., Sherwood, E., Webster, M.T., et al. 2010. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature 464, 587. DOI: 10.1038/nature08832. PMID: 20220755.

Sachidanandam, R., Weissman, D., Schmidt, S.C., Kakol, J.M., Stein, L.D., Marth, G., et al. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409, 928-933. DOI: 10.1038/35057149. PMID: 11237013.

Shastry, B.S., 2009. SNPs: impact on gene function and phenotype. Single Nucleotide Polymorphisms: Methods and Protocols, 3-22. DOI: 10.1007/978-1-60327-411-1_1. PMID: 19768584.

Sun, W., and Yang, J., 2010. Functional mechanisms for human tumor suppressors. J. Cancer. 1, 136-140. PMID: 20922055.

Syvanen, A.C., 2001. Accessing genetic variation: genotyping single nucleotide polymorphisms. Nat. Rev. Genet. 2, 930-942. DOI: 10.1038/35103535. PMID: 11733746.

Taylor, R.L.,Jr, 2004. Major histocompatibility (B) complex control of responses against Rous sarcomas. Poult. Sci. 83, 638-649. DOI: 10.1093/ps/83.4.638. PMID: 15109061.

Vogt, P.K., and Hart, J.R., 2011. PI3K and STAT3: a new alliance. Cancer. Discov. 1, 481-486. DOI: 10.1158/2159-8290.CD-11-0218. PMID: 22348200.

Wee, S., Jagani, Z., Xiang, K.X., Loo, A., Dorsch, M., Yao, Y.M., et al. 2009. PI3K pathway activation mediates resistance to MEK inhibitors in KRAS mutant cancers. Cancer Res. 69, 4286-4293. DOI: 10.1158/0008-5472.CAN-08-4765. PMID: 19401449.

Wright, D., Boije, H., Meadows, J.R., Bed'Hom, B., Gourichon, D., Vieaud, A., et al. 2009. Copy number variation in intron 1 of SOX5 causes the Pea-comb phenotype in chickens. PLoS genetics 5, e1000512. DOI: 10.1371/journal.pgen.1000512. PMID: 19521496.

Zhang, L., Li, P., Liu, R., Zheng, M., Sun, Y., Wu, D., et al. 2015. The identification of loci for immune traits in chickens using a genome-wide association study. PloS one 10, e0117269. doi:10.1371/journal.pone.0117269. PMID: 25822738.

Zhou, H., Mitchell, A., McMurtry, J., Ashwell, C., and Lamont, S.J., 2005. Insulin-like growth factor-I gene polymorphism associations with growth, body composition, skeleton integrity, and metabolic traits in chickens. Poult. Sci. 84, 212-219. doi: 10.1093/ps/84.2.212. PMID: 15742956.

APPENDIX

Table 4.1 Primers used for allele specific PCR

Gene	Primer Name	Oligo Sequence (5'→3')	Annealing Temperature (^{0}C)
BMX	BMX-F1	GAACTTACATACAGATCGTC	55
	BMX-F3	GAACTTACATACAGATCGTT	55
	BMX-R	CTTCCAACCCAAGCCATTAC	55
FAM208B	FAM208B-F1	CCACTCCTTGGTGGAGTATT	55
	FAM208B-F3	CCACTCCTTGGTGGAGTATC	55
	FAM208B-R	AGAAAGATGAGGATCGTGCG	55
IFT140	IFT140-F1	AAATCCATCAAGTTGATTAA	55
	IFT140-F3	AAATCCATCAAGTTGATTAG	55
	IFT140-R	TCTTTCTGAGAACGAAAGGG	55
IGSF	IGSF-F1	CAATGGGACTGTGCTGAGTC	63
	IGSF-F2	CAATGGGACTGTGCTGAGTT	63
	IGSF-R	TCTCAGGCAGAGGTGATGAT	63
LAMB4	LAMB4-F1	TCTCTTATTTGCGTTCAATT	55
	LAMB4-F3	TCTCTTATTTGCGTTCAATC	55
	LAMB4-R	TTGCAGATGAGAGTGTGCCT	55
PIK3R4	PIK3R4-F1	ACTAGGGTGAGATGTTTAAT	55
	PIK3R4-F3	ACTAGGGTGAGATGTTTAAC	55
	PIK3R4-R	GGGGATCATCAGAAGTCTGT	55
THADA	THADA-F1	ACAAACCATGCTGGCATACT	63
	THADA-F3	ACAAACCATGCTGGCATACC	63
	THADA-R	CAGGACATGCTAACCTCTGT	63
TOPAZ1	TOPAZ1-F1	AAGCTCTGGTAGGCTACGGG	63
	TOPAZ1-F3	AAGCTCTGGTAGGCTACGGT	63
	TOPAZ1-R	CAGGCCAGAATACTGCATCT	63

Table 4.2 Data from HiSeq and sequence assembly

Line	# of reads	# of reads aligned	# of reads not aligned	Coverage	Total # of SNP
AP	69,221,284	55,224,050	10,903,306	14×	7,372,778
AR	55,368,344	44,328,649	8,551,173	11×	7,173,788

Table 4.3 The 63 reliable marker SNPs that induced amino acid changes showing ≥10 read depths

Contig ID	C	Ref Pos	Ref Base	Called Base	Impact	SNP %	Feature Name	DNA Change	AA Change	Depth	A Cnt	C Cnt	G Cnt	T Cnt	Del
NC_006088	1	1034107	T	C	N-syn	1	FAM208B [4]	c.4777A>G	p.K1593E	13	0	13	0	-	0
NC_006088	1	14729572	T	C	N-syn	0.8	LAMB4	c.4640A>G	p.K1547R	10	0	8	0	-	0
NC_006088	1	64181733	G	T	N-syn	1	PIK3C2G	c.1585G>T	p.A529S	10	0	0	-	10	0
NC_006088	1	65873551	C	G	N-syn	0.9	LOC101748372	c.960G>C	p.R320S	11	0	-	10	0	0
NC_006088	1	104629701	G	A	N-syn	1	IFNAR2	c.1099G>A	p.A367T	11	11	0	-	0	0
NC_006088	1	112303547	A	C	N-syn	0.8	RPGR	c.2209A>C	p.I737L	10	-	8	0	0	0
NC_006088	1	121349181	C	T|C	N-syn	0.75	BMX	c.[1549G>G] +[1549G>A]	p.D517N, p.D517D	12	0	-	0	9	0
NC_006088	1	127837266	G	T	N-syn	0.89	ARSD	c.449G>T	p.W150L	10	0	0	-	9	0
NC_006088	1	179358329	A	T|A	N-syn	0.76	DDX10	c.[1849T>T] +[1849T>A]	p.Y617N, p.Y617Y	13	-	0	0	10	0
NC_006088	1	185190031	T	C	N-syn	0.8	CCDC67	c.983A>G	p.Q328R	10	0	8	0	-	0
NC_006088	1	193338454	A	T	N-syn	1	ART7B	c.590T>A	p.L197Q	13	-	0	0	13	0
NC_006089	2	19385973	C	T	N-syn	1	LOC420515	c.3400G>A	p.V1134I	10	0	-	0	10	0
NC_006089	2	41279450	G	T	N-syn	1	TOPAZ1	c.2924C>A	p.T975N	10	0	0	-	10	0
NC_006089	2	42201510	T	T|C	N-syn	0.75	PIK3R4	c.[1937A>G] +[1937A>A]	p.D646D, p.D646G	12	0	9	0	-	0
NC_006089	2	63518627	T	C	N-syn	0.89	LOC101751154	c.137T>C	p.I46T	10	0	9	0	-	0
NC_006089	2	96924595	G	A	N-syn	1	PTPN2	c.934G>A	p.A312T	10	10	0	-	0	0
NC_006089	2	105934067	G	C	N-syn	0.9	COPN5L1	c.17C>G	p.A6G	11	1	10	-	0	0
NC_006089	2	120829955	T	C	N-syn	0.91	LOC101751416	c.194T>C	p.L65P	82	0	75	0	-	0
NC_006090	3	24479109	T	C	N-syn	1	THADA	c.859A>G	p.S287G	10	0	10	0	-	0
NC_006090	3	24515241	T	C	N-syn	1	PLEKHH2	c.1522T>C	p.F508L	10	0	10	0	-	0
NC_006090	3	37785831	A	G	N-syn	1	TARBP1	c.1321A>G	p.I441V	10	-	0	10	0	0
NC_006090	3	37785856	C	T	N-syn	1	TARBP1	c.1346C>T	p.T449I	10	0	-	0	10	0
NC_006090	3	38965704	G	T	N-syn	1	DISC1	c.1027C>A	p.L343I	10	0	0	-	10	0
NC_006090	3	104626700	C	G	N-syn	0.85	GVINP1	c.6146G>C	p.R2049T	14	0	-	12	0	0

Table 4.3 (Contd.)

Contig ID	C	Ref Pos	Ref Base	Called Base	Impact	SNP %	Feature Name	DNA Change	AA Change	Depth	A Cnt	C Cnt	G Cnt	T Cnt	Del
NC_006091	4	11530696	C	T	N-syn	0.8	LOC771752	c.62G>A	p.R21Q	10	0	-	0	8	0
NC_006091	4	31100937	A	C	N-syn	0.9	TTC29	c.903T>G	p.D301E	11	-	10	0	0	0
NC_006091	4	56709491	G	G|A	N-syn	0.76	C4H4ORF21	c.[631G>G] +[631G>A]	p.D211N, p.D211D	13	10	0	-	0	0
NC_006091	4	60221486	C	T	N-syn	1	BANK1	c.934C>T	p.P312S	12	0	-	0	12	0
NC_006092	5	24297188	T	C	N-syn	1	MGA	c.7214T>C	p.M2405T	13	0	13	0	-	0
NC_006092	5	45169641	T	C	N-syn	1	LOC100858625	c.2191A>G	p.I731V	10	0	10	0	-	0
NC_006092	5	57915337	A	G	N-syn	0.86	TMX1	c.230A>G	p.D77G	15	-	0	13	0	0
NC_006093	6	130323	G	T	N-syn	1	FAM13C	c.1446G>T	p.E482D	10	0	0	-	10	0
NC_006093	6	23801282	T	C	N-syn	0.81	ITPRIP	c.1528A>G	p.I510V	11	0	9	0	-	0
NC_006094	7	4260582	C	T	N-syn	0.8	ABCA12	c.2110G>A	p.A704T	10	0	-	1	8	0
NC_006094	7	4717753	G	A	N-syn	0.8	RAB17	c.601G>A	p.V201I	10	8	0	-	0	0
NC_006094	7	4717765	G	A	N-syn	0.9	RAB17	c.613G>A	p.V205I	11	10	0	-	0	0
NC_006094	7	14344085	A	G	N-syn	1	ZNF385B	c.131A>G	p.H44R	13	-	0	13	0	0
NC_006094	7	15047308	A	G	N-syn	1	LOC770919	c.1253A>G	p.Y418C	10	-	0	10	0	0
NC_006094	7	15047322	A	G	N-syn	1	LOC770919	c.1267A>G	p.K423E	10	-	0	10	0	0
NC_006096	9	713606	G	A	N-syn	1	SAG	c.151G>A	p.V51M	13	13	0	-	0	0
NC_006096	9	20652272	T	C	N-syn	1	SI	c.2905T>C	p.S969P	11	0	11	0	-	0
NC_006096	9	21722478	G	A	N-syn	0.83	LOC425015	c.1352G>A	p.G451D	12	10	0	-	0	0
NC_006096	9	21722484	C	T	N-syn	0.8	LOC425015	c.1358C>T	p.P453L	10	0	-	0	8	0
NC_006096	9	22653344	T	G	N-syn	0.89	DHX36	c.333T>G	p.Y111.	10	0	0	9	-	1
NC_006097	10	11164703	C	T	N-syn	1	FAM154B	c.287G>A	p.R96K	11	0	-	0	11	0
NC_006098	11	9564860	G	A	N-syn	1	LRP3	c.2206G>A	p.G736R	10	10	0	-	0	0
NC_006099	12	5208897	A	G	N-syn	1	LOC100857401	c.3511A>G	p.K1171E	10	-	0	10	0	0
NC_006100	13	14403319	C	G	N-syn	0.9	LOC101749661	c.260G>C	p.G87A	11	0	-	10	0	0

Table 4.3 (Contd.)

Contig ID	C	Ref Pos	Ref Base	Called Base	Impact	SNP %	Feature Name	DNA Change	AA Change	Depth	A Cnt	C Cnt	G Cnt	T Cnt	Del
NC_006101	14	11909418	T	C	N-syn	1	KIAA0556	c.2794A>G	p.N932D	10	0	10	0	-	0
NC_006101	14	13770765	A	G	N-syn	0.89	IFT140	c.2372A>G	p.K791R	10	-	0	9	0	0
NC_006102	15	10822742	T	C	N-syn	0.89	CCDC157	c.1492A>G	p.S498G	10	1	9	0	-	0
NC_006104	17	6833124	C	T	N-syn	0.83	GBGT1	c.368G>A	p.R123H	12	0	-	0	10	0
NC_006104	17	7711131	A	G	N-syn	1	PPP1R26	c.637A>G	p.I213V	11	-	0	11	0	0
NC_006106	19	5656736	C	T	N-syn	0.8	SLC46A1	c.1174G>A	p.G392S	10	0	-	0	8	0
NC_006107	20	43121	C	T	N-syn	1	IGSF1	c.412C>T	p.R138C	10	0	-	0	10	0
NC_006107	20	478926	C	T	N-syn	0.8	GGT7	c.1472C>T	p.S491F	10	0	-	0	8	0
NC_006107	20	4706188	G	T	N-syn	0.89	LOC101750167	c.619G>T	p.A207S	10	0	0	-	9	0
NC_006108	21	1906952	T	C	N-syn	1	TMEM52	c.2T>C	p.M1T	10	0	10	0	-	0
NC_006108	21	1906991	G	T	N-syn	1	TMEM52	c.41G>T	p.C14F	11	0	0	-	11	0
NC_006108	21	2451062	T	G	N-syn	0.89	SCNN1D	c.1912A>C	p.I638L	10	0	0	9	-	0
NC_006108	21	3241081	C	T	N-syn	1	SLC2A5	c.1246G>A	p.A416T	11	0	-	0	11	0
NC_006108	21	4662973	C	A	N-syn	0.89	EMC1	c.1070G>T	p.S357I	10	9	-	0	0	0
NC_006111	24	130579	C	T|C	N-syn	0.81	FEZ1	c.[1196C>T] +[1196C>C]	p.P399P, p.P399L	16	0	-	0	13	0

Contig ID, chromosome (Chr) numbers, reference position (Ref Pos), reference base (Ref Base), called (SNP) base, impact (kinds of protein mutation), SNP %, feature name (gene name), DNA change, amino acid (AA) change, Depths, and five columns for SNP counts (cnts) are indicated.

Table 4.4 Validation of SNPs using allele-specific PCR in 96 AR and 96 AP line chickens

Chr	Ref Pos	Genes	Ref Base	Called Base	Impact	SNP %	Amino acid change	Results of 96 birds each from AP and AR lines					
								Homozygous RJF		Heterozygote		Homozygous SNP	
								AR	AP	AR	AP	AR	AP
1	193338454	ART7B	A	T	N-syn	1	p.L197Q	5	93	23	3	68	0
1	121349181	BMX	C	T|C	N-syn	0.75	p.D517N, p.D517D	16	96	26	0	54	0
1	1034107	FAM208B	T	C	N-syn	1	p.K1593E	0	94	0	1	96	1
14	13770765	IFT140	A	G	N-syn	0.89	p.K791R	9	95	44	1	43	0
20	43121	IGSF1	C	T	N-syn	1	p.R138C	0	63	0	28	96	5
1	14729572	LAMB4	T	C	N-syn	0.8	p.K1547R	39	84	15	2	42	10
2	42201510	PIK3R4	T	T|C	N-syn	0.75	p.D646D, p.D646G	16	71	4	16	76	9
3	24479109	THADA	T	C	N-syn	1	p.S287G	0	64	0	30	96	2
2	41279450	TOPAZ1	G	T	N-syn	1	p.T975N	0	16	0	78	96	2

Tables 4.5 Gene name and functions of genes containing amino acid changes showing over 10 depth counts in AR chickens

ID	Entrez Gene Name	Location	Type(s)
ABCA12	ATP-binding cassette, sub-family A (ABC1), member 12	Plasma Membrane	Transporter
ARSD	arylsulfatase D	Cytoplasm	Enzyme
BANK1	B-cell scaffold protein with ankyrin repeats 1	Extracellular Space	Other
BMX	BMX non-receptor tyrosine kinase	Cytoplasm	Kinase
CCDC157	coiled-coil domain containing 157	Other	Other
CCDC67	coiled-coil domain containing 67	Other	Other
DDX10	DEAD (Asp-Glu-Ala-Asp) box polypeptide 10	Nucleus	Enzyme
DHX36	DEAH (Asp-Glu-Ala-His) box polypeptide 36	Cytoplasm	Enzyme
DISC1	disrupted in schizophrenia 1	Cytoplasm	Other
EMC1	ER membrane protein complex subunit 1	Plasma Membrane	Other
FAM13C	family with sequence similarity 13, member C	Other	Other
FAM154B	family with sequence similarity 154, member B	Other	Other
FAM208B	family with sequence similarity 208, member B	Other	Other
FEZ1	fasciculation and elongation protein zeta 1 (zygin I)	Cytoplasm	Other
GBGT1	globoside alpha-1,3-N-acetylgalactosaminyltransferase 1	Other	Enzyme
GGT7	gamma-glutamyltransferase 7	Plasma Membrane	Enzyme
GVINP1	GTPase, very large interferon inducible pseudogene 1	Other	Other
IFNAR2	interferon (alpha, beta and omega) receptor 2	Plasma Membrane	transmembrane receptor
IFT140	intraflagellar transport 140	Extracellular Space	Other
IGSF1	immunoglobulin superfamily, member 1	Plasma Membrane	Other
ITPRIP	inositol 1,4,5-trisphosphate receptor interacting protein	Extracellular Space	Other
KIAA0556	KIAA0556	Extracellular Space	Other
LAMB4	laminin, beta 4	Other	Other
LRP3	low density lipoprotein receptor-related protein 3	Plasma Membrane	transmembrane receptor
MGA	MGA, MAX dimerization protein	Nucleus	transcription regulator

Table 4.5 (Contd.)

ID	Entrez Gene Name	Location	Type(s)
PIK3C2G	phosphatidylinositol-4-phosphate 3-kinase, catalytic subunit type 2 gamma	Cytoplasm	Kinase
PIK3R4	phosphoinositide-3-kinase, regulatory subunit 4	Cytoplasm	Kinase
PLEKHH2	pleckstrin homology domain containing, family H (with MyTH4 domain) member 2	Cytoplasm	Other
PPP1R26	protein phosphatase 1, regulatory subunit 26	Nucleus	Other
PTPN2	protein tyrosine phosphatase, non-receptor type 2	Cytoplasm	Phosphatase
RAB17	RAB17, member RAS oncogene family	Cytoplasm	Enzyme
RPGR	retinitis pigmentosa GTPase regulator	Cytoplasm	Other
SAG	S-antigen; retina and pineal gland (arrestin)	Cytoplasm	Other
SCNN1D	sodium channel, non-voltage-gated 1, delta subunit	Plasma Membrane	ion channel
SI	sucrase-isomaltase (alpha-glucosidase)	Cytoplasm	Enzyme
SLC2A5	solute carrier family 2 (facilitated glucose/fructose transporter), member 5	Plasma Membrane	Transporter
SLC46A1	solute carrier family 46 (folate transporter), member 1	Plasma Membrane	Transporter
TARBP1	TAR (HIV-1) RNA binding protein 1	Nucleus	transcription regulator
THADA	thyroid adenoma associated	Other	Other
TMEM52	transmembrane protein 52	Other	Other
TMX1	thioredoxin-related transmembrane protein 1	Cytoplasm	Enzyme
TOPAZ1	testis and ovary specific PAZ domain containing 1	Other	Other
TTC29	tetratricopeptide repeat domain 29	Other	Other
ZNF385B	zinc finger protein 385B	Nucleus	Other
LOC101748372	Uncharacterized	N/A	N/A
ART7B	Uncharacterized	N/A	N/A
LOC420515	Uncharacterized	N/A	N/A
LOC101751154	Uncharacterized	N/A	N/A
COPN5L1	Uncharacterized	N/A	N/A
LOC101751416	Uncharacterized	N/A	N/A
LOC771752	Uncharacterized	N/A	N/A
C4H4ORF21	Uncharacterized	N/A	N/A
LOC100858625	Uncharacterized	N/A	N/A
LOC770919	Uncharacterized	N/A	N/A

Table 4.6 Associated network functions of candidate genes

ID	Molecules in Network	Score	Top Diseases and Functions
1	ABCA12, ACP6, ARSB, ARSD, ARSE, ARSF, ARSG, ARSI, ARSJ, CBWD2, DDX10, DHX36, EMC1, GGT7, GNS, IFT122, IFT140, IFT172, ITPRIP, LAMB4, MGA, PANK2, PLEC, SCNN1D, SGSH, SLC46A1, SRPX, SUMF1, SUMF2, TARBP1, THADA,TTBK2, TTC21B, UBC, WDR19	30	Developmental Disorder, Hereditary Disorder, Metabolic Disease
2	ANO10, BANK1, BCR (complex), CAPN15, CCDC67, CCDC141, CCDC85B, DISC1, FAM208B, FAS, FEZ1, GPR107, IGSF1, KMT2D, LYN, MARVELD1, NT5DC1, OARD1, P2RY8, PLEKHH2, PPP1R26, RELL1, RPGR,S LC13A3, SLC26A11, SLC35F5, SMC3, SYPL2, TGFB1, MEM184B, TMX1, TRMT2B, UBC, ZNF761, ZNF385B	22	Cell Death and Survival, Hematological System Development and Function, Humoral Immune Response
3	ARRB1, BMX, Cgm4/Psg16, CRK-CRKL-CBL, Erbb3 dimer, FAM13C, FRY, GAB1/2, GHR dimer, GSN-PI3K-PIP2-Src, Hcst dimer, IFNAR2, Insulin, IRS2-PI3K, LMNA, LRP, LRP3, LRP10, Met dimer, NFkB (complex), NUPR1, PI3K (complex), PIK3C2G, PIK3R4, Prl4a1, PTPN2, RAB17, SAG, Sh2b3, SHC-GRB2-GAB1, SHP2-PI3K-GAB2, SI, SLC2A5, T3-TRbeta, Vegfr dimer	22	Developmental Disorder, Gastrointestinal Disease, Hereditary Disorder

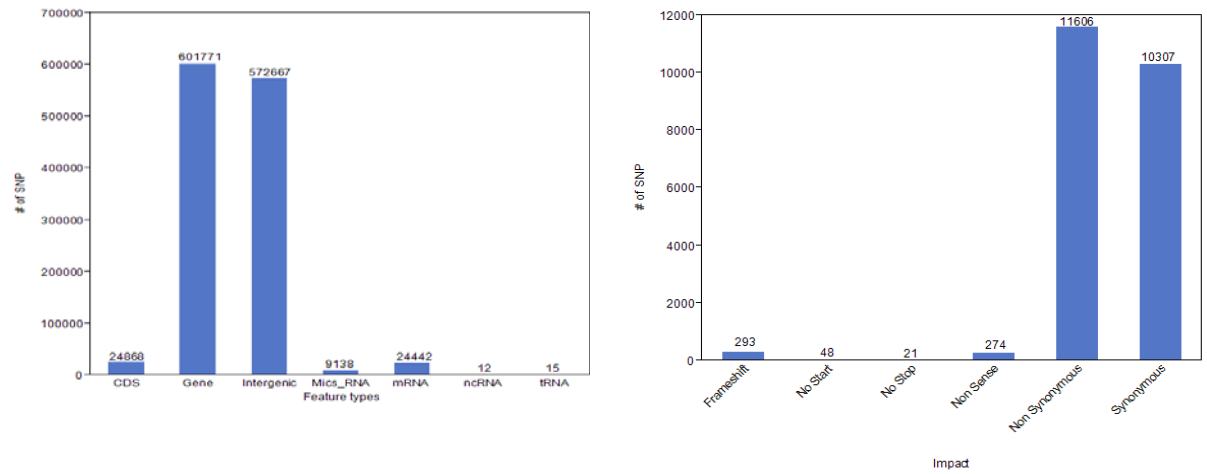

Figure 4.1 Distribution of SNPs in different regions of genome (1a) and SNPs associated with different types of mutations (1b).

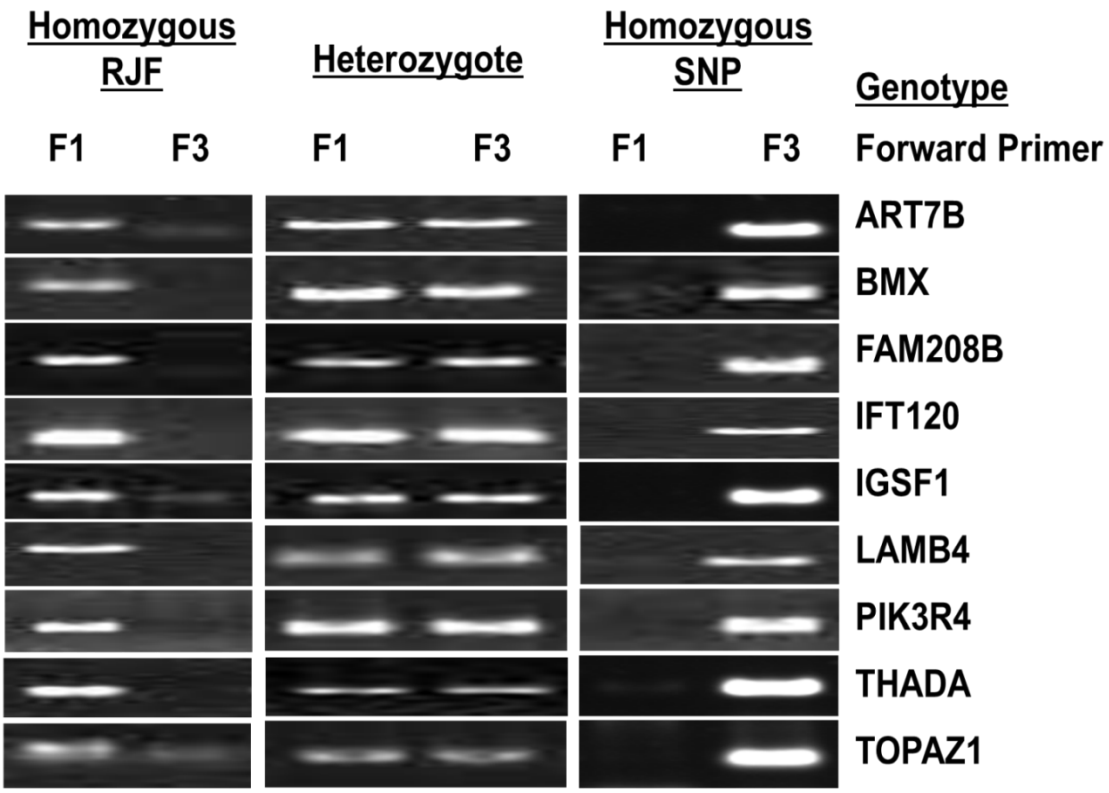

Figure 4.2 Nine candidate markers and different genotypes shown by allele-specific PCR in larger population of AR and AP chicken lines.

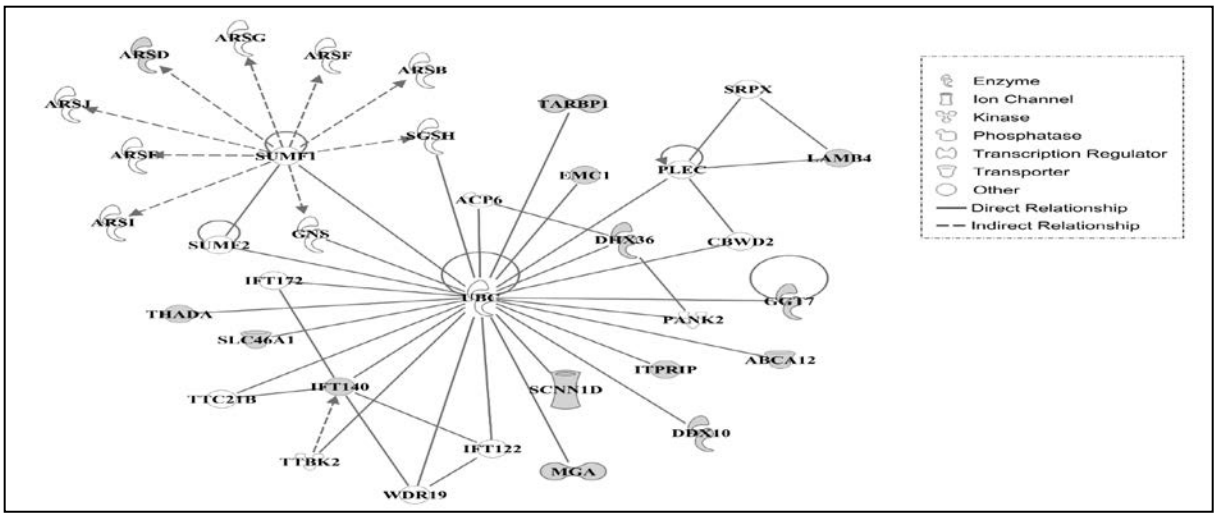

Figure 4.3 Gene network #1. Molecular interactions among the important focus molecules are displayed. Gray symbols show the genes found in the list of SNP while white symbols indicate neighboring genes that are functionally associated, but not included, in the gene list of SNP. Symbols for each molecule are presented according to molecular functions and type of interactions.

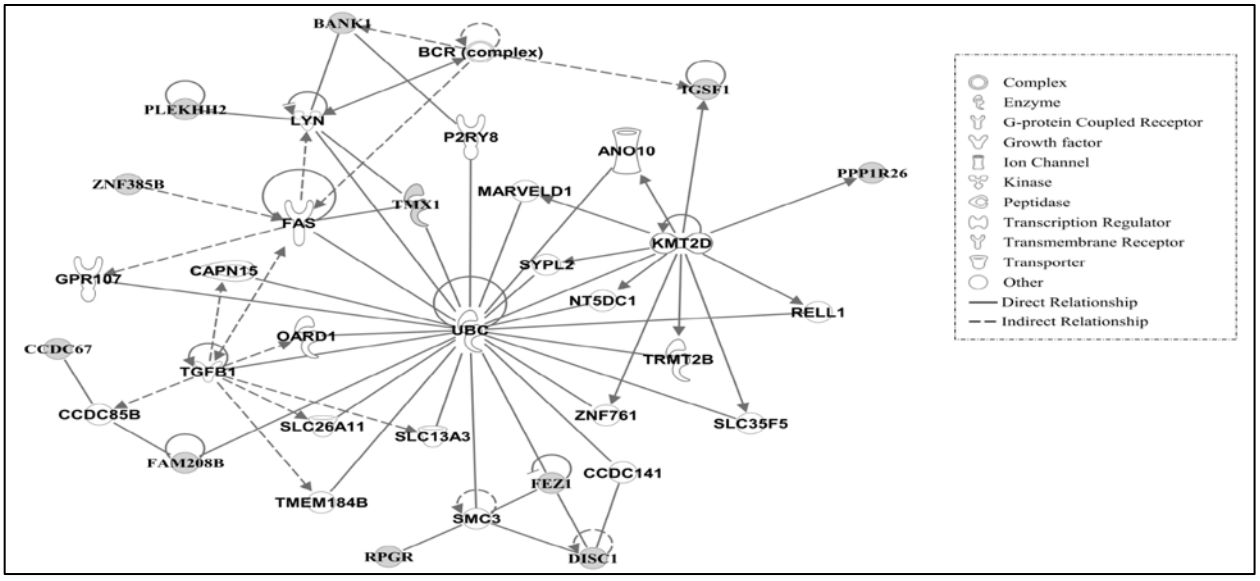

Figure 4.4 Gene network #2. Molecular interaction and symbols are the same as the description in Figure 4.3.

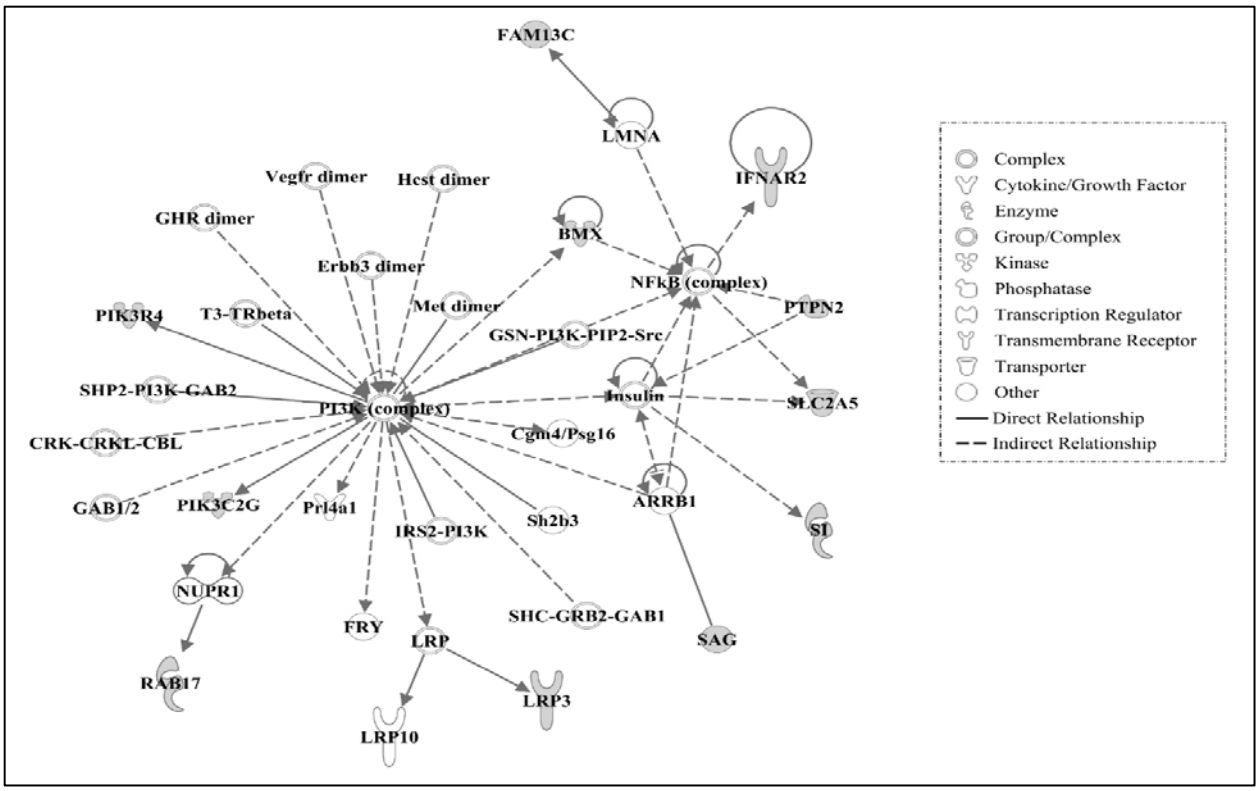

Figure 4.5 Gene network #3. Molecular interaction and symbols are the same as the description in Figure 4.3.

5. CONCLUSION

In this dissertation, modern genomic approaches were used to identify the molecular markers in avian species maintained under different physiological and genetic environmental conditions. In this study, the biomarkers such as microRNA, copy number variation and single nucleotide polymorphism that were associated with different physiological phenomenon such as muscle growth and development, stress response and disease susceptibility which could impact overall growth and development of avian species were identified. More specifically, the main focus of the research in this dissertation was to characterize the breast muscle specific microRNAs in modern broilers and foundational chicken lines, to determine copy number variation in high and low stress quail lines, and to identify single nucleotide polymorphism markers in Rous sarcoma virus susceptible and resistant lines of chickens.

In chapter 2, we identified nine breast muscle specific mature miRNAs differentially expressed in modern broilers compared to the foundational chicken line. These miRNAs could be alternative markers for regulating muscle growth and development in chickens. The IPA predicted the involvement of these miRNAs in several canonical pathways such as P38 MAPK, ERK1/2, PI3K and insulin signaling, calcium signaling, axonal guidance signaling and NRF2-mediated oxidative stress signaling pathways suggesting their roles in muscle specific growth and development in chickens. In the present context of increasing global protein needs for human population, the miRNAs and their functions identified in this study could serve as foundation for understanding molecular basis of their regulatory roles for rapid growth and development of breast muscle in chickens required for human population.

In chapter 3, we identified different sets of genes and regions of DNA affected by CNVs in HS and LS lines of quail. Most importantly the CNV genes were involved in nervous/endocrine systems development and humoral/cell-mediated immune responses. The CNV

genes and CNV regions identified in this study could serve as molecular markers for understanding the effect of stress on overall growth and development of avian and other animal species. In this study, the result supported the hypothesis that CNVs could have impact in increasing genotypic diversity and thereby phenotypic traits observed in quail.

In chapter 4, we identified 63 different reliable SNP markers uniquely present in AR compared to AP chicken lines that could be associated with tumor regression property. The functional study using IPA showed that the SNP containing genes in AR were associated with ubiquitylation, and PI3K and NF-kB signaling pathways, suggesting their roles in tumor regression phenomenon in AR lines. The SNP markers presented in this study could serve as basis for understanding disease resistance and susceptibility in avian and other species.

APPENDIX

Supplementary Table 2.1: Total read counts of abundant mature miRNA from breast muscle of PeM and BPR chickens

SN	Mature miRNA	Total BPR	Total PeM
1	gga-miR-21-5p	15132	9763
2	gga-miR-30d	8196	5217
3	gga-miR-146b-5p	2570	1489
4	gga-miR-199-5p	2102	2044
5	gga-miR-193a-3p	1439	731
6	gga-miR-20a-5p	1178	890
7	gga-miR-148a-3p	1321	796
8	gga-miR-30a-5p	513	530
9	gga-let-7g-5p	466	334
10	gga-miR-133c-3p	601	419
11	gga-miR-10b-5p	712	420
12	gga-miR-30c-5p	900	496
13	gga-miR-146c-5p	351	352
14	gga-miR-2954	218	240
15	gga-miR-1a-3p	464	261
16	gga-miR-222a	163	110
17	gga-miR-103-3p	115	170
18	gga-miR-1454	282	186
19	gga-miR-22-3p	407	317
20	gga-miR-301b-3p	351	190
21	gga-miR-133a-5p	433	315
22	gga-let-7f-5p	401	209
23	gga-miR-26a-5p	281	179
24	gga-miR-125b-5p	227	121
25	gga-miR-16-5p	199	116
26	gga-miR-24-3p	181	141
27	gga-miR-2131-5p	122	77
28	gga-miR-221-3p	141	103
29	gga-miR-17-5p	85	72
30	gga-miR-181a-5p	103	129
31	gga-miR-126-3p	32	63
32	gga-miR-451	133	96
33	gga-let-7c-5p	42	66

Supplementary Table 2.1 (Contd.)

SN	Mature miRNA	Total BPR	Total PeM
34	gga-miR-206	88	46
35	gga-let-7b	145	99
36	gga-miR-128-3p	69	45
37	gga-miR-221-5p	223	79
38	gga-miR-10a-5p	53	41

Supplementary Table 2.2: Target mRNAs of differentially expressed and qPCR validated miRNAs in PeM and BPR comparison

Target Rank	Target Score	miRNA Name	Target Gene	Gene Description	FC(RNASeq)	p-value
5	100	gga-let-7b	IGF2BP3	insulin-like growth factor 2 mRNA binding protein 3	-1.85	0.0014
80	96	gga-let-7b	ACER2	alkaline ceramidase 2	-1.74	0.00644
139	94	gga-let-7b	FGD6	FYVE, RhoGEF and PH domain containing 6	-1.5	0.0342
163	93	gga-let-7b	FOXP2	forkhead box P2	-1.88	0.00815
271	88	gga-let-7b	MGAT4A	mannosyl (alpha-1,3-)-glycoprotein beta-1,4-N-acetylglucosaminyltransferase, isozyme A	-1.54	0.0102
325	86	gga-let-7b	SLC2A12	solute carrier family 2 (facilitated glucose transporter), member 12	-1.34	0.00845
403	83	gga-let-7b	KLHL13	kelch-like 9 (Drosophila)	-1.32	0.000988
457	81	gga-let-7b	GATM	glycine amidinotransferase (L-arginine:glycine amidinotransferase)	-3.81	0.0000822
555	78	gga-let-7b	PDP2	pyruvate dehyrogenase phosphatase catalytic subunit 2	-1.82	0.00558
643	75	gga-let-7b	KIAA1467	KIAA1467	-1.38	0.036
645	75	gga-let-7b	TMEM56	transmembrane protein 56	-2.81	0.00544
676	74	gga-let-7b	SNX16	sorting nexin 16	-1.32	0.0112
1012	64	gga-let-7b	WDR37	WD repeat domain 37	-1.57	0.0122
1064	63	gga-let-7b	PNAT10	N-acetyltransferase, pineal gland isozyme NAT-10	-4.47	0.0000827
1153	61	gga-let-7b	TMEM164	transmembrane protein 164	-1.93	0.00347
1154	61	gga-let-7b	NNT	nicotinamide nucleotide transhydrogenase	-1.75	0.00000402
1196	60	gga-let-7b	GTF3C6	general transcription factor IIIC, polypeptide 6, alpha 35kDa	-1.56	0.00662
1234	59	gga-let-7b	SLC30A1	solute carrier family 30 (zinc transporter), member 1	1.46	0.0253
1307	57	gga-let-7b	AFF2	AF4/FMR2 family, member 2	-1.46	0.00434
1394	55	gga-let-7b	EEF2K	eukaryotic elongation factor-2 kinase	-1.87	0.0068
1395	55	gga-let-7b	AQP4	aquaporin 4	-2.97	0.0179
1443	54	gga-let-7b	SALL4	sal-like 4 (Drosophila)	-2.59	0.0028
415	83	gga-miR-10a-5p	SLC35E1	solute carrier family 35, member E1	-1.33	0.000524

Supplementary Table 2.2 (Contd.)

Target Rank	Target Score	miRNA Name	Target Gene	Gene Description	FC (RNASeq)	p-value
1041	64	gga-miR-10a-5p	HDAC11	histone deacetylase 11	-1.37	0.00311
1044	64	gga-miR-10a-5p	NAA50	N(alpha)-acetyltransferase 50, NatE catalytic subunit	-1.51	0.00299
1140	62	gga-miR-10a-5p	RNPC3	RNA-binding region (RNP1, RRM) containing 3	-1.3	0.00846
1259	59	gga-miR-10a-5p	MFAP3	microfibrillar-associated protein 3	-1.43	0.000123
1475	54	gga-miR-10a-5p	SLC35D1	solute carrier family 35 (UDP-glucuronic acid/UDP-N-acetylgalactosamine dual transporter), member D1	-1.82	0.00041
1532	53	gga-miR-10a-5p	HSPA13	heat shock protein 70kDa family, member 13	-1.45	0.0028
1637	51	gga-miR-10a-5p	KLF11	Kruppel-like factor 11	-1.4	0.016
113	95	gga-miR-125b-5p	KLF13	Kruppel-like factor 13	-2.01	0.0000272
167	93	gga-miR-125b-5p	ZFYVE1	zinc finger, FYVE domain containing 1	-1.31	0.00351
203	91	gga-miR-125b-5p	LRRC8B	leucine rich repeat containing 8 family, member B	-3.03	0.000000287
253	89	gga-miR-125b-5p	TLE3	transducin-like enhancer of split 3 (E(sp1) homolog, Drosophila)	-1.43	0.00136
272	88	gga-miR-125b-5p	ELOVL6	ELOVL fatty acid elongase 6	-2.13	0.0271
304	87	gga-miR-125b-5p	TMEM184B	transmembrane protein 184B	-1.59	0.00379
460	81	gga-miR-125b-5p	TMTC2	transmembrane and tetratricopeptide repeat containing 2	-1.35	0.00622
490	80	gga-miR-125b-5p	TTPA	tocopherol (alpha) transfer protein	-1.95	0.0393
491	80	gga-miR-125b-5p	ZNRF3	zinc and ring finger 3	-2.06	0.00299
948	66	gga-miR-125b-5p	TMEM161B	transmembrane protein 161B	-1.64	0.0381
1454	54	gga-miR-125b-5p	ATMIN	ATM interactor	-1.63	0.0151
285	88	gga-miR-146	PLEKHM3	pleckstrin homology domain containing, family M, member 3	-1.64	0.00373
286	88	gga-miR-146	ZBTB2	zinc finger and BTB domain containing 2	-1.37	0.0168
411	83	gga-miR-146	GIN1	gypsy retrotransposon integrase 1	-1.66	0.0128
699	74	gga-miR-146	GDNF	glial cell derived neurotrophic factor	-1.88	0.0157
759	72	gga-miR-146	KBTBD2	kelch repeat and BTB (POZ) domain containing 2	-1.3	0.00809

Supplementary Table 2.2 (Contd.)

Target Rank	Target Score	miRNA Name	Target Gene	Gene Description	FC (RNASeq)	p-value
818	70	gga-miR-146	PTPN21	protein tyrosine phosphatase, non-receptor type 21	-2.14	0.0206
819	70	gga-miR-146	MYSM1	myb-like, SWIRM and MPN domains 1	-1.51	0.0469
1258	59	gga-miR-146	TM7SF3	transmembrane 7 superfamily member 3	-1.57	0.000276
1421	55	gga-miR-146	FAM168B	family with sequence similarity 168, member B	-1.52	0.045
1685	50	gga-miR-146	ZNRF3	zinc and ring finger 3	-2.06	0.00299
3	100	gga-miR-206	GLCCI1	glucocorticoid induced transcript 1	1.83	0.00000357
72	96	gga-miR-206	CNN3	calponin 3, acidic	1.38	0.0109
74	96	gga-miR-206	EIF2AK3	eukaryotic translation initiation factor 2-alpha kinase 3	1.39	0.0103
103	95	gga-miR-206	SNAI2	snail homolog 2 (Drosophila)	1.66	0.0318
215	90	gga-miR-206	RASSF2	Ras association (RalGDS/AF-6) domain family member 2	1.92	0.000959
222	90	gga-miR-206	MEOX2	mesenchyme homeobox 2	1.51	0.0294
262	88	gga-miR-206	UTRN	utrophin	1.36	0.0277
295	87	gga-miR-206	RRBP1	ribosome binding protein 1 homolog 180kDa (dog)	1.74	0.000101
320	86	gga-miR-206	SULF1	sulfatase 1	1.65	0.0117
343	85	gga-miR-206	SPRED1	sprouty-related, EVH1 domain containing 1	1.32	0.00912
344	85	gga-miR-206	EMP1	epithelial membrane protein 1	2.07	0.00638
345	85	gga-miR-206	RSPO3	R-spondin 3	5.39	0.000156
370	84	gga-miR-206	DAP	death-associated protein	1.61	0.00932
422	82	gga-miR-206	THBS1	thrombospondin 1	5.21	0.0000454
638	75	gga-miR-206	SLC25A22	solute carrier family 25 (mitochondrial carrier: glutamate), member 22	1.8	0.0314
639	75	gga-miR-206	ARAP2	ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 2	2.35	0.00518
670	74	gga-miR-206	FAM150B	family with sequence similarity 150, member B	1.8	0.0344
711	73	gga-miR-206	GSTT1	glutathione S-transferase theta 1	1.88	0.000829

Supplementary Table 2.2 (Contd.)

Target Rank	Target Score	miRNA Name	Target Gene	Gene Description	FC (RNASeq)	p-value
794	70	gga-miR-206	CCBE1	collagen and calcium binding EGF domains 1	1.53	0.00467
829	69	gga-miR-206	TPM3	tropomyosin 3	2.25	0.000142
862	68	gga-miR-206	SYK	spleen tyrosine kinase	2.62	0.00715
866	68	gga-miR-206	ANXA2	annexin A2	4.11	0.0000643
867	68	gga-miR-206	SNX4	sorting nexin 4	1.43	0.00052
936	66	gga-miR-206	FAM126A	family with sequence similarity 126, member A	1.36	0.00702
976	65	gga-miR-206	SOWAHC	sosondowah ankyrin repeat domain family member C	1.51	0.0279
1052	63	gga-miR-206	CORO1C	coronin, actin binding protein, 1C	1.83	0.000261
1144	61	gga-miR-206	PSAT1	phosphoserine aminotransferase 1	1.64	0.000216
1226	59	gga-miR-206	HTRA1	HtrA serine peptidase 1	2.27	0.00758
1227	59	gga-miR-206	OGN	osteoglycin	2.28	0.000403
1389	55	gga-miR-206	PDIA5	protein disulfide isomerase family A, member 5	1.32	0.00909
1429	54	gga-miR-206	GPR137B	G protein-coupled receptor 137B	1.77	0.000358
1430	54	gga-miR-206	OSBPL5	oxysterol binding protein-like 5	1.46	0.0156
1433	54	gga-miR-206	ARSJ	arylsulfatase family, member J	5.21	0.00000147
1490	53	gga-miR-206	BMP6	bone morphogenetic protein 6	1.53	0.0127
1647	50	gga-miR-206	SVEP1	sushi, von Willebrand factor type A, EGF and pentraxin domain containing 1	1.73	0.00134
91	96	gga-miR-2131-5p	H2AFZ	H2A histone family, member Z	-1.34	0.00621
92	96	gga-miR-2131-5p	NCOR1	nuclear receptor corepressor 1	-1.66	0.000706
117	95	gga-miR-2131-5p	KBTBD8	kelch repeat and BTB (POZ) domain containing 8	-1.72	0.0027
171	93	gga-miR-2131-5p	LAPTM4B	lysosomal protein transmembrane 4 beta	-1.77	0.0015
187	92	gga-miR-2131-5p	LIG4	ligase IV, DNA, ATP-dependent	-1.54	0.00213
205	91	gga-miR-2131-5p	EPG5	ectopic P-granules autophagy protein 5 homolog (C. elegans)	-1.47	0.0367

Supplementary Table 2.2 (Contd.)

Target Rank	Target Score	miRNA Name	Target Gene	Gene Description	FC (RNASeq)	p-value
210	91	gga-miR-2131-5p	ISOC1	isochorismatase domain containing 1	-1.32	0.0229
277	88	gga-miR-2131-5p	HSPA13	heat shock protein 70kDa family, member 13	-1.45	0.0028
284	88	gga-miR-2131-5p	RAP1B	RAP1B, member of RAS oncogene family	1.34	0.0000165
313	87	gga-miR-2131-5p	CHN2	chimerin (chimaerin) 2	-1.78	0.0404
334	86	gga-miR-2131-5p	GTF2A1	general transcription factor IIA, 1, 19/37kDa	-1.6	0.0000302
354	85	gga-miR-2131-5p	NR3C2	nuclear receptor subfamily 3, group C, member 2	-1.5	0.00419
429	82	gga-miR-2131-5p	RNGTT	RNA guanylyltransferase and 5'-phosphatase	-1.54	0.00655
432	82	gga-miR-2131-5p	LNPEP	leucyl/cystinyl aminopeptidase	-2	0.0286
468	81	gga-miR-2131-5p	SP4	Sp4 transcription factor	-1.35	0.00323
470	81	gga-miR-2131-5p	SNX2	sorting nexin 2	-1.3	0.0057
494	80	gga-miR-2131-5p	SLC19A2	solute carrier family 19 (thiamine transporter), member 2	-1.84	0.00804
498	80	gga-miR-2131-5p	MSMO1	methylsterol monooxygenase 1	-1.61	0.00916
520	79	gga-miR-2131-5p	WDR51B	WD repeat domain 51B	-1.87	0.000903
527	79	gga-miR-2131-5p	MTMR7	myotubularin related protein 7	-2.06	0.00172
560	78	gga-miR-2131-5p	RRP1B	ribosomal RNA processing 1 homolog B (S. cerevisiae)	-1.41	0.0014
562	78	gga-miR-2131-5p	ZNF770	zinc finger protein 770	-1.53	0.0167
563	78	gga-miR-2131-5p	LGALSL	lectin, galactoside-binding-like	-1.4	0.00442
598	77	gga-miR-2131-5p	PDP2	pyruvate dehyrogenase phosphatase catalytic subunit 2	-1.82	0.00558
620	76	gga-miR-2131-5p	RHOBTB3	Rho-related BTB domain containing 3	-2.31	0.00644
654	75	gga-miR-2131-5p	CHAF1B	chromatin assembly factor 1, subunit B (p60)	-1.4	0.00467
694	74	gga-miR-2131-5p	NUDT6	nudix (nucleoside diphosphate linked moiety X)-type motif 6	-1.4	0.0126
721	73	gga-miR-2131-5p	PKNOX1	PBX/knotted 1 homeobox 1	-1.79	0.00799
725	73	gga-miR-2131-5p	TECPR1	tectonin beta-propeller repeat containing 1	-1.4	0.0294
757	72	gga-miR-2131-5p	PDS5A	PDS5, regulator of cohesion maintenance, homolog A (S. cerevisiae)	-1.36	0.00167

Supplementary Table 2.2 (Contd.)

Target Rank	Target Score	miRNA Name	Target Gene	Gene Description	FC (RNASeq)	p-value
775	71	gga-miR-2131-5p	FOXP2	forkhead box P2	-1.88	0.00815
808	70	gga-miR-2131-5p	FAM98B	family with sequence similarity 98, member B	-1.38	0.0000781
815	70	gga-miR-2131-5p	MLLT10	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 10	-1.38	0.0000451
851	69	gga-miR-2131-5p	GFOD1	glucose-fructose oxidoreductase domain containing 1	-1.45	0.0374
921	67	gga-miR-2131-5p	ROCK2	Rho-associated, coiled-coil containing protein kinase 2	-2.07	0.00366
960	66	gga-miR-2131-5p	TRIM23	tripartite motif containing 23	-1.3	0.0189
989	65	gga-miR-2131-5p	AQP9	aquaporin 9	-3.41	0.00566
1074	63	gga-miR-2131-5p	CASC4	cancer susceptibility candidate 4	-1.4	0.000071
1086	63	gga-miR-2131-5p	ADAMTS20	ADAM metallopeptidase with thrombospondin type 1 motif, 20	-1.65	0.0187
1119	62	gga-miR-2131-5p	ZBTB24	zinc finger and BTB domain containing 24	-1.49	0.0012
1128	62	gga-miR-2131-5p	DENND1B	DENN/MADD domain containing 1B	-1.41	0.0271
1164	61	gga-miR-2131-5p	B9D1	B9 protein domain 1	-1.61	0.0418
1169	61	gga-miR-2131-5p	CLOCK	clock homolog (mouse)	-1.77	0.0419
1175	61	gga-miR-2131-5p	SLC44A5	solute carrier family 44, member 5	-1.89	0.0092
1207	60	gga-miR-2131-5p	ARV1	ARV1 homolog (S. cerevisiae)	-1.56	0.00033
1248	59	gga-miR-2131-5p	LATS2	LATS, large tumor suppressor, homolog 2 (Drosophila)	-1.32	0.0101
1286	58	gga-miR-2131-5p	BMP7	bone morphogenetic protein 7	-1.95	0.0177
1328	57	gga-miR-2131-5p	BTBD3	BTB (POZ) domain containing 3	-2.13	0.00123
1330	57	gga-miR-2131-5p	PFKFB2	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 2	-1.57	0.0431
1362	56	gga-miR-2131-5p	COL9A3	collagen, type IX, alpha 3	-3.23	0.00067
1407	55	gga-miR-2131-5p	CRKL	v-crk sarcoma virus CT10 oncogene homolog-like	-1.32	0.0137
1412	55	gga-miR-2131-5p	CLSPN	claspin	-1.41	0.0389
1415	55	gga-miR-2131-5p	KNTC1	kinetochore associated 1	-2.06	0.00333

Supplementary Table 2.2 (Contd.)

Target Rank	Target Score	miRNA Name	Target Gene	Gene Description	FC (RNASeq)	p-value
1461	54	gga-miR-2131-5p	MOB4	MOB family member 4, phocein	-1.31	0.0000548
1563	52	gga-miR-2131-5p	USP24	ubiquitin specific peptidase 24	-1.82	0.0362
1564	52	gga-miR-2131-5p	SNX16	sorting nexin 16	-1.32	0.0112
1570	52	gga-miR-2131-5p	PDXK	pyridoxal (pyridoxine, vitamin B6) kinase	-2.13	0.00621
1633	51	gga-miR-2131-5p	NFATC1	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 1	-1.67	0.012
1673	50	gga-miR-2131-5p	MBP	myelin basic protein	-2.62	0.000196
264	88	gga-miR-221-5p	SLC7A2	solute carrier family 7 (cationic amino acid transporter, y+ system), member 2	-2.03	0.000519
613	76	gga-miR-221-5p	CRIPT	cysteine-rich PDZ-binding protein	-1.33	0.000371
836	69	gga-miR-221-5p	AFF4	AF4/FMR2 family, member 4	-1.43	0.0396
872	68	gga-miR-221-5p	TK2	thymidine kinase 2, mitochondrial	-1.45	0.000463
1306	57	gga-miR-221-5p	RGMB	RGM domain family, member B	-1.35	0.0301
1392	55	gga-miR-221-5p	SLC12A2	solute carrier family 12 (sodium/potassium/chloride transporters), member 2	-1.8	0.00288
1393	55	gga-miR-221-5p	PDS5A	PDS5, regulator of cohesion maintenance, homolog A (S. cerevisiae)	-1.36	0.00167
1655	50	gga-miR-221-5p	CLOCK	clock homolog (mouse)	-1.77	0.0419

1. IACUC protocol for miRNA, CNV and SNP studies.

UNIVERSITY OF ARKANSAS

Office of Research Compliance

To: Byung-Whi ~~Coon~~ Kong
FR: John Hahn
Date: November 18th, 2016
Subject: IACUC Approval
Expiration Date: December 17th, 2018

The Institutional Animal Care and Use Committee (IACUC) has APPROVED your personnel additions of Bhuwan Khatri and Stephanie Shouse to protocol # 16036 *Proteogenomic integration to characterize economically important genetic traits to improve poultry production.*

In granting its approval, the IACUC has approved only the information provided. Should there be any further changes to the protocol during the research, please notify the IACUC in writing (via the Modification form) prior to initiating the changes. If the study period is expected to extend beyond December 17th, 2018 you must submit a newly drafted protocol prior to that date to avoid any interruption. By policy the IACUC cannot approve a study for more than 3 years at a time.

The IACUC appreciates your cooperation in complying with University and Federal guidelines involving animal subjects.

CNC/aem

2. IBC protocol for miRNA study.

UNIVERSITY OF ARKANSAS | Research & Economic Development
Research Compliance

April 26, 2016

MEMORANDUM

TO:	Dr. Byung-Whi Kong
FROM:	Ines Pinto, Biosafety Committee Chair
RE:	Protocol Modification
PROTOCOL #:	10007
PROTOCOL TITLE:	Genetically engineered vaccine production using virus vectors
APPROVED PROJECT PERIOD:	Start Date 8/13/2009 Expiration Date 8/12/2018

The Institutional Biosafety Committee (IBC) has approved your request, dated February 1, 2016, to modify Protocol # 10007, "Genetically engineered vaccine production using virus vectors".

The IBC appreciates your assistance and cooperation in complying with University and Federal guidelines for research involving hazardous biological materials.

PERSONNEL QUALIFICATIONS & FACILITY INFORMATION:

List all personnel (including PI and Co-PI) to be involved in this project:

Name (First and Last) - Position (Title, academic degrees, certifications, and field of expertise)	Qualifications/Training/Relevant Experience (Describe previous work or training with biohazardous and/or recombinant DNA; include Biosafety Levels)
Byung-Whi Kong - Assistant Professor, PhD - Poultry Science	14 yrs working with avian viruses, 17 yrs working with recombinant DNA, 17 yrs working with avian cell lines
Stephanie Shouse – Program Technician	Will receive training in all areas
Bhuwan Khatri – PhD student – Poultry Science	5 yrs experience with avian cell lines and performing molecular lab work
Click here to enter text.	Click here to enter text.
Click here to enter text.	Click here to enter text.
Click here to enter text.	Click here to enter text.
Click here to enter text.	Click here to enter text.

Page 5 of 10