RNA SEQUENCING TO STUDY DIFFERENTIAL GENE EXPRESSION

AND ALLELE SPECIFIC EXPRESSION IN CHICKENS

by

Zhu Zhuo

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Animal and Food Sciences

Winter 2018

RNA SEQUENCING TO STUDY DIFFERENTIAL GENE EXPRESSION

AND ALLELE SPECIFIC EXPRESSION IN CHICKENS

by

Zhu Zhuo

Approved: _____

Limin Kung, Jr., Ph.D.
Chair of the Department of Animal and Food Sciences

Approved: _____

Mark W. Rieger, Ph.D.
Dean of the College of Agriculture and Natural Resources

Approved: _____

Ann L. Ardis, Ph.D.
Senior Vice Provost for Graduate and Professional Education

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Behnam Abasht, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Susan J. Lamont, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Carl J. Schmidt, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Gregory E. Gonye, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Mark S. Parcells, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Tania L. Roth, Ph.D.
Member of dissertation committee

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to Dr. Behnam Abasht, Dr. Susan Lamont, Dr. Carl Schmidt, Dr. Gregory Gonye, Dr. Mark Parcells, Dr. Tania Roth, Dr. Randall Wisser, Marie Mutryn, Nan Zhou, Weixuan Fu, the Abasht lab, my friends and family. This could not be done without your guidance, help, support, and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Gene expression is an intermediate, quantitative phenotype that bridges genotype, environment and phenotype. Variation in gene expression contributes to phenotypic variation among individuals and populations. Survey of global gene expression can provide valuable insight into the molecular basis and mechanisms of biological phenomena and thus contributes to the understanding of fundamental genomics questions and economically important traits for agricultural animals. Using RNA sequencing, we have studied allele-specific expression and differential gene expression associated with genomic imprinting, heterosis, and feed efficiency in chickens. Firstly, through investigation of allele-specific expression in F1 reciprocal crosses derived from two highly inbred chicken lines, Fayoumi and Leghorn, we found there was no evidence of genomic imprinting in 12-day embryonic brain and liver. Secondly, we identified additivity as the most predominant transgenerational gene expression pattern between F1 crosses and parental lines, and highlighted the differentially expressed genes that could potentially contribute to heterosis. Lastly, we studied the difference of gene expression in abdominal fat between high feed efficiency and low feed efficiency broiler chickens, and determined the functional clusters and pathways that may result in divergent fat deposition between the two groups of chickens.

Chapter 1

INTRODUCTION

Gene expression can be considered as an intermediate, quantitative cellular phenotype [1]. It is determined by multiple factors, including DNA sequence variants, epigenetic modifications, post-transcriptional modifications, cellular environment and external stimuli. The resultant variation in gene expression, reflecting in the RNA stability and the abundance of RNA transcripts, contributes to phenotypic variation among individuals and populations. As most genetic variants are found in non-coding regions that may underlie transcriptional regulation functions [2], it has been concluded that the quantity of RNA transcripts may play a more important role in inheritable phenotypic variation [3]. Therefore, survey of global gene expression level could help understand phenotypes or questions of interest. For example, comparison of gene expression profiles between two groups of individuals that possess distinct phenotypes could pinpoint the key genes that lead to the phenotypic variation.

Chickens are one of the most important livestock and experimental model organisms. Since its domestication 8000 years ago, the chicken has become the most widespread domestic animal [4]. It provides high-quality animal protein in the forms of egg and meat for human consumption. The poultry industry has been undergoing a radical growth [5], and will continue to be a major player to meet the challenges imposed by the growing world population. On the other hand, the chicken is extensively used as a model organism for research in phylogenetics, immunology, embryology, toxicology, human diseases etc. [6–8]. As a result, the benefits of research

in chickens could be three-fold: improving chicken production to feed the growing population, gaining knowledge in biology, and shedding light on human medicine.

Publication of chicken draft genome sequence marked a new era of genomics research in chicken [9]. The reference genome sequence greatly facilitates genome-wide studies in chickens, such as comparative genomics, comparative transcriptomics, genome-wide association studies. Furthermore, with the rapid development of sequencing technologies and bioinformatics tools, next generation sequencing offers an affordable and powerful tool for genomic research.

However, some fundamental genomics questions remain enigma in chickens. One of them is whether genomic imprinting exists in chickens. Genomic imprinting refers to an epigenetic phenomenon that some autosomal genes are expressed monoallelically from either the paternal allele or maternal allele (Figure 1.1). The existence or absence of genomic imprinting in chickens is critical to understand the evolutionary origins of genomic imprinting. In addition, due to its peculiarity of monoallelic expression with parent-of-origin effect, genomic imprinting could be potentially exploited and utilized in animal breeding programs. A widely-accepted hypothesis for genomic imprinting known as “parental conflict” theory maintains that parental genes would pursue their own interest during embryonic development [10,11]. As a result, paternally derived genes promote embryonic growth, while maternally derived genes conserve maternal resources. Based on this hypothesis, genomic imprinting should exist in placental animals but not in oviparous animals. The resource for embryonic development is predetermined, thus, there is no “parental conflict” in the egg-laying animals. Moreover, the contradicting evidence regarding the existence or absence of genomic imprinting in chickens added more mysteries to

the topic. RNA-Seq for global allelic expression provides a new way to survey genomic imprinting at transcriptome level. Using RNA-Seq, it has been reported genomic imprinting is absent in Day 4.5 embryo [12] and in the brain tissue of day-old chicken [13], but may exist in hypothalamus, liver and breast muscle of Day 56 chicken [14]. Those results suggested genomic imprinting may subject to temporal and tissue-specific regulation in chickens. Thus, additional examination of different developmental stages and tissue types is crucial to reach a conclusion about the existence or absence of genomic imprinting in chickens. In Chapter 2, we investigate the imprinting status of genes in chicken liver and brain at embryonic day 12 by examining allele specific expression (ASE) in F1 reciprocal crosses derived from two highly inbred and genetically distant chicken lines - Fayoumi and Leghorn. We also examined allelic imbalance at chromosome-wide and genome-wide levels and checked dosage compensation of sex chromosome in chickens. The manuscript for this work, titled "RNA-Seq Analyses Identify Frequent Allele Specific Expression and No Evidence of Genomic Imprinting in Specific Embryonic Tissues of Chicken," was published in *Scientific Reports* [15].

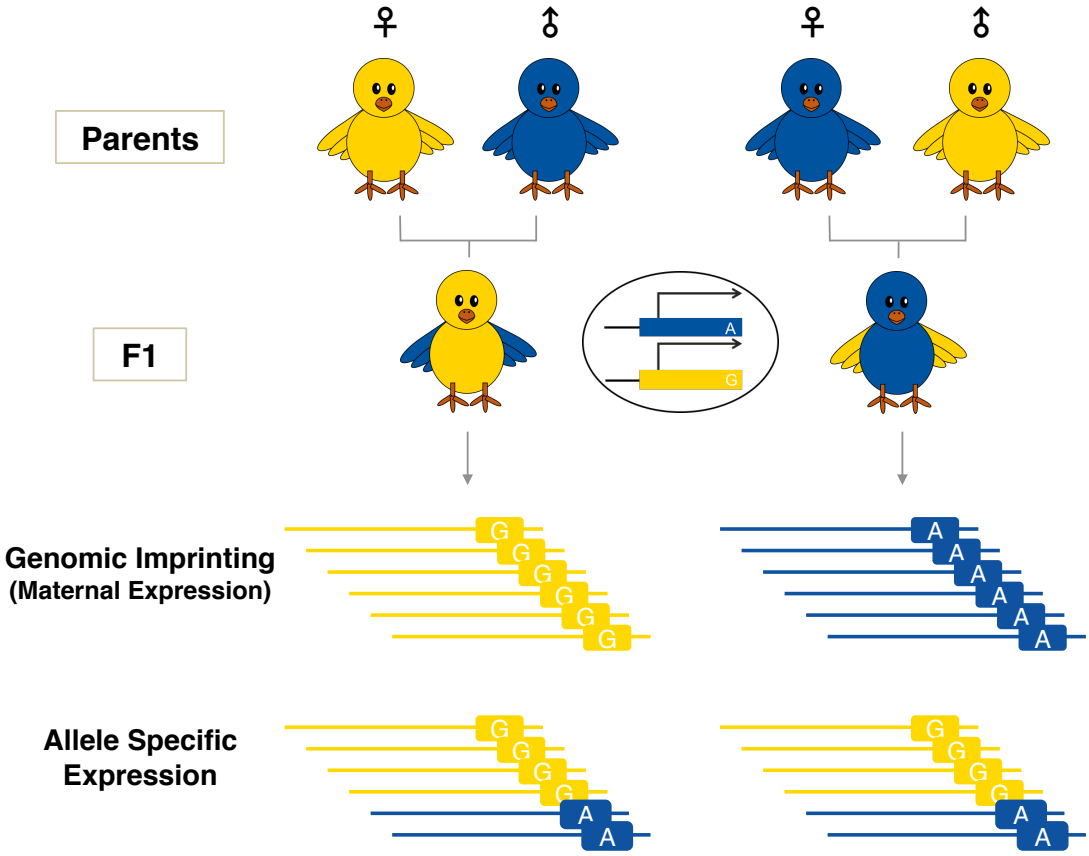

Figure 1.1 Genomic imprinting and allele specific expression illustrated in reciprocal cross design.

Although tremendous success in chicken performance has been achieved through selection and breeding, the underlying genetic basis and molecular mechanisms are not fully understood. In the long run, this may hinder the improvement of poultry production and even potentially cause adverse effects. Heterosis, or hybrid vigor, is one of the key factors that propelled the improvement of chicken performance. It refers to the phenomenon that the hybrid offspring have

superior performance than the parents. Transcriptiomics is gaining importance in comprehending the molecular basis of heterosis, as how gene expression levels are inherited from generation to generation is important to understand heterosis and the superior phenotype in hybrids. Thus, we compared the gene expression profiles between F1 reciprocal crosses and inbred parental lines in Day 12 embryonic brain and liver, and identified the predominant transgenerational gene expression pattern. We also identified the differentially expressed (DE) genes between F1 crosses and parental lines to pinpoint the genes that may contribute to heterotic phenotype. This work is elaborated in Chapter 3.

Feed efficiency (FE) is an important performance index and selection criterion for modern commercial broiler chickens. It describes the efficiency of converting feed intake to body weight gain. As feed cost represents up to 70% of the total economic input for raising broiler chickens, high FE (HFE) chickens are more preferable than low FE (LFE) chickens in the poultry industry. A comprehensive understanding of the biological mechanisms controlling FE is crucial for developing optimal breeding and selection strategies. Adipose tissue is a metabolically active endocrine organ and plays a central role in energy homeostasis. Chicken adipose tissue has some unique features, such as limited *de novo* lipid synthesis capacity [16] and insensitivity to insulin [17,18], which makes it similar to the adipose tissue of obese people and type 2 diabetes patients. In Chapter 4, we studied gene expression in chicken adipose tissue associated with FE in HFE and LFE broiler chickens and identified the key genes that contributed to the divergent fat deposition between HFE and LFE chickens (Figure 1.2). This work was published in *PLoS ONE* with the title "RNA-Seq Analysis of

Abdominal Fat Reveals Differences between Modern Commercial Broiler Chickens with High and Low Feed Efficiencies" [19].

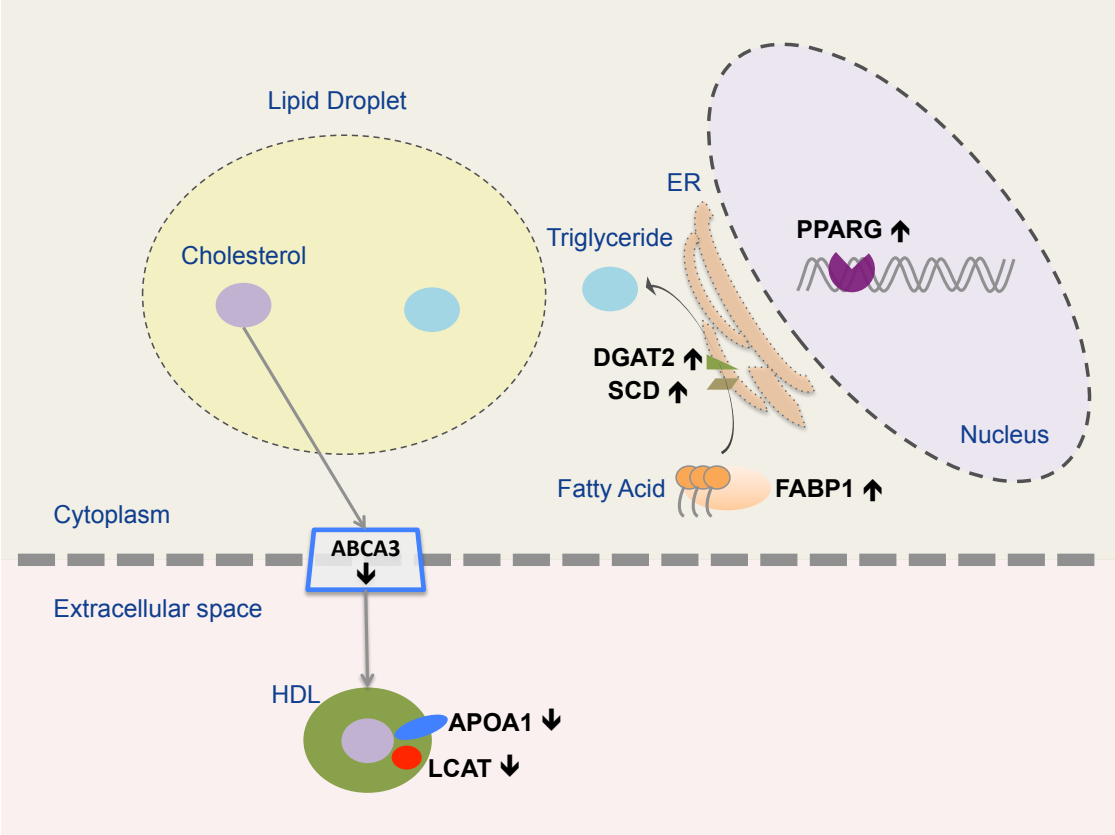

Figure 1.2 The key genes contribute to distinct fat deposition between high feed efficiency and low feed efficiency broiler chickens.

REFERENCES

1. Hubner, N. *et al.* Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nat. Genet. N. Y.* **37,** 243–53 (2005).

2. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337,** 1190–1195 (2012).

3. Buckland, P. R. Allele-specific gene expression differences in humans. *Hum. Mol. Genet.* **13 Spec No,** R255-60 (2004).

4. Pitt, J., Gillingham, P. K., Maltby, M. & Stewart, J. R. New perspectives on the ecology of early domestic fowl: An interdisciplinary approach. *J. Archaeol. Sci.* **74,** 1–10 (2016).

5. Thornton, P. K. Livestock production: recent trends, future prospects. *Philos. Trans. R. Soc. B Biol. Sci.* **365,** 2853–2867 (2010).

6. Lamont, S. J. Perspectives in Chicken Genetics and Genomics. *Poult. Sci.* **85,** 2048–2049 (2006).

7. Burt, D. W. Emergence of the Chicken as a Model Organism: Implications for Agriculture and Biology. *Poult Sci* **86,** 1460–1471 (2007).

8. Bahr, J. M. The Chicken as a Model Organism. in *Sourcebook of Models for Biomedical Research* (ed. Conn, P. M.) 161–167 (Humana Press, 2008).

9. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432,** 695–716 (2004).

10. Haig, D. & Westoby, M. Parent-Specific Gene Expression and the Triploid Endosperm. *Am. Nat.* **134,** 147–147 (1989).

11. Moore, T. & Haig, D. Genomic imprinting in mammalian development: a parental tug-of-war. *Trends Genet. TIG* **7,** 45–9 (1991).

12. Frésard, L. *et al.* Transcriptome-wide investigation of genomic imprinting in chicken. *Nucleic Acids Res.* 1–15 (2014). doi:10.1093/nar/gkt1390

13. Wang, Q. *et al.* Next-Generation Sequencing Techniques Reveal that Genomic Imprinting Is Absent in Day-Old Gallus gallus domesticus Brains. *PLoS ONE* **10,** e0132345 (2015).

14. Pinto, F. L. *et al.* Whole Transcriptome Sequencing in Reciprocal Crosses Suggests Parent-of-Origin Effects on Gene Expression in the Chicken Genome. in *Proceedings, 10th World Congress of Genetics Applied to Livestock Production* (ASAS, 2014).

15. Zhuo, Z., Lamont, S. J. & Abasht, B. RNA-Seq Analyses Identify Frequent Allele Specific Expression and No Evidence of Genomic Imprinting in Specific Embryonic Tissues of Chicken. *Sci. Rep.* **7,** 11944 (2017).

16. O'Hea, E. K. & Leveille, G. A. Lipogenesis in isolated adipose tissue of the domestic chick (Gallus domesticus). *Comp. Biochem. Physiol.* **26,** 111–20 (1968).

17. Simon, J. Insulin in birds: metabolic effects and possible implications in genetically fat and lean chickens. in *Leanness in Domestic Birds* (eds. Leclerq, B. & Whitehead, C. C.) 253–268 (Butterworth-Heinemann, 1988).

18. Simon, J. Chicken as a useful species for the comprehension of insulin action. *Crit. Rev. Poult. Biol. USA* **2,** 121–148 (1989).

19. Zhuo, Z., Lamont, S. J., Lee, W. L. & Abasht, B. RNA-Seq Analysis of Abdominal Fat Reveals Differences between Modern Commercial Broiler Chickens with High and Low Feed Efficiencies. *PLoS ONE* (2015).

Chapter 2

RNA-SEQ ANALYSES IDENTIFY FREQUENT ALLELE SPECIFIC EXPRESSION AND NO EVIDENCE OF GENOMIC IMPRINTING IN SPECIFIC EMBRYONIC TISSUES OF CHICKEN

(Zhuo, Z., Lamont, S. J. & Abasht, B. RNA-Seq Analyses Identify Frequent Allele Specific Expression and No Evidence of Genomic Imprinting in Specific Embryonic Tissues of Chicken. *Sci. Rep.* **7,** 11944 (2017))

2.1 Abstract

Epigenetic and genetic cis-regulatory elements in diploid organisms may cause allele specific expression (ASE) – unequal expression of the two autosomal gene copies. Genomic imprinting is an intriguing type of ASE in which some genes are expressed monoallelically from either the paternal allele or maternal allele as a result of epigenetic modifications. Imprinted genes have been identified in several animal species and are frequently associated with embryonic development and growth. Yet whether genomic imprinting exists in chickens remains debatable, as previous studies have reported conflicting evidence. Albeit no genomic imprinting has been reported in the chicken embryo as a whole, we interrogated the existence or absence of genomic imprinting in 12-day chicken embryonic brain and liver by examining ASE in F1 reciprocal crosses of two highly inbred chicken lines (Fayoumi and Leghorn). We identified 5197 and 4638 ASE SNPs, corresponding to 18.3% and 17.3% of the genes with a detectable expression in the embryonic brain and liver, respectively. But, there was no evidence detected of genomic imprinting in chicken 12-day embryonic brain

and liver. Additionally, while ruling out the possibility of imprinted Z-chromosome inactivation, our results indicated that Z-linked gene expression is partially compensated between sexes in chickens.

2.2 Introduction

Expression difference at the allelic level, termed allele specific expression (ASE), is caused by *cis*-regulatory elements, such as *cis*-acting DNA sequence variants, epigenetic marks, and post-transcriptional modifications. Previous studies identified ASE as a frequent event in animals. More than 50% and 30% of genes show ASE patterns in humans and mice, respectively [1–6]. In agricultural animals, ASE genes have been found related to economically important traits [7–9]. Particularly, ASE SNPs have been observed in response to Marek's disease (MD) virus in chickens [10,11], and selection using those ASE SNPs reduced MD incidence after one generation of selection [12]. Thus, finding ASE is critical for establishing the connection between genotype and phenotype and for application in agriculture.

A unique and intriguing type of ASE is caused by genomic imprinting, in which autosomal genes are monoallelically expressed from either the paternal or maternal allele. Divergent statuses of epigenetic modification between the parental alleles of imprinted genes cause the gene expression to be turned "on" or "off" according to their parental origin. Although the majority of genes are expressed from both parental alleles, a small portion of genes has been classified as imprinted in mammals and these genes are often associated with embryonic growth and development. In agricultural animals, imprinted genes have been identified in cows, pigs and sheep (for references such as [13–15] and reviewed by [16]); however, the evidence regarding

existence of genomic imprinting in chickens seems conflicting. It has been reported that *IGF2* is imprinted in some chicken embryos and the expressed allele can be of either paternal or maternal origin [17]. But several other studies maintained that *IGF2* is biallelically expressed [18–20]. Additionally, the chicken orthologs of imprinted genes in mammals, such as *INS*, *ASCL2/CASH4*, *UBE3A*, *Dlk1*, *GATM*, and *M6P/IGF2R,* were found biallelically expressed in chickens [19–23]. Recently, several genome-wide investigations of genomic imprinting in chickens using RNA-Seq have been conducted. Fresard et al. reported that genomic imprinting is absent in the Day 4.5 chicken embryo [24]. Wang et al. focused on studying brain from the day-old chicken after hatch and also didn't identify evidence of genomic imprinting [25]. In contrast, Pinto et al. recently reported finding thousands of SNPs with parent-of-origin effect in chicken hypothalamus, liver and breast muscle at 56 days of age [26]. Collectively, even though most studies indicated the absence of genomic imprinting in chickens, additional critical examination of different tissue types and developmental stages is necessary to make a conclusive argument.

Here, we performed a genome-wide investigation of ASE using chicken liver and brain samples at embryonic day 12 from a reciprocal cross system. Our main objective was to survey genomic imprinting to answer the question of whether genomic imprinting exists in chickens. We also examined gene expression on the Z chromosome to study dosage compensation in chickens.

2.3 Material and Methods

Sample collection: We utilized two highly inbred experimental chicken lines, Leghorn (*Ghs13*) and Fayoumi (*M5.1*), that are maintained at Iowa State University. Eggs from Leghorn × Fayoumi cross (LF) and from Fayoumi × Leghorn cross (FL), as

well as Fayoumi and Leghorn lines, were collected and kept in an egg cooler at Iowa State University until air-shipped to the University of Delaware. For clarity, F1 cross egg samples were named as paternal origin followed by maternal origin. The eggs were incubated in an egg incubator at 100 °F, 70% humidity. At day 10, the fertility of each egg was checked by candling, and non-fertile eggs were properly disposed. At day 12, brain and liver and a few other tissues were harvested from the embryos and the tissue samples were immediately frozen in liquid nitrogen and stored in a -80 °C freezer. For the purposes of PCR-based sexing and DNA-Seq, the remainder of each embryo was also frozen and preserved. All animal protocols for production of the fertile eggs were conducted with the approval of Iowa State University IACUC Log #4-03-5425-G. No approval of University of Delaware AACUC was required for chicken embryo experiments.

PCR-based sexing: Genomic DNA was isolated from ~25mg of each embryo using a DNeasy Blood & Tissue Kit (Qiagen). A PCR-based method adapted from Clinton et al. was used to determine the sex of each embryo [27]. Two pair of primers were used to amplify a DNA fragment on the W chromosome of female chickens, *XhoI*, and a DNA fragment on ribosome 18S RNA gene as positive control, separately. The sex of embryos was determined based on the gel electrophoresis of PCR products (Figure 2.1 and Table 2.1).

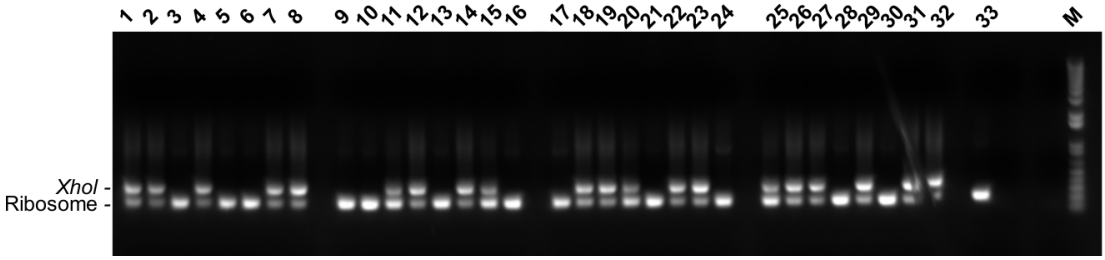

Figure 2.1 PCR-based sexing

Table 2.1 Results of PCR-based sexing

	Dam No.	Sample No.	line / cross	Sex	Note
1	1379	126	F×L	F	
2	1380	127	F×L	F	
3	1378	118	F×L	M	
4	1379	124	F×L	F	
5	1379	123	F×L	M	Included in RNA-Seq
6	1367	106	L×F	M	Included in RNA-Seq
7	1769	142	L	F	
8	1379	122	F×L	F	
9	1367	105	L×F	M	Included in RNA-Seq
10	1769	141	L	M	
11	1784	152	F	F	
12	1379	125	F×L	F	
13	1367	104	L×F	M	Included in RNA-Seq
14	1769	140	L	F	
15	1784	151	F	F	
16	1379	121	F×L	M	Included in RNA-Seq
17	1784	150	F	M	Included in RNA-Seq
18	1774	146	L	F	
19	1380	131	F×L	F	
20	1367	103	L×F	F	

21	1784	149	F	M	Included in RNA-Seq
22	1774	145	L	F	
23	1380	130	F×L	F	
24	1367	102	L×F	M	Included in RNA-Seq
25	1790	153	F	F	
26	1774	144	L	F	
27	1380	128	F×L	F	
28	1766	136	L	M	Included in RNA-Seq
29	1372	113	F×L	F	
30	1378	115	F×L	M	Included in RNA-Seq
31	1768	137	L	F	Included in RNA-Seq
32	1768	138	L	F	
33	1378	117	F×L	M	Included in RNA-Seq

DNA-Seq: The genomic DNA from inbred Fayoumi and Leghorn chickens was sequenced to serve two purposes: first, to identify the DNA polymorphisms that could discriminate the parental origin of alleles in the F1 reciprocal crosses, and second, to create a customized reference genome with parental SNPs masked to reduce reference bias. Two pooled samples were generated by mixing equal molar amounts of genomic DNA from 14 Fayoumi embryos (7 females and 7 males) and 14 Leghorn embryos (7 females and 7 males), separately, and sent to the genomics core facility at Michigan State University for library preparation and sequencing. TruSeq DNA Library Preparation Kit LT (Illumina) was used and the resultant libraries were sequenced on 2 lanes on llumina HiSeq 2500 system with Rapid Run flow cell (v1) using a 150-cycle paired-end sequencing protocol.

The quality of the sequencing data was examined using FastQC v0.11.2 [28]. The reads were mapped to the chicken reference genome Galgal 4 (Ensembl) using the default setting of BWA mem v0.7.12 [29,30]. Unplaced scaffolds of the reference

genome sequence were discarded before read alignment. The SNPs of Fayoumi and Leghorn genomes were identified following the best practice of Genome Analysis ToolKit (GATK) v3.3 with default parameters [31–33]. Briefly, the alignment bam file from each lane was sorted according to chromosome coordinates, and the duplicated reads were marked for exclusion later in the variant-calling step. Realignment was performed around indels, and then the base quality was recalibrated to reduce errors produced by the sequencing machine. The bam files were merged for each sample and processed again through marking duplication, realignment and recalibration. The variants from Fayoumi and Leghorn were identified separately using HaplotypeCaller in GVCF mode and outputted jointly using GenotypeGVCFs. The SNPs were extracted, and an additional criterion of read depth (DP) greater than 10 was applied using custom python scripts. To account for sequencing error, a heterozygous genotype call was reassigned as homozygous when the allele count was less than DP * 1%. A customized reference genome with Fayoumi and Leghorn SNPs replaced by "N" was generated using GATK FastaAlternateReferenceMaker for later RNA-Seq analysis. A list of loci in which both Fayoumi and Leghorn are homozygous but carrying different alleles was generated. At those loci, the genotypes of F1 crosses were predicted to be heterozygous.

RNA-seq: We chose 12 male embryonic samples (2 Fayoumi, 2 Leghorn, 4 FL and 4 LF) for RNA sequencing. Total RNA from brain and liver tissues was extracted using mirVana miRNA Isolation Kit (Thermo Fisher Scientific). The quality of RNA samples was assessed using a Bioanalyzer 2100 (Agilent Technologies), and the RNA integrity numbers (RINs) for all samples were greater than 9.8. Stranded cDNA libraries were prepared using a TruSeq Stranded Total RNA Library Prep Kit (Illumina,

Inc.). During library preparation, cDNA fragments from each sample were ligated with a unique index adapter for further discrimination. The cDNA libraries were validated by Bioanalyzer 2100 analysis (Agilent) and then normalized to 10nM using Tris buffer (Tris-Cl 10mM, 0.1% Tween 20, pH 8.5). Ten ul from each library were pooled into a single sample and sequenced on an Illumina HiSeq 2000 system using a 75-cycle paired-end sequencing protocol.

The RNA-seq data from each sample was demultiplexed utilizing the unique index adapters. A quality check of the RNA-Seq data was performed using FastQC (v0.11.2) [28]. The reads were mapped to the customized chicken reference genome using the STAR (v2.4.1c) 2-pass method with default parameters [34]. Variants relative to the chicken reference genome were discovered using GATK [31–33]. The workflow for variant calling from RNA-Seq data is generally similar to that for DNA-Seq as previously described, except that reads were split to remove "N"s between exon segments and variant calling was conducted using HaplotyperCaller in variant discovery mode with all samples. Low quality calls (QD < 2), or variants with strong strand bias (FS > 30) and SNP clusters (3 SNPs in 35 bp window) were excluded from further analysis. The resultant VCF files were further partitioned to obtain SNPs for each individual sample. Meanwhile, additional filters, including read depth (DP < 10) and genotype quality (GQ < 30), were imposed to obtain high confidence "genotypes". An allelic read count less than DP*1% was considered as a sequencing error and, therefore, was reassigned as 0.

Detection of ASE and genomic imprinting: We performed both within samples analysis and meta-analysis across multiple samples to detect ASE. The loci where the predicted genotypes for F1 crosses (based on the genotypes of parental lines) were

heterozygous and the SNPs were identified in RNA-Seq were designated as testable. Monoallelic expression at testable loci was considered as ASE. If the expression was biallelical, a binomial test was performed on the allele read counts in F1 crosses to test whether the allelic expression is deviated from equal expression. P-values were adjusted using Benjamini-Hochberg method in R, and an adjusted p-value cutoff of 0.1 was applied to claim statistical significance of the ASE SNPs for each sample. The allele with higher expression is referred to as the preferred allele. Additionally, a meta-analysis on samples with the same preferred allele was carried out using Fisher's combining probability test to assess the ASE across the biological replicates [35]. Tissue-specific ASE was evaluated between brain and liver. The genes identified as ASE in one tissue type but not the other were defined as tissue specific ASE. Of note, when a SNP (ASE or not) was detected using RNA-seq data in two or more samples of the same tissue, the SNP-containing gene was declared as being expressed in that type of tissue. SnpEff [36] was used to annotate the ASE SNPs and to identify the ASE genes.

For the ASE SNPs identified from within sample analysis, a parental gene expression ratio (read count of paternal allele or maternal allele / total read count) was calculated. Genomic imprinting was evaluated based on whether the parental origin of the preferred allele is the same in reciprocal crosses. A putative imprinted gene was identified if an ASE SNP was detected for the gene in at least two samples in each reciprocal cross, and the parental origin of the preferred allele was the same in all samples with the detected ASE SNP.

Dosage compensation analysis: To study dosage compensation between autosomes and sex chromosomes, the gene expression levels in Fragments Per

Kilobase of transcript per Million mapped reads (FPKM) were analyzed for each sample using Cufflinks (v2.1.1) [37] after STAR (v2.4.1c) alignment. The FPKM values greater than 0.1 were Log2 transformed and the mean FPKM of Z chromosome and autosomes were calculated for each sample to obtain the difference of gene expression between Z and A by subtracting, i.e., $\Delta_{Z\text{-}A}$ = Mean Log2(FPKM)$_{Z\_genes}$ – Mean Log2(FPKM)$_{A\_genes}$ [38]. Then, the mean expression difference between Z and A for each tissue type was obtained by averaging $\Delta_{Z\text{-}A}$ values of all samples from the same tissue type. Equal expression between Z chromosome and autosomes would result in the value 0.

Verification of ASE genes: Four ASE genes in brain and 6 ASE genes in liver were selected and further verified using Sanger sequencing. The genes were chosen to represent different average expression ratios, depths of RNA-Seq reads, and locations of ASE SNPs. One sample from each F1 cross was randomly selected to perform the verification experiments. PCR primers were designed using Primer 3 [39] to amplify the sequences containing target ASE SNPs and for Sanger Sequencing. Because the genotypes of F1 crosses were established according to the DNA-Seq data from pooled genomic DNA, we sequenced genomic DNA at target SNP loci to confirm the predicted genotypes. Due to a deletion located upstream of the target SNP for an unnamed gene (ENSGALG00000027334), the sequencing trace showed double peaks at the target locus. So, the genotype of this unnamed gene was verified by amplification-refractory mutation system (ARMS) PCR [40] (data not shown). Total RNA was treated with DNA-free DNA remove kit (Thermo Fisher Scientific) to eliminate genomic DNA. First-strand cDNA was synthesized using SuperScript IV First-Strand Synthesis System and Oligo (dT) primers (ThermoFisher Scientific) from the same

RNA samples used for RNA-Seq library preparation. PCR of genomic DNA and cDNA was carried out using AccuPrime *Pfx* SuperMix (Thermo Fisher Scientific). The expression ratio between the two alleles were estimated using Minor Variant Finder (Thermo Fisher Scientific) and ImageJ [41] based on the sequencing traces of cDNA from the F1 samples. Genomic DNA of a Fayoumi sample was also sequenced at the targeted regions to provide a homozygous control to facilitate detection of variants when using Minor Variant Finder. The chromatograms of Sanger Sequencing were viewed using SnapGene software (GSL Biotech).

Data Availability: The datasets generated during and/or analysed during the current study are available in the NCBI Sequence Read Archive (Accession number SRP102082). The ASE SNPs identified in this study are included in the Supplemental datasets.

2.4 Results and Discussion

DNA-Seq of inbred parental lines

A total of 99 GB high-quality sequencing data, consisting of 155.8 million and 137.4 million 150b sequence reads were generated by sequencing DNA pools from Fayoumi and Leghorn samples, respectively. The estimated mean sequencing coverage was 21.79 × for Fayoumi and 19.22 × for Leghorn (Table 2.2). Around 92% of the sequence reads were properly paired and mapped to the chicken reference genome (Ensembl Galgal 4.0). We identified 5,072,830 and 4,632,414 variants - of which 4,442,390 and 4,055,489 were single nucleotide variations - from Fayoumi and Leghorn data, respectively. At 93.6% and 95.8% of the detected SNP loci, Fayoumi and Leghorn lines were homozygous, respectively. Corroborated with a previously reported DNA re-sequencing study using the same chicken lines [42], the results

suggested that the parental chicken lines used in the present study are highly homozygous across the genome. Importantly, at 3,071,441 loci, the two inbred parental lines are homozygous for different alleles, which generate unambiguous heterozygous genotypes in F1 offspring and makes it possible to distinguish the parental origin of the alleles. Those loci were further verified to identify the corresponding genes, which included up to 83.2% of the annotated genes (Galgal4.78).

Table 2.2 Statistics of DNA-Seq data

Sample	Coverage	Mapped	Single Nucleotide Variants *		
			Total	Homozygous	Heterozygous
Fayoumi	21.8 X	98.59%	4,442,390	4,158,536	283,854
Leghorn	19.2 X	98.46%	4,055,489	3,886,221	169,268

* Variants relative to the chicken reference genome

RNA-Seq analysis with customized genome

Sequence read alignments to a reference genome are prone to a bias, because reads carrying a reference allele have a slightly better chance of being mapped to the reference genome than the reads carrying an alternative allele. To minimize this so-called “reference bias” in our RNA-seq reads alignment, a customized reference genome was generated by masking the SNP loci from Fayoumi and Leghorn DNA-seq data. A total of 1.55 billion reads were obtained from 305 GB of RNA-Seq data, averaging 65 million reads per sample. The reads were mapped to

the chicken reference genome Galgal 4 (Ensembl, original reference genome) as well as the customized reference genome. On average, 85% of the reads were mapped, with mapping rates from the customized reference genome being slightly higher than that of the original reference genome (Table 2.3).

Table 2.3 Statistics of RNA-Seq data

Sample	Total Reads	Mapping Rate	
		Ori. Ref.*	Cus. Ref.**
FF_BR_1784_149	55,287,524	84.23%	84.38%
FF_BR_1784_150	58,368,384	85.09%	85.29%
FF_LV_1784_149	65,150,556	87.06%	87.10%
FF_LV_1784_150	54,968,474	88.28%	88.32%
FL_BR_1378_115	55,494,754	80.92%	81.08%
FL_BR_1378_117	69,508,092	79.15%	79.37%
FL_BR_1379_121	71,864,328	80.17%	80.42%
FL_BR_1379_123	59,828,718	80.44%	80.70%
FL_LV_1378_115	71,221,304	88.50%	88.55%
FL_LV_1378_117	58,659,416	86.05%	86.09%
FL_LV_1379_121	70,418,648	86.84%	86.88%
FL_LV_1379_123	57,658,364	82.98%	83.01%
LF_BR_1367_102	70,756,508	85.92%	86.18%
LF_BR_1367_104	71,414,926	80.80%	81.00%
LF_BR_1367_105	71,053,346	85.27%	85.48%
LF_BR_1367_106	57,530,806	79.93%	80.11%
LF_LV_1367_102	67,539,336	83.36%	83.39%
LF_LV_1367_104	68,672,610	86.14%	86.17%
LF_LV_1367_105	72,362,228	85.76%	85.82%
LF_LV_1367_106	50,747,202	85.89%	85.92%
LL_BR_1766_136	74,012,630	87.33%	87.46%
LL_BR_1769_141	66,180,806	83.25%	83.44%
LL_LV_1766_136	72,089,856	86.23%	86.27%
LL_LV_1769_141	64,525,424	88.70%	88.73%

*Ori.Ref. results from original chicken reference genome
**Cus.Ref.: Results from customized reference genome

With the customized reference genome, the average reference ratio in the F1 cross samples was reduced from 50.98% to 49.66%. Thus, a new bias towards alternative alleles was generated after masking the reference genome. To achieve a more accurate estimation of allelic expression, Wang et al. proposed to average the read counts obtained using both the original and the customized reference genomes [43]. However, in our case, the data had greater bias towards the reference allele with the original reference genome than the bias towards the alternative allele with the customized reference genome. Consequently, the reference allele ratio after averaging read counts was 50.78%, which was more biased than that from the customized reference genome. Therefore, we conducted further analyses based on the results from the customized reference genome. It is evident that the results from the customized reference genome had an improved distribution of the reference allele ratio, compared with the right-shifted distribution with the original reference genome (Figure 2.2 a). The allelic read counts for reference allele and alternative allele at heterozygous loci were more evenly dispersed along the diagonal with the customized genome (Figure 2.2 b). In conclusion, using the customized reference genome with SNPs masked instead of the original reference genome is an effective method to reduce reference bias.

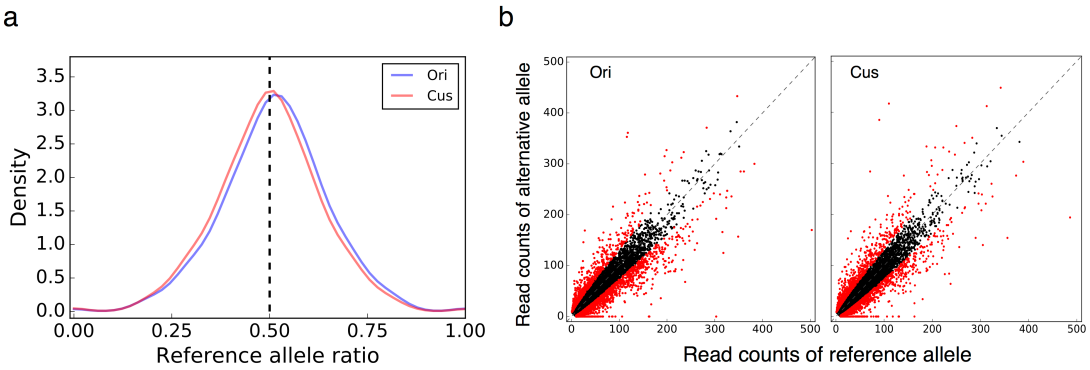

Figure 2.2 Using a customized reference genome reduced reference bias. a. The distribution of the reference allele ratio with the original reference genome and customized reference genome. b. The scatterplot of read counts of reference allele and alternative allele at SNP positions from results of original reference genome and customized reference genome. ASE SNPs are highlighted in red.

SNPs identified in RNA-Seq and the consistency between predicted genotypes and RNA expression

Considering the chicken lines used in the present study are highly inbred, we expected that the DNA-Seq data from the parental genomic DNA pool of 14 individuals could be used to establish high-fidelity predicted genotypes for the F1 crosses. To confirm, we examined the consistency of genotypes between DNA-Seq- and RNA-Seq-based genotype calls. We made two comparisons. First, we compared the DNA-Seq- and RNA-Seq-based genotype calls for inbred lines. On average, 93,137 SNPs and 55,675 SNPs were identified in Fayoumi brain and liver RNA-Seq data; 109,631 SNPs and 71,889 SNPs were identified in Leghorn brain and liver RNA-Seq data, respectively. Of those SNP loci from RNA-Seq data, about 15% were not identified in DNA-Seq data, but the SNP loci identified in both datasets showed 98.6%

consistency of genotypes. Second, we compared the predicted genotypes of F1 embryos based on the DNA-seq data with their observed genotype calls based on RNA-seq data. The numbers of SNPs identified in FL cross versus LF cross did not show a significant difference. On average, 133,834 SNPs and 79,260 SNPs were identified in the brain and liver tissues, respectively. Likewise, 99.75% of the SNP loci found in both datasets showed consistency of genotypes. Except for technical errors, such as sequencing error or artifacts during data handling that cause false prediction, we expect individual differences to be a major contributor for inconsistent genotype calls (< 2%) between RNA-seq and DNA-seq data. In addition, monoallelic expression and RNA editing could be at least partially responsible for the observed inconsistency. In summary, although the gDNA of the F1 sample itself or its biological parents were not directly sequenced, we were able to establish high-confidence predicted F1 genotypes for further analyses based on the DNA-Seq data from gDNA pools.

ASE in embryonic brain and liver from F1 cross chickens

We then examined ASE in embryonic brain and liver from Day 12 F1 embryos. On average, 51,935 and 30,336 loci were identified as being testable in brain and liver datasets per sample (Table 2.4), corresponding to 89.3% and 87.6% of the expressed genes in brain and liver, respectively. Analysis determined about 1.1% of testable loci in brain and 2.9% of testable loci in liver showing ASE pattern, with a total of 2956 ASE SNPs in brain and 3436 ASE SNPs in liver. More than 99% of ASE observed in two or more samples showed higher expression towards the same alleles (Table 2.5). Particularly, ASE was detected for 52 SNPs in brain and 86 SNPs in liver in

all samples from both crosses and with the same preferred alleles (Figure 2.3; Table 2.5). Therefore, the preferred allele of those ASE was line-of-origin dependent. Moreover, the expression ratios of the preferred alleles (reads of preferred allele / total reads) were highly consistent across all the samples, as indicated by low coefficients of variation values (average CV=3.7%; Table 2.6). Meta-analysis across samples identified 4,658 ASE SNPs in brain and 3,974 ASE SNPs in liver, and those ASE SNPs showed high overlapping with the results from within-sample analysis (Table 2.7). Combining the results from within- and across-sample analyses, we identified 5,181 ASE SNPs in brain and 4,617 ASE SNPs in liver.

Table 2.4 Number of testable loci and ASE SNPs in each sample

Sample	No. of testable loci	No. of ASE SNPs with higher expression of paternal allele	No. of ASE SNPs with higher expression of maternal allele	No. of ASE SNPs with higher expression of reference allele	No. of ASE SNPs with higher expression of alternative allele
FL_BR_1378_115	52651	214	249	246	217
FL_BR_1378_117	56314	295	295	315	275
FL_BR_1379_121	51099	291	314	328	277
FL_BR_1379_123	46742	201	221	234	188
LF_BR_1367_102	45020	333	286	324	295
LF_BR_1367_104	55961	464	318	423	359
LF_BR_1367_105	52587	373	326	385	314
LF_BR_1367_106	55108	255	229	247	237
Average	51935				
FL_LV_1378_115	28247	481	455	461	475
FL_LV_1378_117	25970	446	447	434	459
FL_LV_1379_121	34616	578	541	557	562

FL_LV_1379_123	26845	300	300	303	297
LF_LV_1367_102	33675	544	517	536	525
LF_LV_1367_104	33208	504	411	445	470
LF_LV_1367_105	33263	447	457	459	445
LF_LV_1367_106	26861	338	311	309	340
Average	30336				

Table 2.5 Number of ASE SNPs observed in 1 or more samples and consistency of preferred alleles

	Brain		Liver	
Observed in # of samples	No. of ASE SNPs	No. of ASE SNPs with consistent preferred allele	No. of ASE SNPs	No. of ASE SNPs with consistent preferred allele
1	2267	Not Applicable	2156	Not Applicable
2	326	325	467	460
3	131	131	234	232
4	68	67	169	168
5	44	44	116	116
6	32	32	115	114
7	36	36	93	93
8	52	50*	86	84*

* The two inconsistent SNPs are located at mitochondria DNA, so the preferred allele is always maternal

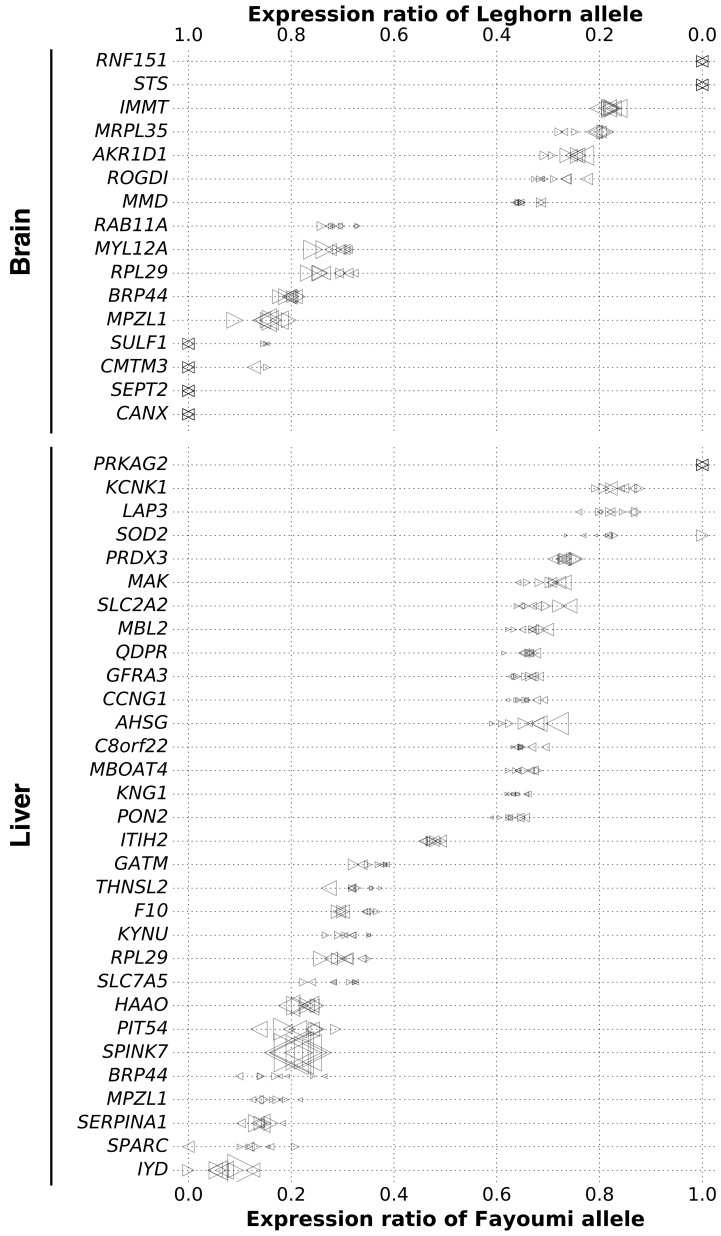

Figure 2.3 Partial list of the genes containing ASE SNPs observed in brain and liver of all samples. The upper X axis is the expression ratio of Leghorn allele and the lower X axis is the expression ratio of Fayoumi allele. Each dot represents a sample (Left triangle: FL cross; Right triangle: LF cross). The size of each marker is proportional to –log10(adjusted p-value).

Table 2.6 CV and mean expression ratio of the preferred allele for the ASE SNPs

Brain

Chromosome	Position	Reference	Average expression ratio of preferred allele	SD	CV
Brain					
1	56721704	C	0.73	0.04	5.0%
1	56731039	G	0.75	0.03	3.6%
1	90871453	A	0.79	0.04	4.8%
1	90871581	T	0.78	0.03	3.4%
1	90871607	C	0.77	0.03	4.6%
1	90871693	T	0.75	0.05	6.3%
1	90871813	C	0.79	0.04	5.1%
1	90871952	G	0.80	0.05	6.5%
1	90872023	A	0.78	0.04	5.4%
1	90872062	A	0.76	0.05	5.9%
1	90872148	T	0.78	0.04	5.6%
1	90872238	C	0.77	0.03	3.4%
1	90872296	T	0.79	0.05	6.4%
1	90872302	A	0.80	0.05	6.0%
1	90872431	T	0.81	0.04	5.2%
1	90872458	T	0.81	0.03	4.3%
1	90872516	C	0.82	0.03	3.9%
1	90872557	A	0.82	0.04	4.6%
1	90872709	T	0.80	0.04	5.1%
1	90872734	A	0.82	0.03	3.9%
1	90872872	G	0.82	0.03	3.3%
1	90873011	G	0.86	0.03	3.4%
1	90873038	T	0.85	0.02	2.7%
1	90873438	G	0.82	0.04	4.9%
1	90873705	C	0.82	0.04	5.1%

1	90873752	A	0.83	0.02	2.6%
1	90875332	G	0.85	0.02	2.1%
1	90875606	A	0.86	0.06	6.9%
1	90891019	G	0.83	0.04	4.6%
1	98653919	G	0.67	0.03	5.1%
1	125763437	C	1.00	0.00	0.0%
2	100545252	C	0.71	0.02	3.4%
2	115993458	G	0.94	0.08	8.3%
2	125910349	A	1.00	0.00	0.0%
3	45157492	T	0.79	0.03	4.2%
4	13273040	A	0.77	0.02	2.4%
4	45797087	A	0.71	0.04	5.8%
4	85062164	A	0.82	0.04	5.0%
4	85062616	A	0.81	0.02	2.7%
4	85063328	G	0.84	0.02	2.5%
4	85064009	G	0.81	0.05	5.7%
4	85064068	C	0.78	0.04	4.7%
9	4497393	T	1.00	0.00	0.0%
10	17662128	A	0.71	0.02	3.5%
11	10904781	T	0.96	0.07	6.8%
12	532343	A	0.72	0.03	4.3%
13	12857392	A	1.00	0.00	0.0%
13	12857409	G	1.00	0.00	0.0%
14	14415138	G	0.71	0.03	4.8%
18	5920301	C	0.65	0.02	3.2%
MT	2207	C	1.00	0.00	0.0%
MT	8058	G	1.00	0.00	0.0%

Liver

1	4110771	T	0.62	0.02	3.7%
1	4120158	C	0.62	0.02	4.0%
1	4121228	C	0.98	0.06	6.0%
1	4126505	T	0.64	0.03	4.1%
1	90872709	T	0.81	0.05	6.5%
1	90872734	A	0.83	0.06	7.0%
1	90891019	G	0.83	0.03	3.6%
1	136923134	C	0.68	0.03	4.7%

1	154785520	C	1.00	0.00	0.0%
1	191988736	C	1.00	0.00	0.0%
2	6378773	A	1.00	0.00	0.0%
2	6380603	C	1.00	0.00	0.0%
2	6380618	G	1.00	0.00	0.0%
2	6380661	A	1.00	0.00	0.0%
2	6380732	G	1.00	0.00	0.0%
2	6380790	G	1.00	0.00	0.0%
2	6381377	C	1.00	0.00	0.0%
2	6382777	T	1.00	0.00	0.0%
2	6382834	A	1.00	0.00	0.0%
2	6382944	C	1.00	0.00	0.0%
2	6383417	T	1.00	0.00	0.0%
2	6383419	C	1.00	0.00	0.0%
2	7817808	C	1.00	0.00	0.0%
2	7817888	C	1.00	0.00	0.0%
2	23986930	A	0.63	0.02	3.6%
2	63336244	A	0.69	0.03	4.5%
2	108191863	C	0.65	0.02	3.1%
3	23462744	T	0.77	0.03	4.1%
3	23462966	C	0.77	0.02	3.0%
3	23462969	T	0.77	0.02	2.9%
3	38008398	T	0.84	0.07	8.0%
3	38008628	T	0.85	0.07	8.1%
3	38009070	A	0.83	0.04	4.9%
3	38009219	T	0.82	0.03	3.6%
3	41123224	T	0.76	0.05	6.4%
3	44930274	T	0.82	0.08	9.5%
3	47729536	T	0.89	0.05	5.9%
3	47730778	A	0.97	0.05	5.2%
4	49039601	G	0.66	0.02	2.8%
4	49039636	G	0.65	0.03	4.4%
4	59600455	G	0.61	0.02	3.8%
4	75553631	G	0.82	0.04	4.5%
4	75582961	A	0.66	0.02	3.0%
4	85518541	T	0.68	0.06	9.0%
4	85519658	A	0.66	0.03	4.7%

5	17675701	G	0.77	0.05	6.1%
5	44822999	T	0.85	0.03	3.2%
5	44823536	G	0.85	0.04	4.4%
5	44824890	C	0.87	0.05	6.2%
5	44825474	T	0.85	0.02	2.9%
6	3981946	T	0.66	0.03	3.9%
6	29560427	C	0.74	0.05	6.5%
6	29560651	C	0.71	0.03	4.2%
6	29560766	A	0.74	0.02	2.1%
6	29560788	A	0.74	0.02	2.0%
6	29560923	T	0.73	0.03	4.2%
6	29560959	T	0.74	0.03	4.1%
7	32197070	T	0.69	0.03	4.1%
9	8999054	C	1.00	0.00	0.0%
9	15612772	C	0.65	0.04	6.5%
9	15659017	A	0.64	0.01	2.3%
9	19180724	G	0.68	0.04	5.3%
10	10413182	C	0.64	0.02	3.8%
11	17464533	T	0.73	0.04	6.0%
11	17465161	G	0.69	0.04	6.4%
12	532343	A	0.70	0.03	4.3%
12	14640323	C	0.63	0.02	3.0%
12	14640486	T	0.69	0.03	4.6%
13	6078755	G	0.67	0.02	3.4%
13	6079944	A	0.63	0.02	3.1%
13	6081530	C	0.67	0.04	5.8%
13	9877048	C	0.79	0.02	3.0%
13	9885038	G	0.78	0.03	3.9%
13	9885062	T	0.78	0.03	3.2%
13	12246793	G	0.88	0.06	6.9%
13	17577272	T	0.66	0.02	3.0%
15	6211426	C	0.69	0.02	2.6%
15	6211517	C	0.73	0.03	4.7%
15	8105432	C	0.61	0.02	3.4%
16	310634	G	0.98	0.06	6.6%
28	794340	T	0.69	0.04	6.4%
LGE64	182232	G	1.00	0.00	0.0%

LGE64	547553	C	0.77	0.06	7.1%
LGE64	548701	C	0.79	0.04	5.0%
MT	2207	C	1.00	0.00	0.0%
MT	8058	G	1.00	0.00	0.0%

Table 2.7 Number of ASE SNPs and genes identified in within-sample and across-sample (meta-analysis) analyses using F1 samples

	Meta analysis (adjust p-value < 0.05)	Within-sample analysis (adjust p-value < 0.1)	Overlapping	Total
Brain				
ASE SNPs	4658	2952	2429	5181
ASE Genes	1998	1523	1279	2242
Liver				
ASE SNPs	3974	3423	2780	4617
ASE Genes	1455	1422	1142	1735

Annotation of ASE SNPs led to the identification of 2,242 ASE genes in brain and 1,735 ASE genes in liver, which represented 18.3% of the expressed genes in Day 12 embryonic brain and 17.3% of the expressed genes in Day 12 embryonic liver. Next, we compared ASE genes in embryonic brain and liver to identify tissue-specific ASE. As a result, there were 832 ASE genes observed in both tissues, yet 1,454 (64.8%) ASE genes in brain and 947 (54.5%) ASE genes in liver were found to be tissue specific (Figure 2.4). Those ASE genes included 334 genes that were only expressed in

brain and 148 genes that were only expressed in liver. Tissue-specific ASE has also been reported previously in human, mice, and cows [44–46]. Taken together, as in mammals, ASE is common in chicken embryonic brain and liver and shows tissue-specificity.

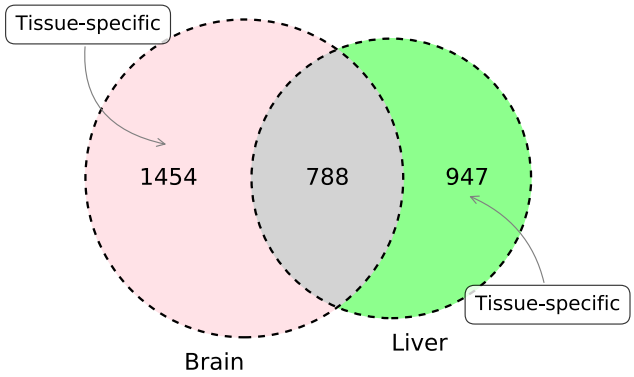

Figure 2.4 Venn diagram of tissue-specific ASE SNPs

Genome imprinting is absent in Day 12 chicken embryonic brain and liver

Based on ASE SNP discovery, we investigated the absence or existence of genomic imprinting in Day 12 chicken embryonic brain and liver. We took advantage of multiple biological replicates in our experiment and applied a set of criteria (see methods) to detect genes subjected to imprinting. As a proof of our approach, we identified SNPs in mitochondrial genes in both brain and liver meeting our criteria, and the expressed alleles of SNPs in mitochondrial genes were all of maternal origin, as expected. As for the nuclear genome, no SNPs in brain or in liver met our criteria.

The debate revolving around whether genomic imprinting exists in chickens has been going on for decades, and it has ignited several efforts to ascertain the answer to the question. RNA-Seq is emerging as an excellent tool for this purpose, as it surveys the global gene expression and therefore is able to examine the phenomenon of genomic imprinting at the genome-wide level. To date, most studies by examining gene expression have resulted in rejection of the existence of genomic imprinting in chickens. A well-known hypothesis of genomic imprinting, the "parent-conflict" theory, does not support the existence of genomic imprinting in egg-laying animals [47,48]. The theory suggests that genes from both parents are competing during embryonic development, with the paternally-derived genes promoting embryonic growth to facilitate the preservation of paternally-derived genes while maternally-derived genes limiting embryonic growth to conserve maternal resources for future offspring and for the fitness of maternally-derived genes. Hence, genomic imprinting should be present mostly in placental but not oviparous animals, as the maternal resource for the offspring is predetermined and the conflict of interest between parental alleles should not occur in oviparous animals. Supporting this hypothesis, the phenomenon of genomic imprinting has been found in eutherians and marsupials but not in monotremes, reptiles, amphibians and fish (summarized by Kaneda [49]). Collectively, our results and the results from Fresard et al. and Wang et al. studies [24,25] suggest that genomic imprinting is missing in chickens, at least at the examined embryonic stages and early post-hatch. However, due to the complexity of gene regulation and the spatiotemporal nature of epigenetic modification, existence of genomic imprinting in other tissue types or developmental stages in chickens cannot be ruled out. Examining the DNA methylome profiles in reciprocal crosses could

provide a new perspective on the conundrum of whether genomic imprinting exists in chickens. It is also important to note that sex difference may be present in ASE and genomic imprinting [50]. Future studies focusing on the female samples would be beneficial to comprehensively understand genomic imprinting in chickens.

Genome-wide and chromosome-wide allelic expression

Next, we queried the data to determine whether there is an imbalance of expression between paternal and maternal alleles at the genome-wide or chromosome-wide levels. We focused on the predicted heterozygous loci in F1 samples and calculated the expression ratios of paternal and maternal alleles (read counts of paternal allele or maternal allele / total read counts) for all samples. The genome-wide distribution of maternal and paternal allelic expression ratios did not show evident skewness (Figure 2.5 a). At the chromosome level, the median values of parental allelic ratio were approximately 0.5 (except for the mitochondrial chromosome) (Figure 2.5 b, Table 2.8). The results suggested that the expression of paternal and maternal alleles is roughly equal at genome-wide and chromosome-wide levels. Of note, in mice and cattle the paternal alleles are more favored in expression [46,51]. In dairy cattle, 54.17% of ASE genes showed preference towards the paternal allele [46]. Our data of within sample analysis suggested the percentages of ASE genes with higher expression of paternal allele ranged from 47.42% to 55.60%, with averages of 51.45% in brain and 50.91% in liver. Cowley et al. hypothesized that genome-wide imbalance between paternal and maternal alleles is prior to classic genomic imprinting in the evolutionary view, because natural selection may intensify the imbalance and develop it into more extreme cases – imprinting [51]. Based on our study, chickens lack genome-wide and chromosome-wide imprinting as well. Taken

together, we demonstrated at multiple levels that genomic imprinting is absent in E12 chicken brain and liver. It is possible that the process of acquiring genomic imprinting did not occur in chickens.

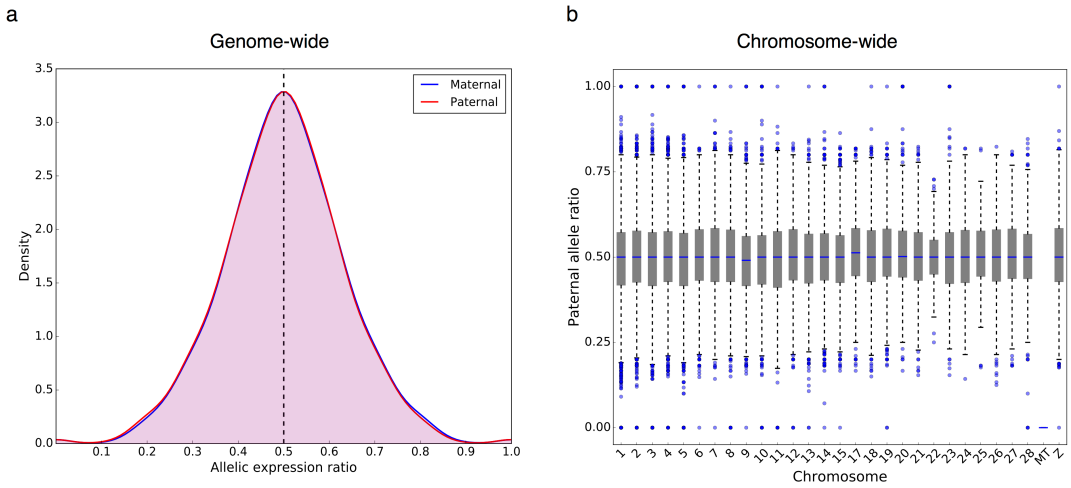

Figure 2.5 Balanced expression at genome-wide and chromosome-wide levels: a. The distribution of paternal and maternal expression ratios at genome-wide level. b. Box plot of paternal allele expression ratio for each chromosome.

Table 2.8 Average median of paternal expression ratio per chromosome

Chromosome	Average median of paternal expression ratio
1	0.50
2	0.50
3	0.50
4	0.50
5	0.50
6	0.50
7	0.50
8	0.50

9	0.50
10	0.50
11	0.50
12	0.50
13	0.50
14	0.50
15	0.50
17	0.50
18	0.50
19	0.50
20	0.50
21	0.50
22	0.50
23	0.50
24	0.50
25	0.50
26	0.50
27	0.50
28	0.50
MT	0.00
Z	0.50

Dosage compensation

Dosage compensation balances gene expression between females and males to compensate for the dosage difference caused by different copy numbers of sex chromosomes. Unlike humans and mice, birds are homogametic for males (ZZ) and heterogametic for females (ZW). Here we compared the mean of log-transformed FPKM values of the expressed genes between Z chromosome and autosomes and between sexes. The average values indicating the difference of gene expression between Z chromosomes and autosomes across all the samples were -0.34 (Z-to-A ratio 0.79) for brain and -0.38 (Z-to-A ratio 0.77) for liver. Additionally, we checked

other available RNA-Seq data in our lab and found the difference of gene expression in breast muscle [52], liver (unpublished) and abdominal fat [53] from Day 47 post-hatch male chickens were -0.31 (Z-to-A ratio 0.81), -0.24 (Z-to-A ratio 0.85) and -0.41 (Z-to-A ratio 0.75), respectively. A recent experiment using both female and male breast muscle samples from our lab (unpublished) showed the expression difference of -0.17 (Z-to-A ratio 0.89) in males and -0.48 (Z-to-A ratio 0.72) in females. Z-linked genes were expressed lower than autosomal genes in all the tissues examined in the current study. Although previous studies of gene expression microarray and protein mass spectrometry showed the Z-to-A ratio close to 1 in males [54,55], RNA-Seq data, including our results, suggested gene expression is not equalized between Z chromosome and autosomes in chickens [55,56]. After between-sample normalization, we found the male-to-female ratio for autosomal genes was 1.00, while the ratio for Z-linked gene expression was 1.35. This result corroborated previous reports that dosage is not fully compensated between females and males and that the Z-linked genes in male chickens express 1.2 to 1.6 times higher than that in female chickens [54,57,58].

The mechanisms of dosage compensation in chickens are not fully understood. As mentioned before, we found that the mean ratio of the paternal expression across all testable loci on the Z chromosome is approximately 0.5 (Figure 2.5 b), indicating an almost equal expression of paternal and maternal alleles on Z chromosome. The results ruled out the possibility of imprinted chromosome inactivation as seen in marsupials, but it is still possible that some genes on paternal and maternal Z chromosomes are subjected to random inactivation. However, recent studies suggested Z chromosome inactivation is absent in chickens based on ASE

analysis and correlation of allelic expressions between parents and crosses [56,59]. In fact, heterogeneity of cells due to possible random inactivation of some genes on paternal and maternal Z chromosomes may complicate examining Z-inactivation if RNA-seq of tissue samples was performed. Future experiments using single cell sequencing might provide a more direct evidence to demystify dosage compensation in chickens.

Verification of ASE Using Sanger Sequencing

To verify the results of ASE analysis from RNA-Seq data, we chose 10 ASE genes and examined the allelic expression using Sanger sequencing in one FL sample and one LF sample (Figure 2.6). The expression of 4 genes was tested in brain samples and 6 genes in liver samples. The Sanger sequencing results of gDNA confirmed our prediction of heterozygous genotypes in both FL and LF samples. The cDNA was sequenced and analyzed to estimate the allelic ratios as described in other studies [60,61]. For eight out of the ten genes, the allelic ratios estimated by Minor Variant Finder agreed with the results from RNA-Seq. *FAM110B* in brain and *PPM1K* in liver showed different preferred alleles and allelic ratios in Sanger sequencing than RNA-Seq. Further examination found that the forward primer for *PPM1K* could only bind to two of the three transcript isoforms of the target region and thus may cause the discrepancy between RNA-Seq and Sanger sequencing. The cause for *FAM110B* discrepancies is unknown, but we speculate the primers for *FAM110B* were only able to amplify one transcript isoform as we only observed one peak in the chromatogram at the target SNP location.

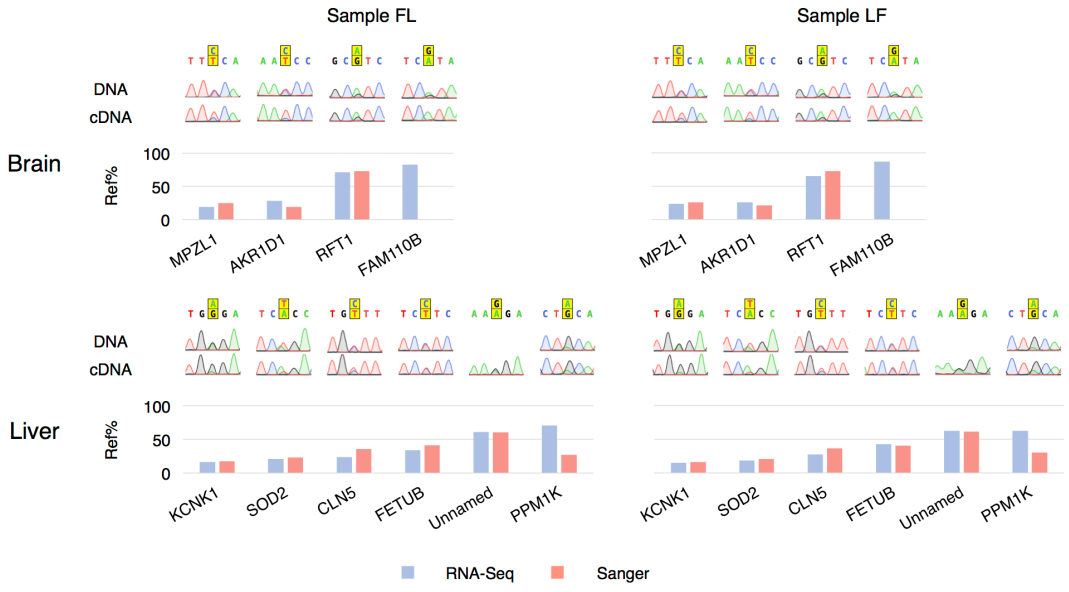

Figure 2.6 Verification of selected ASE genes by Sanger sequencing. The reference allele is displayed on top of the alternative allele at target heterozygous loci, and the barplot shows the expression ratios of the reference allele from RNA-Seq and Sanger Sequencing. Unnamed gene refers to gene ENSGALG00000027334.

2.5 Conclusion

This study set out to investigate whether genomic imprinting exists in chickens by evaluating ASE. Different from previous RNA-Seq studies on genomic imprinting [24,25], our study investigated two tissue types, brain and liver, from Day 12 male chicken embryos, with the aim of scrutinizing the absence or existence of genomic imprinting in chickens. The results showed no evidence for genomic imprinting in Day 12 embryonic brain or liver. Nonetheless, we identified thousands of ASE SNPs and tissue-specific ASE SNPs. We also identified some ASE SNPs with preferred allele showing consistent line-of-origin effect in expression across multiple samples. We observed that the expression of paternal and maternal alleles is

generally balanced at both genome-wide and chromosome-wide levels. Additionally, the gene expression on the Z chromosome is lower than that of autosomes, and the dosage difference of sex chromosome between female and male chickens is not fully compensated by gene expression in chickens. In the future, studies elucidating the relationship between ASE and economically important traits in chickens would be of great importance to fundamental studies as well as for practical application in poultry breeding.

2.6 Acknowledgments

Major funding for this work was provided by the University of Delaware. DNA Sequencing was supported by NRSP-8 Poultry Coordination funding from USDA-NIFA. Sanger Sequencing reported in this project was supported by the Institutional Development Award (IDeA) from the National Institute of Health's National Institute of General Medical Sciences under grant number P20GM103446. The Biomolecular Core Lab at Nemours' Center for Pediatric Research is supported by grants P30GM114736 (COBRE) and P20GM103446 (INBRE). The authors thank Nan Zhou and Marie F. Mutryn for the help with sample collection and cDNA library preparation, and Weixuan Fu for participating in sample collection and discussion on data analyses.

REFERENCES

1. Lo, H. S. *et al.* Allelic variation in gene expression is common in the human genome. *Genome Res.* **13,** 1855–62 (2003).

2. Palacios, R. *et al.* Allele-specific gene expression is widespread across the genome and biological processes. *PloS One* **4,** e4150 (2009).

3. Pant, P. V. K. *et al.* Analysis of allelic differential expression in human white blood cells. *Genome Res.* **16,** 331–9 (2006).

4. Song, M.-Y., Kim, H.-E., Kim, S., Choi, I.-H. & Lee, J.-K. SNP-based large-scale identification of allele-specific gene expression in human B cells. *Gene* **493,** 211–8 (2012).

5. Campbell, C. D., Kirby, A., Nemesh, J., Daly, M. J. & Hirschhorn, J. N. A survey of allelic imbalance in F1 mice. *Genome Res.* **18,** 555–63 (2008).

6. Lagarrigue, S. *et al.* Analysis of allele-specific expression in mouse liver by RNA-Seq: a comparison with Cis-eQTL identified using genetic linkage. *Genetics* **195,** 1157–66 (2013).

7. Olbromski, R., Siadkowska, E., Zelazowska, B. & Zwierzchowski, L. Allelic gene expression imbalance of bovine IGF2, LEP and CCL2 genes in liver, kidney and pituitary. *Mol. Biol. Rep.* **40,** 1189–200 (2013).

8. Muráni, E., Ponsuksili, S., Srikanchai, T., Maak, S. & Wimmers, K. Expression of the porcine adrenergic receptor beta 2 gene in longissimus dorsi muscle is affected by cis-regulatory DNA variation. *Anim. Genet.* **40,** 80–9 (2009).

9. Esteve-Codina, A. *et al.* Exploring the gonad transcriptome of two extreme male pigs with RNA-seq. *BMC Genomics* **12,** 552 (2011).

10. Maceachern, S., Muir, W. M., Crosby, S. D. & Cheng, H. Genome-Wide Identification and Quantification of cis- and trans-Regulated Genes Responding to Marek's Disease Virus Infection via Analysis of Allele-Specific Expression. *Front. Genet.* **2,** (2012).

11. Perumbakkam, S., Muir, W. M., Black-Pyrkosz, A., Okimoto, R. & Cheng, H. H. Comparison and contrast of genes and biological pathways responding to Marek's disease virus infection using allele-specific expression and differential expression in broiler and layer chickens. *BMC Genomics* **14,** 64 (2013).

12. Cheng, H. H. *et al.* Fine mapping of QTL and genomic prediction using allele-specific expression SNPs demonstrates that the complex trait of genetic resistance to Marek's disease is predominantly determined by transcriptional regulation. *BMC Genomics* **16,** 816 (2015).

13. Zaitoun, I. & Khatib, H. Comparative genomic imprinting and expression analysis of six cattle genes. *J. Anim. Sci.* **86,** 25–32 (2008).

14. Li, C. *et al.* Genetic imprinting of H19 and IGF2 in domestic pigs (Sus scrofa). *Anim. Biotechnol.* **19,** 22–27 (2008).

15. Feil, R., Khosla, S., Cappai, P. & Loi, P. Genomic imprinting in ruminants: allele-specific gene expression in parthenogenetic sheep. *Mamm. Genome Off. J. Int. Mamm. Genome Soc.* **9,** 831–834 (1998).

16. Tian, X. (Cindy). Genomic Imprinting in Farm Animals. *Annu. Rev. Anim. Biosci.* **2,** 23–40 (2014).

17. Koski, L. B., Sasaki, E., Roberts, R. D., Gibson, J. & Etches, R. J. Monoalleleic transcription of the insulin-like growth factor-II gene (Igf2) in chick embryos. *Mol. Reprod. Dev.* **56,** 345–52 (2000).

18. O'Neill, M. J., Ingram, R. S., Vrana, P. B. & Tilghman, S. M. Allelic expression of IGF2 in marsupials and birds. *Dev. Genes Evol.* **210,** 18–20 (2000).

19. Nolan, C. M., Killian, J. K., Petitte, J. N. & Jirtle, R. L. Imprint status of M6P/IGF2R and IGF2 in chickens. *Dev. Genes Evol.* **211,** 179–183 (2001).

20. Yokomine, T. *et al.* Sequence polymorphisms, allelic expression status and chromosome locations of the chicken IGF2 and MPR1 genes. *Cytogenet. Cell Genet.* **93,** 109–13 (2001).

21. Colosi, D. C., Martin, D., Moré, K. & Lalande, M. Genomic organization and allelic expression of UBE3A in chicken. *Gene* **383,** 93–8 (2006).

22. Shin, S., Han, J. Y. & Lee, K. Cloning of avian Delta-like 1 homolog gene: the biallelic expression of Delta-like 1 homolog in avian species. *Poult. Sci.* **89,** 948–55 (2010).

23. Jang, H. J. *et al.* Biallelic expression of the L -arginine : glycine amidinotransferase gene with different methylation status between male and female primordial germ cells in chickens. *Poult. Sci.* 760–769 (2013).

24. Frésard, L. *et al.* Transcriptome-wide investigation of genomic imprinting in chicken. *Nucleic Acids Res.* 1–15 (2014). doi:10.1093/nar/gkt1390

25. Wang, Q. *et al.* Next-Generation Sequencing Techniques Reveal that Genomic Imprinting Is Absent in Day-Old Gallus gallus domesticus Brains. *PLoS ONE* **10,** e0132345 (2015).

26. Pinto, F. L. *et al.* Whole Transcriptome Sequencing in Reciprocal Crosses Suggests Parent-of-Origin Effects on Gene Expression in the Chicken Genome. in *Proceedings, 10th World Congress of Genetics Applied to Livestock Production* (ASAS, 2014).

27. Clinton, M., Haines, L., Belloir, B. & McBride, D. Sexing chick embryos: A rapid and simple protocol. *Br. Poult. Sci.* **42,** 134–138 (2001).

28. FastQC [http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc].

29. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* **25,** 1754–60 (2009).

30. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* **26,** 589–95 (2010).

31. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–303 (2010).

32. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43,** 491–8 (2011).

33. Auwera, G. A. *et al.* From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Curr. Protoc. Bioinforma.* 10–11 (2013).

34. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).

35. Edsgärd, D. *et al.* GeneiASE: Detection of condition-dependent and static allele-specific expression from RNA-seq data without haplotype information. *Sci. Rep.* **6,** 21134 (2016).

36. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin)* **6,** 80–92 (2012).

37. Roberts, A., Trapnell, C., Donaghey, J., Rinn, J. L. & Pachter, L. Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol* **12,** R22–R22 (2011).

38. Jue, N. K. *et al.* Determination of dosage compensation of the mammalian X chromosome by RNA-seq is dependent on analytical approach. *BMC Genomics* **14,** 150 (2013).

39. Untergasser, A. *et al.* Primer3--new capabilities and interfaces. *Nucleic Acids Res.* **40,** e115 (2012).

40. Ye, S., Dhillon, S., Ke, X., Collins, A. R. & Day, I. N. M. An efficient procedure for genotyping single nucleotide polymorphisms. *Nucleic Acids Res.* **29,** e88 (2001).

41. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9,** 671–675 (2012).

42. Fleming, D. S., Koltes, J. E., Fritz-Waters, E. R., Reecy, J. M. & Lamont, S. J. SNP Discovery and Genomic Architecture of Highly Inbred Leghorn and Fayoumi Chicken Breeds Using Whole Genome Resequencing. *Anim. Ind. Rep.* **661,** 59 (2015).

43. Wang, X. & Clark, A. G. Using next-generation RNA sequencing to identify imprinted genes. *Heredity* **113,** 156–66 (2014).

44. Zhang, K. *et al.* Digital RNA allelotyping reveals tissue-specific and allele-specific gene expression in human. *Nat. Methods* **6,** 613–8 (2009).

45. Pinter, S. F. *et al.* Allelic Imbalance Is a Prevalent and Tissue-Specific Feature of the Mouse Transcriptome. *Genetics* **200,** 537–549 (2015).

46. Chamberlain, A. J. *et al.* Extensive variation between tissues in allele specific expression in an outbred mammal. *BMC Genomics* **16,** 993 (2015).

47. Haig, D. & Westoby, M. Parent-Specific Gene Expression and the Triploid Endosperm. *Am. Nat.* **134,** 147–147 (1989).

48. Moore, T. & Haig, D. Genomic imprinting in mammalian development: a parental tug-of-war. *Trends Genet. TIG* **7,** 45–9 (1991).

49. Kaneda, M. Genomic imprinting in mammals—Epigenetic parental memories. *Differentiation* **82,** 51–56 (2011).

50. Gregg, C., Zhang, J., Butler, J. E., Haig, D. & Dulac, C. Sex-specific parent-of-origin allelic expression in the mouse brain. *Science* **329,** 682–5 (2010).

51. Crowley, J. J. *et al.* Analyses of allele-specific gene expression in highly divergent mouse crosses identifies pervasive allelic imbalance. *Nat. Genet.* **47,** 353–360 (2015).

52. Zhou, N., Lee, W. R. & Abasht, B. Messenger RNA sequencing and pathway analysis provide novel insights into the biological basis of chickens' feed efficiency. *BMC Genomics* **16,** 195–195 (2015).

53. Zhuo, Z., Lamont, S. J., Lee, W. L. & Abasht, B. RNA-Seq Analysis of Abdominal Fat Reveals Differences between Modern Commercial Broiler Chickens with High and Low Feed Efficiencies. *PLoS ONE* (2015).

54. Itoh, Y. *et al.* Dosage compensation is less effective in birds than in mammals. *J. Biol.* **6,** (2007).

55. Uebbing, S. *et al.* Quantitative Mass Spectrometry Reveals Partial Translational Regulation for Dosage Compensation in Chicken. *Mol. Biol. Evol.* **32,** 2716–2725 (2015).

56. Zimmer, F., Harrison, P. W., Dessimoz, C. & Mank, J. E. Compensation of Dosage-Sensitive Genes on the Chicken Z Chromosome. *Genome Biol. Evol.* **8,** 1233–1242 (2016).

57. Ellegren, H. *et al.* Faced with inequality: chicken do not have a general dosage compensation of sex-linked genes. *BMC Biol.* **5,** 40 (2007).

58. Julien, P. *et al.* Mechanisms and Evolutionary Patterns of Mammalian and Avian Dosage Compensation. *PLOS Biol* **10,** e1001328 (2012).

59. Wang, Q., Mank, J. E., Li, J., Yang, N. & Qu, L. Allele-specific expression analysis does not support sex chromosome inactivation on the chicken Z chromosome. *Genome Biol. Evol.* doi:10.1093/gbe/evx031

60. Wei, Q.-X. *et al.* Germline Allele-Specific Expression of DAPK1 in Chronic Lymphocytic Leukemia. *PLoS ONE* **8,** (2013).

61. Albertson, R. C. *et al.* Genetic basis of continuous variation in the levels and modular inheritance of pigmentation in cichlid fishes. *Mol. Ecol.* **23,** 5135–5150 (2014).

Chapter 3

RNA-SEQ ANALYSES IDENTIFY ADDITIVITY AS THE PREDOMINANT GENE EXPRESSION PATTERN IN F1 CHICKEN EMBRYONIC BRAIN AND LIVER

3.1 Abstract

The superior performance of hybrids to parents, termed heterosis, has been utilized in animal and plant breeding programs for more than a century, but the understanding of the molecular mechanism underlying heterosis remains inadequate and requires further study. RNA-Seq provides a novel way to investigate heterosis at the genome-wide level, because gene expression can be considered as an intermediate phenotype that connects the genetic information to observable phenotypes. We compared embryonic gene expression between the chicken hybrids and their inbred parental lines using RNA-Seq to understand the gene expression basis of heterosis. Two genetically distinct and highly inbred chicken lines, Fayoumi and Leghorn, were crossed reciprocally to obtain F1 fertile eggs. The polyadenylated RNA of brain and liver from Day 12 embryos was converted to cDNA and sequenced on an Illumina HiSeq sequencer. The resultant reads were mapped to the reference genome assembly Gallus gallus 5.0 and the differentially expressed genes were identified, pairwise, among the hybrids, parental lines and synthesized mid-parent expression values. Our results indicated expression of the majority of the genes in F1 crosses are not significantly different from the mid-parental values, suggesting additivity as the predominant gene expression pattern between the F1 and parental lines. The second and third prevalent gene expression patterns are dominance and

over-dominance. Additionally, our results show that only 7-20% of the DE genes exhibit allele-specific expression in the F1, suggesting that differential *trans* – regulation is largely responsible for differential gene expression between the parental lines and is the major regulatory mechanism leading to heterosis in the F1 cross.

3.2 Introduction

Utilization of heterosis has tremendously improved agricultural production in recent centuries. Crossbreeding may take advantage of non-additive genetic effects and produce progeny exhibiting heterosis. This has been intensively exploited by breeders to obtain desirable agronomic traits. However, laborious research effort is needed to identify strains that result in useful heterosis when crossed. Understanding the molecular basis of heterosis may facilitate the identification of such plant or animal varieties. Hypotheses of dominance [1,2], over-dominance [3,4] and epistasis [5,6] have long been proposed to illustrate the mechanistic basis of heterosis, which respectively emphasizes the effect of advantageous allele, heterozygosity, and interaction of genes for multigenic traits. In chickens, heterosis has been previously reported for growth, body composition, egg production and abdominal fatness [7–9]. However, there is limited knowledge about its molecular basis in chickens and other livestock species[9].

Gene expression could be considered as an intermediate phenotype between genotypes and observable characteristics [10]. Variation in gene expression, such as differential gene expression (DE) between parents and hybrid and allele specific expression (ASE) in hybrid may be the key to understanding heterosis. Thus, transcriptomics is gaining importance in comprehending the molecular basis of

heterosis. Using microarray or RNA-Seq, heterosis has been analyzed in species such as rice [11], maize [12], soybeans [13], and silk worm [14]. Transgenerational gene expression patterns that hybrids are similar to one of the parents (dominance), equal to mid-parent values (additivity), above both parents (over-dominance) or below both parents (under-dominance) were identified and the pertaining genes were analyzed to understand heterosis.

Heterosis for economically important traits, such as mortality rate [15] and egg production [16], has been reported in F1 chickens produced from cross between two genetically distant and highly inbred chicken lines, Fayoumi and Leghorn. We previously established reciprocal F1 crosses between Fayoumi and Leghorn to study genomic imprinting and allele-specific expression using RNA-seq data from Day 12 embryonic brain and liver samples of the parental lines and the reciprocal F1 crosses [17]. In the present study, we took advantage of these existing RNA-seq data, and aimed to gain insight into heterosis in chickens by comparing global gene expression patterns across these two parental lines and their reciprocal F1 crosses.

3.3 Material and Methods

Ethical Statement

All animal protocols for production of the fertile eggs were conducted with the approval of Iowa State University IACUC Log #4-03-5425-G. No approval of University of Delaware AACUC was required for chicken embryo experiments.

Animals

Experimental design and sequencing strategy were described in detail previously [17]. Briefly, eggs of parental lines and their reciprocal crosses were

obtained from two highly inbred chicken lines, Fayoumi and Leghorn, and incubated for 12 days. Fertility was checked at Day 10 after incubation and infertile eggs were removed. Egg weights and embryo weights were recorded for 14 FL (Fayoumi ♂ X Leghon ♀), 4 LF (Leghorn ♂ X Fayoumi ♀), 14 F (inbred Fayoumi line) and 16 L (inbred leghorn line) samples. Embryonic efficiency (embryo weight / egg weight) was evaluated [18]. Tissues were sampled at Day 12 and frozen immediately in liquid nitrogen and stored in -80 °C. Statistical analyses of phenotypic data were performed using Tukey-Kramer HSD test in JMP Pro 13.1.0 [19].

Analysis of sequencing data

Twelve samples of male chicken embryos were chosen for further RNA-Seq study, which includes 4 FL, 4 LF, 2 F and 2 L. Total RNA was extracted from brain and liver and cDNA libraries were prepared from polyadenylated RNA. A total of 1.5 billion 75 bp reads were generated on Illumina HiSeq 2000 system, with an average of 64.8 million reads per sample. Reads were aligned to the Gallus gallus 5.0 genome assembly (Eensembl chicken release 89) using HISAT (v2.0.4) [20] with default parameters. Raw gene counts for each sample were obtained using Stringtie v1.3.0 [21]. Identification of new transcripts was disabled to facilitate comparison amongst samples. Differential expression between F, L, FL and LF groups and between F, Cross and L groups were analyzed using DESeq2 (v1.16.1) [22] in a pair-wise fashion. The mid-parent gene expression values (MPV) were calculated by taking the mean of normalized gene count for each combination of the data from paternal lines (four combinations in total), and DE genes between the F1 crosses and MPV were identified. Adjusted p-value less than 0.05 and Log2 (fold change) (abbreviated as

logFC) equal to or greater than 1 were applied to claim DE genes. The regularized log transformed counts (rLog(count)) from DESeq2 analysis were used to compare the relative abundance of gene expression between groups in the clustering analysis. Additionally, FPKM values for each sample were obtained using Cuffnorm (v2.2.1) [23] to normalize for library size and gene length.

Functional analysis of DE genes

A gene list for each tissue type was prepared by compiling all DE genes form pairwise DE analyses of F *vs.* L; Cross *vs.* each of F, L and MPV groups. The genes that were differentially expressed between the Cross and MPV groups were considered as genes showing a non-additive (dominance, over-dominance and under-dominance) expression pattern. Over-dominance, and under-dominance expression patterns were determined if gene expression in the Cross was significantly (q-value < 0.05; logFC ≥ 1) higher than that in the high parent or lower than that in the low parent. Average rLog(count) was calculated for each group for the clustering analysis. The DE genes in brain and liver gene lists were separately analyzed and clustered into 12 clusters using K-means clustering. The optimal numbers of clusters were estimated using a correlation-biased Figure of Merit method implemented in SC^2ATmd tool [24]. K-means clustering was performed in R using Pearson's correlation method of "amap" package [25]. The percentage of the genes showing non-additive expression patterns was used as a criterion to determine the predominant gene expression patterns in each cluster. When the percentage was greater than 50%, the predominate gene expression pattern of this very cluster was considered as non-additive, and additive if otherwise. The DE genes of each gene expression pattern

were further analyzed using DAVID Bioinformatics Resources 6.8 [26]. A threshold Benjamini adjusted p-values of 0.05 and an enrichment score greater than 1.3 were applied to identify the gene annotation (GO) terms and KEGG pathways that are significantly enriched by genes in our gene lists.

3.4 Results and Discussion

Phenotypic data

The fertility rates for the F, FL, LF and L eggs were 50.00%, 55.38%, 44.44% and 38.36%, respectively (Figure 3.1). It seems that the fertility rate is determined by male fertility, for the fertility rates of crosses are close to that of the paternal lines. The average egg weights of the F, FL, LF and L groups after Day 12 of incubation were 39.7 g, 37.3 g, 36.4 g and 33.1 g (Figure 3.2 a). The mean weight of the L group was significantly lower than that of the F group ($p < 0.001$) and FL group ($p = 0.0069$), which might be explained by hen's age or body weight. The average embryo weights of the F, FL, LF and L groups were 3.53, 3.98, 4.24 and 3.75 g (Figure 3.2 b). The average embryo weights of the crosses (the FL and LF groups combined) were higher than that of the F ($p = 0.0144$) and L ($p > 0.1$) groups. Egg weight and embryo weight showed a weak correlation of 0.29. Both the L group and cross groups have higher embryonic efficiency than the F group ($p < 0.001$), and the embryonic efficiency of the cross groups was slightly higher than that of the mid-parent value (Figure 3.3). In summary, with the egg weights not being significantly different between the crosses and parental lines, the crosses showed heterosis effect in embryonic weight and embryonic efficiency.

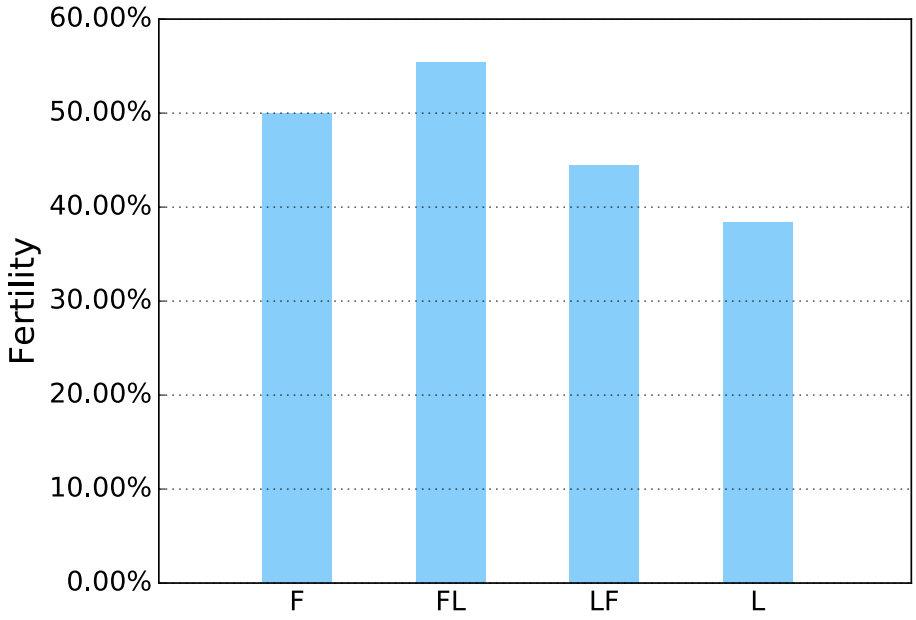

Figure 3.1 Fertility rates for parental lines (Fayoumi (F) and Leghorn (L)) and crosses (Fayoumi X Leghon (FL), Leghorn X Fayoumi (LF))

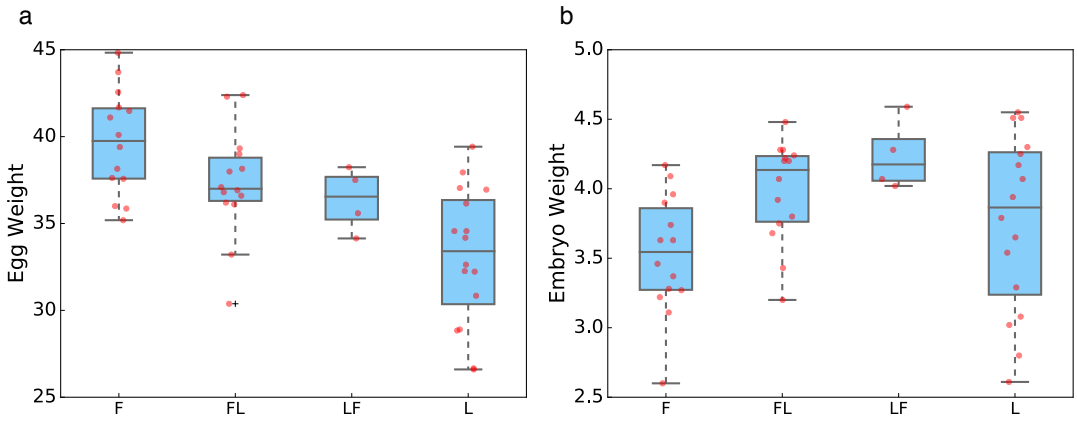

Figure 3.2 Phenotypic data at embryonic day 12: a. egg weight. b. embryo weight.

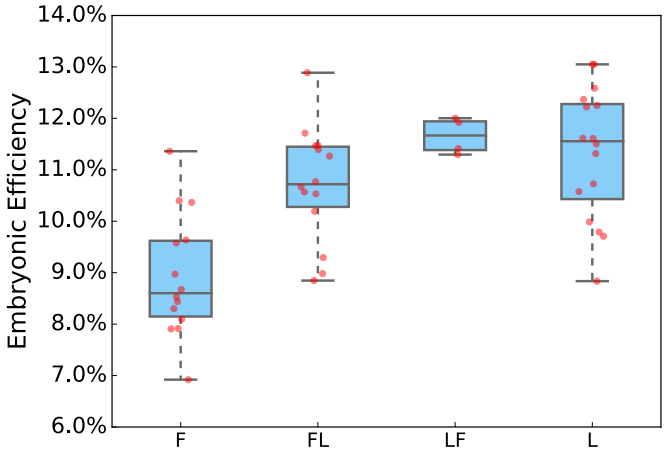

Figure 3.3 Embryo efficiency for parental line and crosses.

The current study terminated the experiment after 12 day of incubation and thus no more phenotypic data was collected, however, a few proceeding studies provided information on heterosis effect in Fayoumi and Leghorn crosses. Nordskog and Phillips observed that the mortality rate of FL cross was lower while LF cross was higher than that of both parental lines [15]. Saadey et al. noticed both reciprocal crossbreds showed negative heterosis in body weight and egg weight, but FL cross achieved positive heterosis in egg production rate and egg number [16]. Based on those results, it is clear crossbreeding of Fayoumi and Leghorn results in lower mortality

and improved egg production in the crosses. It is important to note FL and LF chickens may show heterosis effect for different traits.

DE Genes between Inbred Fayoumi and Leghorn Lines

Fayoumi and Leghorn chickens are genetically distant and have distinct breed traits. Fayoumi chickens have shown high resistance to leuckosis [15], coccidiosis [27], Rous sarcoma virus [28] and Marek's disease [29], while Leghorn chickens have been selected for and are commonly utilized in egg production [30,31]. A recent study reported the genetic variants in inbred Fayoumi and Leghorn chickens and found the genes with fixation index (F_{ST}) of 1 between the two populations are enriched in gene ontology (GO) terms related to their breed traits [32]. Here, we compared difference of gene expression between the two lines at embryonic day 12. Analysis revealed 304 DE genes in brain and 579 DE genes in liver, which are the largest number of DE genes among all pairwise analyses in the current study (Table 3.1). As expected, in principle, the genetic difference between the two lines should be more pronounced than that between cross and one of the parental lines or between reciprocal crosses. Among the DE genes in brain, Fayoumi expression was higher than Leghorn expression in 197 genes, and lower in 107 genes. The top 3 functional annotation clusters were shown in Figure 3.4 a, with most genes expressed higher in Fayoumi. The genes in cluster 1 (*SPIA1*, *SPIA4*, *SPIA5*, *SPIK5*, *ITIH2*, *ITIH3*, *SERPINC1*, *SERPIND1* and *AMBP*) are related to serine proteinase and peptidase inhibitor activity, and the genes in both cluster 2 and cluster 3 are mainly involved in blood coagulation and immune related functions. The DE genes in brain are also enriched in 13 GO terms and 1 KEGG pathway - metabolic pathway (Table 3.2). The most attention-drawing

biological processes are coagulation and fibrinolysis, which play a central role in brain homeostasis and development [33]. For example, inhibitors of plasminogen activator / plasmin system, such as serine proteinase inhibitor, may block or slow neuronal migration [34–36] – a pivotal event in brain development. Both *AvBD9* (ENSGALG00000019845) and *AvBD10* (ENSGALG00000016667) were expressed higher in F than L, implying stronger antimicrobial ability of Fayoumi chickens. Additionally, 45 DE genes were enriched the GO term extracellular exosome. Extracellular exosome vesicles (EV) mediates intercellular interaction in nervous system and thereby impacts brain development (reviewed in [37]). The DE genes between Fayoumi and Leghon suggest a line difference during brain development at embryonic day 12.

Table 3.1 Number of the DE genes.

	Brain	Liver
F vs L	304	579
F vs Cross	310	91
L vs Cross	73	147
FL vs LF	5	6
MPV vs Cross	133	40

Table 3.2 Functional analysis of the DE genes between F and L.

Term	Gene Count	Fold Enrichment	FDR
Brain			
GO:0042730~fibrinolysis	7	47.34	8.34E-07
GO:0031639~plasminogen activation	4	45.08	1.39E-02
GO:0007596~blood coagulation	6	11.27	2.54E-02

GO:0072562~blood microparticle	15	22.35	1.44E-13
GO:0005576~extracellular region	19	4.63	5.89E-06
GO:0070062~extracellular exosome	45	2.08	4.39E-05
GO:0005615~extracellular space	22	2.51	3.63E-03
GO:0004867~serine-type endopeptidase inhibitor activity	9	13.43	5.10E-05
GO:0008392~arachidonic acid epoxygenase activity	4	64.15	1.50E-03
GO:0004252~serine-type endopeptidase activity	9	7.70	1.31E-03
GO:0008395~steroid hydroxylase activity	4	36.66	6.31E-03
GO:0020037~heme binding	8	6.26	1.09E-02
GO:0005506~iron ion binding	8	5.40	2.22E-02
gga01100:Metabolic pathways	27	1.78	4.33E-02
Brain			
GO:0030141~secretory granule	8	8.16	9.66E-03
gga00140:Steroid hormone biosynthesis	8	6.83	1.35E-02
gga01100:Metabolic pathways	56	1.53	2.19E-02

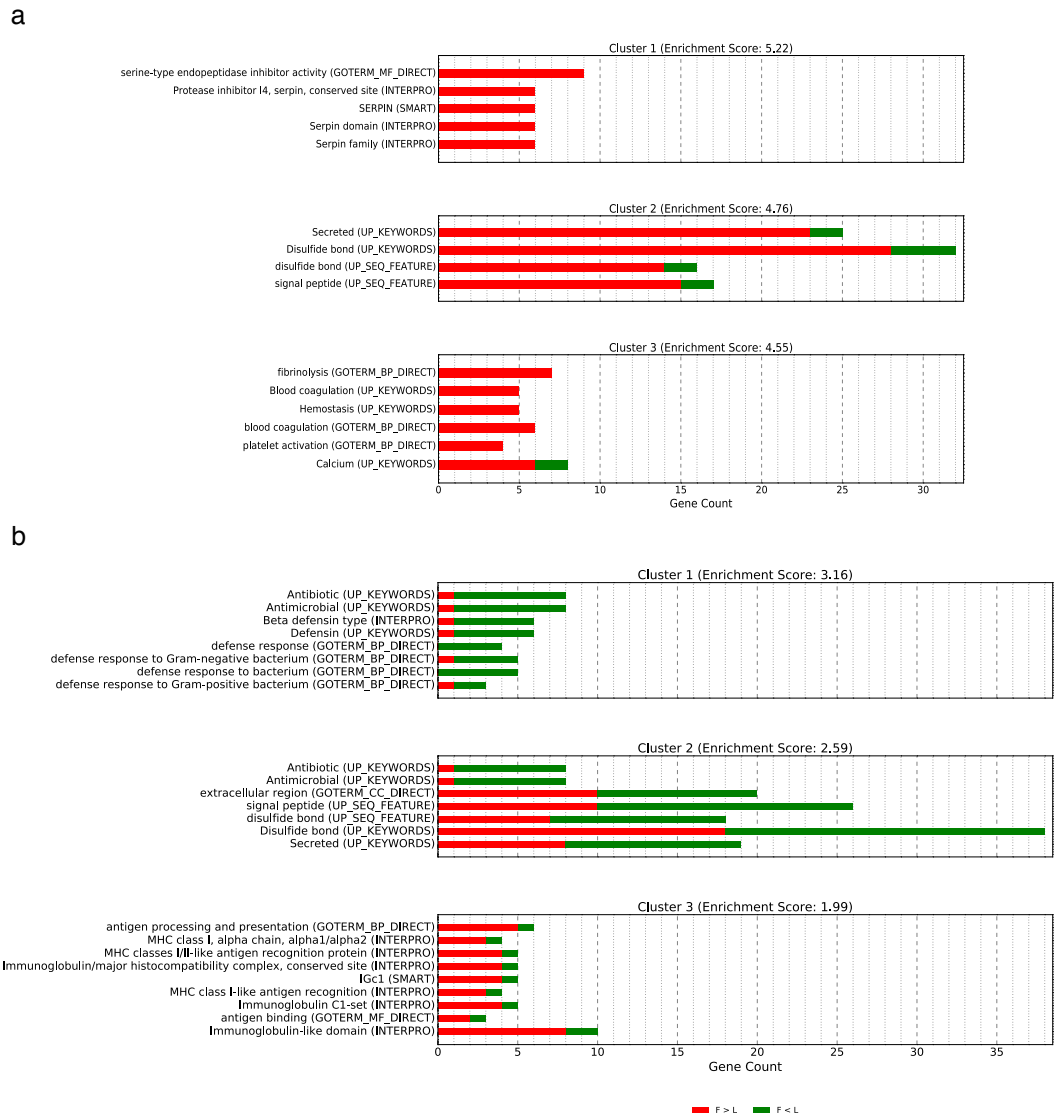

Figure 3.4 Top 3 functional annotation clusters of the DE genes between F and L in brain (a) and liver (b).

The DE genes in liver include 277 genes expressed higher in Fayoumi and 302 genes expressed higher in Leghorn. Some of the DE genes were enriched in immune related functions, such as antimicrobial function, defense response, antigen processing and presentation (Figure 3.4 b). The genes encoding for avian beta defensin (*AvBD1*, *AvBD2*, *AvBD6*, *AvBD8*), and genes encoding for Cathelicidin (*CATH3* and *CAMP*) were all expressed at a lower level in Fayoumi than Leghorn. Only *AvBD10* was expressed higher in Fayoumi. Compared with *AvBD10,* the expression levels of *AvBD1, 2, 6, 8* were relatively low, for liver is known to be the main site for *AvBD10* expression but not for the other 4 genes [38,39]. Synthetic AvBD10 has shown strong inhibition effect on bacteria and fungi *in vitro* [40] , however, there is a lack of study in how liver-expressed beta defensins contribute to innate immunity. It is reasonable to speculate those peptides may exert antimicrobial functions both locally and distantly through blood circulation.

Genes involved in antigen processing and presentation (Figure 2b), such as *BLB2*, *BFIV21*, *YF6*, *MHCIA7,* were also expressed at higher levels in Fayoumi chickens than that of Leghorn chickens. Major histocompatibility complex (MHC) proteins play a central role in adapted immunity. Since the present study doesn't include pathogen challenge, the DE genes involved in both antimicrobial peptides and antigen processing and presenting demonstrate a basal difference of those genes between the two lines. It has been shown that variant frequency at MHC locus is different between Fayoumi and Leghorn [32]. Further, expression of MHC Class Iα in spleen, liver, heart, thymus and bursa was higher in Marek's disease (MD) resistant haplotype than MD-susceptible haplotype [41]. Interestingly, in the chicken spleen, the highly expressed MHC class I proteins on cell surface are more specific in peptide binding

than the lowly expressed ones [42]. Thus, higher gene expression levels of MHC genes in Fayoumi embryos may suggest superior disease resistance and the specificity of antigen recognition than leghorn embryos. This corroborates with previous observation that Fayoumi embryos were found to have higher survival rate than Leghorn embryos following Rous Sarcoma virus infection [15]. It is also important to note liver is immune privilege organ that is inclined to antigen tolerance rather than removal [43], so additional study in other organs of immune system would be helpful to understand the difference of pathogen resistance between the two lines at embryonic stage.

Gene Expression of the Reciprocal Crosses

At embryonic day 12, the reciprocal crosses showed generally similar gene expressions, as only 5 DE genes in brain and 6 DE genes in liver were identified. The 5 DE genes in brain (ENSGALG00000042468, ENSGALG00000046132, ENSGALG00000044374, ENSGALG00000015374, ENSGALG00000036759) are all encoding for long noncoding RNAs (lncRNA), whose roles are unknown. The DE genes in liver include *DOCK3*, *LRRK1*, *KLF5*, *HBE*, *PODN* and *Mt_tRNA*. A related point to consider is that gene expression profiles between FL and LF may exhibit a larger difference in later developmental stages. As we discussed, above FL and LF showed different heterosis effects for traits measured post-hatch.

Comparison of Gene Expression Pattern across Parental Lines and F1 Crosses

Since only a few genes were differentially expressed between the reciprocal crosses, FL and LF samples were combined as the Cross group, and the DE genes

between Cross and parental lines were identified (Table 3.1). The DE genes, including F *vs*. L, F *vs.* Cross, L *vs.* Cross, and MPV *vs.* Cross were separately clustered into 12 K-means clusters for each tissue type (Figure 3.5 and Figure 3.6). Overlapping of the DE genes is shown as Venn diagrams in Figure 3.7. There were 133 genes in brain and 40 genes in liver expressed differentially between the Cross and MPV groups. Those genes, which comprised 26.9% and 6.8% of the DE genes in our gene lists from embryonic brain and liver, were considered as showing non-additive transgenerational gene expression pattern, which lead to the inference that the majority of the DE genes in our gene lists exhibit additive expression pattern. As the percentages of genes with non-additive expression patterns were indicated in parentheses for each K-means cluster (Figure 3.6), the main gene expression pattern for clusters 6, 7, 10 and 12 in brain and clusters 2 and 10 in liver were non-additive, and additive for the remaining clusters. Genes showing non-additive expression patterns were further categorized into dominance, over-dominance and under dominance expression patterns. As a result, there were 5 genes in brain and 2 genes in liver showing over-dominance expression pattern, and 128 genes in brain and 38 genes in liver showing dominance expression pattern. Of genes showing dominance expression pattern, most were of Leghorn dominance in brain and of Fayoumi dominance in liver. No genes showed under-dominance pattern with statistical significance. Overall, our result revealed additivity as the predominant transgenerational gene expression pattern between the F1 cross and parental lines. Dominance and over-dominance gene expression patterns were second and third in frequency, respectively.

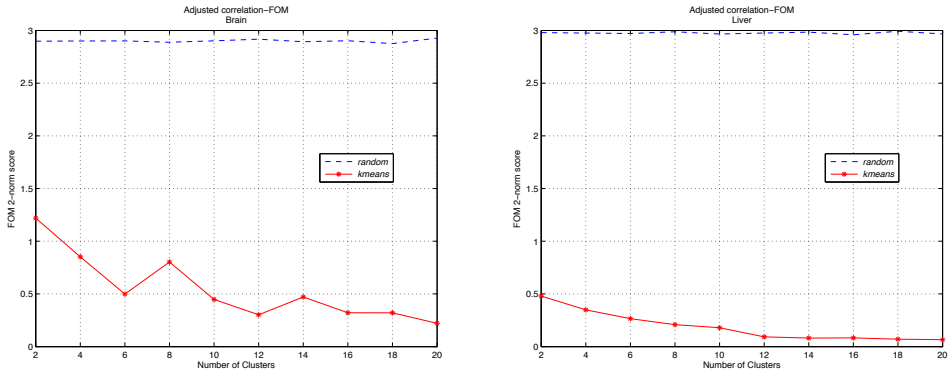

Figure 3.5 Optimal cluster numbers estimation using SC^{2}ATmd software. X axis shows the number of clusters and Y axis shows the estimation of figure of merit.

Transgenerational gene expression pattern has been examined in other species. All gene expression patterns were previously reported between maize F1 hybrids and their parents, with additivity as the main pattern (78%) and dominance as the second main pattern [12]. A study in rice suggested QTLs of additive effect comprised 50%, and QTLs of dominance and over-dominance effect separately comprised about 30% [44]. Rapp et al. found most genes were expressed in additive pattern in allopolyploid cotton [45]. Out of those genes with additive pattern, the authors defined “expression dominance” to describe gene expression pattern in hybrids that is not significantly different from one the parental lines and MPV, but gene expression of this parental line, allopolyploid and MPV required to be all significantly different from the other parental line, and they found it is the most prevalent pattern [45]. Genes falling into “expression dominance” defined by Rapp et

al. are classified as additivity in our study, since gene expression in crosses are not significantly different from MPV.

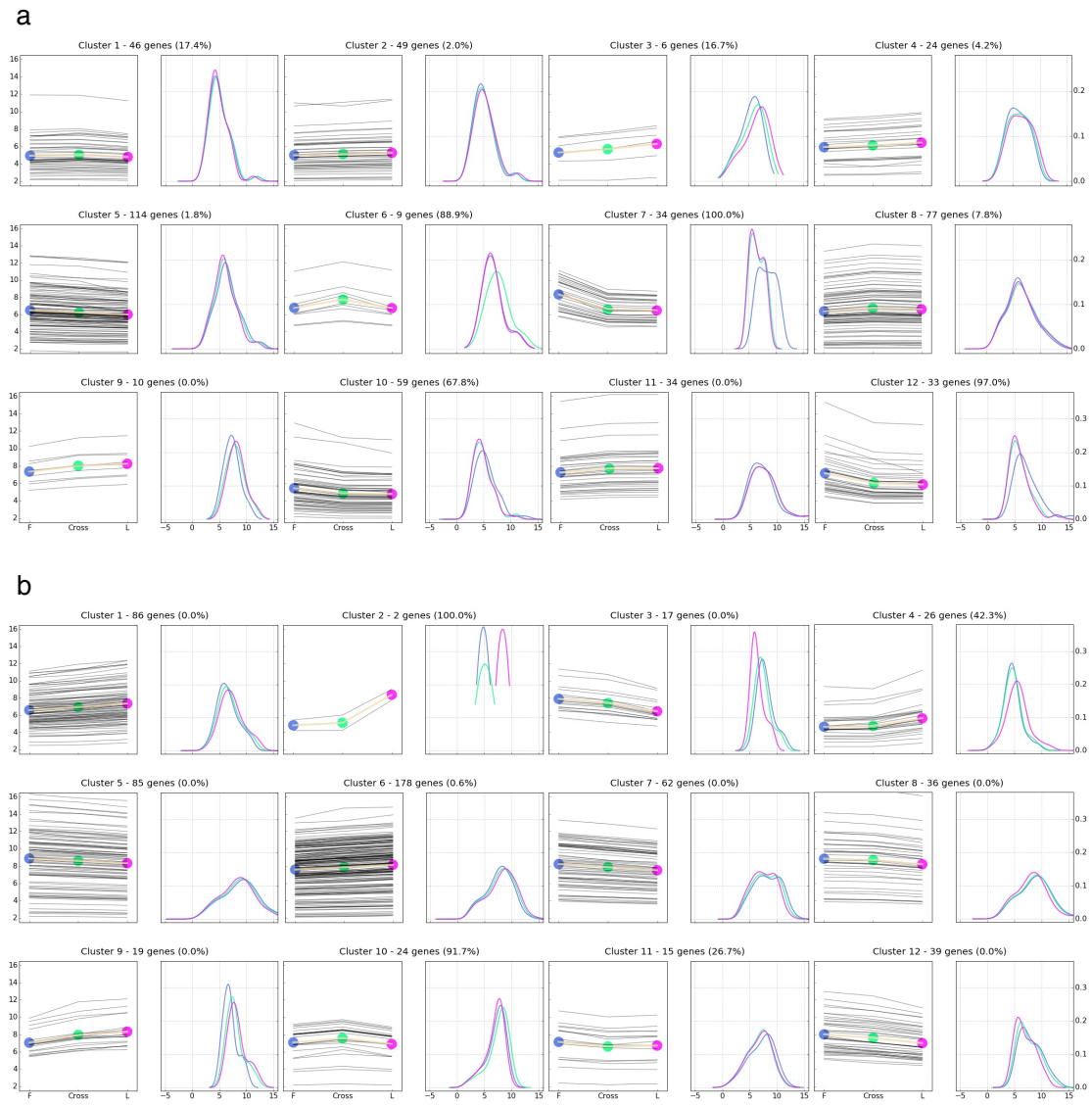

Figure 3.6 K-means clustering of DE genes in brain (a) and liver (b). For each cluster, the left plot shows the average gene expression levels for Fayoumi (F), Cross and Leghorn (L) groups in rLog(count); the right plot shows the kernel density estimation of the three groups; percentages of the genes with non-additive expression pattern were given in parentheses (F: Blue; Cross: Green; L: Fuchsia)

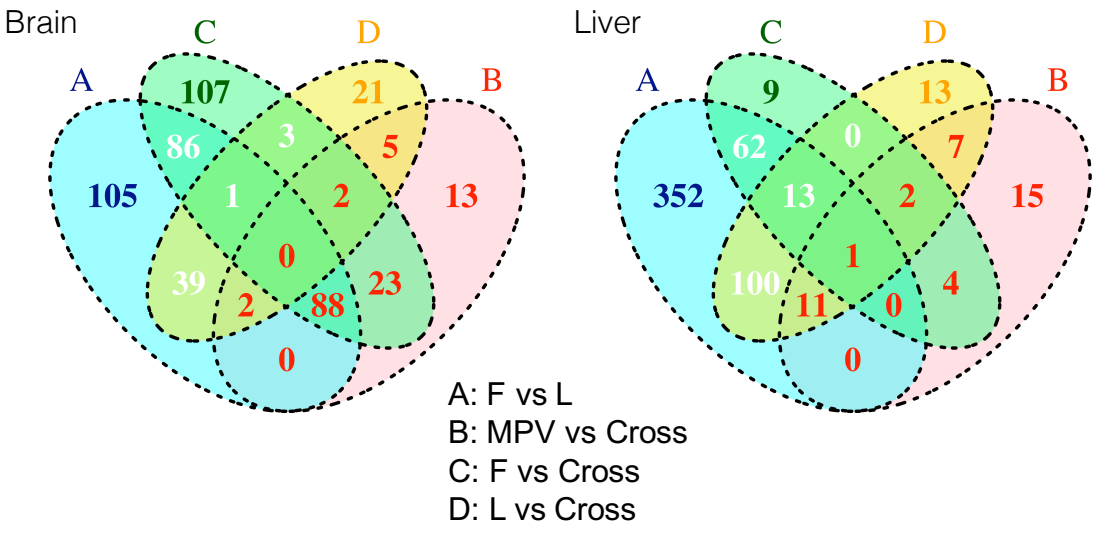

Figure 3.7 Venn diagram of DE genes in F vs. L, Cross vs. MPV, F vs. Cross and L vs. Cross

Implication for the Molecular Mechanisms of Heterosis

Over-Dominance

There were 5 genes expressed in over-dominance pattern in brain, including 2 novel genes (ENSGALG00000042217 and ENSGALG00000031253), a gene encoding lincRNA (ENSGALG00000043625), a gene encoding Mt_tRNA (ENSGALG00000033462), and a gene encoding lactate dehydrogenase D (*LDHD,* ENSGALG00000023828).

Both Mt_tRNA and lactate dehydrogenase D are involved in protein synthesis in mitochondrion and thereby potentially affect energy production. Polyadenylation of tRNA after transcription is critical for its stability and maturation [46,47]. It is also essential for the degradation of incorrectly folded tRNA in bacteria [48], hypomodified initiator tRNA in yeast [49,50] , and tRNA in human mitochondria [51]. Therefore, high polyadenylation may suggest a high turnover rate of this tRNA. The over-dominance pattern of this *Mt_tRNA* gene could result from either overall high gene transcription or high polyadenylation ratio of transcripts in the Cross group. Additionally, polyadenylation was found to function as a discriminator for two tRNAs that are encoded by overlapping mitochondrial tRNA genes in chickens [52]. Despite the cause is uncertain, the results suggested the Cross chickens have elevated protein synthesis activity than both parental lines in brain mitochondria.

LDHD is expressed at a relatively low level in brain (FPKM < 2 in the Cross and < 1 in F and L), which may be related to the low abundance of D-lactate [53]. However, D-lactate could interfere with biological processes that use L-lactate as substrate, like TAC cycle, and impair respiration efficiency in mitochondria [54]. Higher expression of LDHD in the Cross group implies a more efficient energy production in crosses compared with parental lines. In agreement with our inference, *LDHD* expression was previously found up-regulated in breast skeletal muscle of high feed efficiency broiler chickens [55].

There were 2 genes of over-dominance pattern in liver: *RN7SL1*(ENSGALG00000026904) and a novel gene (ENSGALG00000040994). *RN7SL1, or Metazoa_SRP,* is cytosolic lncRNA that mediates translocation of secretory protein across the endoplasmic reticulum membrane [56]. It was also found present in tumor

exosomes [57–59]. Taken together, the over-dominance gene in brain indicated a higher mitochondrial activity in the crossbreds, yet the consequence caused by over-dominance genes in liver remains unclear.

Dominance

In brain, 128 genes showed the dominance expression pattern (Figure 3.8). Gene expression of 13 DE genes in Cross were similar to Fayoumi but significantly different from Leghorn (designated as the Fayoumi dominance pattern), while 115 genes were similar to Leghorn but significantly different from Fayoumi (designated as Leghorn dominance pattern). The genes of the Fayoumi dominance pattern were 5 novel genes (ENSGALG00000003333, ENSGALG00000037297, ENSGALG00000040379, ENSGALG00000041375, ENSGALG00000046355), *OBSL1* (ENSGALG00000011242), *REM1* (ENSGALG00000045602)*, TPBGL* (ENSGALG00000030652), *ATP2C2* (ENSGALG00000035743), *ATP10B* (ENSGALG00000001662), *CHIR-IG1-5* (ENSGALG00000029472)*,* a gene encoding for lectin-like type II transmembrane protein (ENSGALG00000033116) and a gene encoding for C-type lectin domain family 2 member D-like (ENSGALG00000033672). OBSL1 is a cytoskeletal adapter protein [60] and plays a vital role in dendrites morphogenesis [61]. REM1 is a GTPase [62] and potentially involved in cytoskeletal changes that affect neuron morphology and migration, as suggested by similar proteins in the same protein family [63]. Both genes were expressed higher in F and Cross than in L, i.e. the gene expression in Cross elevates to the level of the high parent for these genes (F). We refer to this expression pattern as “enhancing dominance”, and “suppressing dominance” if the gene expression in the Cross

decreases to the level of the low parent for the gene. In fact, 10 out of 13 genes showing the Fayoumi dominance pattern were of the enhancing dominance. Only *ATP2C2*, *ATP10B* and a novel gene (ENSGALG00000046355) were showing the suppressing dominance pattern. ATP2C2, ATP10B, CHIR-IG1-5 and lectin-like type II transmembrane protein are all transmembrane proteins (enrichment score 1.54), but *CHIR-IG1-5* and gene for lectin-like type II transmembrane protein were showing the enhancing dominance pattern.

The majority of the DE genes with a non-additive expression pattern in brain showed the Leghorn dominance pattern. Contrary to the genes of Fayoumi dominance pattern, most (108 out of 135) genes with the Leghorn dominance expression pattern were showing suppressing dominance. For example, defensin genes *AvBD9* and *AvBD10 were* showing the suppressing dominance pattern. Further functional analysis suggested the genes of Leghorn dominance pattern are enriched in functions and biological processes similar to that of DE genes between F and L (Table 3.3), such as coagulation and fibrinolysis. In addition, the genes of Leghorn dominance pattern were enriched in KEGG pathways such as Metabolic pathways, Tyrosine metabolism, Biosynthesis of amino acids (Table 3.3). These results suggest that, in the embryonic brain, the metabolic level in the F1 crosses and Leghorn may be similar, and possibly lower than that in Fayoumi.

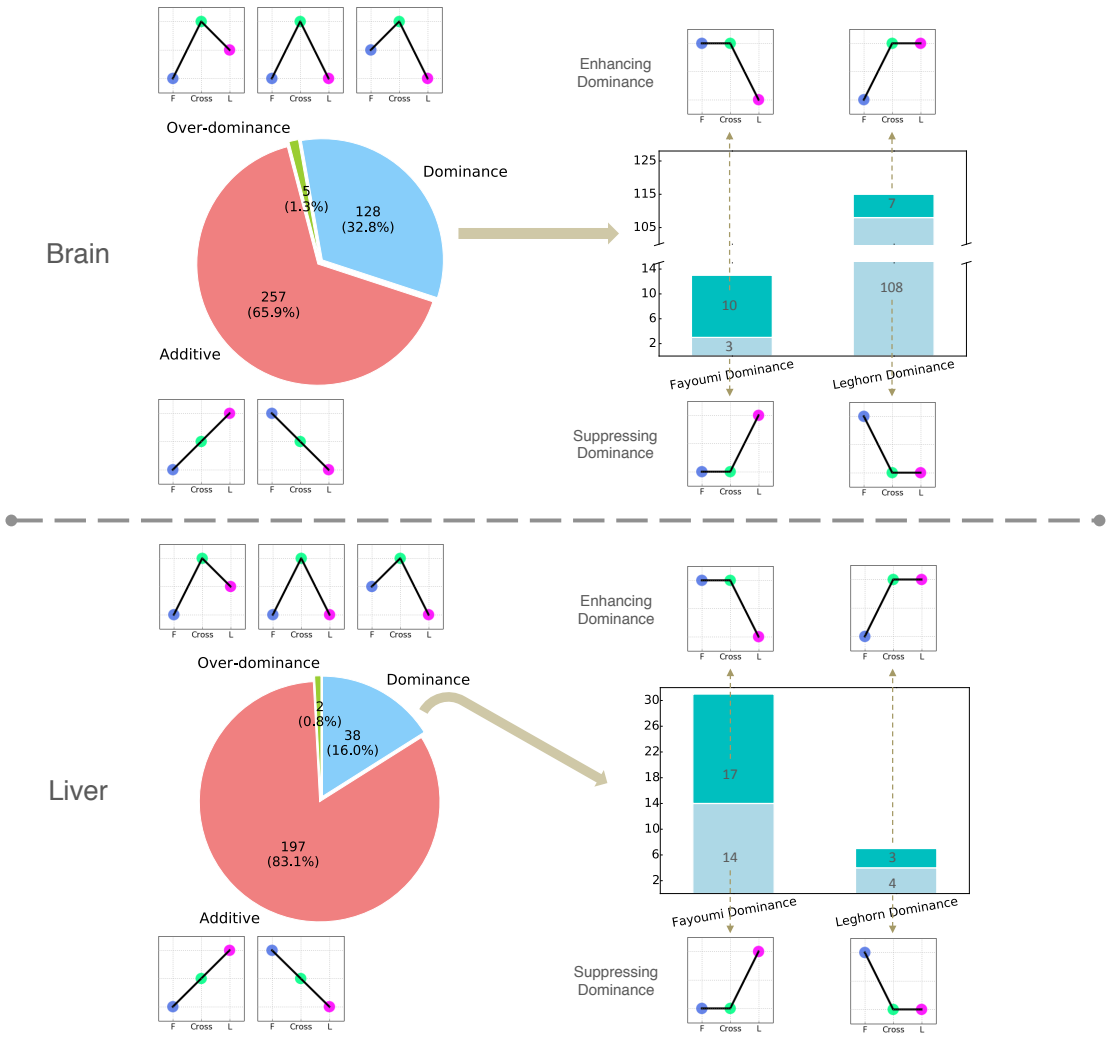

Figure 3.8 The gene expression patterns of DE genes between parental lines and F1 crosses in Day 12 embryonic brain and liver.

Table 3.3 Functional analysis of the genes showing dominance and additive expression patterns in the embryonic brain and liver.

Term	Gene Count	Fold Enrichment	FDR
Genes of Leghorn Dominance in brain			
GO:0042730~fibrinolysis	7	91.24	8.81E-09
GO:0007596~blood coagulation	6	21.72	9.63E-04
GO:0031639~plasminogen activation	4	86.90	7.81E-04
GO:0042632~cholesterol homeostasis	5	17.61	1.17E-02
GO:0034116~positive regulation of heterotypic cell-cell adhesion	3	130.35	9.64E-03
GO:0030195~negative regulation of blood coagulation	3	78.21	2.63E-02
GO:0051258~protein polymerization	3	78.21	2.63E-02
GO:0010873~positive regulation of cholesterol esterification	3	78.21	2.63E-02
GO:0051006~positive regulation of lipoprotein lipase activity	3	65.17	3.35E-02
GO:0072562~blood microparticle	16	44.15	1.30E-19
GO:0070062~extracellular exosome	37	3.17	1.08E-09
GO:0005576~extracellular region	14	6.32	4.19E-06
GO:0005615~extracellular space	19	4.02	8.04E-06
GO:0042627~chylomicron	4	70.95	1.89E-04
GO:0034361~very-low-density lipoprotein particle	4	49.66	5.31E-04
GO:0005577~fibrinogen complex	3	62.08	7.45E-03
GO:0034364~high-density lipoprotein particle	3	53.21	9.08E-03
GO:0031091~platelet alpha granule	3	37.25	1.70E-02
GO:0005623~cell	6	5.17	3.21E-02
GO:0004867~serine-type endopeptidase inhibitor activity	8	19.60	1.52E-05
GO:0004252~serine-type endopeptidase activity	8	11.24	3.69E-04
gga01100:Metabolic pathways	24	2.45	2.12E-04
gga00980:Metabolism of xenobiotics by cytochrome P450	6	21.06	1.62E-04
gga00350:Tyrosine metabolism	5	17.55	2.47E-03
gga00982:Drug metabolism - cytochrome P450	4	15.55	2.35E-02

Genes of Fayoumi Domiance in Liver			
GO:0042742~defense response to bacterium	3	81.76	1.83E-02
GO:0005576~extracellular region	4	12.19	4.44E-02
Genes of Fayoumi Domiance in Liver			
gga00100:Steroid biosynthesis	2	256.06	1.17E-02
Additive genes in Liver			
GO:0019882~antigen processing and presentation	5	19.12	4.04E-02
GO:0090286~cytoskeletal anchoring at nuclear membrane	4	36.71	2.46E-02
GO:0030141~secretory granule	5	16.51	2.10E-02
gga00140:Steroid hormone biosynthesis	6	14.77	3.03E-03
gga00100:Steroid biosynthesis	4	19.70	3.52E-02
gga01100:Metabolic pathways	24	1.89	2.48E-02

In liver, 38 genes were expressed in dominance pattern, 31 genes of Fayoumi dominance pattern and 7 genes of Leghorn dominance pattern (Figure 3.8). There were 17 genes of Fayoumi dominance showing the enhancing dominance pattern, 16 of which, however, are novel genes. The remaining gene (ENSGALG00000029094) also encodes for lncRNA RN7SL1. The genes of suppressing dominance pattern were enriched in keywords such as defensin, antibiotic and antimicrobial in DAVID analysis, because of differential expression of *AvBD1*, *2*, *6*, *and CATH2* between Leghorn and Cross. The genes of Leghorn dominance pattern included 3 genes of enhancing dominance and 4 genes of suppressing dominance. *DHCR24* and *SQLE* are enriched in Steroid biosynthesis pathway (Table 3.3), and both are of suppressing dominance.

The DE genes encoding for antimicrobial peptides in both brain and liver were all expressed in suppressing dominance pattern, which may provide some clue to understand previous observation that Fayoumi and Leghorn crosses and Leghorn pure line showed similar resistance to Rous sarcoma virus [15,28]. Additionally, most genes involved in metabolic pathways in both tissues were also showing suppressing dominance with leghorn dominance pattern. As embryonic weight and embryonic efficiency were found greater in the Cross and Leghorn than that in Fayoumi, additional examination of other metabolic organs and tissues is necessary to understand the molecular basis responsible for this difference. It is possible that in the F1 embryos, Leghorn-derived regulatory elements are more influential in regulating expression of genes involved in metabolism and modulate embryonic development and growth.

Additive

There were 257 DE genes in brain fell into this category (Figure 3.8), but no GO terms or pathway were enriched with a statistical significance in DAVID analysis. Most of the genes encode for transmembrane proteins (enrichment score 2.06), and are involved a variety of process, such as genes encoding for ATPase (*ATP6*, *ATP8B1*), antigen presenting protein (*BF2*, *BLB2*), Cytochromes P450 enzyme (*CYP1A1*, *CYP2C23*, *CYP4V2*), and receptors for growth hormone (*GHRZ*), glycine (*GLRA4*), interleukin 13 (*IL13RA2*), thyroid stimulating hormone (*TSHR*) etc.

In liver, 197 genes belonging to this category were enriched GO terms and pathways such as antigen processing and presentation, steroid hormone biosynthesis, and metabolic pathways (Figure 3.8, Table 3.3). Those pathways were

also enriched in the DE analysis between F and L. The genes involved in antigen processing and presentation (*YF2*, *BF2*, *BLB2*, *MHCIA7)* were expressed significantly higher in Cross than in L, but there was no significant difference in gene expression between Cross and F, a similar pattern observed for *BF2* and *BLB2* expression in brain. These results indicate that the ability of antigen presenting and processing in the crosses is superior than Leghorn to some extent, but it is not comparable to Fayoumi. About half of the DE genes enriched in steroid hormone synthesis and metabolic pathways were DE genes in the F *vs*. Cross analysis while the other half were identified in the L *vs*. Cross analysis. Despite the epistatic relationship was not directly evaluated here, some of those DE genes could potentially be involved in epistatic interactions with other genes and contribute to heterosis.

Cis- vs. Trans-Regulation of Gene Expression

How the gene expression is regulated is important to understand the DE genes between parental lines and between the crosses and the parental lines and, subsequently, the molecular mechanism of heterosis. It was reported both *cis*- and *trans*-regulatory mechanisms contribute to heterosis in plants [64,65]. Another study in rice found more than 40% of DE genes between F1 and parents were potentially caused by ASE [66], suggesting that differential expression in these genes are driven by cis-acting regulatory elements. We previously identified 2,242 ASE genes in brain and 1,735 ASE genes in liver in reciprocal crosses [17]. After excluding the ASE genes with multiple SNPs that showed conflicting preferred alleles, we checked the genes that were identified as both ASE and DE. Less than 10% of DE genes in brain and less than 20% of DE genes in Liver were identified as both ASE and DE, most of which were

showing additive gene expression pattern (Table 3.4). For the majority of DE genes between parental lines and between crosses and parental lines, gene expression of the two parental alleles didn't show significant difference in the F1 crosses. These results suggest, the *cis*-acting regulation of gene expression is not the main mechanism leading to differential expression between parental lines and between the crosses and parental lines. Therefore, *trans*-acting regulation of gene expression or interaction of *cis*- and *trans*- acting mechanisms may be largely responsible for differential gene expression and play more important roles in heterosis.

Table 3.4 Overlapping matrix between DE and ASE genes

	A	OD	D_F	D_L	ASE_F	ASE_L
Brain						
A	362	0	0	0	16	15
OD	0	5	0	0	0	0
D_F	0	0	13	0	0	0
D_L	0	0	0	115	2	1
ASE_F	16	0	0	2	674	0
ASE_L	15	0	0	1	0	736
Liver						
A	549	0	0	0	62	45
OD	0	2	0	0	0	0
D_F	0	0	31	0	0	0
D_L	0	0	0	7	1	0
ASE_F	62	0	0	1	632	0
ASE_L	45	0	0	0	0	608

	DE_F	DE_L	ASE_F	ASE_L
Brain				
DE_F	197	0	15	1
DE_L	0	107	0	15

ASE_F	15	0	674	0
ASE_L	1	15	0	736
Liver				
DE_F	277	0	68	2
DE_L	0	302	2	43
ASE_F	68	2	632	0
ASE_L	2	43	0	608

Abbreviations: A: Additive; OD: Over-dominance; D_F: Fayoumi dominance; D_L: Leghorn dominance; ASE_F: ASE genes with Fayoumi allele expressed higher; ASE_L: ASE genes with Leghorn allele expressed higher; DE_F: DE genes between F and L with expression in F greater than expression in L; DE_L: DE genes between F and L with expression in L greater than expression in F

3.5 Conclusion

To the best of our knowledge, this is the first RNA-Seq study to investigate transgenerational gene expression patterns and heterosis in chickens. We found additivity is the predominant gene expression pattern between F1 crosses and inbred parental lines. Dominance and over-dominance is the second and third, respectively. The DE genes between parental lines and between the crosses and the parental lines were enriched in functions and pathways related to tissue development and immunity in embryonic brain, and immunity and metabolism in embryonic liver. We also found most DE genes did not exhibit allelic imbalance in gene expression in the crosses, which suggested *trans* – acting mechanisms or interaction between *cis*- and *trans-acting* mechanisms may be the main contributors to DE genes between crosses and parental lines and thereby heterosis. At last, as gene expression is tissue and developmental stage dependent, future research of gene expression in other tissues and development stages would provide additional insights into heterosis in chickens.

3.6 Acknowledgments

This work was funded by the University of Delaware.

REFERENCES

1. Bruce, A. B. The Mendelian theory of heredity and the augmentation of vigor. *Sci. N. Y. NY* **32,** 627–627 (1910).

2. Jones, D. F. Dominance of Linked Factors as a Means of Accounting for Heterosis. *Genetics* **2,** 466–79 (1917).

3. Shull, G. H. The composition of a field of maize. (1908).

4. East, E. M. Inbreeding in corn. *Rep Conn Agric Exp Stn* **1907,** 419–428 (1908).

5. Bateson, W. Mendel's Principles of Heredity: Cambridge University Press. *März 1909 2nd Impr* **3,** 1913 (1909).

6. Fisher, R. A. XV.—The correlation between relatives on the supposition of Mendelian inheritance. *Earth Environ. Sci. Trans. R. Soc. Edinb.* **52,** 399–433 (1919).

7. Cahaner, A., Smith, E. J., Swenson, S. & Lamont, S. J. Associations of individual genomic heterozygosity, estimated by molecular fingerprinting, and of dam major histocompatibility complex with growth and egg production traits in layer chickens. *Poult. Sci.* **75,** 1463–1467 (1996).

8. Deeb, N. & Lamont, S. J. Genetic Architecture of Growth and Body Composition in Unique Chicken Populations. *J. Hered.* **93,** 107–118 (2002).

9. Abasht, B. & Lamont, S. J. Genome-wide association analysis reveals cryptic alleles as an important factor in heterosis for fatness in chicken F2 population. *Anim. Genet.* **38,** 491–498 (2007).

10. Hubner, N. *et al.* Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nat. Genet. N. Y.* **37,** 243–53 (2005).

11. Zhai, R. *et al.* Transcriptome analysis of rice root heterosis by RNA-Seq. *BMC Genomics* **14,** 19–19 (2013).

12. Swanson-Wagner, R. A. *et al.* All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents. *Proc. Natl. Acad. Sci.* **103,** 6805–6810 (2006).

13. Zhang, C. *et al.* Comparative transcriptome analysis of flower heterosis in two soybean F1 hybrids by RNA-seq. *PLOS ONE* **12,** e0181061 (2017).

14. Wang, H. *et al.* Heterosis and differential gene expression in hybrids and parents in Bombyx mori by digital gene expression profiling. *Sci. Rep.* **5,** 8750 (2015).

15. Nordskog, A. W. & Phillips, R. E. Heterosis in Poultry 5. Reciprocal Crosses Involving Leghorns, Heavy Breeds and Fayoumi. *Poult. Sci.* **39,** 257–263 (1960).

16. Saadey, S. M., Galal, A., Zaky, H. & El-Dein, A. Z. Diallel crossing analysis for body weight and egg production traits of two native Egyptian and two exotic chicken breeds. *Int. J. Poult. Sci.* **7,** 64–71 (2008).

17. Zhuo, Z., Lamont, S. J. & Abasht, B. RNA-Seq Analyses Identify Frequent Allele Specific Expression and No Evidence of Genomic Imprinting in Specific Embryonic Tissues of Chicken. *Sci. Rep.* **7,** 11944 (2017).

18. Washburn, K. W. & Guill, R. A. Relationship of Embryo Weight as a Percent of Egg Weight to Efficiency of Feed Utilization in the Hatched Chick. *Poult. Sci.* **53,** 766–769 (1974).

19. *JMP Pro, Version 13.1.0*. (SAS Insititute Inc., Cary, NC, 1989-2007).

20. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12,** 357–360 (2015).

21. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11,** 1650–1667 (2016).

22. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).

23. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28,** 511–515 (2010).

24. Olex, A. L. & Fetrow, J. S. SC2ATmd: a tool for integration of the figure of merit with cluster analysis for gene expression data. *Bioinformatics* **27,** 1330–1331 (2011).

25. Lucas, A. *amap: Another Multidimensional Analysis Package*. (2014).

26. Huang, D. W., Sherman, B. T. & Lempicki, R. a. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4,** 44–57 (2009).

27. Pinard-Van Der Laan, M. H., Monvoisin, J. L., Pery, P., Hamet, N. & Thomas, M. Comparison of outbred lines of chickens for resistance to experimental infection with coccidiosis (Eimeria tenella). *Poult. Sci.* **77,** 185–91 (1998).

28. Prince, A. M. Quantitative studies on Rous sarcoma virus. II. Mechanism of resistance of chick embryos to chorio-allantoic inoculation of Rous sarcoma virus. (1958).

29. Lakshmanan, N., Kaiser, M. & Lamont, S. Marek's disease resistance in MHC congenic lines from Leghorn and Fayoumi breeds. *Curr. Res. Marek's Dis. Am. Assoc. Avian Pathol. Kennett Sq. PA* 57–62 (1996).

30. Kerje, S. *et al.* The twofold difference in adult size between the red junglefowl and White Leghorn chickens is largely explained by a limited number of QTLs. *Anim. Genet.* **34,** 264–274 (2003).

31. Cui, J. X. *et al.* Association of polymorphisms in the promoter region of chicken prolactin with egg production. *Poult. Sci.* **85,** 26–31 (2006).

32. Fleming, D. S. *et al.* Single nucleotide variant discovery of highly inbred Leghorn and Fayoumi chicken breeds using pooled whole genome resequencing data reveals insights into phenotype differences. *BMC Genomics* **17,** (2016).

33. Takaishi, T., Ueshima, S. & Matsuo, O. New Aspects of Fibrinolytic Proteins in Brain Development. *Cell Struct. Funct.* **22,** 225–229 (1997).

34. Moonen, G., Grau-Wagemans, M. P. & Selak, I. Plasminogen activator–plasmin system and neuronal migration. *Nature* **298,** 753–755 (1982).

35. Seeds, N. W. *et al.* Plasminogen Activators and Plasminogen Activator Inhibitors in Neural Developmenta. *Ann. N. Y. Acad. Sci.* **667,** 32–40 (1992).

36. Seeds, N. W., Basham, M. E. & Haffke, S. P. Neuronal migration is retarded in mice lacking the tissue plasminogen activator gene. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 14118–14123 (1999).

37. Zappulli, V., Friis, K. P., Fitzpatrick, Z., Maguire, C. A. & Breakefield, X. O. Extracellular vesicles and intercellular communication within the nervous system. *J. Clin. Invest.* **126,** 1198–1207 (2016).

38. Zhao, C. *et al.* Gallinacin-3, an Inducible Epithelial β-Defensin in the Chicken. *Infect. Immun.* **69,** 2684–2691 (2001).

39. Xiao, Y. *et al.* A genome-wide screen identifies a single β-defensin gene cluster in the chicken: implications for the origin and evolution of mammalian defensins. *BMC Genomics* **5,** 56 (2004).

40. Yacoub, H. A. *et al.* Antimicrobial activities of chicken β-defensin (4 and 10) peptides against pathogenic bacteria and fungi. *Front. Cell. Infect. Microbiol.* **5,** (2015).

41. Dalgaard, T., Højsgaard, S., Skjødt, K. & Juul-Madsen, H. R. Differences in Chicken Major Histocompatibility Complex (MHC) Class Iα Gene Expression between Marek's Disease-Resistant and-Susceptible MHC Haplotypes. *Scand. J. Immunol.* **57,** 135–143 (2003).

42. Chappell, P. E. *et al.* Expression levels of MHC class I molecules are inversely correlated with promiscuity of peptide binding. *eLife* **4,** e05345 (2015).

43. Knolle, P. A. & Gerken, G. Local control of the immune response in the liver. *Immunol. Rev.* **174,** 21–34 (2000).

44. Li, L. *et al.* Dominance, Overdominance and Epistasis Condition the Heterosis in Two Heterotic Rice Hybrids. *Genetics* **180,** 1725–1742 (2008).

45. Rapp, R. A., Udall, J. A. & Wendel, J. F. Genomic expression dominance in allopolyploids. *BMC Biol.* **7,** 18 (2009).

46. Tomita, K., Ueda, T. & Watanabe, K. RNA Editing in the Acceptor Stem of Squid Mitochondrial tRNATyr. *Nucleic Acids Res.* **24,** 4987–4991 (1996).

47. Bratic, A. *et al.* Mitochondrial Polyadenylation Is a One-Step Process Required for mRNA Integrity and tRNA Maturation. *PLOS Genet.* **12,** e1006028 (2016).

48. Li, Z., Reimers, S., Pandit, S. & Deutscher, M. P. RNA quality control: degradation of defective transfer RNA. *EMBO J.* **21,** 1132–1138 (2002).

49. Kadaba, S. *et al.* Nuclear surveillance and degradation of hypomodified initiator tRNAMet in S. cerevisiae. *Genes Dev.* **18,** 1227–1240 (2004).

50. Vaňáčová, Š. *et al.* A New Yeast Poly(A) Polymerase Complex Involved in RNA Quality Control. *PLOS Biol.* **3,** e189 (2005).

51. Slomovic, S., Laufer, D., Geiger, D. & Schuster, G. Polyadenylation and Degradation of Human Mitochondrial RNA: the Prokaryotic Past Leaves Its Mark. *Mol. Cell. Biol.* **25,** 6427–6435 (2005).

52. Yokobori, S. & Pääbo, S. Polyadenylation creates the discriminator nucleotide of chicken mitochondrial tRNA(Tyr). *J. Mol. Biol.* **265,** 95–99 (1997).

53. Kondoh, Y., Kawase, M., Kawakami, Y. & Ohmori, S. Concentrations of D-lactate and its related metabolic intermediates in liver, blood, and muscle of diabetic and starved rats. *Res. Exp. Med. Z. Gesamte Exp. Med. Einschl. Exp. Chir.* **192,** 407–414 (1992).

54. Ling, B. *et al.* D-Lactate altered mitochondrial energy production in rat brain and heart but not liver. *Nutr. Metab.* **9,** 6 (2012).

55. Bottje, W. *et al.* Progesterone signalling in broiler skeletal muscle is associated with divergent feed efficiency. *BMC Syst. Biol.* **11,** 29 (2017).

56. Ullu, E., Murphy, S. & Melli, M. Human 7SL RNA consists of a 140 nucleotide middle-repetitive sequence inserted in an Alu sequence. *Cell* **29,** 195–202 (1982).

57. Liu, W. M. *et al.* Supernatants derived from chemotherapy-treated cancer cell lines can modify angiogenesis. *Br. J. Cancer* **106,** 896–903 (2012).

58. Schageman, J. *et al.* The Complete Exosome Workflow Solution: From Isolation to Characterization of RNA Cargo. *BioMed Research International* (2013). doi:10.1155/2013/253957

59. Nabet, B. Y. *et al.* Exosome RNA Unshielding Couples Stromal Activation to Pattern Recognition Receptor Signaling in Cancer. *Cell* **170,** 352–366.e13 (2017).

60. Geisler, S. B. *et al.* Obscurin-like 1, OBSL1, is a novel cytoskeletal protein related to obscurin. *Genomics* **89,** 521–531 (2007).

61. Litterman, N. *et al.* An OBSL1-Cul7Fbxw8 Ubiquitin Ligase Signaling Mechanism Regulates Golgi Morphology and Dendrite Patterning. *PLOS Biol.* **9,** e1001060 (2011).

62. Finlin, B. S. & Andres, D. A. Rem is a new member of the Rad-and Gem/Kir Ras-related GTP-binding protein family repressed by lipopolysaccharide stimulation. *J. Biol. Chem.* **272,** 21982–21988 (1997).

63. Ward, Y. *et al.* The GTP binding proteins Gem and Rad are negative regulators of the Rho-Rho kinase pathway., The GTP binding proteins Gem and Rad are negative regulators of the Rho–Rho kinase pathway. *J. Cell Biol. J. Cell Biol.* **157, 157,** 291, 291–302 (2002).

64. Springer, N. M. & Stupar, R. M. Allele-specific expression patterns reveal biases and embryo-specific parent-of-origin effects in hybrid maize. *Plant Cell* **19,** 2391–402 (2007).

65. Bell, G. D. M., Kane, N. C., Rieseberg, L. H. & Adams, K. L. RNA-Seq Analysis of Allele-Specific Expression, Hybrid Effects, and Regulatory Divergence in Hybrids Compared with Their Parents from Natural Populations. *Genome Biol. Evol.* **5,** 1309–23 (2013).

66. Song, G. *et al.* Global RNA sequencing reveals that genotype-dependent allele-specific expression contributes to differential expression in rice F1 hybrids. *BMC Plant Biol.* **13,** 221 (2013).

Chapter 4

RNA-SEQ ANALYSIS OF ABDOMINAL FAT REVEALS DIFFERENCES BETWEEN MODERN COMMERCIAL BROILER CHICKENS WITH HIGH AND LOW FEED EFFICIENCIES

(Zhuo, Z., Lamont, S. J., Lee, W. L. & Abasht, B. RNA-Seq Analysis of Abdominal Fat Reveals Differences between Modern Commercial Broiler Chickens with High and Low Feed Efficiencies. PLoS ONE (2015))

4.1 Abstract

For economic and environmental reasons, chickens with superior feed efficiency (FE) are preferred in the broiler chicken industry. High FE (HFE) chickens typically have reduced abdominal fat, the major adipose tissue in chickens. In addition to its function of energy storage, adipose tissue is a metabolically active organ that also possesses endocrine and immune regulatory functions. It plays a central role in maintaining energy homeostasis. Comprehensive understanding of the gene expression in the adipose tissue and the biological basis of FE are of significance to optimize selection and breeding strategies. Through gene expression profiling of abdominal fat from high and low FE (LFE) commercial broiler chickens, the present study aimed to characterize the differences of gene expression between HFE and LFE chickens. mRNA-seq analysis was carried out on the total RNA of abdominal fat from 10 HFE and 12 LFE commercial broiler chickens**,** and 1.48 billion of 75-base sequence reads were generated in total. On average, 11,565 genes were expressed (>5 reads/gene/sample) in the abdominal fat tissue, of which 286 genes were

differentially expressed (DE) at q (False Discover Rate) < 0.05 and fold change > 1.3 between HFE and LFE chickens. Expression levels from RNA-seq were confirmed with the NanoString nCounter analysis system. Functional analysis showed that the DE genes were significantly ($p < 0.01$) enriched in lipid metabolism, coagulation, and immune regulation pathways. Specifically, the LFE chickens had higher expression of lipid synthesis genes and lower expression of triglyceride hydrolysis and cholesterol transport genes. In conclusion, our study reveals the overall differences of gene expression in the abdominal fat from HFE and LFE chickens, and the results suggest that the divergent expression of lipid metabolism genes represents the major differences.

4.2 Introduction

Feed efficiency (FE) – the efficiency of converting feed intake to body weight gain – is of great importance to modern commercial broiler chicken production. Feed cost is the major expense for chicken production and represents up to 70% of the total economic input. For a unit of weight gain, HFE chickens consume less feed and produce less excrement [1]. Therefore, improving FE could reduce production cost and lower the cost of waste management and emission of nitrogenous and greenhouse gases. A comprehensive understanding of the biological mechanisms controlling FE is crucial to develop optimal breeding and selection strategies. Previous studies on chicken FE have investigated gene expression in breast muscle by microarray [2–4] and by RNA-Seq [5], but gene expression associated with FE in adipose tissue is still unexamined.

Adipose tissue is now recognized as a metabolically active endocrine organ and plays a central role in energy homeostasis. It serves as the major site for lipid deposition and lipid metabolism. Adipose-derived hormones, proteins, and other biologically active factors regulate metabolic and immune activities locally and systemically (reviewed in [6,7]). Given that obesity and obesity-related conditions are prevalent worldwide, a thorough understanding of adipose biology is needed to prevent and intervene the disease. The chicken has been proposed as a model for adiposity studies, as it possesses unique features relevant to obesity research. Different from rodents adipose tissue, human adipose tissue has a relatively limited lipogenic capacity [8]. Similarly, chicken adipose tissue is not recognized as the major organ for *de novo* lipid synthesis [9]. The majority of lipids accumulated in adipose tissue are synthesized in the liver, circulated in the blood stream, and then absorbed by adipose tissue [10,11]. The chicken adipose tissue is insensitive to insulin [12,13], similar to the physiological behavior of adipose tissue of obese people and type 2 diabetes patients.

A few studies have examined the global gene expression of chicken adipose tissue by using genomic microarrays. By comparing the gene expression of fat line and lean line chickens that were divergently selected for abdominal fat content for seven generations, the gene expression related to adipogenesis and lipogenesis were found to be up-regulated in fat line chickens, but gluconeogenesis or glycolysis genes were down-regulated [14,15]. In commercial broiler chickens, fasting and insulin neutralization affected the expression of adipogenic genes and enhanced lipid oxidization in adipose tissue [16]. Genes involved in immune response were found differentially expressed in different ages of broiler chickens [17]. Compared with

commercial broilers, relatively lean chicken lines, Fayoumi and Leghorn, had higher expression of lipolysis and fatty acid oxidation genes [18].

The present study aimed to investigate gene expression in the adipose tissue associated with FE. Through profiling the gene expression of abdominal fat from selected HFE and LFE chickens using RNA-seq, we identified 286 differentially expressed (DE) genes. We paid special attention to the DE genes and pathways involved in lipid metabolism and interpreted how they contributed to the differences in adiposity between LFE and HFE chickens. Overall, our study provides insights into the relationships between feed efficiency and gene expression in abdominal fat and advances the understanding of the gene expression in chicken adipose tissue.

4.3 Methods and Materials

Experimental animals and tissue collection

A live animal experiment of 2400 commercial broiler chickens was previously conducted and used for studying various aspects of quantitative traits in broiler chickens ([5] and unpublished data). The chickens were sampled from 6 commercial broiler farms (400 chickens per farm) in the Delmarva region (USA) at 29-day age. Then the chickens were transferred to an experimental station, where each chicken was kept in a separate cage for individual feed efficiency measurement and fed *ad libitum*. The cages in the experimental station were arranged in rows at two levels, i.e. top or bottom levels, relative to their distance from the floor, and each row had 100 cages. The weight of feeders and chickens were measured and recorded at the beginning (day 29) and the end (day 46) of the test. Dead (1.5%) and sickly (0.9%) chickens were removed or culled routinely during the test. At day 47, the chickens were euthanized by manual cervical dislocation for tissue sampling. About 1 g of

adipose tissue was harvested and immediately frozen in liquid nitrogen, and kept at –80 °C for further RNA isolation. Fat in abdominal cavity and around gizzard were dissected and weighed after keeping the carcasses at +4 °C for 24 hours. The protocols were approved by the University of Delaware Agricultural Animal Care and Use Committee.

Calculation of feed efficiency and phenotypic correlations

Before estimating FE and correlations between FE and other phenotypic measurements, inaccurate data (1.6% of the total) resulting from artifacts was excluded. In addition, the following criteria were applied to exclude outliers (2.0% of the total) in each group. First, residual weight gain was calculated by adjusting weight gain for cage location effect. Chickens with a residual weight gain that fell outside of the mean ±3 standard deviations (SDs) were excluded. Then, residual feed consumption (RFC) was estimated as a measure of FE by calculating the difference between the actual and expected feed intake using the following equation:

$RFC = FC – (a + b1*BW_{29} + b2*BW_{46} + Level + Row (Level))$,

where FC is the actual feed consumption; BW_{29} and BW_{46} are the body weights at 29 and 46 days of age, respectively; Level represents the fixed effect of row location (top or bottom level) and Row(Level) represents the fixed effect of row nested within row location; and a is the intercept, b1 and b2 are the partial regression coefficients of BW_{29} and BW_{46}, respectively. Chickens with RFC lying outside of the mean ± 3 SDs were excluded to eliminate the data points that might affect the accuracy of estimating RFC. As a result, data from 2254 chickens remained, and new RFC of each bird was calculated using the same model.

Within each experimental group, the birds were ranked by RFC. The chickens with extreme RFC values at both ends, designated as HFE and LFE, respectively, were selected for RNA-seq. The birds with defects (wooden breast muscle [19], leg and wings problem, etc.) were excluded. In total, 12 HFE and 12 LFE chickens were chosen, but two HFE samples did not generate adequate cDNA libraries. Thus, only 10 HFE and 12 LFE were used for RNA-seq, but all the chosen samples (12 HFE and 12 LFE) were used for the NanoString confirmation. The correlation coefficients between FE, feed conversion ratio (FCR), body weight, weight gain, abdominal fat weight, and abdominal fat percentage, as well as the p-values for t-tests between HFE and LFE phenotypes, were estimated using JMP® (Version 11.0.0.). A threshold of p-values less than 0.05 were applied to declare significance in the data analysis.

Total RNA extraction and cDNA library preparation

The fat samples of the selected birds were ground in frozen state in liquid nitrogen. Total RNA was extracted from ~70 mg of samples with the mirVana™ miRNA Isolation Kit (Life Technologies). The concentration of RNA samples was measured using the NanoDrop 1000 (Thermo Scientific). Agilent Bioanalyzer 2100 (Agilent Technologies) was utilized to assess the integrity of the total RNA. The RNA integrity number (RIN) of all samples was greater than 8.

cDNA libraries were constructed using a TruSeq Stranded mRNA LT Sample Prep Kit. Briefly, mRNA was isolated from 2 µg of total RNA using poly-T oligo-attached magnetic beads and fragmented by divalent cation. The first strand cDNA was synthesized using reverse transcriptase (Life Technologies) and random primers, followed by removal of template RNA using RNase H. During the second strand synthesis, dUTPs were used in the reaction instead of dTTPs. The double-stranded

cDNA was recovered using AMPure beads (Beckman Coulter). AMPure beads can purify cDNA from the reaction mix and perform size selection simultaneously. After reverse transcription, a single 'A' nucleotide was added to the 3' ends of the blunt fragments to prevent them from ligating to one another during adapter ligation reaction. Then, adaptors with index were ligated to the fragments, as a corresponding single 'T' nucleotide on the 3' end of the adapter provided a complementary overhang for ligating the adapter to the fragment. Of note, a unique indexing adaptor was used for each sample. After clean up using AMPure beads, DNA fragments with adapter sequences were enriched by PCR. dUTP prevented the second strand cDNA from elongating due to the specificity of the enzyme, leaving only the first-strand cDNA to be amplified. Finally, the concentration of cDNA libraries was measured using a NanoDrop 1000, and the quality of the cDNA libraries was further validated using an Agilent Bioanalyzer 2100.

Sequencing strategy

The concentration of the 22 cDNA libraries was normalized to 10 nm/µl using Tris buffer (Tris-Cl 10mM, 0.1% Tween 20, pH 8.5), as suggested by the manufacturer. Ten microliter of each uniquely-indexed, normalized library was pooled into a single sample, and the resultant pool was sequenced on four lanes of a flow cell for 75 cycles with the paired-end sequencing protocol of the Illumina Hiseq 2000 system. The resultant data was deposited in NCBI's Short Read Archive (SRA) database (Accession SRP058295).

QC and reads alignment

First, the RNA-seq reads of each sample were discriminated (i.e. demultiplexed) based on the indexing adaptors, and then processed with FastQC v0.10.1 to check the quality of raw sequence reads [20]. The reads were mapped to the chicken reference genome Gallus_gallus-4.0 (Ensembl, database version 78.4) using TopHat v2.0.4 [21], a fast splice junction mapper for RNA-seq reads. Parameters of TopHat were set to allow only unique alignment to the reference genome. Reads with more than two mismatches were discarded, and concordant mapping for both reads in a pair was required. To obtain the mapping statistics, the alignment BAM files were further examined using RNA-SeQC v1.1.7 [22]

Differential gene expression and functional analysis

The genes differentially expressed (DE genes) between HFE and LFE groups were identified using Cuffdiff v2.1.1 [23]. FPKM (fragments per kilobase of transcript per million mapped reads) values of each gene for each sample were also reported by Cuffdiff. To identify over-represented pathways and networks, and to predict the activation and inhibition states of upstream regulators, the DE genes were analyzed using the Ingenuity Pathways Analysis (IPA) system [24]. Based on the FPKM values of all genes, a 2-way hierarchical clustering (Ward method) of samples was performed in JMP® (Version 11.0.0.).

Confirmation of RNA-Seq data by nCounter analysis system

Expression results obtained from RNA-Seq were confirmed by the NanoString's nCounter™ analysis system [25]. To gain comprehensive evaluation of the RNA-seq expression data, a set of 204 genes were chosen for nCounter™ probe

design based on multiple ongoing RNA-seq experiments in our laboratory. From the same RNA samples used for RNA-Seq library constructions, 300 ng of total RNA were submitted to NanoString Technologies for hybridization, detection, and scanning. Of the 204 genes chosen, 65 were identified with low number of alignments for performing statistical test in Cuffdiff analysis and thus excluded from data analysis. The other 139 genes, containing 12 designated housekeeping genes, 31 DE genes and 96 non-DE genes based on RNA-Seq, were kept for correlation analysis. No background subtraction was performed since the spike-in negative controls showed a low background noise. Twelve reference genes were chosen based on coefficients of variation among all genes.

The raw gene counts for each transcript were normalized by External RNA Control Consortium (ERCC) spike-in positive controls and by the reference genes. The Pearson correlation coefficients of log2 (fold change) between normalized gene count and FPKM were calculated in JMP® (Version 11.0.0.).

4.4 Results

Phenotypes

In the present study, 2400 commercial broiler chickens were hatched and raised for feed efficiency tests. Weight gain (WG), abdominal fat percentage, FCR, and RFC were calculated based on the records of body weight (BW), feed consumption (FC) and abdominal fat weight. WG had a weak correlation with BW_{29} ($r = 0.23$) and a strong correlation with BW_{46} ($r = 0.81$). Similarly, FC had a moderate correlation with BW_{29} ($r = 0.39$) but a strong correlation with BW_{46} ($r = 0.69$). Further, WG and FC had a strong correlation ($r = 0.85$ and $r^2 = 0.72$), indicating that 72% of the

variability of weight gain can be explained by FC. Moreover, abdominal fat percentage had a moderate correlation with FCR (r = 0.31) and RFC (r = 0.40) (Table 4.1), which is consistent with previous reports that LFE chickens have an overall more fat deposition [26–28].

Table 4.1 Correlation coefficients between WG, BW, FC, Fat%, RFC and FCR

	WG	BW(29)	BW(46)	FC	Fat%	FCR	RFC
WG							
BW_{29}	0.24						
BW_{46}	0.81	0.67					
FC	0.85	0.39	0.78				
Fat%	-0.01	0.02	0.00	0.19			
FCR	-0.54	0.18	-0.30	-0.03	0.31		
RFC	0.00	0.00	0.00	0.47	0.40	0.74	

Average BW_{29}, BW_{46}, FC, WG, breast muscle weight, abdominal fat weight, percentage of abdominal fat, and RFC of selected chickens are summarized in Table 4.2. The RFC of HFE and LFE groups were significantly different, which were the basis used to select chickens for RNA-seq. There were no significant differences of initial body weight (BW_{29}) and final body weight (BW_{46}) between HFE and LFE groups. However, FC ($p < 0.001$), WG ($p = 0.0035$), breast muscle weight ($p = 0.0361$), abdominal fat weight ($p = 0.0012$), and abdominal fat percentage ($p = 0.0040$) were significantly different between HFE and LFE groups. Also of important fact is that HFE and LFE chickens do not necessarily retain the lowest or highest abdominal fat percentage, which is in concert with the moderate correlation between abdominal

fat percentage and RFC. In summary, on average, the LFE birds consumed more feed and deposited less breast muscle but accumulated more abdominal fat.

Table 4.2 Phenotypic data of samples used in RNA-seq (Mean ± S.E.)

	FC (kg)	WG (kg)	Breast muscle percentage (%BW)	Fat percentage (%BW)	FCR	RFC (kg)
HFE	2.91±0.05*	1.81±0.04*	23.46±0.48*	1.52±0.15*	1.61±0.02*	-0.28±0.01*
LFE	3.34±0.05*	1.62±0.04*	21.75±0.44*	2.36±0.13*	2.07±0.02*	0.36±0.01*

* Indicates significant difference (t-test, $p<0.01$) between HFE and LFE groups.

Gene expression profiles of HFE and LFE chickens

In total, 1.48 billion of 75-base sequence reads were generated, and for each sample approximately 64 million (ranging from 45 to 76 million) reads were obtained. Further, 86.8% of the reads from each sample were mapped uniquely to the chicken reference genome (Ensembl Galgal4). Among the mapped reads, 60.5% of the total reads were mapped to exon regions, 17.9% were mapped to the intergenic regions, and 8.3% were mapped to intronic regions (Figure 4.1 a). A total of 11,565 genes with at least five reads mapped per sample were detected in the RNA-seq libraries. The hierarchical clustering of samples based on the expression of all genes is presented in Figure 4.1 b.

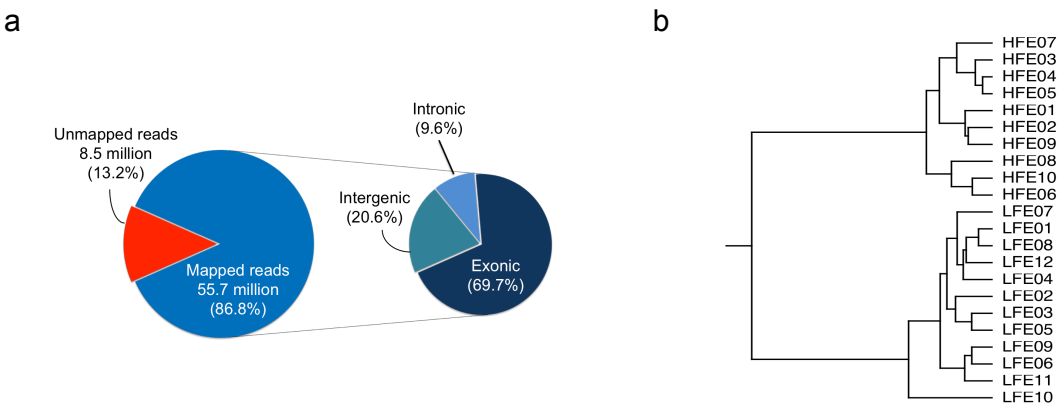

Figure 4.1 Summary of RNA-Seq data. a. The total number of sequence reads for each sample. b. Average mapping statistics. c. Hierarchical clustering of samples based on gene expression profile.

Consistency of samples within groups

To avoid the expression results being affected by outlier samples, we performed a separate systematic evaluation of consistency of all samples in the HFE and LFE groups. The correlation analysis based on the gene expression profiles found that one LFE sample (#LFE10) had the lowest correlation (r = 0.85) with other samples in the LFE group, whereas the correlations among other LFE samples were about 0.94. Sample LFE10 also had a lower correlation with HFE chickens when compared with other LFE samples. Consistently, hierarchical clustering results suggested that sample LFE10 was located on an isolated branch (Figure 4.1 b). The RNA-seq data and Nanostring results showed a correlation of 0.79 for this sample (compared with an average correlation coefficient of 0.78 between RNA-seq and NanoString), confirming that the RNA-seq data for sample LFE10 was reliable. The phenotypic data didn't rule

out sample LFE10 as an outlier; however, sample LFE10 had the lowest WG and breast muscle percentage among all selected chickens. It is likely that chicken LFE10 had a certain morbidity condition with unobservable symptoms, causing the deficiency in gaining weight and building breast muscle. Therefore, sample LFE10 might be an interesting case for further studies to investigate its uniqueness of gene expression patterns and the causes, but for the purpose of the present study, it was excluded from the further DE gene analysis. As a result, a total of 10 samples from the HFE group and 11 samples from the LFE group were used for DE gene analysis.

Identification and functional analysis of DE genes

Differential expression analysis between the HFE and LFE groups was carried out using Cuffdiff software. A total of 286 genes were found to be differentially expressed between HFE and LFE groups with a false discovery rate of 0.05 and a fold change larger than 1.3. Of these genes, 147 were up-regulated and 139 were down-regulated in LFE group. The top ten up- and down-regulated genes are listed in Table 4.3.

Table 4.3 Top 10 up- and down-regulated genes in LFE group

Ensembl gene ID	*Gene name*	Fold Change
Up-regulated genes		$FPKM_{LFE}$ / $FPKM_{HFE}$
ENSGALG00000009118	*PIT 54*	↑11.8
ENSGALG00000009266	*FGA*	↑10.6
ENSGALG00000008601	*AHSG*	↑10.4
ENSGALG00000003957	*APOH*	↑9.9
ENSGALG00000020180	*ALB*	↑9.8
ENSGALG00000019845	*GAL9*	↑9.7

ENSGALG00000008973	*AMBP*	↑9.2
ENSGALG00000011612	*GC*	↑9.0
ENSGALG00000009262	*FGB*	↑8.6
ENSGALG00000016667	*GAL10*	↑7.2

Down-regulated genes		$FPKM_{HFE}$ / $FPKM_{LFE}$
ENSGALG00000002614	*Unnamed*	↓3.4
ENSGALG00000012670	*NRSN1*	↓2.8
ENSGALG00000023622	*AVD*	↓2.6
ENSGALG00000016364	*FAM150B*	↓2.5
ENSGALG00000003212	*TSPO2*	↓2.2
ENSGALG00000019325	*Unnamed*	↓2.2
ENSGALG00000029151	*ISLR2*	↓2.1
ENSGALG00000026075	*AMER3*	↓2.0
ENSGALG00000001417	*CYP11A1*	↓2.0
ENSGALG00000015166	*GCNT1*	↓1.9

↑ indicates up-regulation in LFE group ↓ indicates down-regulation in LFE group

The DE gene list was analyzed using the IPA web application. A summary of IPA results is presented in Table 4.4. The noteworthy networks and functions identified include developmental disorder, hereditary disorder, cell-to-cell signaling and interaction, immune cell trafficking, inflammatory response, and lipid metabolism. There were 17 significant canonical pathways ($p < 0.01$) (The top 10 is shown in Table 4.5), which are involved in lipid metabolism, immune regulation, blood coagulation, and amino acid biosynthesis. The details of networks, functions, pathways, and related genes are further disused in the text.

Table 4.4 Summary of top networks and molecular and cellular functions from IPA. results

Top networks

ID	Associated Network Functions	Score[1]
1	Developmental Disorder, Hematological Disease, Hereditary Disorder	40
2	Cardiovascular System Development and Function, Organismal Development, Cell-to-Cell Signaling and Interaction	35
3	Drug Metabolism, Lipid Metabolism, Molecular Transport	33
4	Organismal Injury and Abnormalities, Tissue Morphology, Reproductive System Development and Function	28
5	Cellular Movement, Immune Cell Trafficking, Inflammatory Response	27

Molecular and cellular functions

Name	p-value[2]	# molecule[3]
Lipid Metabolism	2.31E-07 - 6.53E-03	30
Molecular Transport	2.31E-07 - 6.53E-03	31
Small Molecule Biochemistry	2.31E-07 - 7.46E-03	34
Vitamin and Mineral Metabolism	3.37E-06 - 5.54E-03	12
Cellular Movement	4.52E-06 - 7.17E-03	31

[1]Scores were calculated by IPA to rank the relevancy of DE genes and networks.
[2]p-values were calculated with a Fisher-extract test contingency table by IPA.
[3]# molecule indicates the number of DE genes involved in the molecular and cellular function

Table 4.5 Top 10 canonical pathways

Ingenuity canonical pathways	p-value[1]	Ratio[2]
LXR/RXR activation	1.00E-10	1.01E-01
Acute phase response signaling	5.25E-07	6.63E-02
Cholesterol biosynthesis i	1.07E-05	1.00E-01
Cholesterol biosynthesis ii (via 24,25-dihydrolanosterol)	1.07E-05	1.00E-01
Cholesterol biosynthesis iii (via desmosterol)	1.07E-05	1.00E-01
Superpathway of cholesterol biosynthesis	2.04E-05	5.75E-02
Extrinsic prothrombin activation pathway	2.63E-05	1.82E-01
Zymosterol biosynthesis	2.82E-05	1.36E-01
Intrinsic prothrombin activation pathway	2.63E-04	1.08E-01
Oleate biosynthesis ii (animals)	2.95E-04	1.67E-01

[1]p-values were calculated with a Fisher-extract test contingency table by IPA.
[2]Ratio = number of DE genes mapped to the pathway/total number of genes of the pathway.

Higher accumulation of lipids in LFE birds

The mean abdominal fat weight and percentage of the LFE group were significantly larger than that of the HFE group. This can be attributed to an overall higher accumulation of lipids in LFE birds (Figure 4.2 a). Among the DE genes, a lipid hydrolysis gene [*monoglyceride lipase* (*MGLL*)] and genes involved in high-density lipoprotein (HDL) synthesis [*lecithin-cholesterol acyltransferase* (*LCAT*), *apolipoprotein A-I* (*APOA1*), and lysophosphatidic *acid receptor 1* (*LPAR1*)] and steroid hormone synthesis [*cytochrome P450, family 11, subfamily A* (*CYP11A1*)] were down-regulated, whereas lipid synthesis genes [*1*-acylglycerol-*3-phosphate O-acyltransferase 9* (*AGPAT9*), *stearoyl-CoA desaturase (delta-9-desaturase)* (*SCD*), and *diacylglycerol O-acyltransferase homolog 2 (mouse)* (*DGAT2*)] and a gene that

stimulates the uptake of fatty acids and adipogenesis [peroxisome *proliferator-activated receptor gamma* (*PPARG*)] were up-regulated (indicated in red in Figure 4.2 a) in LFE group. These findings suggest that the up-regulation of genes involved in lipid synthesis and the down-regulation of genes involved in triglyceride hydrolysis and reverse cholesterol transport from adipose tissue were responsible for the higher accumulation of lipids in abdominal fat in LFE group.

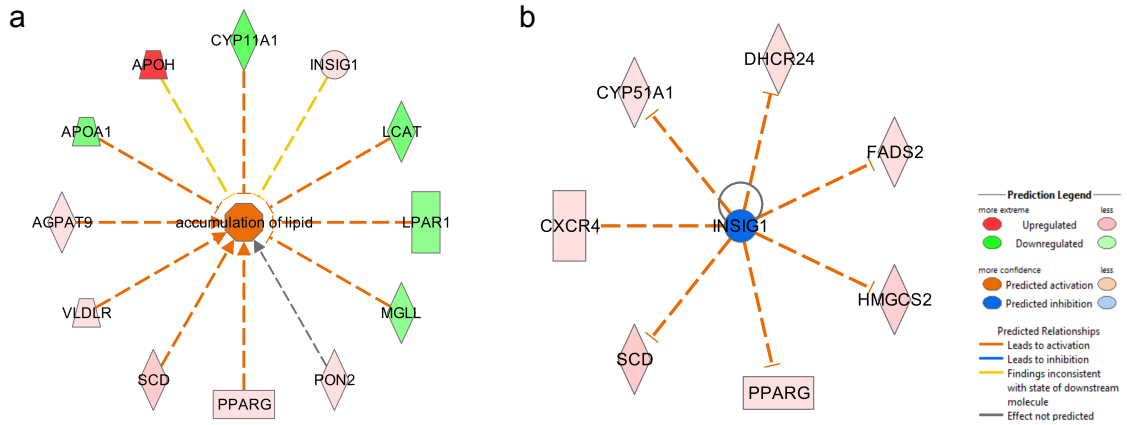

Figure 4.2 The DE genes involved in accumulation of lipid and cell movement of mononuclear leukocytes and upstream regulator INSIG1. a. The accumulation of lipid is predicted to be activated in LFE group. b. The cell movement of mononuclear leukocytes is predicted to be inhibited in LFE group. c. Upstream regulator INSIG1. cholesterol biosynthesis regulator INSIG1 is predicted to be inhibited in LFE chickens.

Over-represented pathways

The IPA identified 17 canonical pathways that were significant with a p-value of less than 0.01. These pathways are involved in lipid metabolism (LXR/RXR

activation, oleate biosynthesis II), cholesterol biosynthesis (cholesterol biosynthesis I/II/III, superpathway of cholesterol biosynthesis, zymosterol biosyntheis), amino acid synthesis (serine biosynthesis, superpathway of serine and glycine biosynthesis), coagulation (intrinsic/extrinsic prothrombin activation pathway, coagulation system), and endocrine functions (estrogen biosynthesis, atherosclerosis signaling, axonal guidance signaling, retinoate biosynthesis I). These pathways will be selectively discussed further later in the text.

Upstream regulators

Based on the DE genes, five transcription regulators [(*HNF 1 homeobox A* (*HNF1A*), *sterol regulatory element binding transcription factor 1* (*SREBF1*), *sterol regulatory element binding transcription factor 2* (*SREBF2*), *E2F transcription factor 1* (*E2F1*), and *fibroblast growth factor 2* (*FGF2*)), Tcf 1/3/4, and SREBP cleavage-activating protein (SCAP) were predicted to be activated in LFE group. The genes *phosphatase and tensin homolog* (*PTEN*), *interleukin 1* (*IL1*), *Tumor protein p53 (TP53),* and *insulin induced gene 1* (*INSIG1*) were predicted to be inhibited in LFE group (Figure 4.2 b).

Confirmation of RNA-seq experiment

We confirmed the gene expression results obtained from RNA-seq data using the Nanostring nCounter analysis system. The normalized Nanostring gene count showed a strong correlation with the FPMK values of RNA-seq. The correlation coefficient between fold change of the gene count and fold change of FPKM values was 0.92 (Figure 4.3). Based on the 31 DE genes from RNA-Seq analysis, the correlation between FPMK and gene count was 0.93 (Figure 4.4). The results showed

a high consistency between the two technologies, and confirmed that the gene expression data of RNA-Seq was reliable.

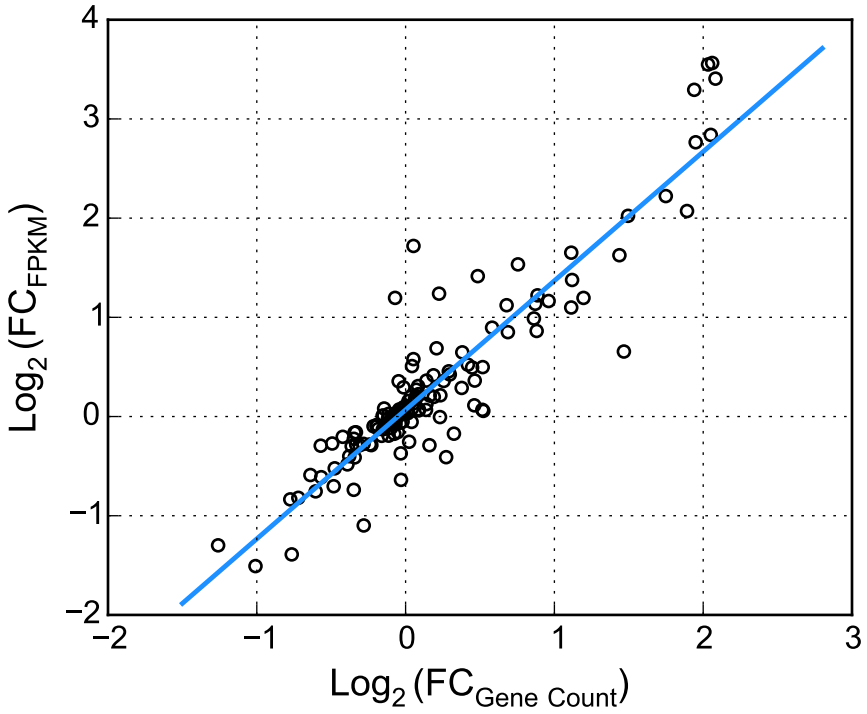

Figure 4.3 Correlations of log2 fold-change between RNA-seq FPKM and Nanostring gene count.

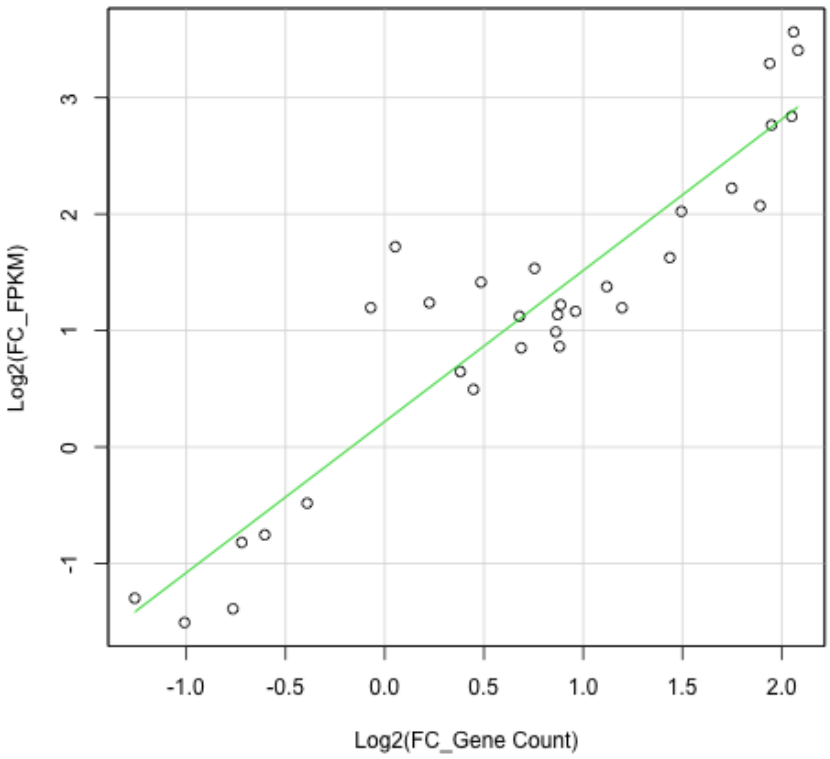

Figure 4.4 Correlations of log2 fold change (FC) of DE genes between RNA-seq FPKM and Nanostring gene count.

4.5 Discussion

Consistent with previous observations [26–28], our data showed a negative correlation between fatness and FE. Comparing the HFE and LFE groups, the BWs were not significantly different, but FC, WG and abdominal fat percentage were significantly different. The LFE chickens had more feed intake (1.15 fold) and deposited less breast muscle (0.93 fold) but more abdominal fat (1.55 fold). The LFE

chickens appeared to partition the energy obtained from diet to accumulate more fat but build less breast muscle than HFE chickens. RNA-seq analysis of the selected chickens revealed that genes involved in *de novo* triglyceride synthesis, cholesterol synthesis, lipid transport, and lipid stabilization were up-regulated, whereas genes involved in lipid hydrolysis and lipid reverse efflux were down-regulated in the abdominal adipose tissue of LFE birds (Table 4.6). Also, several genes related to coagulation, immune system, amino acid metabolism, and carbohydrate metabolism were differentially expressed between LFE and HFE groups.

Table 4.6 Summary of DE genes involved in lipid accumulation.

Functional category	Gene Name	Full Name	RNA-Seq fold change *
Fatty acid transportation	*FABP1*	*Fatty acid binding protein 1*	↑5.7
Stabilization of fatty acid	*ALB*	*Albumin*	↑9.8
de novo triglyceride synthesis	*SCD*	*stearoyl-CoA desaturase 1*-acylglycerol-*3-phosphate*	↑2.6
	AGPAT9	*O-acyltransferase 9*	↑1.4
	DGAT2	*diacylglycerol O-acyltransferase homolog 2*	↑3.1
Triglyceride hydrolysis	*MGLL*	*monoglyceride lipase*	↓1.4
Cholesterol synthesis	*DHCR24*	*24-dehydrocholesterol reductase*	↑1.7
	HSD17B7	*17-beta hydroxysteroid (17-beta) dehydrogenase 7*	↑1.7
	CYP51A1	*cytochrome P450, family 51, subfamily A, polypeptide 1*	↑1.5

	HMGCS2	*3-hydroxy-3-methylglutaryl-CoA synthase 2*	↑2.4
Cholesterol transport	*APOA1*	*apolipoprotein A-I*	↓1.6
	LCAT	*lecithin-cholesterol acyltransferase*	↓1.7
Steroidogenesis	*CYP11A1*	*cytochrome P450, family 11, subfamily A*	↓2.0
	TSPO2	*Translocator protein 2*	↓2.2
Adipogenesis	*PPARG*	*peroxisome proliferator-activated receptor gamma*	↑1.6
	FSTL1	*follistatin-like 1*	↓1.5
	KLF15	*Kruppel-like transcription factors 15*	↑1.5

↑ indicates up-regulation in LFE group, fold change = $FPKM_{LFE}$ / $FPKM_{HFE}$;
↓ indicates down-regulation in LFE group, fold change = $FPKM_{HFE}$ / $FPKM_{LFE}$

Triglyceride and cholesterol metabolism

By comparing the gene expression in the adipose tissue of the HFE and LFE groups, we identified the DE genes that may be responsible for the differences in fatness. IPA predicted the accumulation of lipids in LFE group is activated (activation z-score: 2.14). (Figure 4.2 a). *SCD*, *AGPAT9,* and *DGAT2* are three important genes involved in *de novo* triglyceride synthesis. All of the three genes were expressed at higher levels in LFE birds, with a fold change ($FPKM_{LFE}$ / $FPKM_{HFE}$) of 2.6, 1.4 and 3.1, respectively. SCD is a lipogenic enzyme located on the membrane of the endoplasmic reticulum (ER). It catalyzes the rate-limiting step of mono-unsaturated fatty acid (MUFA) biosynthesis from saturated fatty acids (SAFAs) [29]. The expression of *SCD* is closely associated with adiposity in previous studies [15,30]. AGPAT9 catalyzes the first, and DGAT2 catalyzes the last step of triglyceride synthesis. DGAT2 is located in the

proximity of SCD in the ER membrane, where SCD facilitates substrates transport for triglyceride synthesis [31]. *DGAT2* expression could be affected by available energy sources in cells. In fasted chickens, the expression levels of *DGAT2* in adipose tissue were much lower [16]. According to our FC records, the LFE group consumed 1.15 fold (i.e., ~430 grams) more feed than did HFE group. The relatively more abundant dietary energy resource might promote the *de novo* biosynthesis of triglycerides in the adipocytes of LFE birds through up-regulation of *DGAT2*. Consistent with our results, *SCD* and *DGAT2* were found down-regulated in Leghorn, a relatively lean line, when compared with a relatively fat line, i.e. a commercial broiler line [18].

FABP1 functions as a carrier protein for fatty acids, which transfer the fatty acids across the cell membranes. Increased *FABP1* expression was found in the adipose tissue of obese people who had high acylation stimulating protein and high triglyceride levels in a fasting plasma test [32]. More FABP1 might facilitate the transfer of fatty acid uptake in the adipose tissue and contribute to the accumulation of triglycerides. On the other hand, the expression levels of *ALB* were higher in LFE group. Knockdown or point mutations of the fatty acid binding site of albumin in cultured adipocytes suppressed lipid droplet formation, suggesting the role of albumin is to promote the formation of lipid droplets by binding to fatty acids [33]. The higher expression of albumin in LFE group suggests a similar lipid-stabilizing role of albumin in adipocytes of chickens, as exists in mammals.

Several genes involved in cholesterol metabolism were differentially expressed between HFE and LFE group. Adipose tissue is the largest site for free, un-esterified cholesterol storage [34]. There were three cholesterol biosynthesis pathways over-represented in LFE group, and all four DE genes involved in those pathways

were up-regulated, suggesting a relatively higher cholesterol synthesis activity in the adipocytes of LFE group (Table 4.6). In particular, the expression of *3-hydroxy-3-methylglutaryl-CoA synthase 2 (HMGCS2*) in LFE group was 2.4 fold ($FPKM_{LFE}$ / $FPKM_{HFE}$) higher than that of HFE birds. HMGCS2 catalyzes the production of 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA), a precursor for the rate-limiting step of cholesterol biosynthesis. The expression of *24-Dehydrocholesterol Reductase* (*DHCR24*), which encodes for the final enzyme in the cholesterol biosynthesis pathway, was 1.7 times higher in LFE group. Furthermore, we found several down-regulated genes that may contribute to cholesterol deposition through lower conversion in LFE chickens (Table 4.6). As a major component of HDL, APOA1 starts the formation of HDL by lipidation, and LCAT is responsible for turning the lipidated particles into spherical shapes [35]. Previous studies have shown that *APOA1* expression in liver was higher in a fat line of chickens [36,37]. Down-regulation of the expression of *APOA1* and *LCAT* may affect the formation of HDL, which reduces the capacity of reverse transportation of cholesterol from adipose tissue to liver and muscle, and results in more free cholesterol stored in the abdominal fat of LFE birds. As an endocrine organ, a very important function of adipose tissue is the production of steroid hormones. We found the expression of *CYP11A1* was lower in LFE group. The enzyme encoded by *CYP11A1*, P450scc, is the rate-limiting enzyme for converting cholesterol to pregnenolone (3β-hydroxypregn-5-en-20-one) [38]. Pregnenolone is a neurosteroid and a precursor of several steroid hormones. With a fold change ($FPKM_{HFE}$ / $FPKM_{LFE}$) of 2.0, decreased expression of *CYP11A1* in LFE chickens would reduce the rate of cholesterol conversion to pregnenolone and cause more cholesterol to be stored in the adipocytes in LFE group. Collectively, our data

indicates that more triglycerides and cholesterol were stored in the form of lipid droplets, causing hypertrophic growth of adipocytes.

Upstream regulators of cholesterol synthesis pathway

Sterol regulatory element-binding proteins (SREBPs) and INSIGs are key transcription factors in the regulation of cholesterol metabolism. IPA predicted SREBP1 (z-score = 2.529, overlap p-value = 3.43E-04) and SREBP2 (z-score = 2.449, overlap p-value = 5.23E-05) as being activated but INSIG1 as being inhibited in the abdominal fat tissue of LFE birds (Figure 4.2 b). In mammals, the SREBP1 and SREBP2 genes encode for three different protein isoforms with different target genes [39,40]. Located on the ER membrane, INSIG1 regulates cholesterol biosynthesis by sensing the sterol level. With sterols present, INSIG1 binds to the complex of SREBP and SREBP chaperone (SCAP) and keeps it on the ER membrane. Without sterol, INSIG1 is isolated rather than binding to the SREBP and SCAP complex and thus is subjected to ubiquitination and degradation [41]. The free SERBP migrates to Golgi to be further processed and, subsequently, enters the nucleus and activates genes involved in cholesterol and fatty acid metabolism [42,43], including *INSIG1*. In turn, *INSIG1* expression reduces lipid production and adipogenesis *in vitro* [44]. As a negative regulator of cholesterol synthesis, inhibition of INSIG1 by degradation may trigger activation of SREBP1 and SREBP2, and assist in the higher accumulation of cholesterol. In addition, SREBP proteins activate the expression of *INSIG1* to compensate for the degraded INSIG1 and maintain the level of INSIG1 [41]. Consistent with that, *INSIG1* was up-regulated in LFE chickens ($FPKM_{LFE}/FPKM_{HFE} = 1.4$). Hence, a

self-regulating loop may be present in adipocytes to maintain the cholesterol amount in an appropriate level.

Hyperplastic growth

Adipocyte hypertrophy might be a prominent contributor to abdominal fat mass [45], but adipocyte hyperplasia could also play a role. In particular, the adipose tissue of broiler chickens have hypertrophic and hyperplasic growth until 14 weeks of age [46]. In the present study, the LFE group had a higher expression level of *PPARG* ($FPKM_{LFE}$ / $FPKM_{HFE}$ = 1.56 fold). PPARG is an extremely important regulator in lipid metabolism and adipogenesis. It is required for the development of adipose tissue [47], as it is involved in both differentiation of preadipocytes and proliferation of adipocytes. Previous research has shown that *PPARG* expression in the adipose tissue of chickens is strongly correlated with abdominal fat pad weight [48]. It's possible that higher expression of *PPARG* increases the differentiation and proliferation of adipocytes, causing a multiplication of adipocytes in LFE birds. In agreement, *follistatin-like 1* (*FSTL1*) was expressed lower in LFE group. The expression of *FSTL1* is down-regulated during pre-adipocyte to adipocyte differentiation [49]. Additional support of hyperplasic growth comes from up-regulation of *Kruppel-like transcription factors 15* (*KLF15*). KLF15 is recognized as a regulator of PPARG, reflected by the strong correlation between *KLF15* and *PPARG* expression in our data (r = 0.80). *KLF15* expression is up-regulated during preadipocyte and adipocyte differentiation, and interruption of KLF15 decreases *PPARG* expression and affects differentiation[50]. Based on our collective data, we propose that hyperplasia may also contribute to the higher accumulation of abdominal fat mass in LFE birds.

Amino acid and carbohydrate metabolisms

A few DE genes encodes for key enzymes in amino acid and carbohydrate metabolism. Three genes [*tyrosine aminotransferase* (*TAT*), *phosphoserine phosphatase* (*PSPH*), and *argininosuccinate lyase* (*ASL2*)] were associated with the biosynthesis of tyrosine, serine, and arginine, respectively. Two DE genes were found involved in carbohydrate metabolism. *Amylase alpha 2A* (*AMY2A*) was expressed 2.2-fold ($FPKM_{LFE}$ / $FPKM_{HFE}$) higher in LFE chickens. AMY2A catalyzes the first step in the breakdown of large polysaccharides, including glycogen. The restoration of lipids for lipid-depleted adipocytes requires the accumulation of a certain amount of glycogen, possibly followed by glucose-to-lipid conversion [51]. The higher expression of *AMY2A* possibly indicates that glucose-to-lipid conversion is more active in LFE chickens. In addition, the expression of *AHSG* was found with great difference between HFE and LFE group ($FPKM_{LFE}$ / $FPKM_{HFE}$ = 10.4). Encoded by *AHSG*, alpha-2-HS-glycoprotein is involved in glucose metabolism and the regulation of insulin signaling. Knockout of *AHSG* induces glucose tolerance and decreased body fat [52]. AHSG may affect glucose uptake and lipid oxidation in adipocytes through regulation of adiponectin and may have an impact on fat deposition in LFE chickens.

4.6 Conclusion

In summary**,** our FE tests of commercial broiler chickens suggest a moderate correlation between abdominal fat percentage and feed efficiency. Compared with HFE chickens, LFE chickens had higher feed intake and deposited more abdominal fat but less breast muscle. The higher feed intake may play a role by increasing the lipid

concentration in blood circulation and promote fat deposition in LFE birds, but other triggers of differential gene expression between HFE and LFE chickens remain to be studied. To the best of our knowledge, this is the first study of the relationships between gene expression in adipose tissue and FE. The results of our study provide mechanistic insights into the biological basis of differences in adiposity between HFE and LFE chickens. In addition, as the adipose tissue of human and chicken share certain physiological features and gene homology, our findings regarding chicken adipose tissue could potentially be useful for studies of obesity in humans.

4.7 Acknowledgements

We thank University of Delaware Sequencing and Genotyping Center for sequencing our samples, Dr. Greg Gonye for helping with analyzing Nanostring data and Weixuan Fu for his contribution to screening candidate transcript sequences for Nanostring probe designs.

REFERENCES

1. De Verdal, H. *et al.* Selection for excretion traits in chicken. in *9th World Congress on Genetics Applied to Livestock Production* 1–6 (2010).

2. Kong, B.-W. *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.* **90,** 2535–47 (2011).

3. Bottje, W. G. *et al.* Gene expression in breast muscle associated with feed efficiency in a single male broiler line using a chicken 44K microarray. II. Differentially expressed focus genes. *Poult. Sci.* **91,** 2576–87 (2012).

4. Bottje, W. & Kong, B. CELL BIOLOGY SYMPOSIUM : Feed efficiency : Mitochondrial function to global gene expression. *J. Anim. Sci.* 1582–1593 (2013). doi:10.2527/jas2012-5787

5. Zhou, N., Lee, W. R. & Abasht, B. Messenger RNA sequencing and pathway analysis provide novel insights into the biological basis of chickens' feed efficiency. *BMC Genomics* **16,** 195 (2015).

6. Mohamed-Ali, V., Pinkney, J. H. & Coppack, S. W. Adipose tissue as an endocrine and paracrine organ. *Int. J. Obes. Relat. Metab. Disord. J. Int. Assoc. Study Obes.* **22,** 1145–58 (1998).

7. Kershaw, E. E. & Flier, J. S. Adipose tissue as an endocrine organ. *J. Clin. Endocrinol. Metab.* **89,** 2548–56 (2004).

8. Shrago, E. & Spennetta, T. The carbon pathway for lipogenesis in isolated adipocytes from rat, guinea pig, and human adipose tissue. *Am. J. Clin. Nutr.* **29,** 540–5 (1976).

9. O'Hea, E. K. & Leveille, G. A. Lipogenesis in isolated adipose tissue of the domestic chick (Gallus domesticus). *Comp. Biochem. Physiol.* **26,** 111–20 (1968).

10. Griffin, H. D., Guo, K., Windsor, D. & Butterwith, S. C. Adipose tissue lipogenesis and fat deposition in leaner broiler chickens. *J. Nutr.* **122,** 363–8 (1992).

11. Leveille, G. A., O'Hea, E. K. & Chakbabarty, K. In vivo lipogenesis in the domestic chicken. *Proc. Soc. Exp. Biol. Med. Soc. Exp. Biol. Med. N. Y. N* **128,** 398–401 (1968).

12. Simon, J. Insulin in birds: metabolic effects and possible implications in genetically fat and lean chickens. in *Leanness in Domestic Birds* (eds. Leclerq, B. & Whitehead, C. C.) 253–268 (Butterworth-Heinemann, 1988). doi:10.1016/B978-0-408-01036-8.50028-2

13. Simon, J. Chicken as a useful species for the comprehension of insulin action. *Crit. Rev. Poult. Biol. USA* **2,** 121–148 (1989).

14. Wang, H.-B. *et al.* Profiling of chicken adipose tissue gene expression by genome array. *BMC Genomics* **8,** 193 (2007).

15. Resnyk, C. W. *et al.* Transcriptional analysis of abdominal fat in genetically fat and lean chickens reveals adipokines, lipogenic genes and a link between hemostasis and leanness. *BMC Genomics* **14,** 557 (2013).

16. Ji, B. *et al.* Transcriptomic and metabolomic profiling of chicken adipose tissue in response to insulin neutralization and fasting. *BMC Genomics* **13,** 441 (2012).

17. Hausman, G. J., Barb, C. R., Fairchild, B. D., Gamble, J. & Lee-Rutherford, L. Gene expression profiling in adipose tissue from growing broiler chickens. *Adipocyte* **3,** 297–303 (2014).

18. Ji, B. *et al.* Molecular and metabolic profiles suggest that increased lipid catabolism in adipose tissue contributes to leanness in domestic chickens. *Physiol. Genomics* **46,** 315–27 (2014).

19. Mutryn, M. F., Brannick, E. M., Fu, W., Lee, W. R. & Abasht, B. Characterization of a novel chicken muscle disorder through differential gene expression and pathway analysis using RNA-sequencing. *BMC Genomics* **16,** 399 (2015).

20. FastQC [http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc].

21. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14,** R36 (2013).

22. DeLuca, D. S. *et al.* RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinforma. Oxf. Engl.* **28,** 1530–2 (2012).

23. Trapnell, C. *et al.* Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **31,** 46–53 (2012).

24. Ingenuity® Systems [www.ingenuity.com].

25. Geiss, G. K. *et al.* Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat. Biotechnol.* **26,** 317–25 (2008).

26. Thomas, C. H., Blow, W. L., Cockerham, C. C. & Glazener, E. W. The Heritability of Body Weight, Gain, Feed Consumption, and Feed Conversion in Broilers. *Poult. Sci.* **37,** 862–869 (1958).

27. Washburn, K. W., Guill, R. a & Edwards, H. M. Influence of genetic differences in feed efficiency on carcass composition of young chickens. *J. Nutr.* **105,** 1311–7 (1975).

28. Pym, R. A. E. & Solvyns, A. J. Selection for food conversion in broilers: Body composition of birds selected for increased body-weight gain, food consumption and food conversion ratio. *Br. Poult. Sci.* **20,** 87–97 (1979).

29. Heinemann, F. S. & Ozols, J. Stearoyl-CoA desaturase, a short-lived protein of endoplasmic reticulum with multiple control mechanisms. *Prostaglandins Leukot. Essent. Fatty Acids* **68,** 123–33 (2003).

30. Miyazaki, M., Kim, Y. C., Gray-Keller, M. P., Attie, a D. & Ntambi, J. M. The biosynthesis of hepatic cholesterol esters and triglycerides is impaired in mice with a disruption of the gene for stearoyl-CoA desaturase 1. *J. Biol. Chem.* **275,** 30132–8 (2000).

31. Man, W. C., Miyazaki, M., Chu, K. & Ntambi, J. Colocalization of SCD1 and DGAT2: implying preference for endogenous monounsaturated fatty acids in triglyceride synthesis. *J. Lipid Res.* **47,** 1928–1939 (2006).

32. MacLaren, R. E., Cui, W., Lu, H., Simard, S. & Cianflone, K. Association of adipocyte genes with ASP expression: a microarray analysis of subcutaneous and omental adipose tissue in morbidly obese subjects. *BMC Med. Genomics* **3,** 3 (2010).

33. Yoo, W. *et al.* Albumin expression is required for adipocyte differentiation of 3T3-L1 cells. *Biochem. Biophys. Res. Commun.* **397,** 170–5 (2010).

34. Krause, B. R. & Hartman, A. D. Adipose tissue and cholesterol metabolism. *J. Lipid Res.* **25,** 97–110 (1984).

35. Zannis, V. I., Chroni, A. & Krieger, M. Role of apoA-I, ABCA1, LCAT, and SR-BI in the biogenesis of HDL. *J. Mol. Med. Berl. Ger.* **84,** 276–94 (2006).

36. Douaire, M. *et al.* Identifying genes involved in the variability of genetic fatness in the growing chicken. *Poult. Sci.* **71,** 1911–20 (1992).

37. Daval, S., Lagarrigue, S. & Douaire, M. Messenger RNA levels and transcription rates of hepatic lipogenesis genes in genetically lean and fat chickens. *Genet. Sel. Evol. GSE* **32,** 521–31 (2000).

38. Li, J., Daly, E., Campioli, E., Wabitsch, M. & Papadopoulos, V. De novo synthesis of steroids and oxysterols in adipocytes. *J. Biol. Chem.* **289,** 747–64 (2014).

39. Brown, M. S. & Goldstein, J. L. The SREBP pathway: regulation of cholesterol metabolism by proteolysis of a membrane-bound transcription factor. *Cell* **89,** 331–40 (1997).

40. Espenshade, P. J. & Hughes, A. L. Regulation of sterol synthesis in eukaryotes. *Annu. Rev. Genet.* **41,** 401–27 (2007).

41. Gong, Y. *et al.* Sterol-regulated ubiquitination and degradation of Insig-1 creates a convergent mechanism for feedback control of cholesterol synthesis and uptake. *Cell Metab.* **3,** 15–24 (2006).

42. Kovanen, P. T., Nikkilä, E. A. & Miettinen, T. A. Regulation of cholesterol synthesis and storage in fat cells. *J. Lipid Res.* **16,** 211–23 (1975).

43. Horton, J. D. *et al.* Combined analysis of oligonucleotide microarray data from transgenic and knockout mice identifies direct SREBP target genes. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 12027–32 (2003).

44. Li, J., Takaishi, K., Cook, W., McCorkle, S. K. & Unger, R. H. Insig-1 'brakes' lipogenesis in adipocytes and inhibits differentiation of preadipocytes. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 9476–81 (2003).

45. Cartwright, A. L. Adipose cellularity in Gallus domesticus: investigations to control body composition in growing chickens. *J. Nutr.* **121,** 1486–97 (1991).

46. Hood, R. L. The cellular basis for growth of the abdominal fat pad in broiler-type chickens. *Poult. Sci.* **61,** 117–21 (1982).

47. Rosen, E. D. *et al.* PPAR gamma is required for the differentiation of adipose tissue in vivo and in vitro. *Mol. Cell* **4,** 611–7 (1999).

48. Sato, K. *et al.* Changes in peroxisome proliferator-activated receptor gamma gene expression of chicken abdominal adipose tissue with different age, sex and genotype. *Anim. Sci. J. Nihon Chikusan Gakkaihō* **80,** 322–7 (2009).

49. Wu, Y., Zhou, S. & Smas, C. M. Downregulated expression of the secreted glycoprotein follistatin-like 1 (Fstl1) is a robust hallmark of preadipocyte to adipocyte conversion. *Mech. Dev.* **127,** 183–202 (2010).

50. Mori, T. *et al.* Role of Krüppel-like factor 15 (KLF15) in transcriptional regulation of adipogenesis. *J. Biol. Chem.* **280,** 12867–75 (2005).

51. Birsoy, K. *et al.* Cellular program controlling the recovery of adipose tissue mass: An in vivo imaging approach. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 12985–90 (2008).

52. Mathews, S. T. *et al.* Improved insulin sensitivity and resistance to weight gain in mice null for the Ahsg gene. *Diabetes* **51,** 2450–8 (2002).

Chapter 5

CONCLUSIONS AND FUTURE RESEARCH

In this dissertation, we have studied variation in gene expression associated with fundamental genomic problem and economically important traits in chickens. We reported the absence of genomic imprinting and the relative prevalence of ASE in chicken Day 12 embryonic brain and liver. Additionally, we found there was no significant difference of gene expression between paternal and maternal alleles at chromosome-wide, and genome-wide levels. Examination of the expression ratio between females and males and between sex chromosome and autosomes confirmed that dosage compensation is incomplete in chickens and excluded the possibility of imprinted Z chromosome inactivation. Using the same dataset, we identified additivity as the predominant transgenerational gene expression pattern between F1 crosses and inbred parental lines. The DE genes between F1 crosses and parental lines that may contribute to heterotic phenotype were identified, and most DE genes showed additive and dominant gene expression pattern. Brain and liver had distinct preference of genotype for the dominant genes. At last, we looked into chicken abdominal adipose tissue and identified the DE genes between HFE and LFE broiler chickens. We discussed the enriched functions and pathways and their relationship with divergent fat deposition between HFE and LFE chickens.

One direction for continuing this research is to examine ASE in adipose tissue and identify its relationship with fatness or feed efficiency. Combining of ASE and DE research will be helpful to dissect the main regulatory mechanism controlling gene

expression. Identifying the DNA variants that are responsible for ASE and DE and establishing its relationship with phenotype is also crucial to agricultural improvement, for those causal variants could be used as markers to assist selection and breeding. Meanwhile, since gene expression is resulting from interplay between genetic variants and environmental influence, investigation of epigenetic modifications may generate new knowledge regarding genomic imprinting, ASE and DE, and help to understand the phenomenon itself or its relationship with traits of interest.

Appendix A

REPRINT PERMISSION FOR CHAPTER 2

Dear Zhu,

Thank you for your email. As outlined below you are welcome to use your material in your thesis. **Please make sure that the thesis is not made publicly available (repository, print copies in library, etc.) before the article has been published by us.** You may submit the thesis with the article for academic use only before the publication.

Thank you for contacting Springer Nature. As an author, you have the right to use this manuscript and figures, as per the licence-to-publish you signed:
Ownership of copyright in the article remains with the Authors, and provided that, when reproducing the Contribution or extracts from it, the Authors acknowledge first and reference publication in the Journal, the Authors retain the following non-exclusive rights:

a) To reproduce the Contribution in whole or in part in any printed volume (book or thesis) of which they are the author(s).
b) They and any academic institution where they work at the time may reproduce the Contribution for the purpose of course teaching.
c) To post a copy of the Contribution as accepted for publication after peer review (in Word or Tex format) on the Authors' own web site or institutional repository, or the Authors' funding body's designated archive, six months after publication of the printed or online edition of the Journal, provided that they also give a hyperlink from the Contribution to the Journals web site.
d) To reuse figures or tables created by them and contained in the Contribution in other works created by them.

The above use of the term 'Contribution' refers to the author's own version, not the final version as published in the Journal.

Kind regards,

Oda Siqveland
Permissions Assistant

SpringerNature
The Campus, 4 Crinan Street, London N1 9XW,
United Kingdom
T +44 (0) 207 014 6851
http://www.nature.com
http://www.springer.com
http://www.palgrave.com

Appendix B

REPRINT PERMISSION FOR CHAPTER 4

Dear Ms. Zhuo,

Thank you for your message. PLOS ONE publishes all of the content in the articles under an open access license called "CC-BY." This license allows you to download, reuse, reprint, modify, distribute, and/or copy articles or images in PLOS journals, so long as the original creators are credited (e.g., including the article's citation and/or the image credit). Additional permissions are not required. You can read about our open access license here: http://journals.plos.org/plosone/s/licenses-and-copyright

There are many ways to access our content, including HTML, XML, and PDF versions of each article. Higher resolution versions of figures can be downloaded directly from the article.

Thank you for your interest in PLOS ONE and for your continued support of the Open Access model. Please do not hesitate to be in touch with any additional questions.

Kind regards,

Sue Laborda
Staff EO
PLOS ONE

Case Number: 05418760