

RNaseq Reveals Novel and Differentially Expressed Isoforms In Native and Commercial Poultry

Ayeh Sadat Sadr¹, Mohammadreza Nassiri^{2*}, Seyed Alireza Salami³, Mohammad Reza Bakhtiarizadeh⁴,
Mojtaba Tahmoorespur², Alireza Shafeinia⁵

¹ Department of Animal Science, International Campus, Ferdowsi University of Mashhad, Mashhad, Iran

² Department of Animal Sciences, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran

³ Department of Horticulture, Faculty of Agriculture, University of Tehran, Tehran, Iran

⁴ Department of Animal and Poultry Science, College of Aburaihan, University of Tehran, Pakdasht, Tehran, Iran

⁵ Department of Biotechnology, Faculty of Agriculture, Ramin Agriculture and Natural Resources University, Ahwaz, Iran

Received 15 October 2016

Accepted 24 November 2016

Abstract

The poultry industry occupies an important position in the provision of animal protein. Recently, next generation sequencing technology (RNA-Seq) has become available as a powerful tool to investigate transcriptional profiles for gene expression analysis of many organisms. The main use of RNA-Seq in agriculture species are focusing on finding the immune related genes or pathways by comparison of the whole transcriptome following pathogen challenge. Alternative splicing (AS) is the major fundamental mechanism generating the protein diversity and regulating the gene expression in eukaryotic organism. Identifying genes that are differentially spliced between two groups of RNA-sequencing samples is interesting subject in transcriptome with next-generation sequencing technology in this study used RNA sequencing to comparison isoforms of two breeds. A total of 64,819 transcripts were identified by aligning sequence reads to genome among the evaluated isoforms for expression analysis, 310 were significantly differentially expressed between two breeds, including 251 up-regulated and 59 down-regulated. The KEGG results of up regulated isoforms showed that that no pathway was found significantly different ($FDR \leq 0.05$). However, enrichment analysis suggested that seven were over-represented ($P\text{-value} \leq 0.05$) within the up regulated isoforms. Only one of them functionally related to immune system, natural killer cell mediated cytotoxicity. The results showed genes which are breed-specific expression and the comparative transcriptome analysis help to understand the difference of genetic mechanism.

Keywords: Isoforms, RNASeq, Poultry

Introduction

The poultry industry occupies an important position in the provision of animal protein (meat and egg) to man as well as manure for crops and generally plays a vital role in the national economy as a revenue provider and provides employment (Mohammed and Sunday, 2015). Economic pressure on the modern poultry industry has directed the selection process towards fast-growing broilers that have reduced feed conversion ratio. Selection based on growth characteristics, could adversely affect immune competence leaving chickens more susceptible to disease. Since the innate immune response directs the acquired immune response, efforts to select poultry with an efficient innate immune response would be beneficial (Swaggerty et al., 2009). Indigenous (native) breeds of livestock have higher

disease resistance and adaptation to the environment due to high genetic diversity. Therefore, the conservation of diversity in the existing genetic resource is more important for economic and public health than for the development of new breeds with higher productivity (Jeong et al., 2014). Presently, several major issues confront the poultry industry in meeting the growing demands of consumers. Control of infectious disease and food safety is certainly at or near the top of the list (Cheng et al., 2013). The immune system is an adaptive defense system that evolved in phylogenesis to control an organism's integrity (Muir and Aggrey, 2003)

Mammalian pre-mRNA consists of protein coding regions, exons and intervening sequences, introns. The splicing process joins the exon sequences while

Corresponding authors E-mail:

*nassiry@um.ac.ir

removing the introns. Recent study revealed that most human genes have alternative splicing and can produce multiple isoforms of transcripts. Differences in the relative abundance of the isoforms of a gene can have significant biological consequences. Identifying genes that are differentially spliced between two groups of RNA-sequencing samples is interesting subject in transcriptome with next-generation sequencing technology (Wang et al., 2013). Alternative splicing (AS) is the major fundamental mechanism generating the protein diversity and regulating the gene expression in eukaryotic organism (Black, 2003; Cáceres and Kornblihtt, 2002). The pre-mRNA undergoes splicing in the nucleus where after removal of intronic sequences, exons are joined in different combinations, leading to generation of isoforms with distinct transcript structure. The proteins thus encoded by transcript isoforms vary in their structures as well as functions. This alteration in the protein structure and function resulting from aberrant or AS is commonly associated with diseases (Tazi et al., 2009). Recently, RNA sequencing (RNA-Seq) has considered a powerful tool to investigate transcriptional profiles of thousands of genes in many organisms (Truong et al., 2015)

Gene expression is a widely studied process and a major area of focus for functional genomics. The main use of RNA-Seq in economical aquaculture species are focusing on finding the immune related genes or pathways by comparison of the whole transcriptome following pathogen challenge (Li and Li, 2014).

In this study, we performed comparative gene expression analysis of native and commercial breed poultry to identify differentially expressed isoforms by RNA-Seq technology. Findings revealed significant expression differences in isoform expression levels. In addition, we detected novel splicing events and novel transcript structures that are not described previously.

Materials and Methods

RNA Extraction and Sequencing

In total, six chicken from Esfahani and six chicken from Ross breeds (47 days of age) were chosen for RNA-seq. The chicken raised on the farm of Safi Abad Agriculture and education Center Dezful Iran. These birds were kept under the same environmental and nutritional conditions. Five ml blood samples were collected from Brachial/ulnar wing vein. The total RNA was extracted using Trizol (Invitrogen, USA) following the manufacturer's instructions.

The RNA pool was prepared by mixing together equal quantities of three RNA samples per group to generate a total of 4 pooled RNA samples (two samples in each breed). The four RNA-seq libraries were sequenced based on protocols of Illumina HiSeq 2000 to generate 150 pair-end reads.

Differentially Expressed Isoforms

The quality of the raw data was checked with FastQC vol 0.11.2. Beside on these results, the Trimmomatic (v 0.35) were used to remove Illumina adaptors, trimming of reads as well as quality or filtering reads by removing low-quality reads (Bolger et al., 2014).

The reference genome for chicken (Galgal4) its corresponding annotation GTF file were downloaded from the Ensemble (<http://asia.ensembl.org/info/data/ftp/index.html>).

Clean data of pair-end reads from each sample were mapped to the reference genome using HISAT2 (v2-2.0.3) (Kim et al., 2015). Cufflinks (v2.2.1) were used to process the alignment files and estimate the abundance of the assembled transcripts as FPKM, (fragments per kilobase of exon per million fragments mapped). Cufflinks were used to normalize the number of fragments mapping to individual loci (Trapnell et al., 2010). Cufflinks includes a script called cuffmerge that can use to merge together several Cufflinks assemblies. Merged.gtf file produced by cuffmerge was provided as an input to cuffdiff along with alignment files produced by Hisat2 for differential analysis between two samples. Also cuffdiff labeled genes as significant or not significant based on the *p-value*. Visualization of expression and differential expression results were performed by CummeRbund package which accepted cuffdiff output (Trapnell et al., 2012).

Functional Annotation

To test for enrichment of GO terms of differentially expressed isoforms, David functional annotation tools (<https://david.ncifcrf.gov/>) was used. *P-value* reported by David were corrected to obtain FDR (Dennis et al., 2003).

Novel Isoform Detection

Assembled transcripts were annotated using Cuffcompare from Cufflinks. Cuffcompare simply compares the transcripts that have been assembled through Cufflinks to a reference annotation file. To minimize annotation artifacts, all single exon transcripts were excluded for further analysis. Also, all the transcripts smaller than 200 bp were removed.

Cuffcompare classified each transcript as known or novel and identified transcripts that are potential novel isoforms. The class codes in the Cuffcompare output were used to identify novel isoforms. The transcripts with class code “j” (locus is potentially a novel isoform) were considered as novel isoforms of known genes.

According to investigation of the poultry immune system, the tracking file provided by Cuffcompare was used to infer isoforms and splice variants unique to natural killer cell mediated cytotoxicity isoforms by examining manually which related to immune system. Novel splice variants detected by Cuffcompare were annotated manually by UCSC genome browser.

Results

Quality Analysis and RNA-Seq Data

To identify and compare the transcriptome of different chicken breeds, the blood samples were collected from Esfahani as native and Ross as commercial breed. Total RNA was extracted and the RNA pool was obtained. After sequencing, the average number of reads across all 4 samples (2 samples per breed) was approximately 17 million. After removal of low-quality reads, few of the sequence reads did not pass the quality filtering (Table1). Sequence reads were aligned to the chicken reference genome (Galgal4) using HISAT2, approximately 85% of sequenced fragments were aligned successfully (Table1).

Table 1. Summary of sequencing read alignments

	Native		Commercial	
	Sample1	Sample2	Sample1	Sample2
Total read	18768307	17995632	15923838	15621164
Reads after trimming	18509420	17650558	15677357	15313649
Read aligned to reference genome	86.68	84.89	84.36	84.17

Isoform Expression Analysis

After mapping the reads to the reference genome with HISAT2, transcripts were assembled and the relative expression level calculated by cufflinks. Gene expression intensity was estimated using FPKM method. Then the Cuffdiff was used to

calculate the differentially expressed isoforms.

A total of 64,819 transcripts were identified by aligning sequence reads to genome among the evaluated isoforms for expression analysis, 310 were significantly differentially expressed between two breeds, including 251 up-regulated and 59 down-regulated (figure1). The differentially expressed isoforms in commercial versus native breed with statistically significant fold changes ranged from -3.50894 to 3.75559. Top ten up and down regulated isoforms were presented in figure 2.

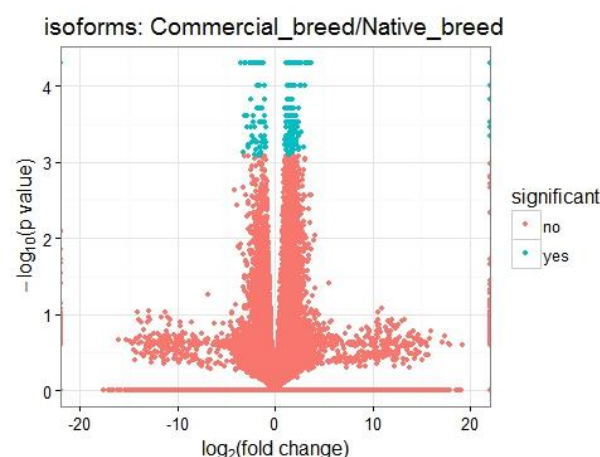


Figure1. Log₂-fold change between the commercial versus native breed. Light blue dots represent significantly differentially expressed isoforms between breeds (P -value < 0.05).

Pathway Analysis of Differentially Expression Isoforms

To gain insights into the biological processes and pathway that are enriched in differentially expressed isoforms, DAVID was used. Results were grouped in cellular component molecular function and biological process. Top five GO terms identified in up and down regulated isoforms are presented in table2. The KEGG results of up regulated isoforms showed that that no pathway was found significantly different ($FDR \leq 0.05$). However, enrichment analysis suggested that seven pathways including focal adhesion, regulation of actin cytoskeleton, lysosome, natural killer cell mediated cytotoxicity, VEGF signaling pathway, glycolysis/gluconeogenesis and ECM-receptor interaction were over-represented (P -value ≤ 0.05) within the up regulated isoforms. Only one of them functionally related to immune system, natural killer cell mediated cytotoxicity. Also, the pathway analysis results showed that no pathway were found significantly enriched for down regulated isoforms (corrected P -value ≤ 0.05).

Figure2. The most up regulated (n=10) and down regulated (n=10) differentially expressed isoforms between native vs commercial breeds. The differentially expressed isoforms were ranked based on their fold change.

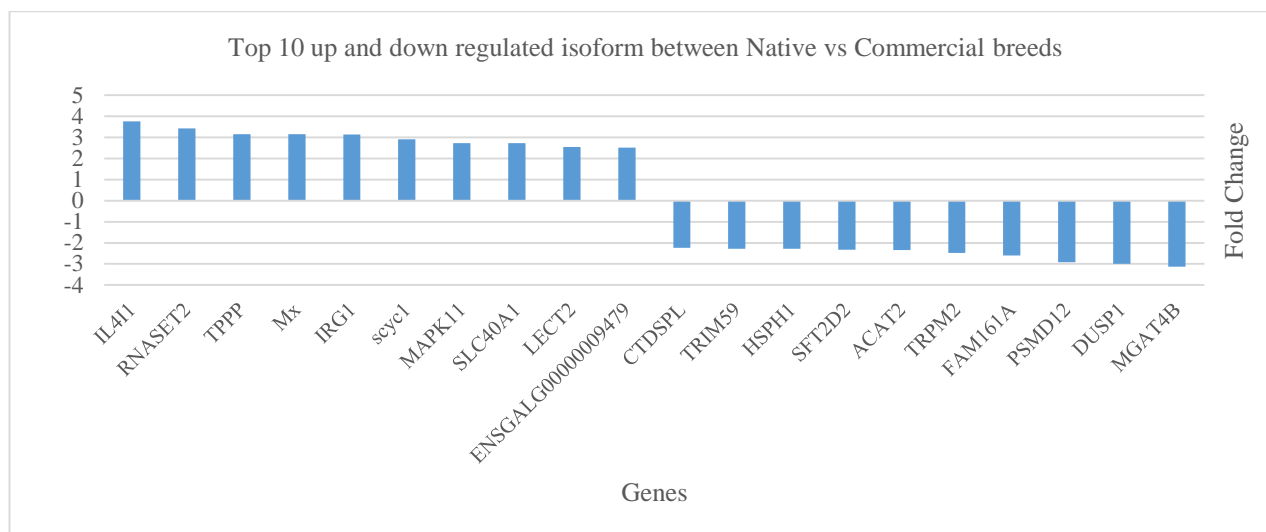


Table2. Top five GO terms significantly enriched in up and down regulated differentially expressed isoforms.

Term	Count	%	P value
Up regulated			
actin cytoskeleton organization	9	4.347826	1.51E-05
biological regulation	63	30.43478	2.48E-05
actin filament-based process	9	4.347826	2.61E-05
signal transduction	30	14.49275	3.54E-05
cytoskeleton organization	11	5.31401	3.65E-05
Down regulated			
histone modification	3	6.382979	0.005163
covalent chromatin modification	3	6.382979	0.005725
chromatin modification	3	6.382979	0.019129
post-translational protein modification	6	12.76596	0.02271
regulation of transcription	7	14.89362	0.031248

Novel Splice Variants

We obtained 64,819 unique alignments that were screened for novel isoforms using Cuffcompare. We found 37,057 isoforms already present in the annotation (corresponding to 10530 distinct genes). Moreover 27,762 novel isoforms were identified (corresponding to 7310 distinct genes), which have not been included in the gene annotation so far. Of these novel splice variants 11,734 were present in native breed, while 13,234 were present in the commercial breed. To gain a better understanding of the genes in immune system the genes in the natural killer cell mediated cytotoxicity pathway were analyzed and four genes with novel isoform were detected (PIK3CG, PLCG2, PTPN6 and PPP3CA) and showed alternative splicing (figure3).

Discussion

In present study, high throughput sequencing was performed to evaluate the transcription of two chicken breeds with emphasis on immune system. The immune system plays a key role in health maintenance and pathogenesis of a wide range of disease.

In this study the Ross as commercial and Esfahani as a native breed were used with difference in same traits. To achieve accurate detection of splicing variants and to optimally decode mechanisms underlying alternative splicing, it is invariably important to accurately map cDNA reads to their genomic counterparts.



Figure3. Blast analysis showing alternative splicing. Splice variants identified in different expressed genes with novel isoform in Natural killer cell mediated cytotoxicity pathway after submission of transcripts assembled by Cuffcompare as custom tracks to UCSC genome browser.

It is also important to note that an aligner for RNA-Seq data should efficiently align reads matching to splice junctions (Li and Homer, 2010). Here, we used an efficient aligner, HISAT2, to identify differentially expressed isoforms as well as novel isoforms. Until now, identification and differential expression analysis of genes involved in immune system has not been reported by comparison between two different breeds in poultry. In our best knowledge, this is the first study to identify differentially expressed isoforms between two chicken breeds. However, the different studies were carried out on poultry which most of them emphasis on disease (Perumbakkam et al., 2013; Sandford, 2011; Wang et al., 2014). As a results of mapping with HISAT2, 85.8% and

84.3% of reads were mapped in native and commercial breed respectively. Through comparison of the transcriptome data of the two breeds, 310 isoforms were significantly differentially expressed, which 251 isoforms were up-regulated in the commercial vs native breed and 59 genes were down-regulated in the native vs commercial breed. KEGG pathway analysis results showed that one immune related pathways was enriched among up-regulated isoforms (P -value ≤ 0.05), Natural killer cell mediated cytotoxicity with seven genes including PIK3CG, PTPN6, VAV3, RAC2, PLCG2, PIK3CD and PPP3CA. Natural killer (NK) cells are lymphocytes of the innate immune system that are involved in early defenses against both allogeneic (no self) cells and autologous

cells undergoing various forms of stress, such as infection with viruses, bacteria, or parasites or malignant transformation. One of the important aspect of RNA-Seq is discovered novel exon and novel exon boundaries. So that for this analysis all genes related to natural killer cell mediated cytotoxicity were analyzed. PIK3CG, PTPN6, PIK3CD and PPP3CA were detected as novel isoform. For splicing variants in these genes, the Cuffcompare as custom tracks to the UCSC genome browser was submitted and the blast was carried out. The blast results showed the different types of alternative splicing like exon skipping, exon insertion, 5 alternative splicing and intron retention which may change the protein sequence. For example the intron retention was observed in PTPN6 gene (figure3). The alternation in the protein sequence may thus lead to either gain of function, loss of function or change in the specify of the protein and its functional diversity. The results demonstrate the ability of RNA-Seq to detection of novel isoform.

Our study is the first investigation of poultry transcriptome to compare two different breeds in normal situation by using high throughput sequencing. The results showed genes which are breed-specific expression and the comparative transcriptome analysis help to understand the difference of genetic mechanism. In conclusion, the results of this study would be used to recognize gene candidates for further breed improvement.

References

- Black D. L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annual Review of Biochemistry* 72:291-336.
- Bolger A. M., Lohse M. and Usadel B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*:btu170.
- Cáceres J. F. and Kornblihtt A. R. (2002) Alternative splicing: multiple control mechanisms and involvement in human disease. *TRENDS in Genetics* 18:186-193.
- Cheng H. H., Kaiser P. and Lamont S. J. (2013) Integrated genomic approaches to enhance genetic resistance in chickens. *Annu. Rev. Anim. Biosci.* 1:239-260.
- Dennis G., Sherman B. T., Hosack D. A., Yang J., Gao W., Lane H. C. and Lempicki R. A. (2003) DAVID: database for annotation, visualization, and integrated discovery. *Genome biology* 4:1.
- Jeong H.-S., Kim D.-W., Chun S.-Y., Sung S., Kim H.-J., Cho S., Kim H. and Oh S.-J. (2014) Native Pig and Chicken Breed Database: NPCDB. *Asian-Australasian journal of animal sciences* 27:1394.
- Kim D., Langmead B. and Salzberg S. L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nature methods* 12:357-360.
- Li E. and Li C. (2014) Use of RNA-seq in aquaculture research. *Poultry Fish Wildlife Sci* 2:e108.
- Li H. and Homer N. (2010) A survey of sequence alignment algorithms for next-generation sequencing. *Briefings in bioinformatics* 11:473-483.
- Mohammed B. R. and Sunday O. S. (2015) An Overview of the Prevalence of Avian Coccidiosis in Poultry Production and Its Economic Importance in Nigeria. *Veterinary Research* 3:35-45.
- Muir W. M. and Aggrey S. E. 2003. *Poultry Genetics, Breeding, and Biotechnology*. CABI.
- Perumbakkam S., Muir W. M., Black-Pyrkosz A., Okimoto R. and Cheng H. H. (2013) 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* 14:1.
- Sandford E. E. (2011) Whole transcriptome response of chicken spleen and peripheral blood leukocytes to avian pathogenic *Escherichia coli*.
- Swaggerty C. L., Pevzner I. Y., He H., Genovese K. J., Nisbet D. J., Kaiser P. and Kogut M. H. (2009) Selection of broilers with improved innate immune responsiveness to reduce on-farm infection by foodborne pathogens. *Foodborne Pathogens and Disease* 6:777-783.
- Tazi J., Bakkour N. and Stamm S. (2009) Alternative splicing and disease. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* 1792:14-26.
- Trapnell C., Roberts A., Goff L., Pertea G., Kim D., Kelley D. R., Pimentel H., Salzberg S. L., Rinn J. L. and Pachter L. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols* 7:562-578.

17. Trapnell C., Williams B. A., Pertea G., Mortazavi A., Kwan G., Van Baren M. J., Salzberg S. L., Wold B. J. and Pachter L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* 28:511-515.
18. Truong A. D., Hong Y. H. and Lillehoj H. S. (2015) RNA-seq profiles of immune related genes in the spleen of Necrotic enteritis-afflicted chicken lines. *Asian-Australasian journal of animal sciences* 28:1496.
19. Wang W., Qin Z., Feng Z., Wang X. and Zhang X. (2013) Identifying differentially spliced genes from two groups of RNA-seq samples. *Gene* 518:164-170.
20. Wang Y., Lupiani B., Reddy S., Lamont S. and Zhou H. (2014) RNA-seq analysis revealed novel genes and signaling pathway associated with disease resistance to avian influenza virus infection in chickens. *Poultry science* 93:485-493.

Open Access Statement:

This is an open access article distributed under the Creative Commons Attribution License (CC-BY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.