



## Data in Brief

## Digital gene-expression of alfalfa saponin extract on laying hens

Wenna Fan<sup>1</sup>, Hongqi Du<sup>1</sup>, Lu Zhou, Pengfei Shi, Chengzhang Wang\*

College of Animal Science and Veterinary Medicine, Henan Agricultural University, Zhengzhou, Henan 450002, China

## ARTICLE INFO

## Article history:

Received 13 November 2014  
 Received in revised form 13 December 2014  
 Accepted 16 December 2014  
 Available online 27 December 2014

## Keywords:

Digital gene-expression  
 Alfalfa saponin  
 Laying hens

## ABSTRACT

Cardiovascular disease is a major cause of death worldwide, so people are advised to limit their intake of dietary cholesterol [1]. Egg consumption has been seriously reduced because of the high levels of cholesterol [2]. The objective of this study was to evaluate the cholesterol metabolism effects of alfalfa saponin extract (ASE) in liver and ovary tissues using digital gene-expression (DGE) profiling analysis. The liver and ovary tissues were isolated from laying hens fed with ASE for RNA sequencing. Here, we provide detailed experimental methods and analysis pipeline in our study to identify digital gene expression of alfalfa saponin extract on laying hens and analysis pipeline published by Singh and colleagues in the *PLOS ONE* [3]. The data generated in our work provide meaningful information for understanding the molecular mechanisms underlying the cholesterol-lowering effects of ASE.

© 2014 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Specifications	
Organism/cell line/tissue	The liver and ovary tissues of laying hens (Hy-Line Brown hens)
Sex	Female
Sequencer or array type	Illumina Hiseq 2000
Data format	Raw data: FASTQ files
Experimental factors	Fed with ASE based on basal diet vs. a basal diet with no ASE
Experimental features	The liver and ovary tissues of laying hens were collected and digital gene expression (DGE) was performed to investigate the mechanisms underlying the cholesterol-lowering effect of ASE.
Consent	N/A
Sample source location	All animal experiments were approved by the Animal Care and Use Committee of Henan Agricultural University. ASE was provided by Hebei Bao'en Biotechnology Co., Ltd. (Shijiazhuang, China)

## Direct link to deposited data

Deposited data can be found here:

[http://www.ncbi.nlm.nih.gov/biosample?Db=biosample&DbFrom=bioproject&Cmd=Link&LinkName=bioproject\\_biosample&LinkReadableName=BioSample&ordinalpos=1&idsFromResult=240129](http://www.ncbi.nlm.nih.gov/biosample?Db=biosample&DbFrom=bioproject&Cmd=Link&LinkName=bioproject_biosample&LinkReadableName=BioSample&ordinalpos=1&idsFromResult=240129)

## Experimental design, materials and methods

## Experimental design

Cardiovascular disease is a major cause of death worldwide, so people are advised to limit their intake of dietary cholesterol [1]. Egg consumption has been seriously reduced because of the high levels of cholesterol [2]. ASE can modulate cholesterol metabolism and decrease cholesterol levels in animals. In this study, the liver and ovary tissues of laying hens were collected and digital gene expression (DGE) was performed to investigate the mechanisms underlying the cholesterol-lowering effect of ASE. 150 Hy-Line Brown hens at the age of 27 weeks were randomly divided into five groups (five replicates were established per treatment, with 6 hens each replicate). Hens in each treatment were randomly fed five different experimental diets: Controls were fed a basal diet with 0 mg/kg ASE, and the other four groups were fed otherwise identical diets supplemented with 60 mg/kg, 120 mg/kg, 240 mg/kg, and 480 mg/kg ASE, respectively. The basal diet was designed to meet, or exceed, the National Research Council (NRC) nutrient requirements. The ingredients and calculated

\* Corresponding author at: No. 95 Wenhua Road, College of Animal Science and Veterinary Medicine, Henan Agricultural University, Zhengzhou, Henan 450002, China. Tel.: +86 371 63558180.

E-mail addresses: [chou0516@163.com](mailto:chou0516@163.com) (W. Fan), [biodhq@126.com](mailto:biodhq@126.com) (H. Du), [zhoulu06@gmail.com](mailto:zhoulu06@gmail.com) (L. Zhou), [bird5682850@163.com](mailto:bird5682850@163.com) (P. Shi), [wangcz@henau.edu.cn](mailto:wangcz@henau.edu.cn) (C. Wang).

<sup>1</sup> Equal contribution.

**Table 1**  
Composition and nutrient content of diet.

Ingredient	%	Nutrient level <sup>b</sup>
Corn	69.70	Crude protein (%) 16.50
Soybean meal	14.45	Calcium (%) 3.44
Limestone power	8.50	Phosphorus (%) 0.59
Peruvian fishmeal	4.96	Available phosphorus (%) 0.42
Dicalcium phosphate	1.00	Lysine (%) 0.83
Salt	0.20	Methionine (%) 0.38
Choline chloride	0.10	Methionine + cysteine (%) 0.63
DL-methionine	0.09	Apparent metabolic energy (MJ/kg) 11.49
Trace mineral and vitamin premix <sup>a</sup>	1.00	

<sup>a</sup> Premix provided the following per kilogram of diet: 60 mg of iron, 80 mg of manganese, 8 mg of copper, 80 mg of zinc, 1 mg of iodine, 0.3 mg of selenium, 12,200 IU of vitamin A, 4125 IU of vitamin D<sub>3</sub>, 30 IU of vitamin E, 4.5 mg of vitamin K, 5 mg of vitamin B<sub>12</sub>, 2 mg of biotin, 5 mg of folic acid, 32.5 mg of niacin, 5.3 mg of pantothenic acid, 8 mg of pyridoxine, 8.5 mg of riboflavin, and 1 mg of thiamine.

<sup>b</sup> Obtained by calculation.

nutrient level of the basal diet are shown in Table 1 [3]. ASE was provided by Hebei Bao'en Biotechnology Co., Ltd. (Shijiazhuang, China) as a standardized product that contained 61.64% saponins, 10.97% flavonoids, 8.12% polysaccharides, 7.11% moisture, and 12.16% unknown compounds. ASE is extracted from *Medicago sativa* L. The main active component is saponins. And ASE played a positive role in lowering yolk cholesterol content [4–6]. According to the cholesterol levels in yolk of feeding experiment, the control group and 120 mg/kg ASE group were chosen for digital gene-expression profiling analysis (Table 2) [3]. The liver and ovary tissues of 120 mg/kg ASE group were isolated, frozen immediately in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until RNA extraction (the liver from the control group (Li\_CK), the liver from the 120 mg/kg ASE group (Li\_Exp), the ovaries from the control group (Ov\_CK), the ovaries from the 120 mg/kg ASE group (Ov\_Exp)).

#### Total RNA isolation

Total RNA was isolated using the Trizol reagent (Invitrogen, Code No.: 15596-018) according to the manufacturer's instructions. RNA degradation and contamination were monitored on 1% agarose gels. RNA was assessed for quantity and quality using a Nanodrop ND-1000 spectrophotometer and Agilent 2100 Bioanalyzer according to the manufacturer's instructions. Contaminating genomic DNA in total RNA samples were treated with DNase-I.

#### Generation of sequencing data

2  $\mu\text{g}$  of total RNA ( $\geq 100\text{ ng}/\mu\text{L}$ ) was pooled from three tissues within each group: Li\_CK, Li\_Exp, Ov\_CK, Ov\_Exp. Four sequencing libraries were generated using NEB Next Ultra Directional RNA Library Prep Kit for Illumina (NEB) in accordance with the manufacturer's recommendations. MRNA was purified from total RNA using poly-T oligo-attached magnetic beads (Life technologies, CA, USA). Fragmentation was carried out using divalent cations under elevated temperature in NEB proprietary fragmentation

**Table 2**  
Dietary ASE and concentrations of yolk cholesterol (mg/g).

Parameter period	ASE in diet (mg/kg)				
	0	60	120	240	480
Day 15	8.89 $\pm$ 0.21	8.20 $\pm$ 0.41	8.27 $\pm$ 0.07	8.46 $\pm$ 0.31	8.49 $\pm$ 0.19
Day 30	9.23 $\pm$ 0.24 <sup>a</sup>	8.22 $\pm$ 0.25 <sup>b</sup>	8.35 $\pm$ 0.59 <sup>b</sup>	8.49 $\pm$ 0.61 <sup>b</sup>	8.49 $\pm$ 0.13 <sup>b</sup>
Day 45	9.33 $\pm$ 0.24 <sup>a</sup>	9.09 $\pm$ 0.26 <sup>ab</sup>	8.78 $\pm$ 0.32 <sup>b</sup>	8.87 $\pm$ 0.11 <sup>b</sup>	9.02 $\pm$ 0.46 <sup>ab</sup>
Day 60	9.57 $\pm$ 0.41 <sup>a</sup>	9.37 $\pm$ 0.25 <sup>ab</sup>	8.89 $\pm$ 0.45 <sup>b</sup>	9.47 $\pm$ 0.39 <sup>ab</sup>	9.31 $\pm$ 0.36 <sup>ab</sup>

<sup>a,b</sup>Means in the same row not sharing a common superscript differ significantly at  $P < 0.05$ .

buffer. First strand cDNA was synthesized using random oligonucleotides and M-Mel Reverse Transcriptase (RNase H-). Second strand cDNA synthesis was subsequently performed using DNA polymerase I and RNase H. Remaining overhangs were converted into blunt ends via exonuclease/polymerase activities and enzymes were removed. After adenylation of 3' ends of DNA fragments, NEB Next adaptor oligonucleotides were ligated to cDNA fragments. In order to select cDNA fragments. DNA fragments with ligated adaptor molecules on both ends were selectively enriched using NEB Universal PCR Primer and Index primer in a 10 cycles PCR reaction. Products were purified and quantified using the Agilent high sensitivity DNA assay on the Agilent Bioanalyzer 2100 system. After the clustering of the indexed samples, each library preparation was sequenced on an Illumina HiSeq 2000 platform and 50 bp single-end (SE) reads were generated.

#### Mapping and analysis of DGE reads

For each DGE library, clean tags were obtained by filtering out adaptor sequences and low-quality tags. Then the clean tags were retained and mapped to the reference genome of *Gallus gallus* at [ftp://ftp.ensembl.org/pub/release-72/fasta/gallus\\_gallus/dna](ftp://ftp.ensembl.org/pub/release-72/fasta/gallus_gallus/dna) by TopHat v1.4.0 [7]. HTSeq v0.5.3 was used to count the reads mapped to each gene [8]. Then reads per kilobase transcriptome per million mapped reads (RPKM) of each gene were calculated based on the length of the gene and reads count mapped to this gene. It is currently the most common method of for estimating levels of gene expression [9]. In order to balance the number of false positives and false negatives a RPKM threshold value of 0.3 was established to determine whether or not given genes was expressed, as in several other studies [10–12]. Differential expression analysis of two conditions was performed using the DEGSeq R package v1.12.0 [13]. P-values were adjusted using Q-value. [14]. The Q-values of 0.05 and  $\log_2$  (Fold\_change) with no limitations were served as the threshold of significance for differential expression. Here, 110 significantly expressed genes were identified in liver samples; 66 were up-regulated and 44 were down-regulated. In ovary samples, 107 significantly expressed genes were obtained; 63 were up-regulated and 44 were down-regulated (Table S1) [3]. Gene Ontology (GO) enrichment analysis of differentially expressed genes was implemented by the Goseq R package [15]. We used KOBAS software to test the statistical enrichment of differential expression genes in KEGG pathways [16,17]. KEGG pathway analysis of differently expressed genes was also performed (Table S2) [3]. Protein interaction networks of differentially expressed genes were acquired based on the information in STRING database (<http://string-db.org/>). Cytoscape software was used to visualize the interaction networks [18]. The process of DGE analysis was depicted in Fig. 1.

#### Discussion

Based on the supplementing 120 mg/kg ASE played a positive role in lowering yolk cholesterol content [3], the group with 120 mg/kg ASE was used for digital gene-expression profiling analysis. In this study,

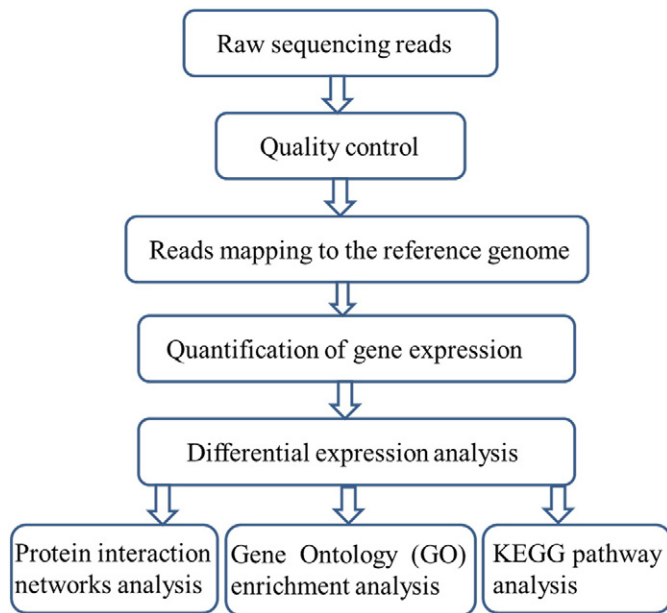


Fig. 1. The process of DGE analysis.

some genes were found to be associated with the cholesterol metabolism. In order to shed more light on the functional roles of differentially expressed genes responsible for the cholesterol-lowering effect of ASE, functional annotation was investigated using GO classification and KEGG pathway analysis between Ov\_CK and Ov\_Exp samples [3]. Our results provided that ASE were partially mediated by enhancement of cholesterol efflux in the liver and this reduced of cholesterol deposition in the ovary.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.gdata.2014.12.003>.

#### Conflict of interest

The authors declare no conflict of interest.

#### Acknowledgments

This work was funded by the Earmarked Fund for Modern Agro-industry Technology Research System (CARS-35).

#### References

- [1] N.H. Lung, B. Institute, Third report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III): final report. *Circulation* 106 (2002) 3143.
- [2] J.D. Spence, D.J. Jenkins, J. Davignon, Egg yolk consumption and carotid plaque. *Atherosclerosis* 224 (2012) 469–473.
- [3] L. Zhou, Y. Shi, R. Guo, M. Liang, X. Zhu, C. Wang, Digital gene-expression profiling analysis of the cholesterol-lowering effects of alfalfa saponin extract on laying hens. *PLoS ONE* 9 (2014) e98578.
- [4] M. Malinow, P. McLaughlin, L. Papworth, C. Stafford, G. Kohler, A. Livingston, P. Cheeke, Effect of alfalfa saponins on intestinal cholesterol absorption in rats. *Am. J. Clin. Nutr.* 30 (1977) 2061–2067.
- [5] M. Malinow, W. Connor, P. McLaughlin, C. Stafford, D. Lin, A. Livingston, G. Kohler, W. McNulty, Cholesterol and bile acid balance in *Macaca fascicularis*. Effects of alfalfa saponins. *J. Clin. Invest.* 67 (1981) 156.
- [6] C.-H. Yu, G. Xie, R.-R. He, Y.-J. Zhai, Y.-F. Li, B. Tsoi, H. Kurihara, D.-P. Yang, Effects of a purified saponin mixture from alfalfa on plasma lipid metabolism in hyperlipidemic mice. *J. Health Sci.* 57 (2011) 401–405.
- [7] C. Trapnell, L. Pachter, S.L. Salzberg, TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25 (2009) 1105–1111.
- [8] S. Anders, HTSeq: analysing high-throughput sequencing data with Python. <http://www-huber.embl.de/users/anders/HTSeq/doc/overview.html>2010.
- [9] A. Mortazavi, B.A. Williams, K. McCue, L. Schaeffer, B. Wold, Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5 (2008) 621–628.
- [10] M. Beyer, M.R. Mallmann, J. Xue, A. Staratschek-Jox, D. Vorholt, W. Krebs, D. Sommer, J. Sander, C. Mertens, A. Nino-Castro, High-resolution transcriptome of human macrophages. *PLoS ONE* 7 (2012) e45466.
- [11] L.T. Sam, D. Lipson, T. Raz, X. Cao, J. Thompson, P.M. Milos, D. Robinson, A.M. Chinnaiyan, C. Kumar-Sinha, C.A. Maher, A comparison of single molecule and amplification based sequencing of cancer transcriptomes. *PLoS ONE* 6 (2011) e17305.
- [12] H.L. Wright, H.B. Thomas, R.J. Moots, S.W. Edwards, RNA-Seq reveals activation of both common and cytokine-specific pathways following neutrophil priming. *PLoS ONE* 8 (2013) e58598.
- [13] S. Anders, W. Huber, Differential expression analysis for sequence count data. *Genome Biol.* 11 (2010) R106.
- [14] J.D. Storey, The positive false discovery rate: a Bayesian interpretation and the q-value. *Ann. Stat.* (2003) 2013–2035.
- [15] M. Young, M. Wakefield, G. Smyth, A. Oshlack, Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 11 (2010) R14.
- [16] X. Mao, T. Cai, J.G. Olyarchuk, L. Wei, Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* 21 (2005) 3787–3793.
- [17] M. Kanehisa, S. Goto, KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28 (2000) 27–30.
- [18] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13 (2003) 2498–2504.