Reprod Dev Biol 37(4) : 185-192 (2013)
http://dx.doi.org/10.12749/RDB.2013.37.4.185
ISSN : 1738-2432 (Print)
ISSN : 2288-0151 (Online)

Gene Co-Expression Network Analysis of Reproductive Traits in Bovine Genome

Dajeong Lim[1,†], Yong-Min Cho[1], Seung-Hwan Lee[2], Han-Ha Chai[1] and Tae-Hun Kim[1]

[1]*Division of Animal Genomics and Bioinformatics, National Institute of Animal Science, RDA, Suwon 441-706, Korea*
[2]*Hanwoo Experiment Station, National Institute of Animal Science, RDA, Pyungchang, 232-950, Korea*

ABSTRACT

Many countries have implemented genetic evaluation for fertility traits in recent years. In particular, reproductive trait is a complex trait and need to require a system-level approach for identifying candidate genes related to the trait. To find the candidate gene associated with reproductive trait, we applied a weighted gene co-expression network analysis from expression value of bovine genes. We identified three co-expressed modules associated with reproductive trait from bovine microarray data. Hub genes (ZP4, FHL2 and EGR4) were determined in each module; they were topologically centered with statistically significant value in the gene co-expression network. We were able to find the highly co-expressed gene pairs with a correlation coefficient. Finally, the crucial functions of co-expressed modules were reported from functional enrichment analysis. We suggest that the network-based approach in livestock may an important method for analyzing the complex effects of candidate genes associated with economic traits like reproduction.

(Key words : Reproduction, Gene Co-expression Network (GCN), Cattle)

INTRODUCTION

A good reproductive performance is crucial for economic as well as ethical reasons. Without reproduction there will be no animal production. Therefore, In recent years many other countries have also implemented genetic evaluation for reproductive traits. The cattle breeding objectives has gradually shifted from production to functional traits such as pregnancy. From this point of view it is important to increase knowledge of genes and their regulation. Reproduction trait is also a complex trait, which is contributed from many genes. Therefore, it is the agreement that no single factor determines a large proportion of the trait variation in the population (Hwang *et al*., 2008). From this reason, systems biology approach has been useful to identify genes that underlie complex trait from network of genetic interactions among all possible genes. Furthermore, patterns of covariation in the expression of multiple loci can be used to build networks that show relationships between genes, and between genes and functional traits. These networks provide information on the genetic control of complex traits and can help identify causal genes that affect gene function rather than gene expression (Haley and de Koning, 2006). System-oriented approaches have been applied by animal geneticists to investigate livestock traits (Nobis *et al*., 2003; Donaldson *et al*., 2005; Smith and Rosa, 2007), resulting in the identification and characterization of economically important causal *trans*-acting genes within QTL regions. These *trans*-QTL regions share a common biological function (e.g., similar gene ontology function, metabolic pathway, transcriptional co-regulation) (Schadt *et al*., 2003; Gibson and Weir, 2005; Subramanian *et al*., 2005). In the case of bovines, several countries identify quality challenges. Three genes were identified as being significantly correlated with bovine skeletal muscle based on microarray data from a gene network (Reverter *et al*., 2006). Jiang *et al*. (2009) (Jiang *et al*., 2009) reported that the genetic network was associated with 19 economically important beef traits. This report suggested the three candidate gene approach as targets. Therefore, we need to systemic approach in order to identify candidate genes

* To whom correspondence should be addressed.
* This work was supported by Agenda (PJ906956) of the National Institute of Animal Science, Rural Development Administration (RDA), Republic of Korea.
† Corresponding author : Phone: +82-31-290-1606, E-mail: lim.dj@korea.kr

in the network analysis among many genes related to reproduction within QTL intervals. A gene co-expression network (GCN) is a gene correlation network created from expression profiling, with each gene having several neighbors and is useful for identifying genes that control quantitative phenotypes. In this study, we introduce a systemic approach involving network analysis of reproduction-related genes and suggest co-expressed pattern of hub genes in the network.

MATERIALS AND METHODS

Our analysis involved three main steps: finding candidate genes in the Animal QTL database and analyzing the results of microarray experiments from the Gene Expression Omnibus (GEO) database, using these genes and co-expression information to construct co-expression networks related to the "pregnancy rate" trait and analyzing the network topology and functional enrichment for modules using DAVID tool and CytoScape.

Identification of Candidate Genes Associated with the Pregnancy Rate

To determine candidate genes associated with the reproduction within QTL intervals, we obtained genomic positions of the "pregnancy rate" trait using "QTL location by bp" information from the Animal QTL database (http://www.genome.iastate.edu/cgi-bin/QTLdb/BT/index). Most of QTLs are identified in the different regions in a chromosome. There are rare regions of overlap. Therefore, we select the genes associated with pregnancy rate from Animal QTL database with QTL IDs that have marker information in term of 'pregnancy rate' within Animal Trait Ontology (ATO) category. In the GEO database (http://www.ncbi.nlm.nih.gov/geo/), all data from microarray experiments related to bovines were used: GEO series (GSE) 15544, GSE 15342, GSE 13-725, GSE 6918, GSE 10695, GSE 12327, GSE 9256, GSE 12688, GSE 11495, GSE 11312, GSE 7360, GSE 9344, and GSE 8442. All arrays were processed to determine the robust multiarray average (RMA) (Irizarry *et al*., 2003) using the "affy" software package (Gautier *et al*., 2004). Expression values were computed in detail from raw CEL files by applying the RMA model of probe-specific correction for perfect-match probes. These corrected probe values were then subjected to quantile normalization, and a median polish was applied to compute one expression measure from all probe values. Resulting RMA expression values were log2-transformed. We determined mean intensity for an expression intensity of each gene matching to at least two probes. Finally, we used 221 genes associated with pregnancy rate for network construction.

Gene Co-Expression Network Construction and Network Module Identification

In co-expression networks, we refer to nodes as genes whose degrees indicate the number of links connected by the node. We extracted expression values for 221 genes and evaluated pairwise correlations between the gene expression profiles of each pair of genes using Pearson's correlation coefficients (denoted as *r*). The unweighted network encoded gene co-expression as binary information (connected=1, unconnected=0) using a "hard" threshold. In contrast, the weighted network represented "soft" thresholding that weighed each connection as a continuous number [0, 1]. The power adjacency function $\alpha_{i,j}=|\mathrm{cor}(x_i, x_j)^{\beta}$ was used to construct a weighted network as the connection strength between two genes. We investigated soft thresholding with the power adjacency function and selected a power of beta (β)=7. A major aim of co-expression network analysis is to determine subsets of nodes (modules) that are tightly connected to each other. To organize genes into modules, we used a module identification method based on a topological overlap dissimilarity measure (Ravasz *et al*., 2002) in conjunction with a clustering method, which detected biologically meaningful modules. The topological overlap of two nodes refers to their relative interconnectedness. The topological overlap matrix (TOM) $\Omega=[\omega_{ij}]$ provides a similarity measure, which has proven useful in biological networks (Ye and Godzik, 2004), where $l_{ij}=\sum_u \alpha_{iu}\alpha_{uj}$ and $k_i=\sum_u \alpha_{iu}$ is the node connectivity as follows:

$$\omega_{ij} = \frac{l_{ij} + a_{ij}}{\min\{k_i, k_j\} + 1 - a_{ij}}$$

In the case of our network, l_{ij} equals the number of nodes to which both i and j are connected. To identify modules, we used TOM-based dissimilarity d_{ij}^{w} ($d_{ij}^{w}=1-\omega_{ij}$) in a hierarchical cluster analysis. Each module represents a group of genes with similar expression profiles across the samples and the expression profile pattern is distinct from those of other modules. The weighted gene co-expression analysis (WGCNA) software packages for R were used to identify co-expression values associated with pregnancy rate (Peter and Steve, 2008). The degree distribution of a network has a generalized power-law form $p(k)\sim k^{-r}$, which is the defining property of a scale-free network (Barabasi and Albert, 1999). The genes of highly connected nodes to nodes with few connections (hubs) play an important role as a local property in a network (Barabasi and Oltvai, 2004). A node with high BC has great influence over

what flows in the network; BC may play a major role as a global property since it is a useful indicator for detecting bottlenecks in a network. For node k, BC is the fraction of the number of shortest paths that pass through each node (Brandes, 2001) and is defined as

$$b(k) = \sum_{i,j} b_{i \to j}(k) = \sum_{i,j} \frac{g_{i \to j}^{k}}{g_{i \to j}},$$

where $g_{i \to j}$ is the number of the shortest geodesic paths from node i to node j, $g_{i \to j}^{k}$ is the number of geodesic paths among $g_{i \to j}$ from node i to node j that pass through node k. We calculated BC as global properties according to all nodes in a network. From the results of the network topology analysis, we selected high-degree nodes and high-centrality nodes as key drivers that are most associated with our trait of interest in the network.

Functional Enrichment Analysis

We performed functional enrichment analysis against the 221 genes that were associated with pregnancy rate in the Gene Ontology terms using the Database for Annotation Visualization and Integrated Discovery (DAVID) tool (http://david.abcc.ncifcrf.gov/). To construct network for functional network, the *Enrichment Map Cytoscape plugin* (Merico *et al.*, 2010) was used to build networks of inter-related terms based on the number of genes shared between gene ontology (GO) terms. Terms were indicated as nodes (circles). Edges linking nodes represented gene sharing, and their thickness, the degree of gene set overlap.

RESULTS AND DISCUSSION

Identification of the Global Co-Expression Network

We constructed a weighted gene co-expression network associated with the pregnancy rate using soft threshold. we found that the results of the weighted network analysis were highly robust to the selection of the soft parameter β when it was used for module identification, connectivity definition, and to identify the relationship between intra-modular connectivity, as in previous studies (Ghazalpour *et al.*, 2006; Keller *et al.*, 2008; Presson *et al.*, 2008).

The hard threshold may cause a loss of information and sensitivity because of the choice of threshold and artifact from clustering coefficient result. The nodes represent candidate genes obtained from the animal QTL database and microarray data, and the links between the nodes represent the association between expression profiles across all microarray samples. The absolute value of Pearson's correlation coefficient was calculated for all pairwise comparisons. This correlation matrix was transformed into a matrix of adjacency using "soft" threshold (β, 8.0), producing a gene co-expression network. The network follows a power-law ($P(k) \sim k^{-r}$) degree distribution, where r is the degree exponent and ~ indicates "proportional to." We examined whether the co-expression network followed a power-law distribution with an exponent of approximately −1.75 (Stelzl

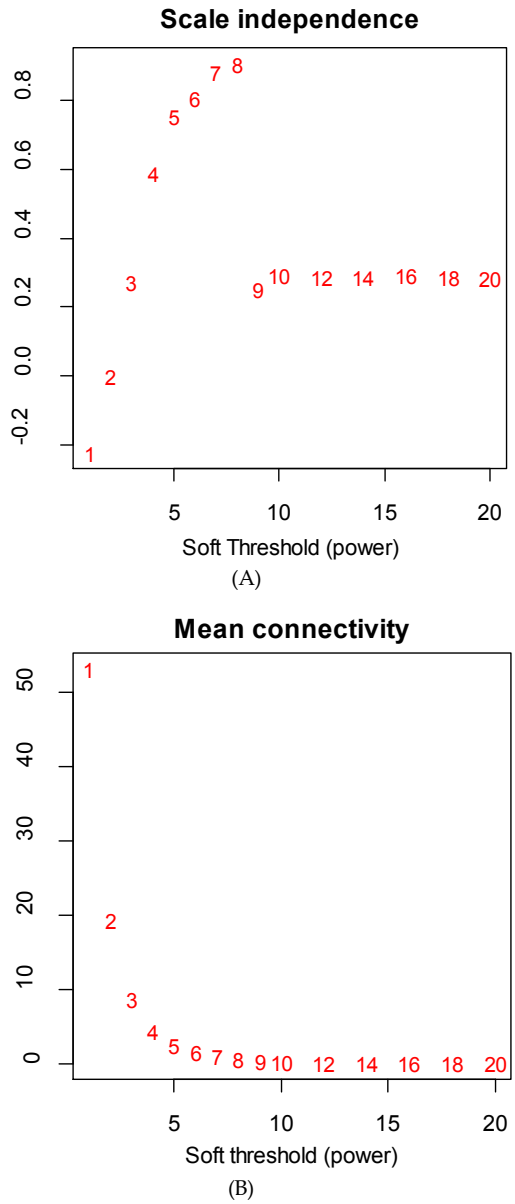
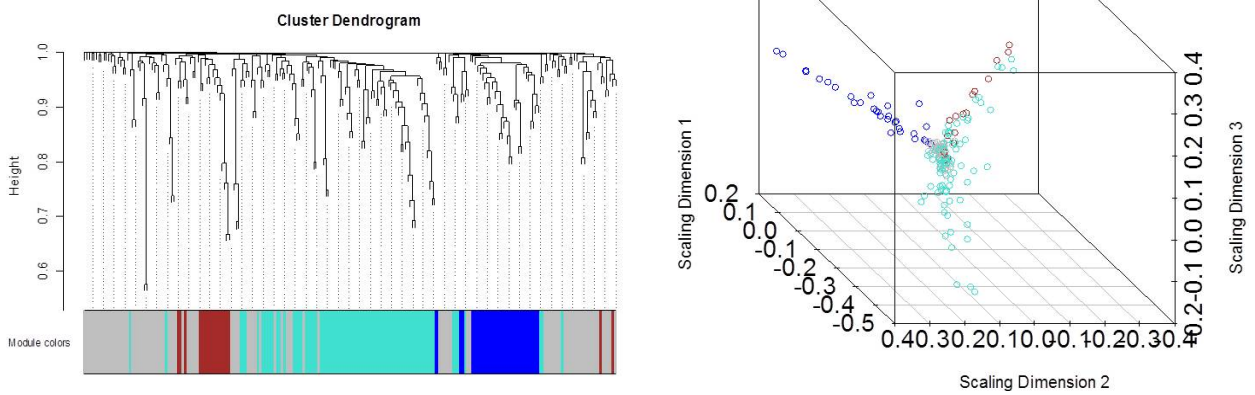


Fig. 1. The scale free topology model fitting index (R^2) versus the cut-off tau. The fitting index (denoted as R^2) quantify the extent of how well a network satisfies a scale-free topology. In the plot, tau=8 indicates R^2=0.9. It means that the resulting slope looks good.

Table 1. The top 5 genes information in each module

Module	Probe ID	RefSeq ID	Gene symbol	Gene description	*P*-value
Turquoise	Bt.7207.1.S1_at	NM_173975	ZP4	Zona pellucida glycoprotein 4	1.16E-77
	Bt.17354.1.A1_at	NM_001075801	ARMC2	Armadillo repeat containing 2	4.00E-64
	Bt.26104.1.A1_at	NM_001034017	WASF1	WAS protein family, member 1	1.66E-53
	Bt.16756.1.A1_at	NM_001103093	KLRAQ1	Protein phosphatase 1, regulatory subunit 21	1.48E-51
	Bt.20749.1.A1_at	NM_001077869	OOEP	Oocyte expressed protein	1.04E-49
Brown	Bt.14213.1.A1_at	NM_001034260	CES2	Carboxylesterase 2	3.07E-87
	Bt.20241.1.S1_at	NM_001075458	HAAO	3-Hydroxyanthranilate 3,4-dioxygenase	3.39E-75
	Bt.11126.1.S1_at	NM_001046069	LCAT	Lecithin-cholesterol acyltransferase	3.08E-64
	Bt.9416.1.S1_at	NM_001081582	MICAL1	Microtubule associated monooxygenase, calponin and LIM domain containing 1	1.62E-37
	Bt.628.1.S1_at	NM_001075988	TMED6	Transmembrane emp24 protein transport domain containing 6	2.32E-35
Blue	Bt.29515.1.A1_at	NM_001105347	INO80B	INO80 complex subunit B	1.32E-60
	Bt.29772.1.S1_at	NM_176632	SLC8A1	Solute carrier family 8 (sodium/calcium exchanger), member 1	1.75E-60
	Bt.91.1.S1_at	NM_173983	AGRP	Agouti related protein homolog	1.04E-46
	Bt.3066.1.A1_at	NM_001105422	DDX28	DEAD (Asp-Glu-Ala-Asp) box polypeptide 28	2.05E-45
	Bt.16590.1.A1_at	NM_001075914	SFXN5	Sideroflexin 5	4.66E-45



(A) (B)

Fig. 2. (A) Hierarchical clustering of pregnancy rate related genes and visualization of gene modules, (B) Multi-dimensional scaling plot of the weighted network. The colored bars (below) directly consistent with the module (color) for the clusters of genes. Distance between genes is shown as height on the y-axis. Genes are represented by a dot and colored by module membership. The distance between each gene is indicated by their topological overlap. This representation provides that how the module is related to the rest of the network, and how closely two modules are linked.

et al., 2005) using $log((p(k))$ and $log(k)$. Thus, our network has characteristics of a scale-free network whose degree distribution approximates a power law. Highly connected nodes are statistically more significant in a scale-free network than in a random graph. Most biological networks are characterized by a small number of highly connected nodes, while most of the other nodes have few connections (Brandes, 2001).

Detection of Co-Expression Gene Modules Related to the Pregnancy Rate

To find clusters (gene modules) of highly correlated genes, we used average linkage hierarchical clustering, which uses TOM as dissimilarity. We choose a height cutoff of 0.99 to identify modules using a dynamic cut-tree algorithm. Connectivity is the number of nearest neighbors of a node and the effective mean degree is the average degree of all nodes except isolated nodes. We able to identify seven distinct modules (except for the "grey" module, which is not grouped into any module) for groups of genes with high topological overlap): turquoise, black, yellow, brown, blue, green and red. Figure 2 shows the visualization of the modules in the weighted network. It consisted of ranges of gene modules, 18 (brown), 31 (blue) and 78 genes (turquoise). ZP glycoprotein (ZP4) is the most significant (*p*-value=1.16E-77) in the turquoise module. The zona pellucida (ZP) is an extracellular matrix that surrounds the oocyte and early embryo. It is composed primarily of three or four glycoproteins (ZP1, ZP2, ZP3 and ZP4) with various functions during oogenesis, fertilization and preimplantation development (Wassarman and Litscher, 2008). Also, ZP4 was differentially expressed in bovine preimplantative embryos (Hoelker *et al.*, 2009). Recently, sequence variation ZP4 gene has a role as potential modifier of zona pellucid architecture in pregnancy women (Pökkylä *et al.*, 2011). From our network analysis, ZP4 and oocyte Expressed Protein (OOEP) gene are strongly correlated (r=0.48) in the network (4.30E-20$\leq r \leq$6.83E-01). OOEP is one of reproduction-related genes (Tian *et al.*, 2009). ZP4 affects reproductive process and may act as hub gene in the turquoise module. Four and a half LIM domains 2 (FHL2) has statistically significant (*p*-value=2.42E-15) in the brown module. Matulis (2012) reported that knockdown of FHL1 and FHL2 decreases the expression of genes involved in mammalian reproduction. Also, FHL2 has functions as a novel coactivator of NR5A nuclear receptors in ovarian granulosa cells and its involvement in regulating target genes important for mammalian reproduction (Matulis and Mayo, 2012). In the brown module, FHL2 and carbohydrate sulfotransferase 10 (CHST10) are moderately correlated (r=0.007) in the network (5.18E-26$\leq r \leq$8.79-E-02). CHST10 sulfates glucuronidated steroid hormone to regulate steroid hormone *in vivo* and has a potential role for reproduction (Suzuki-Anekoji *et al.*, 2013). In the blue module, the early growth response 4 (EGR4) showed a significant result (*p*-value=7.19E-31) in the network. EGR4 plays a critical role in human and mouse fertility. Recently, the study of microarray showing differential gene expression in the ovaries between Erhualian and Pietrain pigs has revealed that EGR4 expression is increased 120-fold in ovaries from Erhualian sows. From this study, EGR4 gene was suggested as a candidate gene for porcine reproductivity (Wang *et al.*, 2013). EGR4 and par-6 partitioning defective 6 homolog alpha (PARD6A) showed co-expression (r=0.02) in the brown module (4.63E-22$\leq r \leq$1.04E-01). PARD6A was found to locate asymmetrically at the animal pole in mouse oocytes. Also, PARD6a is expressed in oocytes during the early stages of folliculogenesis, and then associated with oocyte and follicle development (Pepling and Spradling, 2001).

Gene modules are important for identifying genes related to the trait of interest because it may be highly correlated in biological pathways. Therefore, each modu-

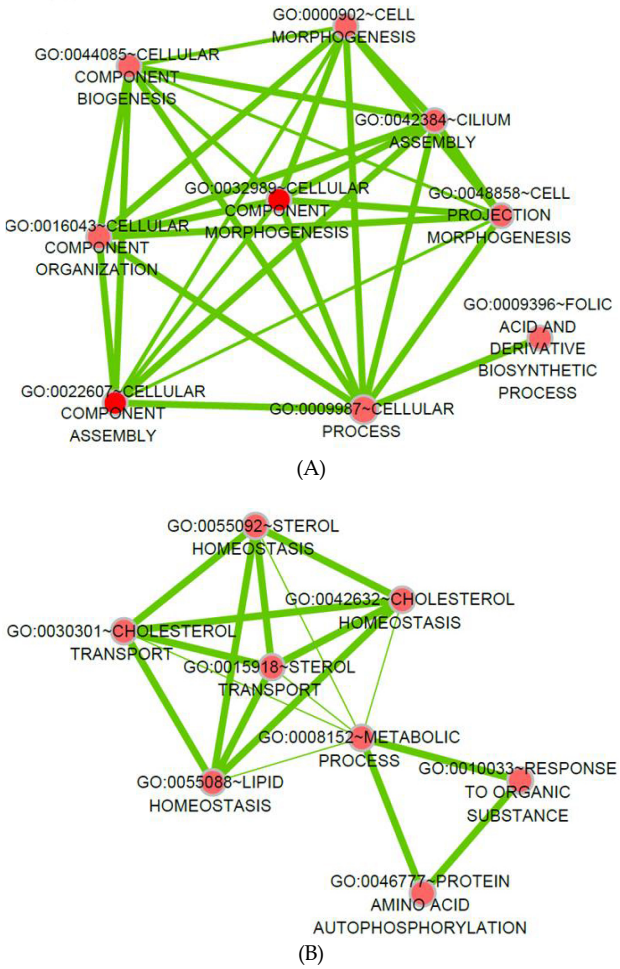


Fig. 3. Networks of functional terms of candidate genes. (A) Network of functional terms in the turquoise module. (B) Network of functional terms in the brown module. Nodes (red circles) are annotated functional terms. Edges connecting nodes represent gene share, being thickness proportional to the number of genes shared between terms (i.e., the degree of gene set overlap).

Table 2. The results of functional enrichment analysis for each module

Module	GO term	*p*-value	Genes information
Turquoise	GO:0022607~cellular component assembly	0.003	TSPYL1, NUP133, E2F4, OOEP, DYNC2LI1, TUBE1, ASF1A, SPAST
	GO:0032989~cellular component morphogenesis	0.005	NUP133, E2F4, OOEP, DYNC2LI1, TBCE
	GO:0016043~cellular component organization	0.006	NUP133, E2F4, OOEP, DYNC2LI1, TBCE, ELMO3, TSPYL1, NCAPH, TUBE1, ASF1A, SPAST, GFOD2, RANBP10
	GO:0044085~cellular component biogenesis	0.009	TSPYL1, NUP133, E2F4, OOEP, DYNC2LI1, TUBE1, ASF1A, SPAST
	GO:0009987~cellular process	0.010	PSMB10, ARV1, E2F4, DYNC2LI1, PPM1B, HNRPLL, TSPYL1, MTHFD2, SFRS7, NCAPH, PIGF, CDC40, B3GALNT2, BUB1, VPS4A, TUBE1, THAP11, ASF1A, GFOD2, NUP133, PLA2G15, OOEP, MSH2, TBCE, MRPS5, TXNDC9, SOCS5, ELMO3, PCGF1, TRADD, CIAO1, CCT7, TRMT11, FBXL8, PLN, MTR, DUS2L, SPAST, RANBP10
	GO:0000902~cell morphogenesis	0.019	E2F4, OOEP, DYNC2LI1, TBCE
	GO:0042384~cilium assembly	0.021	E2F4, DYNC2LI1
	GO:0009396~folic acid and derivative biosynthetic process	0.047	MTHFD2, MTR
	GO:0048858~cell projection morphogenesis	0.048	E2F4, DYNC2LI1, TBCE
Brown	GO:0010033~response to organic substance	0.023339	FYN, FHL2, EIF2AK2
	GO:0030301~cholesterol transport	0.026414	ABCG8, LCAT
	GO:0042632~cholesterol homeostasis	0.026414	ABCG8, LCAT
	GO:0055092~sterol homeostasis	0.026414	ABCG8, LCAT
	GO:0015918~sterol transport	0.026414	ABCG8, LCAT
	GO:0055088~lipid homeostasis	0.036306	ABCG8, LCAT
	GO:0046777~protein amino acid autophosphorylation	0.039993	FYN, EIF2AK2
	GO:0008152~metabolic process	0.042654	MGC152281, GALM, CHST10, FYN, LCAT, HAAO, FHL2, CTCF, EIF2AK2, FAHD2A

le was analyzed through functional enrichment analysis using gene ontology terms to understand the biological significance of the module genes and to determine putative functions. The three modules and their representative pathway terms were turquoise, folic acid and derivative biosynthetic process (GO:0009396, *p*-value=0.04); brown, sterol transport (GO:0015918, *p*-value=0.02) and blue, regulation of system process (GO:0044057, p-value=0.01). In particular, folic acid and derivative biosynthetic proces is important for animal reproduction. folic acid metabolism is linked to reproduction. Supplementation of folic acid to sows has been described to increase litter size and embryo survival in pig (Lindemann and Kornegay, 1989). Therefore, these functions appears to regulate folate metabolism in the embryo development in the network.

In this study, we extracted gene list related to the pregnancy rate from the Animal QTL database and microarray experiments from the GEO database. We subsequently constructed a weighted gene co-expression network based on pearson's correlation matrix that displayed degrees using a power-law distribution, with an exponent of approximately −2. Second, we identified hub genes ZP4, FHL2 and EGR4 in each module with significant values in the network-topology analysis. They were topologically centered with significant values in the network. The pair of co-expression (ZP4-OOEP, FHL2-CHST10, EGR4-PARD6A) also determined in each module. Finally, we constructed the network of functional terms using functional enrichment analysis. Further study should be conducted to identify biological mechanism of the genes in the network associated with

reproduction traits.

REFERENCES

1. Barabasi AL, Albert R (1999): Emergence of scaling in random networks. Science 286:509-512.
2. Barabasi AL, Oltvai ZN (2004): Network biology: understanding the cell's functional organization. Nat Rev Genet 5:101-113.
3. Brandes U (2001): A faster algorithm for betweenness centrality. J Math Soc 25:163-177.
4. Donaldson L, Vuocolo T, Gray C, Strandberg Y, Reverter A, McWilliam S, Wang Y, Byrne K, Tellam R (2005): Construction and validation of a Bovine innate immune microarray. BMC Genomics 6: 135.
5. Gautier L, Cope L, Bolstad B, Irizarry R (2004): Affy-analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 20:307.
6. Ghazalpour A, Doss S, Zhang B, Wang S, Plaisier C, Castellanos R, Brozell A, Schadt E, Drake T, Lusis A (2006): Integrating genetic and network analysis to characterize genes related to mouse weight. PLoS Genet 2:e130.
7. Gibson G, Weir B (2005): The quantitative genetics of transcription. Trends Genet 21:616-623.
8. Haley C, de Koning DJ (2006): Genetical genomics in livestock: potentials and pitfalls. Anim Genet 37 Suppl 1:10-12.
9. Hocquette J, Brandstetter A (2002): Common practice in molecular biology may introduce statistical bias and misleading biological interpretation. J Nutrit Biochem 13:370-377.
10. Hoelker M, Rings F, Lund Q, Ghanem N, Phatsara C, Griese J, Schellander K, Tesfaye D (2009): Effect of the microenvironment and embryo density on developmental characteristics and gene expression profile of bovine preimplantative embryos cultured *in vitro*. Reproduction 137:415-425.
11. Hwang S, Son SW, Kim SC, Kim YJ, Jeong H, Lee D (2008): A protein interaction network associated with asthma. J Theor Biol 252:722-731.
12. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP (2003): Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res. 31, e15.
13. Jiang Z, Michal JJ, Chen J, Daniels TF, Kunej T, Garcia MD, Gaskins CT, Busboom JR, Alexander LJ, Wright RW, Jr, Macneil MD (2009): Discovery of novel genetic networks associated with 19 economically important traits in beef cattle. Int J Biol Sci 5: 528-542.
14. Keller M, Choi Y, Wang P, Belt Davis D, Rabaglia M, Oler A, Stapleton D, Argmann C, Schueler K, Edwards S (2008): A gene expression network model of type 2 diabetes links cell cycle regulation in islets with diabetes susceptibility. Genome Res 18: 706.
15. Lindemann MD, Kornegay ET (1989): Folic acid supplementation to diets of gestating-lactating swine over multiple parities. J Anim Sci 67:459-464.
16. Matulis CK, Mayo KE (2012): The LIM domain protein FHL2 interacts with the NR5A family of nuclear receptors and CREB to activate the inhibin-α subunit gene in ovarian granulosa cells. Molecular Endocrinology 26:1278-1290.
17. Merico D, Isserlin R, Stueker O, Emili A, Bader G. D (2010): Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS One 5:e13984.
18. Nikitin A, Egorov S, Daraselia N, Mazo I (2003. Pathway studio-the analysis and navigation of molecular networks. Bioinformatics 19:2155.
19. Nobis W, Ren X, Suchyta SP, Suchyta TR, Zanella AJ, Coussens PM (2003): Development of a porcine brain cDNA library, EST database, and microarray resource. Physiol Genomics 16:153-159.
20. Peter L, Steve H (2008): WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9:599.
21. Pökkylä RM, Lakkakorpi JT, Nuojua-Huttunen SH, Tapanainen JS (2011): Sequence variations in human ZP genes as potential modifiers of zona pellucida architecture. Fertility and Sterility 95:2669-2672.
22. Presson A, Sobel E, Papp J, Suarez C, Whistler T, Rajeevan M, Vernon S, Horvath S (2008): Integrated weighted gene co-expression network analysis with an application to chronic fatigue syndrome. BMC Systems Biol 2:95.
23. Ravasz E, Somera A, Mongru D, Oltvai Z, Barabasi A (2002): Hierarchical organization of modularity in metabolic networks. Science 297:1551.
24. Reverter A, Hudson N, Wang Y, Tan S, Barris W, Byrne K, McWilliam S, Bottema C, Kister A, Greenwood P (2006): A gene coexpression network for bovine skeletal muscle inferred from microarray data. Physiol Genomics 28:76.
25. Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, Linsley PS, Mao M, Stoughton RB, Friend SH (2003): Genetics of gene expression surveyed in maize, mouse and man. Nature 422:297-302.
26. Smith GW, Rosa GJ (2007): Interpretation of microarray data: trudging out of the abyss towards elucidation of biological significance. J Anim Sci 85: E-20-23.
27. Stelzl U, Worm U, Lalowski M, Haenig C, Brem-

beck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S, Timm J, Mintzlaff S, Abraham C, Bock N, Kietzmann S, Goedde A, Toksoz E, Droege A, Krobitsch S, Korn B, Birchmeier W, Lehrach H, Wanker EE (2005): A human protein- protein interaction network: a resource for annotating the proteome. Cell 122:957-968.

28. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005): Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 102:15545-15550.
29. Suzuki-Anekoji M, Suzuki A, Wu SW, Angata K, Murai KK, Sugihara K, Akama TO, Khoo KH, Nakayama J, Fukuda MN (2013): *In vivo* regulation of steroid hormones by the Chst10 sulfotransferase in mouse. Journal of Biological Chemistry 288:5007-5016.
30. Tian X, Pascal G, Fouchécourt S, Pontarotti P, Monget P (2009): Gene birth, death, and divergence: the different scenarios of reproduction-related gene evolution. Biology of Reproduction 80:616-621.
31. Wang J, Zhao Y, Gu K, Yu P, Zhang B, Wang W, Yang J, Xu Y (2013): The novel porcine gene early growth response 4 (Egr4) is differentially expressed in the ovaries of Erhualian and Pietrain pigs. Reproduction, Fertility and Development.
32. Wassarman PM, Litscher ES (2008): Mammalian fertilization: the egg's multifunctional zona pellucida. Int J Dev Biol 52:665-676 .
33. Ye Y, Godzik, A (2004): Comparative analysis of protein domain organization. Genome Res. 14:343.

(Received: 7 November 2013/ Accepted: 19 November 2013)