International Scholars Journals

African Journal of Pig Farming ISSN 2375-0731 Vol. 6 (7), pp. 001-008, July, 2018. Available online at www.internationalscholarsjournals.org © International Scholars Journals

Full Length Research Paper

Identification of pig reproductive QTL genes based on gene set enrichment analysis of mouse microarray dataset

Kan He[1,2], Fan Yang[1,2], Minghui Wang[1,2], Qishan Wang[1,2], Yuchun Pan[1,2] and Yufang Ma[1,2]*

[1]School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai, China.
[2]Shanghai Key Laboratory for Veterinary Biotechnology, Shanghai, 200240, China.

Accepted 23 January, 2018

Nowadays, abundantly different transcripts related to reproductive traits using mouse model have been identified by microarray analysis with the robustness, which were needed of re-evaluation with quantitive real time PCR (qPCR). Meanwhile, pig QTL databases were established but with large spans in the confidence interval for QTL location. Therefore, a genome-wide comparative and functional analysis of the association between mouse microarrays analysis results and reported pig QTLs is needed to better elucidate the genetic and physiological background of pig enhanced reproductive performance in genomics. Here, we employed a microarray dataset of a high fertility mouse line and applied gene set enrichment analysis method to identify significant mouse genes in a pathway level. Next we used a comparative genomic approach to find the homologous pig genes and locate them to the interval of the reported QTL for pigs' traits based on AnimalQTLdb. Finally, we identified some pathways participating in pig high fertility, and some pig genes were further identified within the reported QTL location related to reproductive performance. Combining microarray analysis of mouse high fertility dataset by GSEA with pigQTLdb will substantially help us to identify candidate genes in reported QTL regions related to pig reproduction that are eventually responsible for increased fertility performance in pig and are also helpful for understanding the genetic information of pig reproduction in genomics. Furthermore, we can also provide some novel pig genes in identified pathways with relationship of reproduction.

Key words: Reproductive, microarray, pig, pathway, endocrine.

INTRODUCTION

It is well known that the number of ovulation is an important parameter for reproductive performance in polytocous species as mouse and pig. Folliculogenesis, ovulation, and luteinization are regulated by some important endocrine factors including the hypothalamic gonadotropin releasing hormone, follicle stimulating hormone (FSH) and luteinizing hormone from the pituitary gland, and steroid and peptide hormones generated within the ovary as estrogen, progesterone, inhibin, activin, and follistatin (Bastida et al., 2005; Rice et al., 1993; Emmen et al., 2005). Nowadays, mice have been successfully used as models for swine selection systems. Especially for studying the regulatory mechanism of reproduction in genomics, the expression profiles of animal model mouse reproductive organs have extensively been used to detect genetic factors involved in reproductive performance. However, abundantly different transcripts were identified by microarray analysis with the robustness, which were needed of re-evaluation with quantitive real-time PCR (qPCR). Whereas DNA microarray technology has extended to all fields of genomic research and has become practically the primary tool for gene expression analysis, the platforms of livestock are still not perfect, especially for pigs. Therefore, comparative analysis of mouse model microarray datasets may help us to better understand the regulatory mechanism of pig reproductive performance.

Furthermore, there is a wealth of information that has been collated from many QTL studies over the last

*Corresponding author. E-mail: mayufang@sjtu.edu.cn Tel: 86-021-3420-6147.

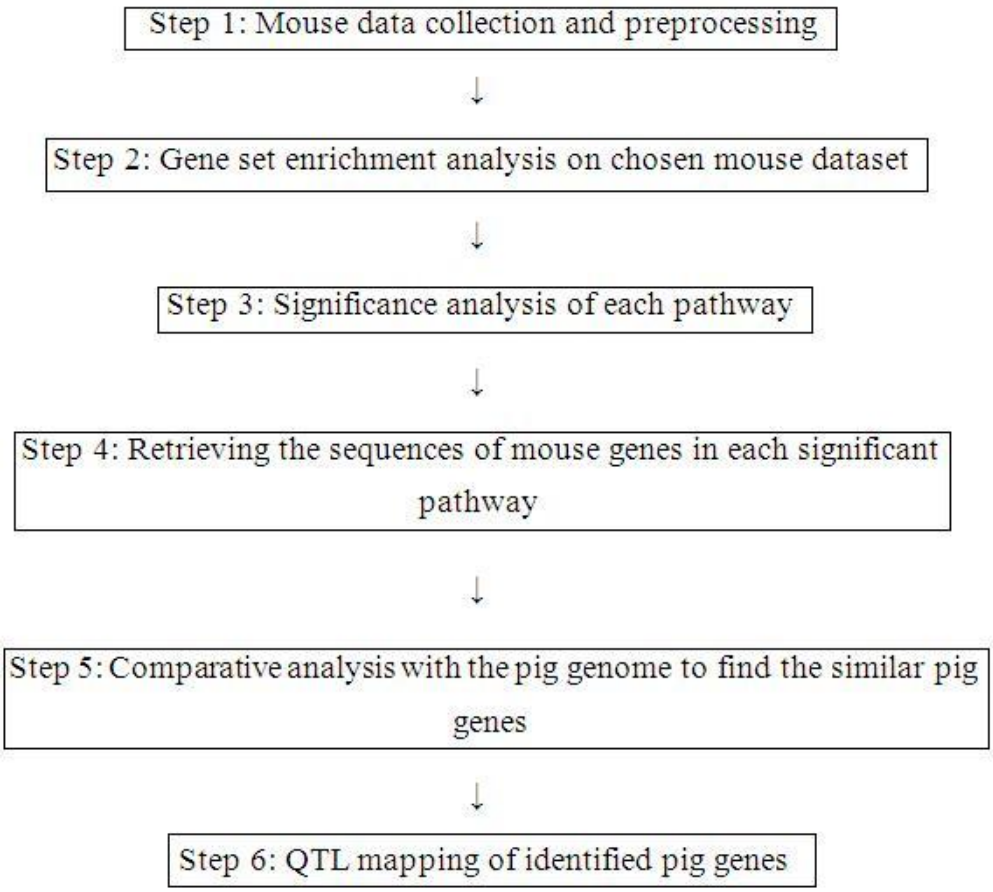

Figure 1. Six major steps in our analysis strategy for pig QTL genes identification.

decade, summarized in resources such as AnimalQTLdb (Hu et al., 2005). However, even for relatively large effects with a saturated marker map, the confidence interval for QTL location by linkage analysis spans tens of map units, or hundreds of genes. This requires further linkage disequilibrium mapping to target their gene of interest (Meuwissen et al., 2004). Therefore, a genome-wide comparative and functional analysis of the association between mouse microarrays analysis results and pig QTLs is needed to better elucidate the genetic and physiological background of enhanced reproductive performance, including reproductive traits, reproductive organ, litter size and endocrine.

Currently the most well-known Gene set enrichment analysis (GSEA) method has been widely used to analysis gene expression profiles, especially to identify pre-defined gene sets which exhibited significant differences in expression between samples from control and treated (Keller et al., 2007; Subramanian et al., 2005). The algorithms calculate the statistical significance of the expression changes across groups or pathways rather than individual gene, thus allowing identification of groups or pathways most strongly affected by the observed expression changes. The analysis based on a group of relevant genes instead of on an individual gene increases the likelihood for investigators to identify the critical functional processes under the biological phenomena under study. GSEA is likely to be more powerful than conventional single-gene methods in the study of complex diseases in which many genes make subtle contributions. The animal quantitative trait locus (QTL) database (AnimalQTLdb, http://www.animalgenome.org/ QTLdb) is designed to house all publicly available QTL data on livestock animal species for easily locating and making comparisons within and between species (Hu et al., 2005).

Here, we employed a microarray dataset of a high-fertility mouse line and applied gene set enrichment analysis method to identify significant pathways and genes in mouse. Next we used a comparative genomic approach to find the homologous pig genes and locate them to the interval of the reported QTLs for pigs' traits based on AnimalQTLdb.

MATERIALS AND METHODS

There were six major steps in our analysis strategy showing in Figure 1.

Data collection and preprocessing

We searched the public database GEO (www.ncbi.nlm.nih.gov/geo/) for the gene expression profiling studies related to reproductive performance. Finally, we chose the dataset GSE11113 for our re-analysis, which was contributed by Vanselow et al. (2008) for mouse high fertility performance study. In this dataset, ovaries from 30 animals of line FL1 (selected for high fertility performance) and of the control line DUKsi were collected at the metestrous stage, combined to 6 pools each, and processed for microarray analysis. The samples of GSM280489 to GSM280500 were from line FL1 and GSM280501 to GSM280512 were from line DUKsi, which were hybridized onto Affymetrix Mouse Expression 430A Array. For the assessment of the influence of preprocessing on the comparison, data preprocessing was performed using software packages developed in version 2.6.0 of Bioconductor and R version 2.10.1 (Gentleman et al., 2004). Each Affymetrix dataset was background adjusted, normalized and log2 probe-set intensities calculated using the Robust multichip averaging (RMA) algorithm in Affy package (Gautier et al., 2004).

During the previous study a high fertility mouse line FL1, which was generated by index trait selection over more than 130 generations was comparatively analyzed with the non-selected control line, DUKsi, and both lines have been derived from the same genetic pool (Spitschak et al., 2007). The mean number of corpora lutea at the first day of FL1 pregnancy was 18.2 ± 6.8 (mean ± std, n = 10) compared to 12.2 ± 5.5 (n = 10) in the control line ($p = 0.0437$). The mean litter size and litter weight at birth increased during the selection period to 17.3 ± 7.7 (n = 60) and 27.8 ± 4.6 g, respectively in FL1 compared to 9.81 ± 2.03 (n = 75) and 15.1 ± 6.2 g in the control line ($p < 0.0001$ for litter size and weight). It suggested that the body weight of individual newborn pups however was not significantly reduced in FL1 despite of the strongly increased number of pups per litter.

Gene set enrichment analysis

Our gene set enrichment analysis of pathways and genes were performed using category package in version 2.6.0 of Bioconductor (Chiaretti et al., 2004). The goal of GSEA is to determine whether the members of a gene set S randomly distributed throughout the entire reference gene list L or are primarily found at the top or bottom in a pathway level. One of the advantages of GSEA is the relative robustness to noise and outliers in the data. The gene sets represented by less than 10 genes were excluded. The t-statistic mean of the genes was computed in each pathway. Using a permutation test with 1000 times, the significantly changed pathways were identified with p-value less than 0.001. Accordingly, the significant pathways and genes between line FL1 and line DUKsi were then identified.

Significance analysis of each pathway and Retrieving the sequences of significant genes

Based on GSEA, we have identified the significantly up- or down-regulated pathways and ranked genes list in each pathway. Then we retrieved each significant gene's Entrez ID and used biomaRt package (http://www.biomart.org/) to mine the sequence information of significantly identified genes in each significant pathway.

Comparative analysis with the pig genome and QTL mapping

To identify pig genes homologous to the aforementioned mouse pathway targets and to predict their map locations in Sus scrofa, the mouse sequences were Blast searched against the pig genome using Blast 2.2.23. The homologous gene was identified when the pig's transcript or gene with the homology higher than 85% and alignment region longer than 300 bp. Then we wrote a Perl script to relatively locate genes to the interval of the QTL for pigs' traits within PigQTLdb.

RESULTS AND DISCUSSION

Significantly up or down regulated pathways by GSEA of mouse dataset

The previous GSEA method was applied to the dataset of GSE11113. Based on the cutoff of the permutation p-values, there were 5 up-regulated and 69 down-regulated pathways under the comparison of FL1 to DUKsi. The analysis results including significant genes in each identified pathway were summarized in Appendices 1 and 2. Mapping these ovarian genes to such pathway terms could elucidate the line-specific differences due to the genetic consequences of selection for reproductive performance. In total, the pathway categories of steroid/lipid metabolism, immune response, signaling pathways and other metabolic processes were significantly different in the selected high fertility mouse line FL1 compared to a non selected control line. On the other hand, numerous of the differently expressed genes in each significant pathway had also been identified by GSEA which were known to be involved in female reproductive processes as folliculogenesis, ovulation, atresia, or luteinization and thus might play a role in increasing the ovulation number (in brown color in Appendix 2). For example, we identified *Casp3* in the pathway of natural killer cell mediated cytotoxicity or Huntington's disease or Parkinson's disease, *Casp7* in Alzheimer's disease, *Casp8* in RIG-I-like receptor signaling pathway or Toll-like receptor signaling pathway and *Casp9* in VEGF signaling pathway. Those members of *Casp* family are well known to play a fundamental role during follicular atresia (Valdez et al., 2005). Members of *Gpx* family including from *Gpx1* to *Gpx7* were mostly identified in the pathway of glutathione metabolism, which may be important factors in protecting cells of the cumulus-oocyte complex and luteal cells from oxidative stress inducing cell death (Al-Gubory et al., 2005; Luciano et al., 2006). Otherwise, *H2-Q7* was identified in the pathway of antigen processing and presentation whose functional role for reproductive performance is not yet clear.

Identification of pig reproductive QTL genes

Based on the previous identified mouse genes in each significant pathway, we then applied comparative genomic analysis to the pig genome with PigQTLdb in order to locate homology genes to the interval of the QTL for pigs' traits of reproduction. The results are shown in Appendices 3 and 4, respectively for up or down regulated pathway genes to pig QTLs. The information

Table 1. The number of pig down or up QTL genes related to traits of pig reproduction in PigQTLdb.

Trait type	Trait name	Number of pig up QTL genes	Number of pig down QTL genes
Reproduction traits	Age at puberty	8	101
	Ejaculation duration	0	5
	Ejaculation times	0	3
	Gestation length	0	29
	Semen pH	0	1
	Sperm abnormality rate	0	2
	Sperm concentration	0	2
	Sperm per ejaculate	0	1
Litter size	Embryo survival	0	1
	Mummified pigs	0	4
	Number of viable embryos	0	1
	Ovulation rate	6	50
	Total number born	1	8
	Total number born alive	0	7
Reproductive organ	Epididymis weight	0	4
	Left teat number	0	10
	Nonfunctional nipples	10	115
	Number of corpora lutea	0	14
	Right teat number	0	5
	Seminiferous tubule diameter	0	2
	Teat number	3	192
	Testicular weight	2	20
	Uterine horn length	7	31
Endocrine	Plasma FSH concentration	0	2
	Testosterone level	0	11

In PigQTLdb, the traits of pig reproduction were classified into four types including reproduction traits, litter size, reproductive organ and endocrine. In each type, different traits were named. The numbers of pig up or down QTL genes were based on Additional file 3 and 4. Up or down QTL genes here represent that the genes located in reported pig reproduction QTL regions within identified up or down pathways.

about the number of pig down or up QTL genes related to traits of pig reproduction is shown in Table 1.

In PigQTLdb, the traits of pig reproduction were classified into four types including reproduction traits, litter size, reproductive organ and endocrine. In each type, different traits were named. Obviously, more pig down QTL genes were found than up QTL genes as more down-regulated pathways were identified. For up QTL genes, most of them were related to age at puberty (one of the reproduction traits in AnimalQTLdb), ovulation rate (litter size), nonfunctional nipples and uterine horn length (reproductive organ). By contrast, most of down QTL genes were related to teat number and nonfunctional nipples (reproductive organ), age at puberty (reproduction traits) and ovulation rate (litter size), each number was over 50. On the other hand, 13 down QTL genes were identified to be related to pig trait type of endocrine, including 11 genes for testosterone level and 2 genes for plasma FSH concentration (in green color in Appendix 4), these related traits were not found in up QTL genes. For examples, *UQCRC1* (ubiquinol-cytochrome c reductase core protein I) in pathways of Alzheimer's disease and Parkinson's disease and Huntington's disease and oxidative phosphorylation, *UBA7* (ubiquitin-like modifier activating enzyme 7) in the pathway of Parkinson's disease, *GMPPB* (GDP-mannose pyrophosphorylase B) in the pathway of fructose and mannose metabolism, they were located in chromosome 13 and identified to be related to testosterone level (Ren et al., 2009). *OAT* (ornithine aminotransferase) in the pathway of arginine and proline metabolism and *GK* (glycerol kinase) in the pathway of glycerolipid metabolism were both located in chromosome X and identified to be related to plasma FSH concentration.

Totally, less pig genes were located in reported QTL regions than mouse identified pathway genes, which can ensure the accuracy of targets for pig reproduction.

Interestingly, there were more homology genes in latterly ranked pathway located to the interval of pig reproductive QTLs especially for five up-regulated pathways (Appendix 3).

For example, there was only one pig QTL gene *MPST* (mercaptopyruvate sulfurtransferase) identified in the pathway of cysteine and methionine metabolism, which was located in chromosome 5 related to reproductive organ of teat number.

METTL6 (methyltransferase like 6) might play a functional role in transferase activity, which was found in the pathway of selenoamino acid metabolism and located in chromosome 13 and with relationship of litter size of ovulation rate and reproductive trait of age at puberty and reproductive organ of nonfunctional nipples (Rathje et al., 1997; Jonas et al., 2008). Another gene found in the same pathway was *SEPHS1* (selenophosphate synthetase 1), located in chromosome 15 and related to both ovulation rate and nonfunctional nipples, which was well known in function of ATP binding (Rohrer et al., 1999).

In the pathway of heparan sulfate biosynthesis, *GLCE* (glucuronic acid epimerase) in chromosome 1 and *EXT1* (exostoses 1) in chromosome 4 were identified to be related to age at puberty and ovulation rate respectively (Jonas et al., 2008).

In the last two up-regulated pathways including hedgehog signaling pathway and the pathway of arrhythmogenic right ventricular cardiomyopathy (ARVC), there were almost the same pig QTL genes identified. For examples, *CSNK1E* (casein kinase 1, epsilon) in chromosome 5 was found to be related to reproductive organs of both teat number and nonfunctional nipples. Members of wingless-type MMTV integration site family including *WNT1* and *WNT10B* and *DHH* in chromosome 5 were all found to be related to reproductive organ of uterine horn length (Wilkie et al., 1999). *BMP5* (bone morphogenetic protein 5) and a member of RAS oncogene family *RAB23* in chromosome 7 were both identified to be related to age at puberty (Cassady et al., 2001). *BTRC* (beta-transducin repeat containing) and *SUFU* (suppressor of fused homolog, drosophila) in chromosome 14 were found to be related to nonfunctional nipples (Jonas et al., 2008). *GLI2* (GLI family zinc finger 2) in chromosome 15 was identified to be related to both ovulation rate and nonfunctional nipples (Jonas et al., 2008; Rohrer et al., 1999). Besides there were some novel genes in chromosome 1 and 5 related to age at puberty or uterine horn length.

By contrast, there were more important pig QTL genes found in down-regulated pathways (the brown colors in Appendix 4). For examples, in the pathway of natural killer cell mediated cytotoxicity or Huntington's disease or Parkinson's disease, *Casp3* (caspase 3, apoptosis-related cysteine peptidase) was also identified as one of pig QTL genes, which was located in chromosome 15 and related to the traits of nonfunctional nipples, number of corpora lutea, and gestation length (Jonas et al., 2008; Wilkie et al., 1999). In some signaling pathways such as VEGF signaling pathway, Toll-like receptor signaling pathway, T cell receptor signaling pathway, B cell receptor signaling pathway and Fc epsilon RI signaling pathway, the gene of *MAP2K1* (mitogen-activated protein kinase kinase 1) in chromosome 1 was identified to be almost related to the traits of teat number, testicular weight and age at puberty (Ren et al., 2009; Rohrer, 2000; Guo et al., 2008). Besides there were also some novel genes identified to be related to pig traits of reproduction. For examples, in the pathway of fatty acid metabolism, a novel gene described as acyl coenzyme A synthetase long-chain 1 fragment (EC 6.2.1.3) in chromosome 15 was identified to be related to nonfunctional nipples, number of corpora lutea and gestation length (Jonas et al., 2008; Wilkie et al., 1999).

Nowadays, mice have been used as models for swine selection systems, the genetic parameters as the number of ovulation could be useful in determining alternative selection criteria for increasing the number born in mice, and potentially in swine (Christenson et al., 1987; Long et al., 1991). Combining microarray analysis of mouse high fertility dataset by GSEA with pigQTLdb will substantially help us to identify candidate genes in reported QTL regions related to pig reproduction that are eventually responsible for increased fertility performance in pig and are also helpful for understanding the genetic information of pig reproduction in genomics. Furthermore, we can also provide some novel pig genes in identified pathways with relationship of reproduction.

ACKNOWLEDGEMENTS

Our work is funded by National Natural Science Foundation of China (grant No. 31072003 and 30871782) and the National High Technology Research and Development Program of China (863 project) (grant No. 2008AA101002 and 2006AA10Z1E3).

REFERENCES

Bastida CM, Cremades A, Castells MT, Lopez-Contreras AJ, Lopez-Garcia C, Tejada F, Penafiel R (2006). Influence of ovarian ornithine decarboxylase in folliculogenesis and luteinization. Endocrinology, 146: 666-674.

Rice VC, Zusmanis K, Malter H, Mitchell-Leef D (1993). Pure FSH alone induces ovulation and subsequent pregnancy in the mouse resulting in fetal development. Life Sci., 53: 31-39.

Emmen JM, Couse JF, Elmore SA, Yates MM, Kissling GE, Korach KS (2005). In vitro growth and ovulation of follicles from ovaries of estrogen receptor (ER){alpha} and ER{beta} null mice indicate a role for ER{beta} in follicular maturation. Endocrinology, 146: 2817-2826.

Hu ZL, Dracheva S, Jang W, Maglott D, Bastiaansen J, Rothschild MF, Reecy JM (2005). A QTL resource and comparison tool for pigs: PigQTLDB. Mamm Genome, 16: 792-800.

Meuwissen TH, Goddard ME (2004). Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multitrait data. Genet. Sel. Evol., 36: 261-279.

Keller A, Backes C, Lenhof HP (2007). Computation of significance scores of unweighted Gene Set Enrichment Analyses. BMC Bioinf., 8: 290.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Nat. Acad. Sci. USA, 102: 15545-15550.

Vanselow J, Nurnberg G, Koczan D, Langhammer M, Thiesen HJ, Reinsch N (2008). Expression profiling of a high-fertility mouse line by microarray analysis and qPCR. BMC Genomics, 9: 307.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol., 5: R80.

Gautier L, Cope L, Bolstad BM, Irizarry RA (2004). affy--analysis of Affymetrix GeneChip data at the probe level. Bioinformatics, 20: 307-315.

Spitschak M, Langhammer M, Schneider F, Renne U, Vanselow J (2007). Two high-fertility mouse lines show differences in component fertility traits after long-term selection. Reprod. Fertil. Dev., 19: 815-821.

Chiaretti S, Li X, Gentleman R, Vitale A, Vignetti M, Mandelli F, Ritz J, Foa R (2004). Gene expression profile of adult T-cell acute lymphocytic leukemia identifies distinct subsets of patients with different response to therapy and survival. Blood, 103: 2771-2778.

Valdez KE, Cuneo SP, Turzillo AM (2005). Regulation of apoptosis in the atresia of dominant bovine follicles of the first follicular wave following ovulation. Reproduction, 130: 71-81.

Al-Gubory KH, Ceballos-Picot I, Nicole A, Bolifraud P, Germain G, Michaud M, Mayeur C, Blachier F (2005). Changes in activities of superoxide dismutase, nitric oxide synthase, glutathione-dependent enzymes and the incidence of apoptosis in sheep corpus luteum during the estrous cycle. Biochim. Biophys. Acta, 1725: 348-357.

Luciano AM, Goudet G, Perazzoli F, Lahuec C, Gerard N (2006). Glutathione content and glutathione peroxidase expression in in vivo and in vitro matured equine oocytes. Mol. Reprod. Dev., 73: 658-666.

Ren DR, Ren J, Xing YY, Guo YM, Wu YB, Yang GC, Mao HR, Huang LS (2009). A genome scan for quantitative trait loci affecting male reproductive traits in a White Duroc x Chinese Erhualian resource population. J. Anim. Sci., 87: 17-23.

Rathje TA, Rohrer GA, Johnson RK (1997). Evidence for quantitative trait loci affecting ovulation rate in pigs. J. Anim. Sci., 75: 1486-1494.

Jonas E, Schreinemachers HJ, Kleinwachter T, Un C, Oltmanns I, Tetzlaff S, Jennen D, Tesfaye D, Ponsuksili S, Murani E (2008). QTL for the heritable inverted teat defect in pigs. Mamm. Genome, 19:127-138.

Rohrer GA, Ford JJ, Wise TH, Vallet JL, Christenson RK (1999). Identification of quantitative trait loci affecting female reproductive traits in a multigeneration Meishan-White composite swine population. J. Anim. Sci., 7: 1385-1391.

Wilkie PJ, Paszek AA, Beattie CW, Alexander LJ, Wheeler MB, Schook LB (1999). A genomic scan of porcine reproductive traits reveals possible quantitative trait loci (QTLs) for number of corpora lutea. Mamm. Genome, 10: 573-578.

Cassady JP, Johnson RK, Pomp D, Rohrer GA, Van Vleck LD, Spiegel EK, Gilson KM (2001). Identification of quantitative trait loci affecting reproduction in pigs. J. Anim. Sci., 79:623-633.

Rohrer GA (2000). Identification of quantitative trait loci affecting birth characters and accumulation of backfat and weight in a Meishan-White Composite resource population. J. Anim. Sci., 78: 2547-2553.

Guo YM, Lee GJ, Archibald AL, Haley CS (2008). Quantitative trait loci for production traits in pigs: a combined analysis of two Meishan x Large White populations. Anim. Genet., 39: 486-495.

Christenson RK, Leymaster KA, Young LD (1987). Justification of unilateral hysterectomy-ovariectomy as a model to evaluate uterine capacity in swine. J. Anim. Sci., 65: 738-744.

Long CR, Lamberson WR, Bates RO (1991). Genetic correlations among reproductive traits and uterine dimensions in mice. J. Anim. Sci., 69: 99-103.

APPENDICES

Appendix 1

The list of significantly up or down regulated pathways by GSEA of mouse dataset

The data provided represent ranked list of significantly up or down regulated pathways by GSEA of dataset GSE11113. The numbers in the bracket after each pathway name represent the numbers of genes in each identified pathway.

Appendix 2

The list of significant genes in each identified pathway in mouse

The data provided represent the list of significant genes in each identified pathway, including the information of probe set IDs and gene symbols. Up-regulated pathways are in red color and down-regulated pathways are in green color.

Appendix 3

Pig QTL genes in up-regulated pathways

The data provided represent the list of pig QTL genes in up-regulated pathways based on mouse data analysis and PigQTLdb.

Appendix 4

Pig QTL genes in down-regulated pathways

The data provided represent the list of pig QTL genes in down-regulated pathways based on mouse data analysis and PigQTLdb.