October 2003

Generation and sequence characterization of a normalized cDNA library from swine ovarian follicles

Alexandre R. Caetano
Embrapa Recursos Genéticos e Biotecnologia, Parque Estação Biológica, Final Av. W/S Norte, Brasília-DF, C.P. 02372, 70770-900, Brazil

Rodger K. Johnson
University of Nebraska-Lincoln, rjohnson5@unl.edu

Daniel Pomp
University of Nebraska-Lincoln, dpomp1@unl.edu

Published in *Mammalian Genome* 14 (2003), pp. 65–70; DOI: 10.1007/s00335-002-2220-5

Submitted July 23, 2002; accepted August 25, 2002.

The nucleotide sequence data reported in this paper have been submitted to GenBank and are assigned the accession numbers BI180971–BI186447 and BI233963–BI234008.

Generation and sequence characterization of a normalized cDNA library from swine ovarian follicles

Alexandre R. Caetano*, Rodger K. Johnson, and Daniel Pomp§

Department of Animal Science, University of Nebraska–Lincoln, Lincoln, Nebraska, 68583-0908, USA

* Present address: Embrapa Recursos Genéticos e Biotecnologia, Parque Estação Biológica, Final Av. W/S Norte, Brasília-DF, C.P. 02372, 70770-900, Brazil.

§ Corresponding author; Email: dpomp@unl.edu

Abstract: Ovulation rate is a major factor determining litter size in swine and is, therefore, a trait of economic importance to the pork industry. The dynamics of follicle development, which in turn are dictated by a balance between follicle recruitment, maturation, selection, and atresia, are a major determining factor of ovulation rate. The role of several genes expressed in the ovaries during these processes has been described, but studies utilizing large-scale genomic approaches have yet to be conducted to examine gene expression in this tissue more globally. We have developed a normalized cDNA library from swine ovarian follicles in various stages of development, ranging from 2.0 to 10.0 mm in diameter, collected from gilts from divergent genetic lines selected for high and low ovulation rates, during the 7 initial days of the follicular phase of the estrous cycle. EST sequences were obtained from 5231 distinct clones derived from this library. In total, 3479 unique sequence clusters were obtained, of which 2661 singletons (76.5%) were observed. BLASTN searches with the primary sequences from the clusters obtained resulted in 1037 sequences not matching ($E < 1.0^{-06}$) any of the sequences in the nt database (29.8% novelty rate). This resource will facilitate the use of cDNA microarrays in functional genomics studies aiming at unraveling the genetic and physiological mechanisms underlying follicle maturation and ovulation rate in swine.

Introduction

Significant progress is being made by research groups working on detecting, finding, and isolating genes involved in the regulation of important traits in humans, mice, and other species. A key factor for these achievements has been the generation of comprehensive resources for genetic analysis, such as high-density linkage and physical maps, saturated with known genes, ESTs, and polymorphic DNA markers.

Although public resources for completely sequencing the genomes of livestock species are not currently available, more modest efforts to sequence cDNA libraries from tissues most likely to express genes that affect economically important traits are well under way in cattle and pigs (see The Institute for Genomic Research [TIGR] Gene Indices; Quackenbush *et al.* 2001). These sequences, associated with comparative mapping data generated in silico (Ma *et al.* 1998) and/or by other means, compose a substantial resource of major value to the scientific community working with agriculturally important domestic animals.

Recent advances in gene expression analysis, mostly in the form of transcription profiling studies with microarrays (Schena *et al.* 1995; DeRisi *et al.* 1996), have provided an important alternative to research aimed at identifying genes involved in the genetic/physiological regulation of complex traits. However, the efficient development and application of cDNA microarray technology in these studies is dependent on the availability of large collections of non-redundant cDNA/EST clones, and, in many instances, these resources still remain to be generated.

In general, gene expression can vary considerably in a particular tissue depending on the developmental and/or physiological state of an organism. However, many EST discovery efforts are based on cDNA libraries generated from tissues collected at a single point in time. The ovaries of reproductively active female pigs are a good exam-

ple of organs that undergo major physiological changes during a relatively short time period. Of particular interest to animal scientists is the follicular phase of the porcine estrous cycle, when a highly dynamic process of ovulatory follicle selection takes place to determine the ovulation rate of the animal, a major factor influencing litter size (Johnson *et al.* 1984, 1999).

We generated a normalized cDNA library from ovarian follicles in various stages of development during the follicular phase of the porcine estrous cycle and obtained sequences from 5231 clones derived from this library. This resource will facilitate the use of functional genomics approaches (i.e., cDNA microarrays) to study the genetic mechanisms that are involved in determining ovulation rate in pigs. In addition, these resources will be useful for elucidating the physiological mechanisms involved in the process of follicle maturation in swine.

Materials and methods

Tissue collection, RNA extraction, and pooling. Gilts from UNL swine lines selected for high ($n = 7$) and low ($n = 7$) ovulation rates (Lamberson *et al.* 1991) were injected with PGF2α on day 13 following estrus detection. Ovaries were harvested by ovariectomy on days 0 to 6 following PGF2α injection, and follicles were rapidly dissected, measured (circumference), and snap frozen in liquid nitrogen. Total RNA was isolated from individual follicles with Trizol® (Gibco-Life Technologies) according to the protocol provided by the manufacturer. Equal amounts of RNA extracted from four to nine follicles of different sizes was used to make 25 µg RNA pools for each animal. The 14 pools of RNA were then combined into a single tube and used for isolating poly(A)$^+$RNA. In total, 83 ovarian follicles, ranging from 2.0 to 10.0 mm in diameter, were used.

cDNA library construction and normalization. Procedures to construct and normalize the cDNA library have been previously described in detail (Soares *et al.* 1994; Bonaldo *et al.* 1996). Briefly, poly(A)$^+$RNA was isolated from the pooled total RNA by two rounds of purification. In total, 1 µg of poly(A)$^+$RNA was mixed with 2 µg of *Not*I-dT_{18} oligonucleotide (TGTTACCATTCTGATGTTGGAGCGGCCGCCACAC-T_{19}), containing the sequence CACAC as a library tag, and reverse transcribed with SuperScriptII (Gibco Life Technologies). Second-strand synthesis was performed with T4 DNA polymerase in the presence of DNA ligase and RNase H. The size-selected cDNAs were ligated to *Eco*RI adapters and restriction digested with *Not*I. The purified cDNA fragments were directionally ligated to the pT7T3-Pac vector (Bonaldo *et al.* 1996) and electroporated into DH10B *E. coli* bacteria. Library quality [i.e., small poly(A)$^+$ tails, absence of clones with rRNA inserts] was assessed by sequencing 96 clones.

The library normalization process was based on the reassociation of an excess of cDNA inserts (generated by PCR) to the cDNA library in the form of single-stranded circles (Soares *et al.* 1994; Bonaldo *et al.* 1996). Single-stranded plasmid DNA generated in vivo was purified via chromatography in a hydroxyapatite (HAP) column. In total, 2 ng of the single-stranded plasmid library was used as template in high fidelity PCR (EHF; Boehringer Mannheim) to amplify cDNA inserts with T7/T3 primers annealing to regions flanking the vector-cloning site. An excess of PCR products was mixed with 50 ng of single-stranded circular library DNA along with 5′, 3′, and polyA$^+$ blocking oligos and allowed to hybridize for 24 h (Cot = 5). The single-stranded circles remaining were purified by HAP chromatography, converted to double-stranded plasmids, and electroporated into DH10B bacteria. Bacterial colonies were picked and grown in 96-well plates. Glycerol stocks were made in duplicate and stored at –80 °C at UNL.

Sequencing. Plasmid DNA was prepared from clones arrayed in 96-well plates according to methods described by Marra *et al.* (1999). Reactions to sequence the 30 ends of the cDNA inserts were performed with the Thermo Sequenase Fluorescent Labeled Primer Cycle Sequencing Kit with 7-Deaza-dGTP (Amersham-Pharmacia) in 384-well plates, according to the supplied protocols. Sequencing products were run in Li-Cor DNA Analyzer 4200 automated sequencers with the parameters recommended by the manufacturer. Sequencing gels were analyzed with e-Seq™ software (Li-Cor; Lincoln, Neb.).

Sequence processing. Sequence files with quality values were produced and processed locally. Programs were used to trim raw sequences based on sequence quality, to find the initial vector sequence, restriction site, library tag, poly(A)$^+$ tail, polyA signal, and trim vector sequences. Sequences were screened for contamination with bacterial, vector and mitochondrial sequences, masked (RepeatMasker; http://ftp.genome.washington.edu/cgi-bin/RepeatMasker) and clustered (UIcluster; http:// genome.uiowa.edu/pubsoft/clustering/). BLASTN searches of the nt database (version posted on 12/03/2000) were performed with the primary sequences from each cluster. Sequences were annotated and submitted to GenBank.

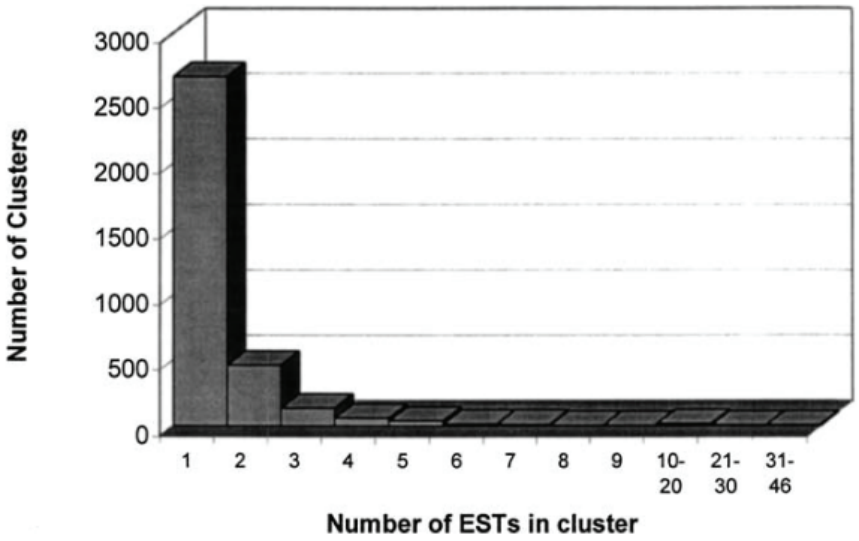

Figure 1. Overall redundancy of the swine ovarian follicle cDNA sequences. A total of 5231 sequences obtained from unique cDNA clones was assembled into 3479 clusters. The x-axis indicates the number of sequences contained in each cluster, while the y-axis indicates the number of clusters of a particular size. A total of 2661 singletons (76.5%) was observed, and only 33.5% of the sequences were redundant.

Quality control and clone re-arraying. A subset of 3636 clones representing the primary sequences of each cluster with an additional 157 clones were re-arrayed into new, 96-well plates with a Packard MultiPROBE® II automated liquid handling system. Clones from the top row from each re-arrayed plate were sequenced to provide a measure of data tracking quality control.

Results

The initial start-up cDNA library contained 5×10^6 cfu, with an average insert size of 1.2 kb. Average polyA$^+$ tail length in the clones was determined to be 24 bp. The success of the subsequent normalization process was assessed by the clustering results obtained with the sequences from the normalized cDNA library. Sequences of sufficient quality to be processed and annotated were obtained from a total of 5231 unique cDNA clones. These sequences were assembled into 3479 clusters (33.5% of the sequences were redundant). Figure 1 shows a histogram of the number of sequences contained in the clusters obtained from these data. In total, 2661 singletons (76.5%) were observed. On nine clusters contained more than 22 sequences. The total number of sequences submitted to GenBank, including multiple sequences of the same clone, was 5523 (accession numbers: BI180971–BI186447, BI233963–BI234008).

BLASTN searches with the primary sequences from the 3479 clusters obtained resulted in 1037 sequences not matching ($E < 1.0^{-06}$) any of the sequences in the nt database (29.8% novelty rate). Figure 2 illustrates the frequency of different classes of genes found within this set of sequences. A significant number of matches were observed between our EST clusters and human sequences derived from large-insert genomic libraries (6%). In total, 7% and 9% of our clusters matched sequences from human hypothetical proteins and cDNAs of unknown function, respectively. Matches between sequences from 65 clusters and different ribosomal proteins were found.

In total, 5490 sequences were incorporated into the TIGR Porcine Gene Index (SsGI, version 4.0, 02/01/02). Sequences from the start-up and normalized libraries were incorporated into 2115 Tentative Consensus (TC) sequences. Of these TCs (containing 526 sequences), 249 are unique to these libraries. Figure 3 shows a histogram of the number of sequences from our libraries in the Porcine Gene Index TCs.

Discussion

Ovulation rate is a complex trait, having high economic importance to the swine industry because it is a major component determining litter size (Johnson *et al.* 1984, 1999). Even though several hormonal signals involved in regulating ovarian follicle maturation and ovulation rate are known to be produced by the hypothalamus, anterior pituitary, and ovaries, many of the specific underlying mechanisms regulating ovulation rate remain to be elucidated. This project was conceptualized to generate a significant resource for utilizing cDNA microarrays to carry out functional genomics experiments to study the genetic and physiological architecture of the biological processes that take place in the swine ovary to determine ovulation rate.

Improvement of swine reproductive traits is of major importance to the pork industry. Attempts to genetically improve reproduction traits in pigs by direct selection have been relatively slow, hindered by low realized

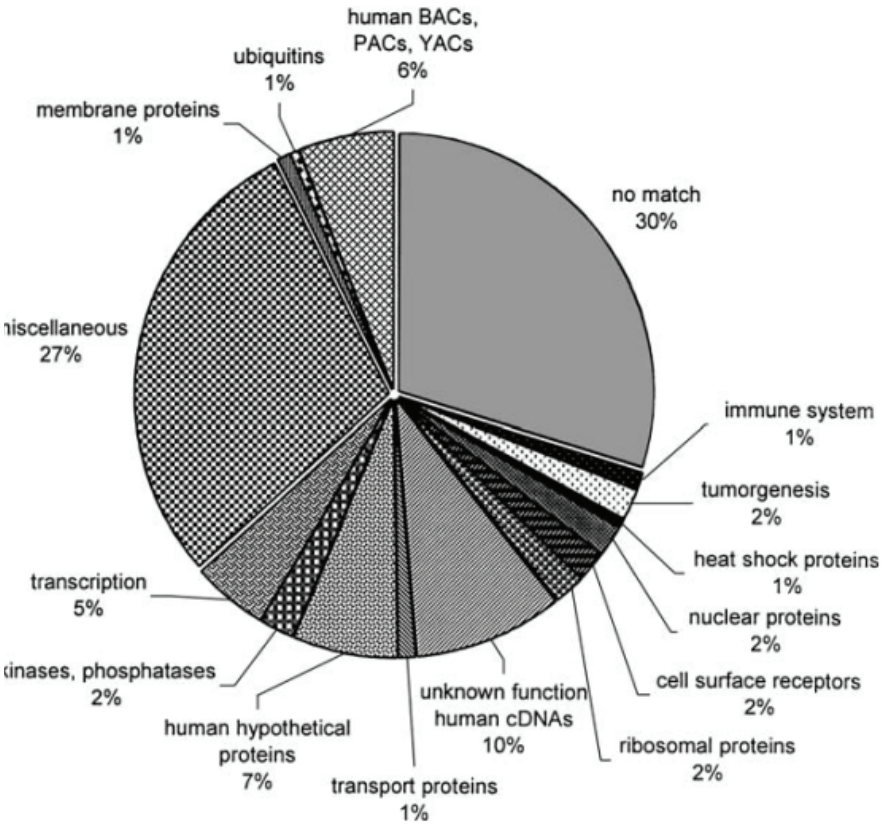

Figure 2. Functional classification of the swine ovarian follicle cDNA sequences. BLASTN searches of the GenBank nt database were performed with the primary sequences from the 3479 clone clusters. Distribution of matches in gene classes is shown. A total of 1037 sequences did not match ($E < 1.0^{-06}$) any of the sequences in the nt database (29.8% novelty rate).

heritabilities. Conversely, experiments utilizing index selection for components of litter size have been more successful (Johnson *et al.* 1984, 1999). Attempts to identify genetic factors responsible for the ovulatory advantage found in certain swine breeds (Rohrer *et al.* 1999; Wilkie *et al.* 1999) and selection lines (Cassady *et al.* 2001; Linville *et al.* 2001), with quantitative trait locus (QTL) and candidate gene analysis approaches, have produced varying results but have added little to our understanding of the genetic and physiological control of reproduction. Lack of power to resolve locus location and to identify the specific underlying genes has limited the utility of these approaches in regard to genetic improvement of reproductive traits. Furthermore, these efforts have yielded very little information about the biological phenomena underlying the substantial phenotypic variation observed.

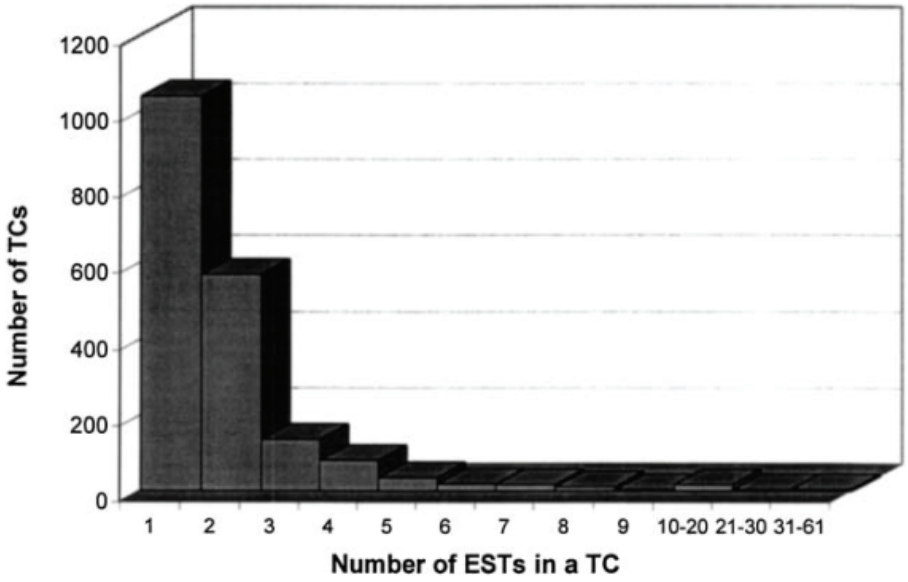

Figure 3. Distribution of the ovarian follicle cDNA sequences in the TIGR Swine Gene Index. EST sequences submitted to GenBank were incorporated into the Porcine Gene Index at The Institute for Genomic Research (TIGR). The figure shows the distribution of the sequences that were incorporated into Tentative Consensus (TC) gene sequences. Similar to what was observed with our local clustering, the number of EST sequences within the TCs indicates a low rate of redundancy of the normalized swine ovarian follicle cDNA library.

Functional genomics approaches that integrate genetic and physiological aspects of a particular biological event have the potential to provide discoveries that will lead to alternative ways to improve reproduction in swine and enhance our understanding of the physiological mechanisms involved.

Expression profiling studies with cDNA micro-arrays are being used to dissect the genetic/physiological architecture of complex traits such as obesity (Liang and Tall 2001) and cancer (Alizadeh *et al.* 2000) and will be highly useful to study the biology of follicle maturation and ovulation in swine. Even though it is possible to use cross-species and cross-tissue probes in microarray hybridizations, the most informative experiments are usually conducted by maximizing the number of cDNA probes for the genes expressed in the biological system being studied (i.e., cDNAs derived from the same species/same tissue).

When this project was initiated, no cDNA libraries were available from porcine ovarian follicles. Several cDNA libraries from ovarian tissue were available from other mammalian species, but these were usually derived from material obtained from a single time point, representing a specific physiological state. The mammalian ovary of fertile females is very dynamic, periodically undergoing major physiological changes at every estrous cycle, where gene expression levels change significantly. The follicular phase is of particular importance in this process because it is when ovulatory follicles are selected and matured. Ovarian follicles are more actively involved in this process than the many other tissues that compose the ovary.

Our strategy to construct a cDNA library from a pool of RNA extracted from ovarian follicles at different stages of development, obtained at different stages of the follicular phase of the porcine estrous cycle, was designed to maximize the complexity of the library and its utility for future functional genomics studies targeted to understanding the ovulation rate in pigs. Animals from two genetically divergent swine lines, previously selected for high and low ovulation rates (Lamberson *et al.* 1991), were used to increase the chances that gene messages putatively associated with this trait would be well represented in the library. Other efforts to sequence cDNA clones from swine ovarian tissue (Clouscard-Martinato *et al.* 1998; Tuggle *et al.* 2001; Fahrenkrug *et al.* 2002) are currently under way and are complementary to this work. The tissues used for producing those ESTs were obtained from different phases of the swine estrous cycle, or *in vitro* cell cultures, and therefore are very likely to have a different composition of transcribed messages.

Some differences can be observed in the sequence clustering results obtained here and in the results presented at the TIGR Porcine GI. Many of the sequences included in our final clustering analysis were excluded from the Porcine GI because of more stringent sequence quality parameters. Even though the same low sequence redundancy can be observed in both cases (Figures 1 and 3), inspection of the TIGR swine gene index revealed that some sequences that belong to different clusters in our sequence analysis have been put into the same tentative consensus (TC) sequences. Large contigs, containing 5′ and 3′ sequences from multiple cDNA clones derived from different libraries, were generated at the Porcine GI. This allowed for sequences with no homology, obtained from cDNA clones derived from different regions from the same gene, to be aligned into the same cluster.

It is likely that the high level of novelty observed in the results we obtained by performing BLASTN searches of the nt database (29.8% of our sequences did not match any sequences in the database) is an overestimate. Our sequences were derived from the 3′ end of the cDNA clones and therefore contain untranslated gene regions that are not necessarily highly conserved across taxa. However, this strategy was highly useful to achieve our main goal of identifying cDNA clones containing unique sequences. This will also facilitate enhanced success of physical mapping of these clones with porcine–rodent somatic cell and/or radiation hybrid panels.

In total, 59 ribosomal proteins of the 80 components of the ribosome were found among the clones we sequenced, indicating that we have a significant sample of the messages expressed in the ovarian follicle. Conversely, the frequency of redundant clones of each of these sequences (2.3 ± 0.40, data not shown) indicates that further sequencing of the ovarian follicle normalized library will yield additional unique sequences at a high rate. Several EST sequences for genes involved in the synthesis of estrogen, as well as other biological processes relatively specific to the ovarian follicle, were found. However, several other genes known to be expressed in the ovarian follicle (i.e., follistatin, inhibin, STAR, etc) were not found among the sequenced clones.

Acknowledgments

We are grateful to Joe Ford for assistance with the training of ovarian follicle dissection. We thank Denny Aherin, Christine Gladney, and Derek Petry for assistance with animal care, surgeries, and/or dissection of ovarian tissues. We also thank Dr. Bento Soares and Dr. Maria de Fati-

ma Bonaldo for training A.R. Caetano in cDNA library construction and normalization procedures, and Dr. Tom Casavant for consultations in setting up the EST data processing pipeline. This project was partially funded by a Postdoctoral Training Grant from Cotswold Pig Development Company (UK), by a special Hatch Grant allocation from UN-L (IANR), and by the UN-L Center for Biotechnology (Animal Molecular Biology Focus Group). This work is published as paper number 13760 of the Journal Series, Nebraska Agricultural Experiment Station.

References

1. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, *et al.* (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature **403**, 503–511
2. Bonaldo MF, Lennon G, and Soares MB (1996) Normalization and subtraction: two approaches to facilitate gene discovery. Genome Res **6**, 791–806
3. Cassady JP, Johnson RK, Pomp D, Rohrer GA, Van Vleck LD, *et al.* (2001) Identification of quantitative trait loci affecting reproduction in pigs. J Anim Sci **79**, 623–633
4. Clouscard-Martinato C, Mulsant P, Robic A, Bonnet A, Gasser F, *et al.* (1998) Characterization of FSH-regulated genes isolated by mRNA differential display from pig ovarian granulosa cells. Anim Genet **29**, 98–106
5. DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, *et al.* (1996) Use of a cDNA microarray to analyze gene expression patterns in human cancer. Nat Genet **14**, 457–460
6. Fahrenkrug SC, Smith TPL, Freking BA, Cho J, White J, *et al.* (2002) Porcine gene-discovery by normalized cDNA-library sequencing and EST cluster assembly. Mamm Genome **13**, 475–478
7. Johnson RK, Zimmerman DR, and Kittok RJ (1984) Selection for components of reproduction in swine. Livest Prod Sci **11**, 541–558
8. Johnson RK, Nielsen MK, and Casey DS (1999) Responses in ovulation rate, embryonal survival, and litter traits in swine to 14 generations of selection to increase litter size. J Anim Sci **77**, 541–557
9. Lamberson WR, Johnson RK, Zimmerman DR, and Long TE (1991) Direct responses to selection for increased litter size, decreased age at puberty, or random selection following selection for ovulation rate in swine. J Anim Sci **69**, 3129–3143
10. Liang CP, and Tall AR (2001) Transcriptional profiling reveals global defects in energy metabolism, lipoprotein, and bile acid synthesis and transport with reversal by leptin treatment in ob/ob mouse liver. J Biol Chem **276**, 49066–49076
11. Linville RC, Pomp D, Johnson RK, and Rothschild MF (2001) Candidate gene analysis for loci affecting litter size and ovulation rate in swine. J Anim Sci **79**, 60–67
12. Ma RZ, van Eijk MJ, Beever JE, Guerin G, Mummery CL, *et al.* (1998) Comparative analysis of 82 expressed sequence tags from a cattle ovary cDNA library. Mamm Genome **9**, 545–549
13. Marra MA, Kucaba TA, Hillier LW, and Waterston RH (1999) High-throughput plasmid DNA purification for 3 cents per sample. Nucleic Acids Res **27**, e37
14. Quackenbush J, Cho J, Lee D, Liang F, Holt I, *et al.* (2001) The TIGR gene indices: analysis of gene transcript sequences in highly sampled eukaryotic species. Nucleic Acids Res **29**, 159–164
15. Rohrer GA, Ford JJ, Wise TH, Vallet JL, and Christenson RK (1999) Identification of quantitative trait loci affecting female reproductive traits in a multigeneration Meishan-White composite swine population. J Anim Sci **77**, 1385–1391
16. Schena M, Shalon D, Davis RW, and Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science **270**, 467–470
17. Soares MB, Bonaldo MF, Jelene P, Su L, Lawton L, *et al.* (1994) Construction and characterization of a normalized cDNA library. Proc Nat Acad Sci USA **91**, 9228–9232
18. Tuggle CK, Green JA, Fitzsimmons C, Woods R, Prather R, *et al.* (2001) Production of 17 cDNA libraries and successful EST sequencing of 10,124 clones from porcine female reproductive tissues. J Anim Sci **79**, Suppl. 1, 338
19. Wilkie PJ, Paszek AA, Beattie CW, Alexander LJ, Wheeler MB, *et al.* (1999) A genomic scan of porcine reproductive traits reveals possible quantitative trait loci (QTLs) for number of corpora lutea. Mamm Genome **10**, 573–578