Reproduction Supplement **58**, 277–292

Applying functional genomics research to the study of pig reproduction

D. Pomp, A. R. Caetano, G. R. Bertani, C. D. Gladney and R. K. Johnson

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

Functional genomics is an experimental approach that incorporates genome-wide or system-wide experimentation, expanding the scope of biological investigation from studying single genes to studying potentially all genes at once in a systematic manner. This technology is highly appealing because of its high throughput and relatively low cost. Furthermore, analysis of gene expression using microarrays is likely to be more biologically relevant than the conventional paradigm of reductionism, because it has the potential to uncover new biological connections between genes and biochemical pathways. However, functional genomics is still in its infancy, especially with regard to the study of pig reproduction. Currently, efforts are centred on developing the necessary resources to enable high throughput evaluation and comparison of gene expression. However, it is clear that in the near future functional genomics will be applied on a large scale to study the biology and physiology of reproduction in pigs, and to understand better the complex nature of genetic control over polygenic characteristics, such as ovulation rate and litter size. We can look forward to generating a significant amount of new data on differences in gene expression between genotypes, treatments, or at various temporal and spatial coordinates within a variety of reproductively relevant systems. Along with this capability will be the challenge of collating, analysing and interpreting datasets that are orders of magnitude more extensive and complex than those currently used. Furthermore, integration of functional genomics with traditional genetic approaches and with detailed analysis of the proteome and relevant whole animal phenotypes will be required to make full use of this powerful new experimental paradigm as a beneficial research tool.

General background

Improvement of reproductive traits in pigs remains of great importance and interest to the pork industry. The sow is the unit of production, and increases in litter size and preweaning viability have been highlighted as major factors for reducing the costs of producing pork (Tess *et al.*, 1983a,b). Although geneticists have made much progress in the improvement of production efficiency and carcass quality traits in pigs with the use of traditional quantitative genetic methods (Clutter and Brascamp, 1998), reproductive traits have been more difficult to

Email: dpomp@unl.edu

© 2001 Society for Reproduction and Fertility

enhance due to low heritability (10–15%). For this reason, several studies have focused on genomic technologies to study pig reproduction, primarily in the form of quantitative trait loci (QTL) and candidate gene identification for traits such as ovulation rate, litter size and age at puberty (see, Rothschild, 2000; Table 5 in Cassady *et al.*, 2001). Although there have been some successes, large variation in results and lack of power to localize the underlying genes have limited the use of DNA marker-assisted selection for improvement of reproduction. In addition, these efforts have yielded very little basic information regarding the biology of reproduction. Detected QTL represent only broad regions containing potentially hundreds or even thousands of genes, and even evidence for candidate gene associations (for example, oestrogen receptor; Rothschild *et al.*, 1996) may be caused by other genes (known or unknown) in linkage disequilibrium with the candidate locus.

Thus, new methods and models for identification of genes are required to understand better the genetics underlying reproduction in pigs. Furthermore, an approach that better integrates genetic and physiological investigation is well overdue, and promises to provide synergistic discoveries leading to alternative ways to improve these traits through genetic selection, therapeutics and management practices. Functional genomics, along with its 'sister technology' proteomics, is currently the most appropriate technology to achieve this aim.

What is functional genomics?

The term 'functional genomics' is widely used, but is often poorly understood given a lack of common definition. In its broadest sense, functional genomics encompasses all methods that enable a better understanding of the function of a gene, and includes a wide variety of techniques including DNA sequencing, evaluation of gene expression, and even mutational and transgenic analysis. However, a more useful definition of functional genomics incorporates the genome-wide or system-wide experimentation, expanding the scope of biological investigation from studying single genes or proteins to studying all genes or proteins at once in a systematic manner. This definition will be used in this review.

In some organisms (including humans and mice), functional genomics is built upon the framework of knowing the near-complete sequence of the genome, and is used to narrow the gap between sequence and function. In livestock species, including pigs, data are limited to fairly extensive collections of partially sequenced regions of expressed genes. Although this may limit the scale (and progress) of functional genomic studies, it does not change the nature or overall goals of investigation.

Large-scale functional genomic studies that attempt to understand pig reproduction are fairly new and ongoing. Thus, this review must be based primarily on progress reports, abstracts and updates rather than on a large body of published data. The aims of this review are to define the strategies of functional genomics applied to understanding the genetic and physiological control of pig reproduction, to describe the primary existing resources and research efforts, and to speculate on the future utility and benefit of this approach in increasing our understanding of the genetics and physiology of reproduction in pigs. Although it is recognized that analysis of the proteome is critical and mandatory for understanding the connection between genotype and phenotype, the focus will be on analysis of gene expression at the mRNA level, or the transcriptome.

Functional genomics – transcriptome analysis

The development of a new approach to investigate the expression of thousands of genes, in parallel, in complex biological samples has produced a paradigm shift in biology (Lander,

1999). It is now possible to look at patterns of expression of the entire genome of an organism (for example, yeast) during normal cell processes, or in response to environmental changes (Schena *et al.*, 1995, 1996; Brown and Botstein, 1999). This technology has also been used to gain new insights into human gene expression in cancerous compared with normal cells (DeRisi *et al.*, 1996), or the response of fibroblasts to serum starvation (Iyer *et al.*, 1999). Thus, this exploratory paradigm can be used in the absence of complete genome information without losing its intrinsic power, because of the high throughput and relative low cost of the technology.

Expression microarrays can be thought of as reverse northern blots, in which thousands of DNA 'probes' are individually spotted (printed) on to a solid support (glass slide) in defined patterns or grids. The probes are usually derived from partially sequenced cDNA clones or from sets of oligonucleotides representing expressed genes. Subsequently, fluorescent targets are produced from mRNA isolated from two distinct biological samples (for example, cancerous cells and normal cells, or tissue samples from two genetic lines). Each mRNA population is labelled with a different fluorescent dye (usually Cy3 and Cy5), mixed in equimolar proportions and hybridized to the microarray. After stringent washes, detection of the remaining fluorescent material at each spot is made with a microarray laser scanner, which can accurately estimate the intensity of each fluorescent dye at each spot. Thus, it is possible to determine the expression of each mRNA in both samples at the same time. The data are usually represented as the ratio of the expression detected between the two mRNA samples used. These experiments produce large sets of data, which require specialized computer software to manage and analyse them (Ermolaeva *et al.*, 1998; Bassett *et al.*, 1999). Indeed, bioinformatics may be the most important, complex and challenging aspect of using microarrays to dissect complex traits, such as reproduction at the transcription level.

Reports of changes in gene expression associated with reproductive phenotypes in pigs have been mostly gene specific. The evaluation of differences in expression profiles of thousands of genes in parallel offers powerful new opportunities for gene discovery. Although the advent of differential display PCR has enabled preliminary investigations of changes in gene expression in whole tissue, the methods are laborious and not facilitative of large-scale analysis of various time points, tissue sources and treatments. Thus, the ability to analyse the expression of thousands of genes simultaneously makes microarray technology a valuable tool for studying biological events in domestic species. The advantage of the technology is its high throughput and relatively low cost. Furthermore, microarray analysis is likely to be more biologically relevant than the conventional paradigm of reductionism because it has the potential to uncover new biological connections between genes and biochemical pathways. This ability may be especially important for biological events that are specific to pigs and that have not been studied in model species.

Although several studies are underway to develop microarrays containing thousands of pig genes, no research data have been published to date. However, functional genomic analysis of gene expression using microarrays has been applied to the study of reproduction in other species, providing an example of what might be forthcoming for scientists working with pigs. These studies range from identification of sex-regulated transcriptional changes in *Caenorhabditis elegans* (Jiang *et al.*, 2001), the evaluation of gene expression profiles during folliculogenesis (Liu *et al.*, 2001) and mid-gestation embryo and placenta (Tanaka *et al.*, 2000) in mice, to identification of differentially expressed genes in human myometrium during pregnancy and labour (Aguan *et al.*, 2000).

The ability to examine the transcriptome on a system-wide or genome-wide basis enables not only a thorough comparison between treatments, developmental stages or genotypes, but also the opportunity to identify and link pathways that play critical roles in regulating

important reproductive phenotypes. This bridge between genetics and physiology will be critical for implementing a fully integrated research programme combining quantitative genetics, genomics, proteomics, metabolomics and phenomics. It is likely that such an approach will be required to dissect fully the complex and polygenic nature of reproductive traits in pigs, and lead to discoveries that will have a strong impact on improvement of reproduction in the pork industry.

Functional genomics in pigs – early studies

For many years, researchers have focused on analysis of expression of individual genes involved in pig reproductive processes. For example, Li *et al.* (1997) evaluated the expression of follistatin and inhibin/activin alpha, beta (A) and beta (B) subunit genes in pig ovarian follicles during the follicular phase of the oestrous cycle, whereas Yelich *et al.* (1997a,b) focused on gene expression during the period of rapid trophoblastic elongation in the pig conceptus. Although much has been learned from these (and many other) single gene investigations, the use of functional genomic approaches promises to carry transcriptome analysis further by providing access to many genes in parallel. Although current and future efforts are focused on microarrays, functional genomic analysis of transcription differences has been taking place in the absence of genomic sequence or even large, defined sets of expressed sequence tags (ESTs). Differential display PCR (Liang and Pardee, 1992) has been used successfully to isolate genes that have important roles in the reproductive biology of pigs.

Li *et al.* (1996) used differential display PCR to isolate transcripts that are differentially expressed in the anterior pituitary gland of Meishan and White Composite sexually mature boars. Several transcripts showing differential amplification were isolated including the gene for the β-subunit of thyroid stimulating hormone (TSH-β). Several attempts have been made to isolate genes that are expressed by the uterus and the conceptus during early pregnancy using differential display PCR. Green *et al.* (1996) used differential display PCR to isolate putative transcripts differentially expressed in the peri-implantation (days 11–12) endometrium of unilaterally pregnant pigs, including the pig gene homologous to human spermidine/spermine N1-acetyltransferase (SSAT). Chang *et al.* (2000) conducted differential display PCR experiments to isolate transcripts from the epithelium of oviducts from gilts carrying embryos at various stages of early development. Wilson *et al.* (2000) used RNA arbitrarily primed-PCR (RAP–PCR) to search for genes differentially expressed by trophectodermal cells during elongation of the pig conceptus. The pig heterogeneous nuclear ribonucleoprotein (hnRNP) A2/B1 was found to be differentially expressed by trophectodermal cells during the elongation process.

Differential display PCR has also been used to isolate genes that are differentially expressed by cell cultures derived from pig gonadal tissue. Clouscard-Martinato *et al.* (1998) used differential display PCR to isolate transcripts differentially expressed by pig primary granulosa cells in culture in the presence or absence of FSH. A total of 16 differentially amplified differential display PCR fragments were isolated and sequenced, of which 12 were regulated by FSH. Tosser-Klopp *et al.* (1997) isolated 238 clones by differential hybridization from a pig granulosa cell cDNA library, using probes prepared from RNA extracted from either untreated or FSH-treated cells. In a follow-up study (Tosser-Klopp *et al.*, 2001), several of these clones were functionally evaluated in granulosa cells (with or without FSH treatment) and in fresh isolated ovarian follicles using comparative RT–PCR analysis.

In addition, differential display PCR has been used in attempts to dissect the genetic response to long-term selection for components of litter size in pigs, using both anterior

pituitary gland (Bertani *et al.*, 2000a,b) and ovarian follicle tissue (Gladney *et al.*, 2000a,b). This work is described in more detail below.

EST development

The raw material for large-scale functional genomics research in pigs is DNA sequence. Although whole genome sequencing in pigs is unlikely to occur in the near future (at least in the public domain), several large-scale studies to develop extensive catalogues of expressed sequence tags from a variety of relevant pig tissues are in progress.

Scientists at the Roman L. Hruska US Meat Animal Research Center (MARC) are sequencing (5′ single pass reads) from two normalized cDNA libraries constructed from pooled tissues (library 1: embryos from day 10 to day 30; and library 2: testes, ovary, pituitary, hypothalamus, placenta and endometrium). Over 43 000 EST sequences have been submitted to Genbank (B. A. Freking, personal communication). Sequencing within these libraries is continuing, and a third pooled-tissue library is currently being processed for creation of a new normalized cDNA library.

The Animal Science Department at the University of Missouri (principal investigators: R. S. Prather and B. N. Day), through primary funding from Monsanto Company, is focusing EST discovery efforts on female reproductive tissues (oocyte, preimplantation stage embryo, oviduct and uterus). To date, this project has generated 29 658 EST sequences (ascswine.rnet.missouri.edu).

The Pig Reproduction EST Project is a publicly funded (USDA-NRICGP) consortium comprising Iowa State University (Coordinating Institution; pigest.genome.iastate.edu), the University of Iowa, the University of Missouri, the University of Nebraska and the National Center for Genomic Resources (NCGR; Albuquerque, NM). The consortium has produced a total of 17 libraries from the following tissues: whole embryo (day 14, day 20 and day 45 of gestation); term placenta; anterior pituitary (day 0, day 5 and day 12 of oestrus); hypothalamus (day 0, day 5 and day 12 of oestrus); ovary (day 0, day 5 and day 12 of oestrus); and uterus (day 12 and day 14 of gestation). By January 2001 (Tuggle *et al.*, 2001), a total of 10 124 EST sequences (out of an overall goal of 20 000) had been produced and most have been submitted to Genbank. As assessed by cluster analysis of 9911 sequences, these data represent 6655 different genes for a novelty rate of 67.2%. In addition, selected ESTs ($n = 700$) will be localized physically using radiation hybrid mapping to improve the density of genes in the pig genetic maps and to facilitate more powerful comparative mapping. Approximately 10% of this mapping has been completed. Such comparative maps will be critical for capitalizing on the large-scale studies undertaken to sequence the complete human and mouse genomes.

Our group is focusing on ESTs from a normalized library developed from ovarian follicles (2–10 mm in diameter) at various stages of folliculogenesis (Caetano *et al.*, 2001). A total of 5231 processed sequences were assembled into 3479 clusters (33.5% of the sequences were redundant). BLAST searches of the Genbank nucleotide database (non-redundant; does not include EST database) were performed with sequences from each cluster. A total of 1037 sequences did not match any of the sequences in the nucleotide database (42% novelty rate). All sequences have been released into Genbank.

At the University of Illinois, libraries have been made from whole fetuses at various stages of early gestation, and from skeletal muscle from fetuses at the mid- and late stages of gestation (J. Beever, personal communication). Over 5000 clones have been sequenced from these libraries.

In addition, several smaller studies have contributed numerous ESTs to Genbank from reproductively relevant tissues, including granulosa cells (Tosser-Klopp *et al.*, 1997, 2001;

Clouscard-Martinato *et al.*, 1998), ovarian follicles (Gladney *et al.*, 2000a,b) and anterior pituitary (Bertani *et al.*, 2000a,b). All of these ESTs have resulted from projects focusing directly on functional analysis using differential display PCR or differential hybridization.

A pig gene index has been generated through collaboration between MARC and The Institute for Genomic Research (TIGR; Rockville, MD) and this index clusters all public EST sequences. As of August 2001, 13 017 tentative consensus clusters comprising 53 082 ESTs have been identified and there are an additional 30 014 single ESTs (Release 3.0; tigr.org/tdb/ssgi). Thus, it is likely that publicly available ESTs represent over 50% of the genes in the expressed pig genome.

The Nebraska functional genomics effort – a case study

Unique University of Nebraska (UN-L) genetic model

The UN-L selection lines represent a unique biological resource that will be valuable for identifying genes that control variation in ovulation rate, litter size and embryo survival in pigs, and for discovering pathways and mechanisms regulating reproduction. The high index (I) line has advantages of 7.4 ova, three fully formed pigs and 1.4 live pigs per litter compared with the randomly selected contemporary control (C) line (Johnson *et al.*, 1999). This research model differs considerably from others such as the Meishan paradigm, because the pigs are highly viable in a commercial setting. Understanding the genetics and biology of reproduction in these pigs can thus lead to immediate technology transfer to the industry, and the I line is now an important component of commercial pig production.

The observed difference in ovulation rate in the UN-L selection lines is primarily the result of changes in the dynamics of follicular development (Yen, 1999), which in turn have been shown by independent studies to be controlled by the expression of several genes. Our work to understand the genetic and physiological control of reproduction in these lines began with a genome-wide QTL screen (Cassady *et al.*, 2001) and analysis of selected candidate genes (Linville *et al.*, 2001). We next focused on gene expression analysis using differential display PCR methods (Bertani *et al.*, 2000a,b; Gladney *et al.*, 2000a,b). Recently, we have created and partially characterized a normalized ovarian follicle cDNA library (Caetano *et al.*, 2001), and production of follicle-specific expression microarrays is nearly completed, allowing for comprehensive investigation of the transcriptional factors that have changed as correlated responses to selection for components of litter size. We have thus created the resources necessary to monitor simultaneously the expression of thousands of genes normally expressed in ovarian follicles, and to use these resources to determine the expression profile of developing ovarian follicles in the UN-L selection lines.

QTL and candidate gene analysis

A three-generation resource population was developed by crossing low indexing animals from the randomly selected C line with high indexing animals from the high I selection line (Cassady *et al.*, 2001). Phenotypic data were collected in F_2 females for ovulation rate, age at puberty, litter size and number of nipples. Litter size data included number of fully formed, live, stillborn and mummified pigs. The genotypes of grandparent, F_1 and F_2 (n = 423) animals were determined for 151 microsatellite markers distributed across all 19 chromosomes. Average spacing between markers was approximately 19 cm. LOD scores were calculated by least squares, including fixed effects of sire–dam combination and replicate. Genome-wide significance thresholds of 5% and 10% were calculated using an empirical permutation approach. There was evidence at the 5% significance level that QTL affects ovulation rate on SSC9, age at puberty on SSC7 and SSC8,

number of stillborn pigs on SSC5 and SSC13, and number of fully formed pigs on SSC11. At the 10% significance level there was evidence of additional QTL affecting age at puberty on SSC7, SSC8, and SSC12, and number born live on SSC11.

In addition, a candidate gene approach was used to determine whether specific loci explain responses in ovulation rate and number of fully formed, live, stillborn and mummified pigs at birth observed in two lines selected for ovulation rate and litter size compared with a randomly selected C line (Linville *et al.*, 2001). The IOL line was selected for an index of ovulation rate and embryonic survival for eight generations, followed by eight generations of two-stage selection for ovulation rate and litter size. The C line was selected at random for 16 generations. The COL line, derived from the C line at generation eight, underwent eight generations of two-stage selection. IOL and C lines differed in mean estimated breeding value by 6.1 ova and 4.7 fully formed pigs, whereas COL and C lines differed by 2.2 ova and 2.9 fully formed pigs (Ruiz-Flores and Johnson, 2001). The genotype of pigs was determined for the retinal-binding protein 4 (*RBP4*), epidermal growth factor (*EGF*), oestrogen receptor (*ESR*), prolactin receptor (*PRLR*), follicle stimulating hormone β (*FSHβ*) and prostaglandin-endoperoxide synthase 2 (*PTGS2*) loci. On the basis of a chi-squared analysis for homogeneity of genotypic frequencies, distributions for *PRLR*, *FSHβ* and *PTGS2* were significantly different among lines. Differences in gene frequencies between IOL versus C and COL versus C were: 0.33 ± 0.25 and 0.16 ± 0.26 for *PRLR*, 0.35 ± 0.20 and 0.15 ± 0.24 for *FSHβ*, and 0.16 ± 0.16 and 0.08 ± 0.18 for *PTGS2*. Although these differences are consistent with a model of selection acting on these loci, estimates of additive and dominance effects at these loci did not differ from zero and several of them had signs that were inconsistent with the changes in allele frequencies. We were not able to find significant associations between the polymorphic markers and phenotypes studied; however, it cannot be ruled out that other genetic variation within these candidate genes has an effect on the traits studied.

The QTL studies in the UN-L selection lines, taken together with the other primary pig reproduction QTL experiments (Rohrer *et al.*, 1999; Wilkie *et al.*, 1999), demonstrate relatively inconsistent results and poorly resolved localizations of putative genes. The small number of detected reproduction QTL and the low detection resolution attained may be a consequence of the highly polygenic control of these traits, by loci with small effects that interact with each other and with the environment. Given current resources and power of investigation, it is extremely unlikely that the underlying polygenes controlling the inheritance of reproductive phenotypes will be identified.

Functional genomics, differential display PCR studies

Reproductive tissues (anterior pituitary gland and ovarian follicles) were collected from I and C sows from the UN-L selection lines on day 2 and day 4 after $PGF_{2\alpha}$ administration between day 12 and day 14 of the oestrous cycle (that is, follicular phase). Differential display PCR was conducted in both ovarian follicles and in anterior pituitaries, using 200 primer combinations (10 anchor oligo-dT 3′ primers and 20 arbitrary 5′ 10-mers), to compare gene expression between the selection lines.

Anterior pituitary. Differential display PCR yielded 372 putative differentially expressed bands; 151 of these were cloned and sequenced yielding anterior pituitary ESTs that were deposited into Genbank (Bertani *et al.*, 2000a,b). Confirmation of differential expression by northern blotting was undertaken for several of the most promising results, confirming that FSHβ, ferritin (heavy chain) and G-beta-like protein are expressed differently in the two lines (Fig. 1).

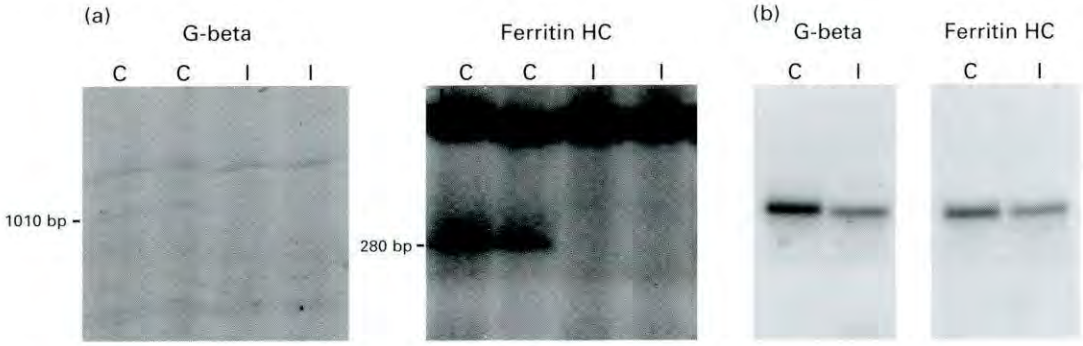


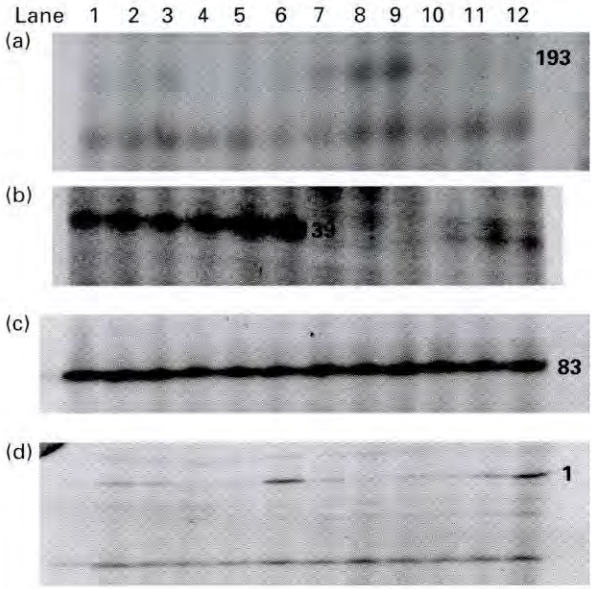
Fig. 1. (a) Differentially expressed genes in anterior pituitaries from the Nebraska high index and control selection lines of pigs. cDNA samples were amplified using PCR with 200 combinations of differential display PCR primers. (b) Differentially expressed bands were isolated, sequenced and used as probes in northern blots for confirmation of differences. Two confirmed differentially expressed genes are shown: G-beta-like protein and ferritin heavy chain (HC). C: control line; I: index line.

Ovarian follicles. Twelve pools of mRNA from follicles of different size classes dissected on day 2 or day 4 (Gladney *et al.*, 2000a,b) were used in the differential display PCR, resulting in 282 putative differentially expressed bands. Of these, 107 were cloned and sequenced, yielding 84 unique follicle ESTs that were deposited in Genbank. Northern blot hybridization confirmed expression results for calpain I light subunit, cytochrome C oxidase subunit III, cytochrome P450 aromatase and cytochrome P450 side chain cleavage. In addition to differences among selection lines, changes in gene expression were found between days of the follicular period, and between sizes of follicle (Fig. 2).

Further confirmation. Unfortunately, differential display PCR is labour intensive and is prone to a high rate of false positives. Given that each putative change in gene expression must be confirmed individually, we have printed all of the clones representing putative differentially expressed genes (n = approximately 250) from the above two experiments on to glass slides as microarrays. These microarrays are being used to confirm simultaneously each potential change in gene expression. In this manner, it is likely that many additional genes will be confirmed to have differential expression between the high I and C selection lines.

These two differential display PCR studies demonstrate changes in pituitary and follicular gene expression as the result of long-term selection for reproduction. Integration of QTL and expression studies will help to determine whether these changes represent actual genetic variation within predisposition genes (QTL), or whether they are the result of regulation of physiological genes that represent correlated responses to selection manifested downstream from the causative genetic variation. On the basis of comparison of the map locations of these differentially expressed genes and the QTL identified for reproductive traits in the same populations, it appears that the latter explanation is more likely.

In these studies, although interesting results were obtained that could result in potentially new mechanisms to investigate as contributors to genetic control of reproduction, it is difficult to examine simultaneously correlated genes and multiple members of pathways. Given that a plethora of EST data are now available in the pig, microarray analysis will be a much more powerful method for analysis of gene expression and large-scale functional genomics experimentation.



Lane	Line	Day	Size(mm)
1	I	2	4.0–4.25
2	I	2	4.5–4.75
3	I	2	5.0–5.25
4	I	4	4.0–4.5
5	I	4	5.0–5.5
6	I	4	6–7
7	C	2	4.0–4.25
8	C	2	4.5–4.75
9	C	2	5.0–5.25
10	C	4	4.0–4.5
11	C	4	5.0–5.5
12	C	4	6–7

Fig. 2. Differential display gel images of ovarian follicle mRNA showing (a) differences between days of folliculogenesis (day 2 versus day 4; band 193), (b) differences between lines (high index versus control; band 39), (c) bands demonstrating equal expression (band 83) and (d) differences between different follicle sizes (band 1). C: control line; I: index line.

Functional genomics, microarrays – cross-species hybridization

Before the development of pig ESTs and microarrays, the only opportunity to use microarray technology was to attempt cross-species hybridization with arrays developed for humans or mice. Incyte UniGEM human gene chips were used to examine changes in follicular gene expression between the UN-L selection lines (C. D. Gladney and D. Pomp, unpublished). For microarray analysis, two mRNA pools, containing follicles (day 2; 4.50–4.75 mm) from two I or two C line sows were evaluated on the Incyte UniGEM V1.0 human chip (approximately 7000 gene probes). An additional evaluation on the Incyte UniGEM V2.0 human chip (approximately 9100 gene probes) was performed using follicles (day 2; 4.50–5.00 mm) from four sows selected from the I line and five sows selected from the C line sows. UniGEM V1.0 results indicated significant differences between I and C lines (statistical significance defined as two-fold or greater relative expression when comparing I and C lines, following adjustment for expression of control probes) for 33 genes. UniGEM V2.0 results indicated 21 significant differences between C and I lines. Results were inconsistent between the two array hybridizations. However, expression differences for two genes, *follistatin* and *nuclear receptor subfamily 4, group A, member 1*, were confirmed using northern blot hybridization. This study represents the first application of microarray techniques to livestock populations and to study the underlying effects of long-term selection on gene expression. However, the current development of significant EST resources in pigs will enable pig-specific microarray experimentation, thus eliminating the need for future cross-species hybridization studies.

Functional genomics, microarrays – pig follicle microarray

Although the ultimate tool for transcriptome analysis will be a genome-wide microarray encompassing all pig genes, smaller, more directed projects could now be initiated with existing techniques and resources. Given our focus on understanding the changes that have occurred in the ovarian follicle as the result of selection for litter size, we have established resources to investigate this in a high throughput and powerful manner. Our experimental paradigm is presented as a flow-chart (Fig. 3).

cDNA library construction. Gilts from high and low UN-L ovulation rate selection lines (Lamberson *et al.*, 1991) were injected with $PGF_{2\alpha}$ on day 13 (day 0 of treatment) of the oestrous cycle. Ovaries were harvested by ovariectomy on days 0–6 after treatment, and follicles were dissected. mRNA isolated from ovarian follicles (2.0–10.0 mm in diameter) was used to construct a cDNA library (5×10^6 c.f.u.) with an average insert size of 1.2 kb. This cDNA library was subsequently normalized to decrease the relative number of redundant messages derived from highly transcribed genes (Soares *et al.*, 1994; Bonaldo *et al.*, 1996).

Plasmid DNA was isolated from clones derived from the normalized ovarian follicle cDNA library for sequencing. Raw sequences obtained with Li-Cor 4200 (Lincoln, NE) sequencers were processed with an automated sequence-processing software pipeline. A total of 5231 processed sequences were assembled into 3479 clusters (33.5% of the sequences were redundant). BLAST searches of the nucleotide database (all non-redundant GenBank + EMBL + DDBJ + PDB sequences, but no EST, STS, GSS, or HTGS sequences) were performed with sequences from each cluster (for reference, see www.ncbi.nlm.nih.gov/BLAST). A total of 1037 sequences did not match any of the sequences in the nucleotide database (42% novelty rate).

Microarray construction. Clones representing the 3479 unique EST clusters, the 107 follicle ESTs isolated in the differential display PCR study and several other genes known to be expressed in follicles but not represented in the clone list were arrayed on glass slides in duplicate. A total of 120 duplicated arrays have been produced to support an experimental design to compare gene expression between the I and C selection lines (Fig. 4).

Expression studies. The UN-L selection lines have different numbers of follicles of different size classes during the follicular phase. Labelled mRNA will be hybridized to the follicle microarrays from whole ovaries and dissected follicles from gilts (generation 18; high I and C selection lines). Studies with whole ovaries will provide a complete, overall picture of follicular gene expression, by including follicles of all sizes and health status. Studies within pools of healthy follicles of similar size are designed to search for expression differences within and between standardized conditions, to provide insights into the mechanisms underlying differences in ovulation rate in the I and C selection lines.

The merging of genetics and reproductive biology

The chromosomal regions that have been found to harbour reproduction QTL do not appear to coincide with many of the most prominent candidate genes, selected based on physiological evidence. Consider QTL to represent 'redisposition genes' or those within which heritable genetic variations (DNA polymorphisms) explain phenotypic variation in traits. Alternatively, consider genes the protein products of which are key regulators of reproduction to represent 'physiological genes'. Merging of existing databases and genetic maps indicates a

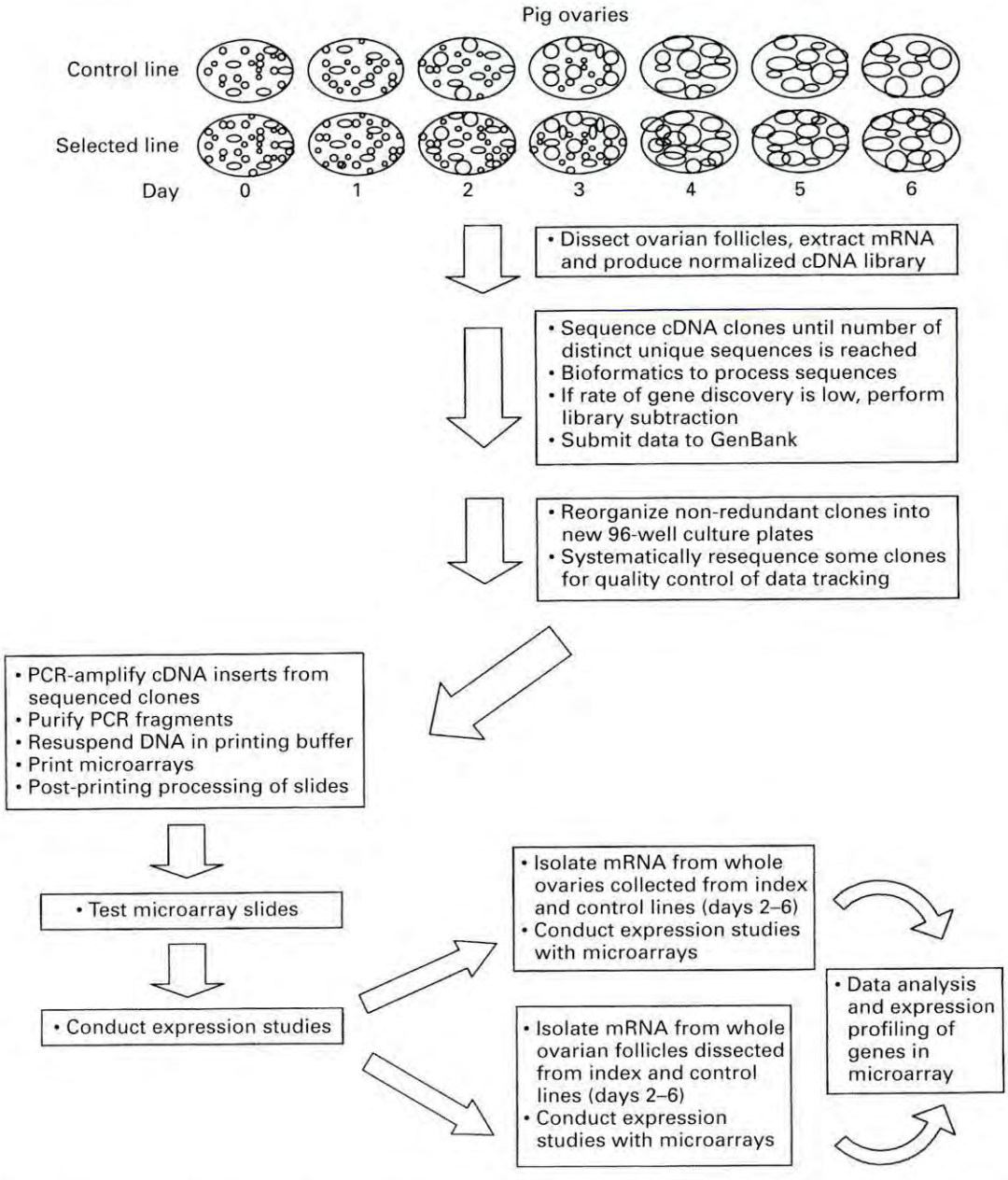


Fig. 3. Flow chart of Nebraska Functional Genomics Project. Initially, a normalized library was created and characterized by sequencing genes. Microarrays were printed on glass slides and expression analysis was used in an attempt to determine the transcriptional changes caused by selection for enhanced reproduction.

potential 'polygenic paradox', whereby predisposition and physiological genes appear to represent two distinct subsets of genes. In other words, evidence for an important role of a protein in regulation of a phenotype does not necessarily implicate the underlying gene as a QTL. This may represent our relatively poor knowledge of existing genes (that is, many

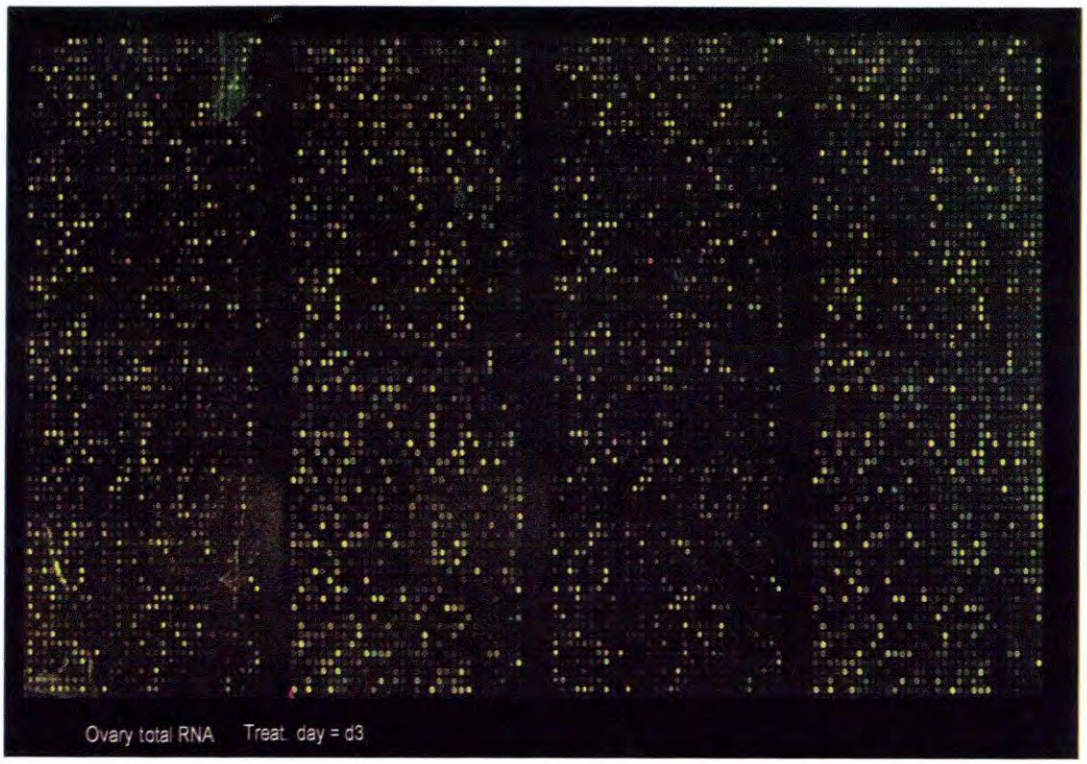


Fig. 4. Representative image of hybridization of control (Cy3) and index (Cy5) line mRNA from a whole ovary to the ovarian follicle microarray. The array is organized into four quadrants. The two left quadrants contain 4608 spots with clones representing 3479 unique gene clusters, several hundred random duplicates, the 107 follicle-expressed sequence tags (ESTs) isolated in the differential display PCR study, several other candidate genes known to be expressed in follicles and a panel of control yeast transcripts. The two right quadrants contain the same clones, providing replication of results within each array. For this hybridization, green and red indicate over-expression in the control and index lines, respectively, whereas yellow indicates relatively equal expression.

candidate genes have yet to be discovered) and of how genes function and interact. However, it can also be hypothesized that QTL primarily represent regulatory elements or initiation factors in a cascade of events, culminating in expression of physiological genes. Fuller understanding of this 'polygenic paradox' will require broad evaluation at the DNA, mRNA, protein and detailed phenotypic levels. The merging and integrating of quantitative genetics, functional genomics, proteomics, metabolomics and phenomics will establish linkages between predisposition and physiology, enhancing our understanding of how economically important traits in domestic animals are controlled and regulated, and facilitating improvement via genetics, intervention and management.

In particular, functional genomics will be a very powerful tool for studying the quantitative polygenic control of reproduction, as well as for understanding the underlying biology and physiology. Gene expression profiles between selection lines (or for example between divergent breeds) will allow for the dissection of selection response (or genetic variation) into two major categories: (i) loci that have been selected for (by definition, the QTL); and (ii) genes the expression of which (quantity or quality of mRNA) has changed as the result of direct or

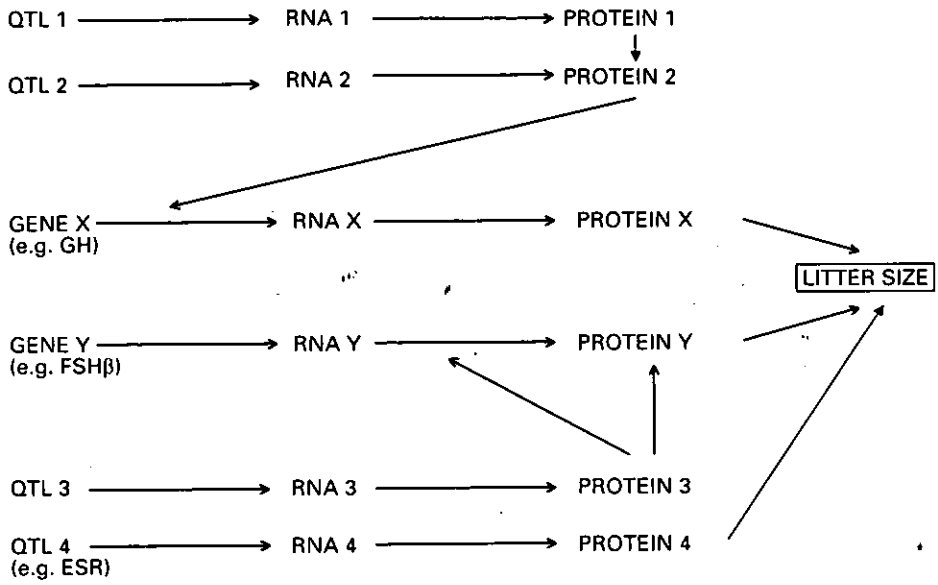


Fig. 5. Schematic diagram representing a simplified hypothesis of the genetic architecture of a complex polygenic trait, such as litter size or ovulation rate in pigs. Quantitative trait loci (QTL) represent predisposition genes, or those within which heritable genetic variations (DNA polymorphisms) explain phenotypic variation in traits. Alternatively, genes that produce proteins that are key regulators of reproduction, but within which may be no important genetic variation, represent 'physiological genes'. In this hypothesis, some QTL may directly influence a reproductive phenotype (for example, oestrogen receptor (ESR)), but more often (for example QTL 1, 2, 3) will exert effects by interacting with and regulating expression of 'physiological genes' (for example, FSHβ, growth hormone (GH)) or other QTL.

indirect interaction with QTL (Fig. 5). It is within this second category that we are most likely to identify the key components underlying the biology of reproduction.

These results support the traditional hypothesis that quantitative traits are controlled by a large number of genes, each with a small or moderate effect. In other words, there are many QTL that operate in tandem and with potentially complex interactions to control reproduction genetically. Although the heritable genetic variation must, by definition, reside within QTL, it is likely that these genes are regulatory and initiate critical changes in the transcription or translation of other genes, within which heritable sequence variation does not occur (Fig. 5). For example, FSH is clearly a critical rate-limiting protein in the determination of ovulation rate in pigs, and has been implicated definitively as an important correlated response to selection for increased ovulation rate and litter size. However, there is no evidence that the FSHβ locus is a QTL for reproduction. Thus, reproductive QTL must stimulate changes in FSHβ mRNA profiles through direct interaction, or through a cascade of regulatory events, which may also be manifested at the mRNA or protein levels. Further support for this contention is provided by recent work searching for QTL for FSH concentrations in Meishan–White composite boars (G. Rohrer, personal communication). Although several chromosomal regions were found to harbour genes with effects on FSH concentrations, the FSHβ locus itself was not implicated as a QTL.

Conclusion

One important issue in relation to the use of functional genomics as a tool to study reproduction is that, by nature, the use of microarrays for discovery is more 'question-driven' than 'hypothesis-driven'. At least initially, large-scale and high throughput determination of differentially expressed genes will potentially uncover vast databases of putative genes involved in reproductive processes. Many of these may not have been considered to have important roles in reproduction. In addition, new pathways and links between pathways may be uncovered. Indeed, it is this potential for new discovery that makes functional genomics such an appealing paradigm.

However, it is still critical that the design of functional genomics experiments is based on relevant scientific hypotheses, which will be required for determining which samples to study, and the temporal and spatial coordinates to be used. Furthermore, perhaps the most important outcome of functional genomics is that it will be 'hypothesis-generating', revealing significant new opportunities for experimentation as the plethora of collected expression data becomes organized and better understood.

Finally, this discussion would be remiss if it were not emphasized that evaluation of the transcriptome is but one important component of understanding how the genome influences reproduction. Amounts of mRNA are themselves phenotypes, under the potentially strong influence of environmental factors and interactions with other genetic components. Importantly, they may not be directly correlated with concentrations or activity of their respective translated proteins, or with the economically relevant end-point phenotypes such as litter size or age at puberty. Although functional genomics may uncover a large amount of information, significant efforts will be required to confirm and corroborate the influences of changes in gene expression within the broader, complex genetic and physiological models that are currently used.

References

Aguan K, Carvajal JA, Thompson LP and Weiner CP (2000) Application of a functional genomics approach to identify differentially expressed genes in human myometrium during pregnancy and labour *Human Reproduction* **6** 1141–1145

Bassett DE, Jr, Eisen MB and Boguski MS (1999) Gene expression informatics – it's all in your mine *Nature Genetics* **21** (Supplement 1) 51–55

Bertani GR, Gladney C, Johnson RK and Pomp D (2000a) Differentially expressed genes in anterior pituitary of pigs selected for reproduction *Plant and Animal Genome VIII, San Diego* 201 (Abstract)

Bertani GR, Gladney C, Johnson RK and Pomp D (2000b) Pig anterior pituitary EST's isolated by differential display in gene expression study of two lines selected for fertility *27th International Conference on Animal Genetics, Minneapolis, July* 73 (Abstract)

Bonaldo MF, Lennon G and Soares MB (1996) Normalization and subtraction; two approaches to facilitate gene discovery *Genome Research* **6** 791–806

Brown PO and Botstein D (1999) Exploring the new world of the genome with DNA microarrays *Nature Genetics* **21** (Supplement 1) 33–37

Caetano AR, Johnson RK and Pomp D (2001) Characterization of a normalized cDNA library of pig ovarian follicles *Plant and Animal Genome IX, San Diego* 77 (Abstract)

Cassady JP, Johnson RK, Pomp D, Rohrer GA, Van Vleck LD, Spiegel EK and Gilson KM (2001) Identification of quantitative trait loci affecting reproduction in pigs *Journal of Animal Science* **79** 623–633

Chang HS, Cheng WT, Wu KH and Choo KB (2000) Identification of genes expressed in the epithelium of porcine oviduct containing early embryos at various stages of development *Molecular Reproduction and Development* **56** 331–335

Clouscard-Martinato C, Mulsant P, Robic A, Bonnet A, Gasser F and Hatey F (1998) Characterization of FSH-regulated genes isolated by mRNA differential display from pig ovarian granulosa cells *Animal Genetics* **29** 98–106

Clutter AC and Brascamp EW (1998) Genetics of performance traits. In *The Genetics of the Pig* pp 427–462 Eds MF Rothschild and A Ruvinsky. CAB International, Wallingford

DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, Chen Y, Su YA and Trent JM (1996) Use of a cDNA microarray to analyse gene expression patterns in human cancer *Nature Genetics* **14** 457–460

Ermolaeva O, Rastogi M, Pruitt KD *et al.* (1998) Data

management and analysis for gene expression arrays *Nature Genetics* **20** 19–23

Gladney C, Bertani GR, Johnson RK and Pomp D (2000a) Evaluation of gene expression in ovarian follicles of pigs selected for reproduction using microarray and differential display PCR technologies *Plant and Animal Genome VIII, San Diego* **201** (Abstract)

Gladney C, Bertani GR, Johnson RK and Pomp D (2000b) Differential display PCR and microarray evaluation of ovarian follicle gene expression in pigs selected for reproduction *27th International Conference on Animal Genetics, Minneapolis, July* **73** (Abstract)

Green ML, Blaeser LL, Simmen FA and Simmen RC (1996) Molecular cloning of spermidine/spermine N1-acetyltransferase from the periimplation porcine uterus by messenger ribonucleic acid differential display; temporal and conceptus-modulated gene expression *Endocrinology* **137** 5447–5455

Iyer VR, Eisen MB, Ross DT *et al.* (1999) The transcriptional program in the response of human fibroblasts to serum *Science* **283** 83–87

Jiang M, Ryu J, Kiraly M, Duke K, Reinke V and Kim SK (2001) Genome-wide analysis of development and sex-regulated gene expression profiles in *Caenorhabditis elegans. Proceedings National Academy of Sciences USA* **98** 218–223

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 *Journal of Animal Science* **77** 541–557

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 *Journal of Animal Science* **69** 3129–3143

Lander ES (1999) Array of hope *Nature Genetics* **21** (Supplement) 3–4

Li MD, Matteri RL, Macdonald GJ, Wise TH and Ford JJ (1996) Overexpression of beta-subunit of thyroid-stimulating hormone in Meishan swine identified by differential display *Journal of Animal Science* **74** 2104–2111

Li MD, DePaola LV and Ford JJ (1997) Expression of follistatin and inhibin/activin subunit genes in porcine follicles *Biology of Reproduction* **57** 112–118

Liang P and Pardee AB (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction *Science* **257** 967–971

Linville RC, Pomp D, Johnson RK and Rothschild MF (2001) Candidate gene analysis for loci affecting litter size and ovulation rate in swine *Journal of Animal Science* **79** 60–67

Liu H, He Z and Rosenwaks Z (2001) Application of complementary DNA microarray (DNA chip) technology in the study of gene expression profiles during folliculogenesis *Fertility and Sterility* **75** 947–955

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 *Journal of Animal Science* **77** 1385–1391

Rothschild MF (2000) Advances in pig molecular genetics, gene mapping and genomics. *X Reunion Nacional de Mejora Genetica Animal, Caldes de Montbui, June 8–9* (http://www.etsia.upv.es/acteon/docs/final.pdf)

Rothschild MF, Jacobson C, Vaske D *et al.* (1996) The estrogen receptor locus is associated with a major gene influencing litter size in pigs *Proceedings National Academy of Sciences USA* **93** 201–205

Ruíz-Flores A and Johnson RK (2001) Direct and correlated responses to two-stage selection for ovulation rate and number of fully formed pigs at birth in swine *Journal of Animal Science* **79** 2286–2297

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

Schena M, Shalon D, Heller R, Chai A, Brown PO and Davis RW (1996) Parallel human genome analysis; microarray-based expression monitoring of 1000 genes *Proceedings National Academy of Sciences USA* **93** 10 614–10 619

Soares MB, Bonaldo MF, Jelene P, Su L, Lawton L and Efstratiadis A (1994) Construction and characterization of a normalized cDNA library *Proceedings National Academy of Sciences USA* **91** 9228–9232

Tanaka TS, Jaradat SA, Lim MK *et al.* (2000) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15 000 mouse development cDNA microarray *Proceedings National Academy of Sciences USA* **97** 9127–9132

Tess MW, Bennett GL and Dickerson GE (1983a) Simulation of genetic changes in life cycle efficiency of pork production I. A bioeconomic model *Journal of Animal Science* **56** 336–353

Tess MW, Bennett GL and Dickerson GE (1983b) Simulation of genetic changes in life cycle efficiency of pork production II. Effects of components on efficiency *Journal of Animal Science* **56** 354–368

Tosser-Klopp G, Benne F, Bonnet A, Mulsant P, Gasser F and Hatey F (1997) A first catalog of genes involved in pig ovarian follicular differentiation *Mammalian Genome* **8** 250–254

Tosser-Klopp G, Bonnet A, Yerle M and Hatey F (2001) Functional study and regional mapping of 44 hormono-regulated genes isolated from a porcine granulosa cell library *Genetics Selection Evolution* **33** 69–87

Tuggle CK, Green JA, Fitzsimmons C *et al.* (2001) Production of 17 cDNA libraries and successful EST sequencing of 10 124 clones from porcine female reproductive tissues *Journal of Animal Science* **79** 338 (Abstract)

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

Wilson ME, Sonstegard TS, Smith TP, Fahrenkrug SC and Ford SP (2000) Differential gene expression during elongation in the preimplantation pig embryo *Genesis* **26** 9–14

Yelich JV, Pomp D and Geisert RD (1997a) Detection of

transcripts for retinoic acid receptors, retinol-binding protein, and transforming growth factors during rapid trophoblastic elongation in the porcine conceptus *Biology of Reproduction* 57 286–294

Yelich JV, Pomp D and Geisert RD (1997b) Ontogeny of elongation and gene expression in the early developing porcine conceptus *Biology of Reproduction* 57 1256–1265

Yen HW (1999) *Follicular Development, Maturation and Atresia during the Estrous Cycle in Gilts Expressing High and Low Ovulation Rates* PhD Thesis, University of Nebraska-Lincoln