

Porcine gene discovery by normalized cDNA-library sequencing and EST cluster assembly

Scott C. Fahrenkrug,^{1*} Timothy P.L. Smith,¹ Brad A. Freking,¹ Jennifer Cho,² Joseph White,² Jeff Vallet,¹ Tommy Wise,¹ Gary Rohrer,¹ Geo Perteau,² Razvan Sultana,² John Quackenbush,² John W. Keele¹

¹USDA, ARS, U.S. Meat Animal Research Center, P.O. Box 166, Clay Center, Nebraska 68933, USA

²The Institute of Genomic Research (TIGR), Rockville, MD, USA

Received: 4 April 2001 / Accepted: 5 April 2002

Abstract. Genetic and environmental factors affect the efficiency of pork production by influencing gene expression during porcine reproduction, tissue development, and growth. The identification and functional analysis of gene products important to these processes would be greatly enhanced by the development of a database of expressed porcine gene sequence. Two normalized porcine cDNA libraries (MARC IPIG and MARC 2PIG), derived respectively from embryonic and reproductive tissues, were constructed, sequenced, and analyzed. A total of 66,245 clones from these two libraries were 5'-end sequenced and deposited in GenBank. Cluster analysis revealed that within-library redundancy is low, and comparison of all porcine ESTs with the human database suggests that the sequences from these two libraries represent portions of a significant number of independent pig genes. A Porcine Gene Index (PGI), comprising 15,616 tentative consensus sequences and 31,466 singletons, includes all sequences in public repositories and has been developed to facilitate further comparative map development and characterization of porcine genes (<http://www.tigr.org/tdb/ssgi/>). The clones and sequences from these libraries provide a catalog of expressed porcine genes and a resource for development of high-density hybridization arrays for transcriptional profiling of porcine tissues. In addition, comparison of porcine ESTs with sequences from other species serves as a valuable resource for comparative map development. Both arrayed cDNA libraries are available for unrestricted public use.

Introduction

Livestock genomics has made significant progress in recent years with the development of comprehensive genetic linkage maps (Kappes et al. 1997; Rohrer et al. 1996). These maps have been instrumental in genome scans to detect quantitative trait loci (QTL) for production characteristics in several livestock species (Andersson-Eklund et al. 1998; de Koning et al. 1999; Kirkpatrick et al. 2000; Knott et al. 1998; Rohrer and Keele 1998a, 1998b; Stone et al. 1999; Zhang et al. 1998). However, identification of genetic determinants influencing production trait variation has been constrained by a lack of livestock sequence data. Information about the sequence and location of expressed sequences in the porcine genome will enhance the selection and evaluation of candidate genes.

Two porcine cDNA libraries were developed and analyzed in the current study. Library MARC IPIG was produced from pig embryos at several developmental stages selected to correspond with early muscle development and organogenesis (Butler and Juurlink 1987), which are associated with significant embryo mortality in U.S. pig breeds (Ford 1997; Vallet 2000). Library MARC 2PIG was produced with pooled RNA from tissues with important roles in porcine reproductive physiology. Over 30,000 clones from each library have been sequenced, representing a significant contribution to the number of porcine expressed sequences in public databases. Cluster analysis of all porcine sequences suggests the discovery of up to 34,000 expressed pig genes that have been assembled and annotated in a Porcine Gene-Index (PGI).

Material and methods

Tissue collection. To produce the tissues for the MARC IPIG embryonic library, Landrace-Yorkshire boars were mated to gilts having at least one estrous cycle of normal length (17–23 days). Gilts were slaughtered at 11, 13, 15, 20, and 30 days, and the uterus was recovered. On days 11–15, each uterine horn was flushed with 20 ml 0.9% saline, and the complete conceptus for each embryo (including trophoctoderm) was recovered from the flushings and frozen in liquid nitrogen. On days 20 and 30, the uterus was opened along the antimesometrial border, and embryos were dissected from the placentas and snap frozen in liquid nitrogen. To produce MARC 2PIG, tissues were collected from multiple Landrace-Yorkshire female animals at various stages of sexual development. Pituitary and hypothalamus were collected from d10 and d15 of the midluteal phase of adult animals. Ovaries were collected from the following physiological stages: prepubertal; d10 and d15 midluteal; undergoing estrous; and anestrus animals. Animals were defined as anestrus if they did not display estrous within 14 days postweaning. In addition, endometrium and placenta (days 20 and 30 of pregnancy) were collected from the animals that provided embryos for MARC IPIG. Testes (110 days of gestation and adult) were collected from Meishan × Chester White-Landrace-Yorkshire boars. For both libraries, matched tissues from several individuals were pooled before mRNA isolation to provide a potential resource for polymorphism discovery.

RNA extraction and pooling. Tissues were shipped to Life Technologies, Inc. (Rockville, Md.) for contracted primary library production essentially as described (Smith et al. 2001). To construct the pooled-tissue libraries, poly(A) + RNA was purified from each source tissue and pooled by weight to produce the desired proportional representation. MARC IPIG was produced with RNA from embryos at 11, 13, 15, 20, and 30 days post coitum (dpc), with 20% of the total weight of RNA from each stage. MARC 2PIG was 25% ovary (equal parts prepubertal, d10, d15 cycle, estrous, and anestrus), 21% hypothalamus (equal parts d10, d15 cycle, estrous, and anestrus), 21% pituitary (equal parts d10, d15 cycle, estrous, and anestrus), 11% placenta (equal parts 20 & 30 dpc), 11% endometrium (equal parts days 15 and 30 dpc), 11% testis (equal parts day 110 gestation and adult).

*Present address: University of Minnesota, Department of Animal Science, 1988 Fitch Ave., St. Paul, MN 55108-60143, USA.

Correspondence to: T.P.L. Smith; E-mail: Smith@email.marc.usda.gov

Library construction and normalization. Library construction and normalization was performed as a service by a commercial provider (Life Technologies, Inc.), who provided the pertinent data described in this section. The vector, adapter primer sequences, and general procedure used were described previously (Smith et al. 2001). The pooled RNA was reverse transcribed by using a tailed poly(T) primer, and the resulting cDNA was used to construct a primary library in the pCMV-SPORT6 vector (inserted between the *NotI* and *Sall* sites of the vector). The number of transformants in primary libraries was determined by dilution titration of bacterial cells onto ampicillin plates.

Average insert size of the libraries was determined by PCR amplification of inserts from 22 clones chosen at random from each library, by using vector-specific primers, and by size estimation on agarose/ethidium bromide gels. Contamination of the primary libraries with ribosomal RNA and mitochondrial sequences was determined by colony hybridization with probe directed against 28S rRNA or cytochrome B, to a plated aliquot of each library. Primary cDNA libraries were expanded for 45 h at 30°C by semi-solid amplification as described (Kriegler 1990). The libraries were normalized by self-subtraction to a C_0t 500 (Bonaldo et al. 1996; Li et al. 1994; Swaroop et al. 1991; Smith et al. 2001). The average insert size after library normalization was determined by PCR analysis of 500 clones chosen at random from each library as described above. The normalization of the libraries was assessed by colony-lift hybridization (Sambrook et al. 1989) with the selected cDNA clones as probes.

EST sequencing. Individual colonies were robotically picked and arrayed into 384-well plates by BACPAC Resources (Oakland, Calif.). Single-pass sequencing was conducted with PCR amplification of inserts for template preparation as described (Smith et al. 2000). Sequences were processed to remove low-quality sequence (with cutoff Phred score of 20) and vector (using cross-match with the -minscore 18 and 23 -minmatch 12 options), and the trimmed sequences were submitted to GenBank.

Cluster analysis and porcine gene index assembly. The Pig Gene Index database was assembled at The Institute for Genomic Research (TIGR) as described (Quackenbush et al. 2001). Porcine EST sequences were downloaded from dbEST (NCBI, 12/28/01) and subjected to a second round of quality control screening to more stringently remove the remaining vector, poly(A)/poly(T) tails, adapter, and bacterial sequences. An additional set of expressed transcripts or pig genes (NP) were also obtained by extraction and curation of sequences through Entrez from GenBank. All sequences were concatenated and compared pair-wise by FLAST (similar to BLAST), a rapid sequence comparison program based on the DOS algorithm (Huang et al. 1997). Sequences with >95% identity over at least 40 bp and with unmatched overhangs less than 20 bp were placed into clusters, leaving unclustered sequences as singletons. For each cluster, its component sequences were assembled by using CAP3 to produce aligned sequence assemblies (Huang and Madan 1999). These assemblies were assigned Tentative Consensus (TC) numbers, which together comprise the Pig Gene Index. TCs that included annotated sequence from the GenBank non-redundant nucleotide database were assigned the annotation of that gene. TCs and ESTs without assigned annotation were compared with a non-redundant protein database by using the algorithm DPS (Huang and Madan 1999); high-scoring hits (score \geq 350) were assigned a putative orthologous identity. The completed Porcine Gene Index was released through the TIGR Gene Index web site at <http://www.tigr.org/tdb/ssgi/>.

Results

MARC libraries were made from mRNAs from multiple tissues to increase the number of unique genes represented (Smith et al. 2001), normalized by subtractive hybridization to increase the efficiency of sequencing unique genes (Bonaldo et al. 1996; Li et al. 1994; Swaroop et al. 1991), and sequenced from the 5' end to increase the probability of collecting unique sequence in the case where two independent clones contain fragments of the same porcine gene. Library construction and normalization were contracted as a service from Life Tech-

nologies, Inc. The complexity of MARC IPIG was expected to be very high as a result of the breadth of tissues sampled by using whole embryos (representing many tissues) at multiple stages. Complexity of MARC 2PIG was increased by pooling mRNA from several tissue types, in this case tissues important to porcine reproduction and growth.

The first quality control analysis was performed at the primary library stage. This control indicated minimal contamination from ribosomal RNA or mitochondrial sources (0.03% and 0.12% 28S rRNA clones, 0.15% and 0.26% cytochrome B clones in MARC IPIG and MARC 2PIG, respectively). Average insert size of both primary libraries was 1.9 ± 0.97 kbp, with 96% of clones producing amplicons larger than predicted from vector lacking insert.

Initial sequencing of each pre-normalization library demonstrated that clones corresponding to EF1 α were present in relatively high abundance, and this gene was selected to monitor normalization efficiency. In addition, α 1-antitrypsin was found to be in high abundance in the MARC IPIG library. Subsequent hybridization analysis with these cDNAs as probes indicated that clones encoding EF1 α and α 1-antitrypsin represented 0.65% and 1.67% of the cDNAs, respectively, in the MARC IPIG primary library, confirming that these genes are among the most abundant expression class in the pooled mRNA from which the library was developed (Hastie and Bishop 1976; Soares et al. 1994). After normalization, clones representing expression from these genes were reduced in abundance by a factor of 38 and 20, respectively, as assayed by hybridization. In the MARC 2PIG library, porcine EF1 α represented 0.53% of the clones in the primary library but was reduced by a factor of 26.5 after normalization. The extent of clone subtraction, as well as the relationship of degree of subtraction to original clone representation, was consistent with effective subtractive normalization (Bonaldo et al. 1996; Soares et al. 1994). Average insert size of the clones in the normalized libraries showed significant reduction from 1.9 kbp to 1.1 kbp (SD = 0.58 kbp) for MARC IPIG, and 1.2 kbp (SD = 0.64 kbp) for MARC 2PIG, a phenomenon previously noted (Bonaldo et al. 1996).

On the basis of the apparent success of normalization, clones from each library were arrayed in 138 plates (384 clones per plate) and used as templates for collection of EST sequence. Clones were sequenced from the 5' end of the directionally cloned cDNAs to facilitate identification of porcine sequences orthologous to open reading frames represented in public sequence repositories, and to increase the information derived from partially redundant clones representing various portions of porcine genes. Culling of vector, low quality, or low complexity sequences resulted in selection and submission of 32,760 sequences from MARC IPIG and 33,569 sequences from MARC 2PIG to the GenBank dbEST database. Average length of high-quality sequence read for both libraries was 400 bp. Analysis of the sequences by BLASTN to the GenBank non-redundant (nr) database revealed that 37 and 15 clones from MARC IPIG and MARC 2PIG, respectively, represented porcine EF1 α , corresponding to an occurrence of 0.001% and 0.0004% and directly confirming the success of the normalization step.

The complete set of 97,948 expressed sequences present in GenBank (12/28/01) were used to build the PGI, of which 66,245 (68%) were from the MARC libraries. Application of an additional, higher-stringency screen for vector sequences and other artifacts (see materials and methods) removed 1045 sequences (1%), including 322 from MARC IPIG and 159 from MARC 2PIG. An additional 30,070 ESTs (including those from other libraries in GenBank) were affected by trimming of contaminating sequences at either end of the sequence read. In addition to EST sequences, the PGI includes

Library Clone Representation

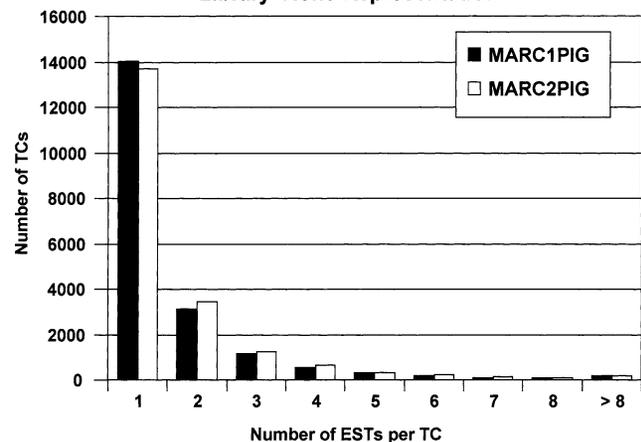


Fig. 1. Distribution of sequences within clusters of the PGI. The number of TCs (or TCs + singletons for the "1 EST per TC" interval) for the individual MARC libraries are plotted with respect to the number of sequences within each consensus sequence cluster.

2,171 pig transcript sequences present in the non-redundant portion of the GenBank database.

The assembly process produced 15,616 TC sequences, of which a subset of 13,356 contained at least 1 EST from a MARC library. The median was 2 porcine ESTs per TC, and the mean was 3.3 ESTs per TC (range, 2 to 59; standard deviation, 1.8). In addition, there were 31,924 singleton sequences for a total of 47,540 unique sequences (15,616 TCs + 31,924 singletons) in the TIGR PGI, of which 34,410 (13,356 TCs + 21,054 singletons) contained at least 1 EST from either MARC library. Overall, the proportion of unique sequences contributed by the MARC libraries is 72% (34,410/47,540).

The distribution of MARC 1PIG sequences in the PGI was 21,874 ESTs assembled into 9,160 TCs (average of 2.4 ESTs per TC), leaving 10,570 singletons (33%). Distribution in the MARC 2PIG library was 22,926 ESTs assembled into 9,510 TCs (average of 2.4 ESTs per TC), leaving 10,484 singletons (31%). Results of clustering in the PGI were used to evaluate success of the library construction approach and normalization, by examining the distribution of the number of clones from these two libraries present in each TC cluster (Fig. 1). The number of singletons added to the number of TCs having only one EST from a particular MARC library represents the largest proportion of the distribution (indicated by the bar in the interval of one EST per TC) and illustrates that the preponderance of sequence collected within each library was non-overlapping, unique sequence. The number of TCs containing higher numbers of sequences from either MARC library drops rapidly, demonstrating that the pooled tissue approach resulted in sampling a high percentage of sequence from unique genes. If library normalization had been less successful, deep sequencing would have resulted in an increase in the number of ESTs per library assembling into single TC. Furthermore, between-library redundancy was relatively low, as only 5314 TCs had ESTs from both libraries. Library MARC 1PIG had 3,846 TCs (28.8%) not found in library MARC 2PIG, and library MARC 2PIG had 4,196 TCs (31.4%) not found in library MARC 1PIG. Overall, approximately 60% of the singletons and TCs were specific to each library.

From the 13,356 TCs containing MARC ESTs, 506 (4%) have at least one or more NP sequences as part of the assembly. In addition, predicted translations for 5,564 TCs (42% of 13,356 TC) have significant matches in the TIGR NRAA database (NRAA is a complete set of amino acid sequences

Table 1. Significant matches (DPS score ≥ 350) of MARC ESTs with proteins from other species. The number of tentative consensus (TC) sequences containing MARC 1PIG or MARC 2PIG ESTs that match specific organisms are indicated. The number of MARC singletons matching proteins from other species is also presented. In all cases, only the highest ranking match was scored, i.e., the data are not redundantly presented.

Organism	TC with matches	Singletons with matches
<i>H. sapiens</i>	3,719	2,748
<i>S. scrofa</i>	418	25
<i>M. musculus</i>	809	574
<i>B. taurus</i>	233	86
<i>R. norvegicus</i>	148	141
<i>M. fascicularis</i>	49	61
<i>D. melanogaster</i>	10	18
<i>C. familiaris</i>	37	25
<i>O. aries</i>	17	9
<i>C. elegans</i>	6	3
<i>G. gallus</i>	11	7
<i>C. griseus</i>	7	7
<i>X. laevis</i>	5	7
Others	95	68
Total	5,564	3,779

from EGAD, GenPept, PIR, and SwissProt). Comparison of 21,054 singletons from MARC libraries to NRAA revealed 18% (3,779) match proteins in the TIGR NRAA database with a score exceeding our stringent threshold (DPS score ≥ 350 ; Huang and Madan 1999). This analysis detected a high degree of identity for approximately 24% of the unique MARC sequences to proteins from other species, a feature important for comparative mapping efforts. Table 1 delineates the matches to sequences from different organisms. The two most frequent genes represented in the MARC 1PIG library were N-acylaminoacyl-peptide hydrolase (APEH; 44 sequences) and EF1 α (37 sequences), and in the MARC 2PIG library they were collagen type III alpha 1 (COL3A1) and nucleolar mitotic apparatus protein 1 (NUMA1; 18 sequences for each). Sequences in PGI not yet annotated by virtue of amino acid identity may be predicted as orthologous to other sequences based upon similarity at the nucleotide level. However, to avoid the accumulation of erroneous sequence annotation, we will reserve such predictions until TC have grown or condensed to include enough open reading frame to detect significant identity at the amino acid level.

Discussion

The objective of these two pig cDNA libraries was to maximize collection of unique sequences and sample the widest possible diversity of genes. Since a single tissue is expected to express 10,000 to 20,000 genes (Davidson and Britten 1979; Hastie and Bishop 1976), complexity was increased by a mixed-tissue approach to library construction. The libraries were normalized to a C_0t 500 in order to equalize the representation of clones corresponding to genes in the abundant, middle-repetitive, and rare-abundant expression classes (Hastie and Bishop 1976), thereby enhancing collection of non-redundant expressed sequence data.

The normalized libraries described proved to be an excellent resource for collection of unique sequence expressed in porcine embryos and reproductive tissues. Even after sequencing of 30,000 clones from a library, the genes sampled most often were represented only 44 and 18 times for MARC 1PIG and MARC 2PIG, respectively. Further sequencing would be predicted to continue in the discovery of expressed porcine genes, although the rate of new sequence acquisition has decreased at the end of the relatively deep sequencing employed, such that approximately 38% of new sequences show no overlap with data present in the PGI.

The maximum number of genes identified can be estimated from the analyses conducted here by summation of the number of TCs and singletons, but the minimum gene count cannot be similarly derived. Co-linearity of sequences deriving from different portions of the same mRNA cannot be established without significant overlap, thereby inflating the estimate of gene number owing to variation in the origin or extent of reverse transcription during first-strand cDNA synthesis. In addition, alternative mRNA processing may result in a failure to ascertain that transcripts were derived from the same gene. Sequencing of approximately 60,000 ESTs from the two libraries resulted in the identification of a maximum of 34,410 unique pig genes by summation of TCs and singletons.

Other porcine EST sequencing efforts ongoing at USDA MARC and at other academic institutions (Seo and Beaver 2001; Tuggle et al. 2001; Yao et al. 2001) focus on collection of sequence from the 3'-end of cDNA libraries. These sequence data will identify additional porcine genes and assist in condensation of current TCs into clusters representing longer stretches of cDNA sequence. The evolving Porcine Gene Index will continue to provide a useful sequence resource for PCR primer design in physical and genetic mapping efforts. The arrayed libraries provide a resource for construction of EST-based microarrays and other functional genomics studies, and individual clones or complete library copies are publicly available through BACPAC resources (www.chori.org/bacpac).

Acknowledgments. We wish to acknowledge the expert secretarial assistance by Sherry Kluver; essential technical support by Steve Simcox, Renee Godtel, and Robert Lee; assistance in the collection of porcine tissues by Terri Alberts and Dr. Joe Ford; and bioinformatic support from Yudan Lee, Svetlana Karamycheva, Babak Parvizi, and Jennifer Tsai. We thank Christian Gruber, Lois Mohelski, and Paul Nisson of Life Technologies, Inc. for their participation in the construction of these libraries. Helpful comments regarding this manuscript were provided by Dr. Christopher Tuggle and Dr. Jim Fox.

References

- Andersson-Eklund L, Marklund L, Lundstrom K, Haley CS, Andersson K et al. (1998) Mapping quantitative trait loci for carcass and meat quality traits in a wild boar \times Large White intercross. *J Anim Sci* 76, 694–700
- Bonaldo MF, Lennon G, Soares MB (1996) Normalization and subtraction, two approaches to facilitate gene discovery. *Genome Res* 6, 791–806
- Butler H, Juurlink BH (1987) In: *An Atlas for Staging Mammalian and Chick Embryos*, 1st edn. CRC Press, Boca Raton, FL. 155–168
- Davidson EH, Britten RJ (1979) Regulation of gene expression, possible role of repetitive sequences. *Science* 204, 1052–1059
- de Koning DJ, Janss LL, Rattink AP, van Oers PA, de Vries BJ et al. (1999) Detection of quantitative trait loci for backfat thickness and intramuscular fat content in pigs (*Sus scrofa*). *Genetics* 152, 1679–1690
- Ford SP (1997) Embryonic and fetal development in different genotypes in pigs. *J Reprod Fertil* 52, 165–176
- Hastie ND, Bishop JO (1976) The expression of three abundance classes of messenger RNA in mouse tissues. *Cell* 9, 761–774
- Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Res* 9, 868–877
- Huang X, Adams MD, Zhou H, Kerlavage AR (1997) A tool for analyzing and annotating genomic sequences. *Genomics* 46, 37–45
- Kappes SM, Keele JW, Stone RT, McGraw RA, Sonstegard TS et al. (1997) A second-generation linkage map of the bovine genome. *Genome Res* 7, 235–249
- Kirkpatrick BW, Byla BM, Gregory KE (2000) Mapping quantitative trait loci for bovine ovulation rate. *Mamm Genome* 11, 136–139
- Knott SA, Marklund L, Haley CS, Andersson K, Davies W et al. (1998) Multiple marker mapping of quantitative trait loci in a cross between outbred wild boar and Large White pigs. *Genetics* 149, 1069–1080
- Kriegler M (1990) Expression cloning. In: *Gene Transfer and Expression*, A Laboratory Manual. (New York: Stockton Press), pp 114–135
- Li WB, Gruber CE, Lin JJ, Lim R, D'Alessio JM et al. (1994) The isolation of differentially expressed genes in fibroblast growth factor stimulated BC3H1 cells by subtractive hybridization. *BioTechniques* 16, 722–729
- 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
- Rohrer GA, Keele JW (1998a) Identification of quantitative trait loci affecting carcass composition in swine. I: Fat deposition traits. *J Anim Sci* 76, 2247–2254
- Rohrer GA, Keele JW (1998b) Identification of quantitative trait loci affecting carcass composition in swine. II: Muscling and wholesale product yield traits. *J Anim Sci* 76, 2255–2262
- Rohrer GA, Alexander LJ, Hu Z, Smith TP, Keele JW et al. (1996) A comprehensive map of the porcine genome. *Genome Res* 6, 371–391
- Schultz RD, Wang JT, Dunne HW (1971) Development of the humoral immune response of the pig. *Am J Vet Res* 32, 1331–1336
- Seo K, Beaver JE (2001) Monitoring gene expression in swine skeletal muscle. *Proc. Plant and Animal Genome IX*, San Diego, Calif., Abstr. W207, p. 59
- Smith TPL, Godtel RA, Lee RT (2000) PCR-based setup for high-throughput cDNA library sequencing on the ABI 3700 automated DNA sequencer. *BioTechniques* 29, 698–700
- Smith TPL, Grosse WM, Freking BA, Roberts AJ, Stone RT et al. (2001) Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle. *Genome Res* 11, 626–630
- Soares MB, Bonaldo MF, Jelene P, Su L, Lawton L et al. (1994) Construction and characterization of a normalized cDNA library. *Proc Natl Acad Sci USA* 91, 9228–9232
- Stone RT, Keele JW, Shackelford SD, Kappes SM, Koochmaria M (1999) A primary screen of the bovine genome for quantitative trait loci affecting carcass and growth traits. *J Anim Sci* 77, 1379–1384
- Swaroop A, Xu JZ, Agarwal N, Weissman SM (1991) A simple and efficient cDNA library subtraction procedure, isolation of human retina-specific cDNA clones. *Nucleic Acids Res* 19, 1954
- Tuggle CK, Green J, Fitzsimmons CJ, Woods R, Prather P et al. (2001) Development of resources for functional genomics in the pig, production of 14 cDNA libraries and sequencing of over 7,000 clones from female reproductive tissues. *Proc Plant and Animal Genome IX*, San Diego, CA, Abstr. P67, p. 77
- Vallet JL (2000) Fetal erythropoiesis and other factors which influence uterine capacity in swine. *J Appl Anim Res* 17, 1–26
- Yao J, Coussens P, Burton J, Ernst C (2001) Development of high quality cDNA libraries and landmarks for cDNA microarrays to study complex disease and production traits in cattle and swine. *Proc Plant and Animal Genome IX*, San Diego, Calif., Abstr. P69, p. 77
- Zhang Q, Boichard D, Hoeschele I, Ernst C, Eggen A et al. (1998) Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree. *Genetics* 149, 1959–1973