Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Genomics 86 (2005) 618 – 625
GENOMICS
www.elsevier.com/locate/ygeno

Method

Validation of a first-generation long-oligonucleotide microarray for transcriptional profiling in the pig

Shu-Hong Zhao [a,1], Justin Recknor [b], Joan K. Lunney [c], Dan Nettleton [b], Daniel Kuhar [c], Sarah Orley [d], Christopher K. Tuggle [a,*]

[a] Department of Animal Science, 2255 Kildee Hall, Iowa State University, Ames, IA 50011, USA
[b] Department of Statistics, Iowa State University, Ames, IA 50011, USA
[c] Animal Parasitic Diseases Laboratory, ANRI, ARS, U.S. Department of Agriculture, Beltsville, MD 20705, USA
[d] Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011, USA

Received 14 April 2005; accepted 1 August 2005

Abstract

A first-generation porcine oligonucleotide set, representing 13,297 cDNAs and ESTs, has been designed by Qiagen–Operon for transcriptional profiling. To validate this set, microarrays containing each 70-mer oligonucleotide, referred to as the Qiagen–NRSP8 array, were hybridized with targets from porcine adult liver, lung, muscle, or small intestine. Transcriptome analyses showed that 11,328 of the oligonucleotides demonstrated expression in at least one tissue. Statistical analyses revealed that 1810 genes showed differential expression among tissues (Bonferroni adjusted $p < 0.05$). Biological pathways identified by DAVID/EASE analysis using a list of 423 tissue-selective genes matched archetypal pathways in the corresponding human or mouse tissue. Real-time quantitative PCR confirmed expression patterns for 9 of 11 genes tested. Our results demonstrate that this first-generation porcine oligonucleotide array is informative and the specificity is high. This is essential validation for investigators using the Qiagen–NRSP8 array for porcine functional genomics and for using the pig in modeling important physiological problems.
Published by Elsevier Inc.

Keywords: Pig; Long oligonucleotide microarray; Transcriptional profiling; Tissue

Since the first description of high-density DNA microarrays in 1995 [1], DNA microarrays have been widely used in genomics research for those organisms with sufficient genomic resources. The long oligonucleotide microarray, which is composed of gene-specific oligonucleotides of 40–70 nt in length spotted on glass slides, has become a powerful tool for globally detecting differential gene expression. However, the specificity of the oligonucleotides on the array is crucial. A useful oligonucleotide should have low cross-hybridization to other transcripts and should hybridize efficiently so that the set is able to detect expression levels of many genes in the target. There are a large number of genes that have tissue-specific or tissue-selective expression [2]; thus using targets independently synthesized from different tissues to hybridize with candidate oligonucleotide sets is a good way to test the quality of the oligonucleotide array platform.

Different breeds or types of pigs have been widely used as biomedical research models for many years. The pig's organ sizes and its anatomy and physiology make it an ideal comparative human model for normal physiology as well as disease research [3,4]. For example, functional studies of the lung showed the pig is a good model for human asthma disease [5,6]. Porcine liver function is metabolically similar to that of humans and can be used for liver disease and transplant studies [3]. Further, pigs are omnivores and the physiology of porcine digestion is similar to that of humans, making pigs a good model for studying the human digestive system and obesity [4,7]. As well, significant efforts are under way to

* Corresponding author. Fax: (515) 294 2401.
E-mail address: cktuggle@iastate.edu (C.K. Tuggle).
[1] Current address: Laboratory of Molecular Biology and Animal Breeding, College of Animal Science and Technology, Huazhong Agricultural University, Wuhan 430070, China.

0888-7543/$ - see front matter. Published by Elsevier Inc.
doi:10.1016/j.ygeno.2005.08.001

modify pigs genetically so that pig tissues such as kidney, islets, liver, intestine, or lung may be used for human organ transplantation [8–10]. Understanding the expression and regulation of genes in physiologically relevant pathways extant in pig organs is very important in the above research areas. However, up to now, there have been only a few reports on transcriptional profiling in pig biology. We have developed arrays to study muscle biology and embryo development [11,12], while others have used human arrays with porcine targets [13,14] or tissue-specific custom porcine cDNA arrays [14–18]. None of these studies involved the use of long oligonucleotide arrays.

To develop broadly useful tools in porcine functional genomics, Qiagen–Operon and the USDA-NRSP-8 Swine Genome community collaborated to develop a novel 13,297-oligonucleotide set [19]. The Qiagen oligonucleotide set represents porcine cDNAs and ESTs, designed from The Institute of Genome Research (TIGR) Tentative Consensus cDNA sequences. Neither the utility nor the specificity of these elements has been experimentally validated. To develop information on the utility of the array, referred to as the Qiagen–NRSP8 array, hybridization tests of glass microarray slides containing these oligonucleotides were conducted by using target from four adult tissues (liver, lung, muscle, and small intestine) from six pigs (24 tissue samples).

Results

GO annotation of the Qiagen oligonucleotide set

The Qiagen oligonucleotide set was designed primarily from TIGR tentative consensus (TC) sequences (Porcine Gene Index, version 5.0, October 2002) with structural similarity to human sequences confirmed through BLAST analyses with the entire porcine TC as query. To assess the current state of orthology for this set, we used two methods. Initially, we developed Perl scripts to extract the TIGR GO term annotations for the TCs that contained oligonucleotides in the set, queried the GO database (http://www.godatabase.org) to determine the global categories to which these specific terms are assigned, and then sorted the terms to enumerate the number of oligonucleotides in each major GO annotation. However, this resulted in only 3973 annotations across the entire set; these annotations are available in the online Supplementary data. To attempt to improve the level of annotation, we used recently completed BLAST analyses available from Qiagen. Initially, we found there are 12,303 current TIGR TC sequences to which the Qiagen oligonucleotides now match; i.e., there are 994 oligonucleotides that match 2 (or more) TCs (7.5% redundant). By using a minimum criterion of 75% similarity over 100 bp to human or mouse sequence in BLAST analyses with these 12,303 TCs, we found 11,349 hits to human or mouse RefSeq or pig annotated gene NCBI accession numbers (e.g., there are 8356 unique human RefSeq top hits, 7350 unique mouse RefSeq top hits, and 1249 unique pig known gene top hits). For a limited number of oligonucleotides, their corresponding TC matched the same human or mouse RefSeq or pig known gene as other oligonucleotides. If we remove these redundancies there are 8541 unique human or mouse RefSeq or pig annotated gene NCBI accession numbers. We then used these RefSeq accession numbers to annotate the oligonucleotides by using the "Build Simplified Ontology" option in GeneSpring 6.1. Thus, 6244 oligonucleotides were assigned GO terms. A summary of the molecular function GO annotation for the AROS (Array Ready Oligo Set), as annotated by GeneSpring, is shown in Fig. 1. The sequence information of all oligos, corresponding human RefSeq IDs, and complete GO annotations are presented in Supplementary Table 1.

Utility of the array—number of genes expressed in tissues

Obtaining information on the number of genes present on the array that are expressed in different tissues can help to evaluate the utility of the array in studying biological questions based on different cell/organ types. To declare a gene expressed in one or more of the tissues tested (liver, lung, skeletal muscle, or small intestine), we used gene-specific mixed linear model analyses to identify genes whose signal intensities were significantly greater than threefold above the median signal of *Arabidopsis* gene oligonucleotides included on the array as negative controls. When controlling the false discovery rate separately for each tissue at 0.01, the number of oligonucleotides exhibiting signal significantly greater than threefold above the median of the *Arabidopsis* negative controls was 8682 in liver, 8358 in muscle, 10,328 in small intestine, and 10,556 in lung. Cumulatively, there were 11,328 oligonucleotides with signal indicating gene expression in at least one of the four tissues; among these, 7290 indicated expression in every tissue. Based on these results, we would expect the array to be useful for studying gene expression in many different types of tissues.

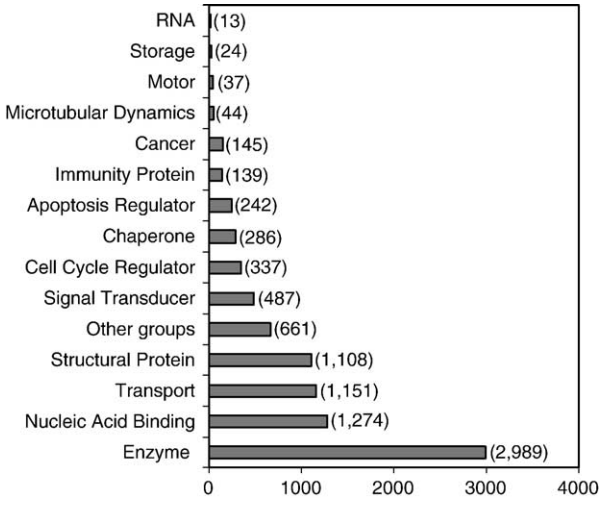

Fig. 1. Gene Ontology (GO) molecular function classification of the Qiagen 13K oligonucleotide set. Numbers in parentheses indicate the exact number of oligonucleotides assigned to each molecular function GO term across the AROS set, as annotated by GeneSpring analysis.

Specificity of the array—identification of genes with tissue-selective expression

In addition to identifying genes expressed in one or more tissues, we used mixed linear model analyses to identify genes that were differentially expressed across tissues. Determining the tissue-specific gene expression detected by these oligonucleotides can help to evaluate their specificity. All four tissues were collected from each of six adults pigs, of which three were 15 weeks of age and three were 17 weeks of age. Thus each gene-specific mixed linear model included fixed effects for tissue, age, and age-by-tissue interaction. The terms involving age were not significant. Thus all our subsequent analyses focused on expression differences among the tissues, averaging over varying animal age. Mixed model analysis of the data across the tissues revealed a large number of genes that were differentially expressed in the tissues surveyed. There were 1810 (13.6%) genes that showed differential expression between at least two tissues ($p < 0.05$) after Bonferroni correction (based on 13,122 tests) to account for multiple testing [20]. Without Bonferroni correction, there were 4406 (34%) genes that showed differential expression at $p < 0.001$ (estimated false discovery rate (FDR) = 0.00032) and 8200 (62%) genes at $p < 0.05$, estimated FDR = 0.082 (Supplementary Table 2). Genes expressed at a very high level in one tissue compared to all other tissues are called "tissue-selective" genes [21]. The numbers of genes with tissue-selective expression across all four tissues are listed in Table 1 for the three different p value criteria. These results depict the genes that demonstrate both high expression in one tissue and significantly lower expression in the other three tissues.

To determine if these differentially expressed genes have orthologs in other species that are already known to be tissue-selectively expressed, we compared our data to mouse or human expression results for a limited number of genes. The results are shown in Supplementary Table 3 for the 60 highest and 60 lowest expressed genes selected from a list of 857 genes (Table 1, $p < 0.0001$, FDR < 0.00004). Affymetrix gene expression data for 110 human or mouse orthologs of these 120 pig genes in our study were found in the SymAtlas database (http://symatlas.gnf.org/SymAtlas/) [22]. Of these 110 genes, 96 (87%) had a qualitatively similar expression pattern to either the mouse or the human (or both) orthologous genes (Supplementary Table 3).

Table 1
Number of tissue-selective genes under different p value criteria[a]

Tissue	$p \leq 0.0000038$[b] (FDR < 2.82×10^{-6})	$p \leq 0.0001$ (FDR < 0.00004)	$p \leq 0.001$ (FDR < 0.0003)
Liver	135	270	405
Lung	69	153	266
Muscle	190	356	538
Small intestine	29	78	147

[a] If the expression level of a gene in one tissue is higher than in any other tissues and the p value in pair-wise comparison is also significant according to the criteria, this gene is called a tissue-selective gene for that particular tissue.
[b] Equivalent to Bonferroni corrected $p < 0.05$.

Cluster analyses of the tissue expression data

As mentioned above (Table 1), genes exhibiting tissue-selective expression patterns were identified when the expression level of a gene in one tissue was significantly higher than in any other tissue (Bonferroni corrected pair-wise $p < 0.05$). We found that 423 genes fit these criteria; there are 29, 69, 135, and 190 genes selectively expressed in small intestine, lung, liver, and muscle, respectively (Table 1, listed in Supplementary Table 4). To test these groups further, the "Find classes" function in GeneCluster 2 was used to do a cluster analysis using the normalized expression data for these 423 genes. The results showed four distinct clusters of expression patterns, and the genes in each tissue cluster are exactly the same as the results based on p value, indicating that GeneCluster 2 analyses was correctly assigning genes to expression classes with likely biological relevance (data not shown). We then used the software to perform analyses of the 1810 genes differentially expressed between at least two tissues, to determine if specific patterns of expression could be identified among this larger and more complex group of differentially expressed genes. The results identified 16 clusters with distinct expression patterns across these tissues (Fig. 2; details available in Supplementary Table 5) and gene groups with tissue-selective expression patterns (i.e., clusters 1, 3, 5, 10, 14, 16) as well as more complex patterns of interest (i.e., clusters 2, 9, 12, 13).

Biological pathways identified by EASE

To gather biological information related to the tissue-selective gene lists we developed above, we used the RefSeq IDs from human orthologs of the 423 tissue-selective genes (Supplementary Table 4) to find available gene pathway annotations using the EASE software package [23]. The list of tissue-selective genes for each tissue was used as input. A number of biological pathways were identified, with significant EASE scores, for each tissue (data not shown). For example, muscle cell differentiation, muscular dystrophy, and muscle contraction pathways were identified from the muscle-selective gene list. Liver-selective genes were found in bile acid metabolism, coagulation, alcohol metabolism, and cholesterol biosynthesis pathways, and many additional pathways representative of liver metabolic functions. The human RefSeq IDs corresponding to all oligonucleotides on the array and our tissue-selective lists are available in Supplementary Tables 1 and 3, respectively, for readers who are interested in using this array and EASE software to analyze these data further.

Results of quantitative RNA analyses using real-time PCR (Q-PCR) of selected genes for tissue survey study

We selected 12 genes (11 test genes plus the housekeeping control gene RPL32) to use Q-PCR to confirm microarray expression patterns. These were selected for three purposes: (1) to validate the microarray results across tissues; (2) to

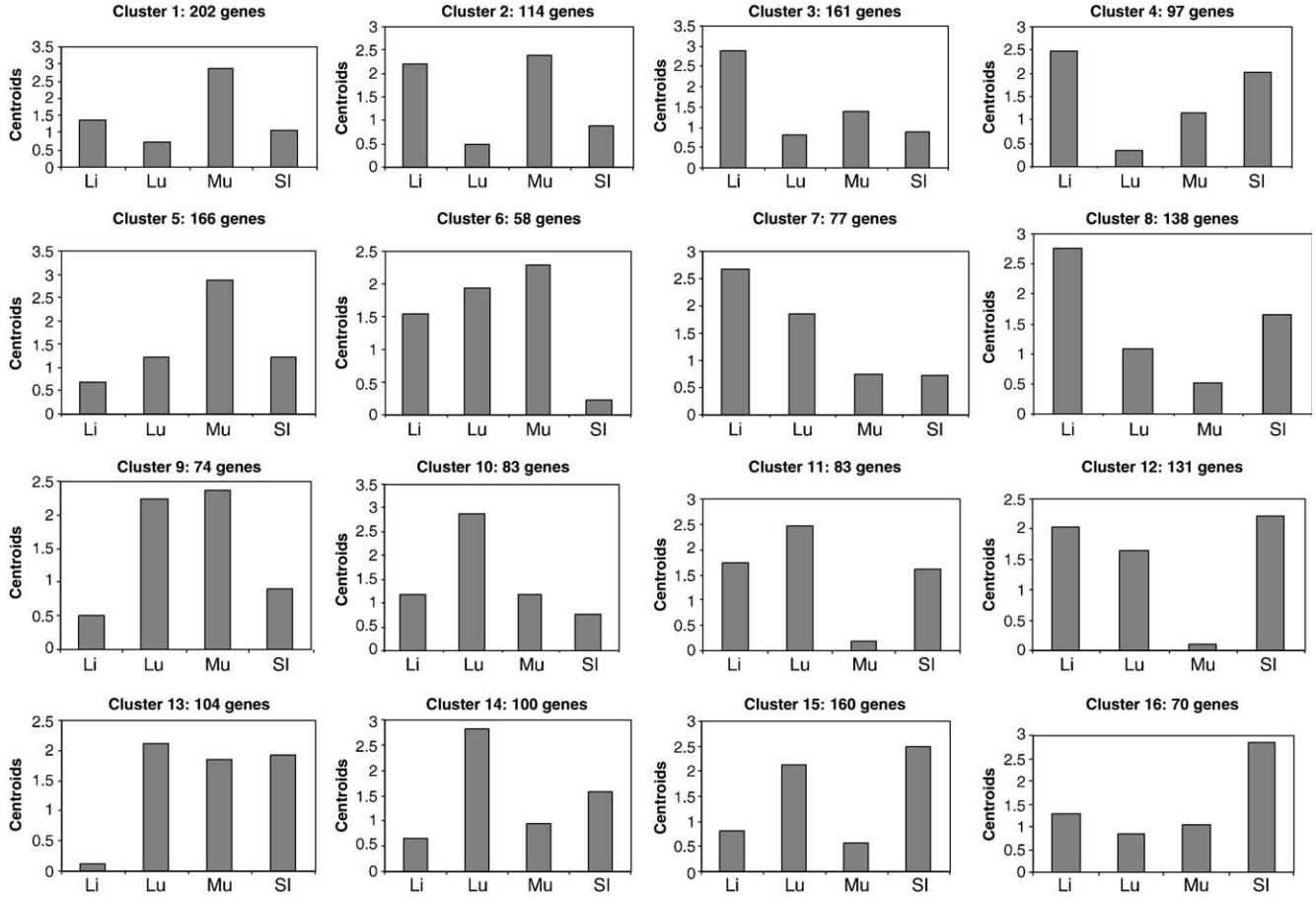

Fig. 2. Unsupervised cluster analysis of differentially expressed genes. Bar graphs were generated by using centroid values from GeneCluster 2 analysis for each tissue by cluster. Cluster analysis was based on 1810 differentially expressed genes after Bonferroni correction. Li, liver; Lu, lung; Mu, muscle; SI, small intestine.

validate the expression pattern, comparing that to available mouse experimental results; and (3) to validate the gene specificity of the oligonucleotides within gene families by comparing gene expression patterns for these family members in other species. The results showed that 9 of the 11 test genes had a statistically significant ($p < 0.05$) expression patterns that were in agreement with the microarray results (Table 2).

Table 2
Quantitative PCR results of 11 selected genes in four tissues

Oligonucleotide ID	Gene symbol	Tissue expression level (ΔC_t*)**				Microarray results***	Agree with microarray results?
		Liver	Lung	Muscle	Small intestine		
SS00002529	NOS2A	20.7 ± 0.6^{a}	15.8 ± 0.5^{c}	18.7 ± 0.4^{b}	19.8 ± 1.1^{ab}	Lu high	Yes
SS00010183	ICAM1	13.4 ± 1.9^{ab}	11.6 ± 1.0^{b}	12.7 ± 0.8^{ab}	14.9 ± 0.6^{a}	Lu high	Yes
SS00000872	CASP1	6.0 ± 0.2^{b}	3.7 ± 0.3^{c}	8.1 ± 0.3^{a}	3.5 ± 0.2^{c}	Lu + SI high	Yes
SS00006633	INDO	10.8 ± 1.2^{a}	4.4 ± 0.4^{b}	11.0 ± 0.8^{a}	10.9 ± 0.5^{a}	Lu high	Yes
SS00004427	STAT6	10.6 ± 1.9^{a}	10.5 ± 1.3^{a}	10.5 ± 1.1^{a}	12.2 ± 0.6^{a}	Mu low	No
SS00002396	IRF1	5.1 ± 0.6^{b}	3.7 ± 0.2^{c}	6.2 ± 0.4^{a}	4.4 ± 0.3^{bc}	Lu + SI high	Yes
SS00002273	IRF2	6.7 ± 0.8^{ab}	6.2 ± 0.2^{b}	7.6 ± 0.5^{a}	7.0 ± 0.3^{ab}	Lu > Mu	Yes
SS00007514	MAKP14	3.1 ± 0.5^{c}	3.9 ± 0.3^{b}	4.4 ± 0.3^{b}	5.2 ± 0.1^{a}	Li high	Yes
SS00008774	MAKP1	3.9 ± 0.3^{ab}	3.5 ± 0.3^{b}	4.2 ± 0.1^{a}	4.1 ± 0.9^{a}	Lu > Li	Yes****
SS00000832	TGFB1	11.3 ± 0.5^{a}	8.8 ± 0.8^{b}	11.3 ± 0.7^{a}	10.4 ± 0.5^{a}	Mu high	No
SS00000662	TGFB2	7.6 ± 0.7^{a}	3.9 ± 0.4^{b}	4.8 ± 0.1^{b}	6.9 ± 0.2^{a}	Lu + Mu high	Yes

* C_t is the cycle threshold, the cycle number at which amplification crosses the threshold set in the geometric portion of the amplification curve. Lower C_t means higher expression level. ΔC_t is the target transcript C_t − RPL32 C_t, the normalization of C_t for target gene relative to RPL32 RNA C_t.

** ΔC_t levels not connected by the same letter are significantly different at $p \leq 0.05$ across tissues.

*** Lu, lung; SI, small intestine; Li, liver; Mu, muscle.

**** Results generally agree that lung is highest, but the lowest expression in liver in the microarray results was not confirmed by Q-PCR test.

Discussion

Annotation of array oligonucleotides

The oligonucleotides tested in this study were designed by Qiagen–Operon from specific TC sequences assembled at TIGR. A Gene Ontology functional annotation assigned GO terms to only 6244 oligonucleotides, indicating there is limited knowledge from the current GO database to assign function to many of the mammalian genes represented on the array. For the oligonucleotides that were assigned molecular function GO terms, significant numbers of genes belong to enzyme, nucleic acid binding, structural protein, and transport groups (Fig. 1). There are also a number of genes in the signal transducer, apoptosis, chaperone, and immunity protein categories. Thus, although not all of the oligonucleotides on the array had GO terms, the available annotation indicates that this array may be used to investigate many different biological pathways.

Specificity of the array

Specificity is an important property of arrays, especially for mammalian genomes with large structurally related gene families. This can be evaluated by checking the expression pattern of tissue-specific known genes, e.g., the expression changes obtained from these genes should follow the pattern observed in closely related species. In this study, we were able to find hundreds of selectively expressed genes. For example, genes involved in the coagulation pathway, such as plasminogen, anti-thrombin III, and fibrinogen, were selectively expressed in liver, as expected. As well, alcohol sulfotransferase related to alcohol metabolism, apolipoprotein B for lipid processing, acetyl-coenzyme A for bile metabolism, serum amyloid protein P, α1-antitrypsin, and genes involved in xenobiotic metabolism (cytochrome P450 2E1, P450 c27, P450 mono-oxygenase, and P450 2C33) were detected selectively in liver. Liver is the main source of complement component proteins; complement regulator factor H, complement component C1s, and complement cytolysis inhibitor were selectively expressed in liver. Finally, liver is the main source of mannose-binding lectin biosynthesis. We found higher expression of mannose-binding lectin and mannan-binding lectin serine protease 2 in porcine liver compared to other tissues, which is similar to the human and mouse Affymetrix data [2,21,22,24].

Genes expressed in a muscle tissue-selective manner were identified (Supplementary Table 2) and corresponded to other species results from Affymetrix or cDNA array data [21,25–27]. A number of tissue-selective genes in lung or small intestine were expressed as anticipated, e.g., surfactant-associated proteins A and B and matrix Gla protein associated with extracellular matrix were selectively expressed in lung; smooth muscle α-tropomyosin and villin, a major structural component of the brush border cytoskeleton, were found highly expressed in small intestine. More interestingly, a number of immune-related genes showed tissue-selective expression in lung and small intestine, further confirming that these two organs, with exposure to the external environment, are important in host defense against pathogens (Supplementary Table 2).

In addition to the specific gene expression level, we also investigated whether the tissue-selective gene lists could be used to identify expected pathways for tissue-selective genes using EASE software. Results showed that biological pathways identified from these genes are quite similar to the canonical pathways found in the corresponding human or mouse tissues. These results demonstrate that the lists of genes created using this profiling approach describe porcine tissue functions that are quite similar to those in human and mouse, confirming the validity of the array and affirming the utility of the pig as a model for many aspects of mammalian physiology.

In addition to the genes described above, there are a number of genes in each tissue-selective list that have not been annotated. We created a number of clusters for those differentially expressed genes revealed by statistical analyses (Fig. 2). With the functionally annotated “known genes” in the cluster as a signature, our microarray data provide information on coexpression between known and unannotated genes. Such unannotated genes with coexpression data to known genes may be new targets for further understanding the molecular basis of the function of these organs.

Comparison of Q-PCR and microarray data

To validate further our porcine oligonucleotide chip data, we performed Q-PCR on the same RNA samples that were used for the microarray experiments. In total, 12 genes were tested by Q-PCR, including RPL32, which was used as a housekeeping control [28,29]. Similar differences in RNA levels were identified by Q-PCR and by oligonucleotide array analyses for 9 of the 11 test genes. For example, as predicted by microarray results, statistically significant differences determined by Q-PCR were seen for ICAM1 (also called CD54) and INDO, which are highly expressed in lung, while apoptosis-related cysteine protease CASP1 was highly expressed in both lung and small intestine. Members from three gene families that had differential expression among tissues were selected. In the microarray data, MAPK1 showed significantly higher expression in lung, while MAPK14 showed highest expression in liver; these results were all confirmed by Q-PCR. However, significantly higher expression of TGFB1 in muscle in microarray data was not confirmed by Q-PCR, which showed that this gene is most highly expressed in lung. Another member from the TGF family, TGFB2, showed high expression in both muscle and lung, which was confirmed by Q-PCR results (Table 2). It is worth noting that the microarray expression level of most of the genes tested in the Q-PCR analyses was lower than the average microarray expression level across all genes and that the selected genes were all near the p value cutoff for significance (only 1 gene, TGFB2, had $p < 0.05$ after Bonferroni correction). Thus the selected genes were most likely to test strongly the statistical validity of our results. So Q-PCR confirmation of the microarray data for most of the

genes selected provides strong evidence for the quality of the data produced and the statistical model used to declare differential expression. It is interesting to note further that 4 genes (INDO, NOS2A, TGFB1, and TGFB2) showed a difference in expression pattern between porcine microarray and mouse Affymetrix data [22]. Yet Q-PCR results confirmed the porcine microarray expression pattern for 3 of these genes, providing evidence that some genes may have a "species-specific" expression pattern. A species-specific expression pattern for TGFB1 in fibroblast cells was also found among human, bonobo, and gorilla [30].

In summary, understanding gene expression patterns in the pig is a key for using this species as a biomedical model in research on cardiovascular and digestive diseases, cancer, diabetes, and obesity, as well as tissue xenotransplantation [3,4]. High-throughput gene expression studies can also generate molecular data useful in pig breeding and genetics to improve pork production and food quality and safety. The present study provides evidence that this first-generation porcine long oligonucleotide microarray is a useful resource for both these research purposes.

Materials and methods

Porcine oligonucleotides and microarray characteristics

The set of 13,297 oligonucleotides (porcine AROS 1.0/1.0 Extension; www.qiagen.com) represents porcine cDNAs and ESTs and was designed from TIGR TC cDNA sequences (SsGI release 5.0, http://www.tigr.org/tigr-scripts/tgi/T_release.cgi?species=pig). The Qiagen AROS 1.0 has 10,665 oligonucleotides with human/mouse/pig known transcript hits, while the AROS 1.0 Extension has 2632 oligonucleotides with no hit but a clear 3′ untranslated region. In total, there are 13,297 pig-specific ~70-mer oligonucleotides. All oligonucleotides were designed within 1000 bp of an annotated 3′ end, the cross-oligonucleotide percentage identity <70%. No oligonucleotide has 20 contiguous bases in common with any other oligonucleotide. No oligonucleotide has repeats >8 bases or a potential hairpin stem >9 bp. Overall, >98% of the designed oligonucleotides fulfilled these criteria.

The synthesized oligonucleotides were spotted at the University of Minnesota microarray facility. Each oligonucleotide was spotted (~0.5–1 nl at 20 μM) on Corning GAPS II slides with 240-μm spacing. Oligonucleotides were UV cross-linked to the slides after spotting. Each of the pig-specific oligonucleotides was spotted 1 time and each control gene was spotted 16 times. The Qiagen–NRSP8 array has 48 subgrids spotted on the slide; 24 subgrids contain positive controls, and the other 24 contain negative controls.

Gene Ontology terms were annotated by Build Simplified Ontology in GeneSpring 6.1 or self-developed Perl scripts (S. Orley et al., unpublished data). In addition, there are 10 positive control genes and 12 negative controls including 5 *Arabidopsis* genes known to have minimal cross-hybridization with mammalian transcripts [31]. For more information, see www.qiagen.com. For simplicity, we refer to this oligonucleotide set as the Qiagen oligonucleotide set and the microarray produced (see below) as the Qiagen–NRSP8 array.

Tissue collection

Lung, liver, muscle (semitendinosus), and small intestine (duodenum; approximately 5 cm from stomach) were collected from six cross-bred white pigs (three at 15 weeks of age and three at 17 weeks of age).

Experimental design

Tissue samples from the six animals were assigned to six loops (three for younger animals and three for older animals). One sample from each of the four tissues from a single animal was used in each loop. The order of the tissues in each loop was varied so that all pairs of tissues were represented together on a slide an equal number of times. Dye balance was used throughout so that each tissue was measured an equal number of times with each dye. Data from 48 measurements from 24 slides were collected, with 12 measurements for each tissue.

RNA and target preparation for hybridization

Total RNA from ~200 mg of tissue was isolated by a RNeasy Midi kit with on-column RNase-free DNase digestion (Qiagen, Valencia, CA, USA) according to the manufacturer's protocol for all samples. Briefly, frozen tissues were homogenized in liquid nitrogen using a mortar and pestle. RNA purity and integrity were determined by UV spectroscopy at 260 and 280 nm and by denaturing gel electrophoresis.

For generating amino-allyl–dUTP-labeled single-strand cDNA, 30 μg RNA from the four tissue samples was reverse transcribed for 2–3 h at 42°C in the presence of 6 μg random and 1 μg oligo(dT)$_{18}$ primers (Invitrogen/Life Technologies), 1× first-strand buffer, 10 mM DTT, dNTPs (25 mM dCTP, 25 mM dGTP, 25 mM dATP, 15 mM dTTP, 10 mM aa-dUTP), and 400 units Superscript II reverse transcriptase (Invitrogen/Life Technologies). RNA was hydrolyzed with 1 M NaOH and then neutralized with 1 M Hepes (pH 7.0). First-strand cDNA was precipitated, air dried, and resuspended with 4 μl 0.1 M Na_2CO_3 buffer (pH 9.0), mixed with either Cy3 or Cy5 NHS–ester (Amersham Pharmacia), and incubated 1 h at room temperature. Cy3- and Cy5-labeled cDNA was purified on QIAquick PCR purification columns (Qiagen) according to the manufacturer's instructions, except that QIAquick wash buffer was replaced by phosphate buffer [31]. Eluted cDNA was precipitated with EtOH and hybridized to the microarray at 42°C for 12–16 h in hybridization solution containing 50% formamide, 5× SSC, 0.1% SDS, and 0.2 μg/μl sheared salmon sperm DNA. After hybridization, microarray slides were washed with 2× SSC, 0.1% SDS for 5 min at 42°C, 0.1× SSC for 1 min at room temperature four times, and 0.01× SSC for 10 s at room temperature and dried by centrifugation.

Preliminary data obtained from labeling the same liver RNA sample with Cy3 and Cy5 and hybridization to the same slide by the above protocol showed that very few oligonucleotides detected more than a twofold difference in expression between two identical samples (0.38%, data not shown). This technical replication demonstrated the reproducibility of the hybridization data. Thus, we used this protocol in data collection for all subsequent slides.

Image processing

Slides were scanned at 10-μm resolution using a ScanArray 5000 scanner under conditions to limit saturation to <1% and were saved as TIFF images. The intensities of spots on each image were quantified by ImaGene 5.1 software, and data were saved as .txt files for further analyses.

Normalization of microarray data

Local background values were subtracted from signal means, and a small constant was added to all differences to allow for log transformation of the background-corrected signals. Following log transformation, LOWESS normalization [32] was applied to remove intensity-dependent dye bias from each slide. The resulting values were adjusted so that the median normalized signal for each gene would be constant across all slide and dye combinations.

Transcriptome and differential gene expression analyses

For transcriptome analyses in the four tissues, gene-specific mixed linear model analyses were used to identify genes that had signal intensity significantly higher than 3 times the signal median of five *Arabidopsis* genes spotted on the array as negative controls. Each *Arabidopsis* gene was spotted on the array either 12 or 16 times. These replicated spots were summarized for each slide, dye, and negative control gene by computing the median of the normalized signal intensities. For each slide and dye combination, the median of these five medians was used as a baseline

measure of normalized signal intensity for genes not expressed in our samples. These baseline values were subtracted from the normalized signal intensities of each pig oligonucleotide on the array. A separate mixed linear model was fit to these differences for each pig oligonucleotide. The mixed linear model included fixed effects for tissue, age, and tissue-by-age interaction as well as random effects for slide, animal, and animal-by-tissue interaction. Means were estimated for each tissue, and a t test was conducted as part of each mixed linear model analysis to determine whether each estimated mean was greater than log 3. Estimated means significantly greater than log 3 using the normalized data correspond to estimated fold increases greater than 3 on the original scale. A p value was obtained for each pig oligonucleotide and tissue. The set of p values for each tissue was converted to a set of q values using the algorithm proposed by Storey and Tibshirani [33]. To control FDR, genes with q values less than or equal to 0.01 for a given tissue were considered to be expressed in that tissue.

To identify genes differentially expressed across tissues, the same mixed linear model described above was fit to the normalized signal intensities for each gene. Genes with expression significantly higher in one tissue than in each of the others are referred to here as tissue-selective genes. To identify such genes, pair-wise comparisons of normalized signal intensities between tissues were conducted as part of our mixed model analyses for each gene. To control FDR, q values were calculated [33] separately for each pair-wise comparison. Microarray results from this study were submitted to the NCBI GEO database (Accession Nos.: platform GPL1881, samples GSM43151–GSM43174, series GSE2335).

Cluster analyses

An unsupervised learning procedure in "Find classes" in Gene Cluster 2 [34] was used to do the cluster analyses based on the normalized expression levels (log values) for significantly differentially expressed genes.

Pathway identification by EASE

EASE software was downloaded from http://david.niaid.nih.gov/david/ease.htm. We obtained the related pathway information by querying EASE using the human RefSeq IDs for known porcine orthologs that have a tissue-selective pattern of expression. The list of genes for each tissue was used separately to generate four different outputs.

Q-PCR

The purity and integrity of RNA purified as above were confirmed using the Agilent Bioanalyzer 2100 and RNA 6000 Labchip kit (Agilent Technologies, Palo Alto, CA, USA), and synthesis of cDNA was performed using Superscript reverse transcriptase (Invitrogen) and oligo(dT) with 5–10 μg RNA [30,35]. The same liver, lung, muscle, and small intestine RNA samples from four animals in the microarray study were used in Q-PCR. All samples in Q-PCR were measured in duplicate. Using the Stratagene Brilliant kit (La Jolla, CA, USA), Q-PCR was performed on 100 ng RNA equivalents at 25 μl/reaction/well on an ABI Prism 7700 sequence detector system (Applied Biosystems, Foster City, CA, USA). Amplification conditions were 50°C for 2 min, 95°C for 10 min, 40 cycles of 95°C for 15 s and 60°C for 1 min, and then 4°C. All probes and primers were designed using the Primer Express (Applied Biosystems) software and nucleotide sequences obtained from GenBank or the TIGR porcine EST database ([28,29]; http://www.ba.ars.usda.gov/nrfl/nutri-immun-db/nrfl_query1new.html). Gene names and abbreviations are human gene nomenclature based on the International Society for Animal Genetics guidelines. The relative quantitative gene expression level across tissues was evaluated using the comparative C_t method [29]. The ΔC_t values were calculated by subtracting the RPL32 C_t value for each sample from the target C_t value of that sample [35]. Liver, lung, muscle, and small intestine RNA samples from each of four animals were measured in duplicate. The two measures for each sample were averaged, and a linear model analysis of these averages was conducted using JMP5.1 (Computer Associates, Cary, NC, USA). The linear model included animal as a blocking factor and tissue as the factor of interest.

Acknowledgments

Slides used in this study were spotted in collaboration with Dr. V. Kapur of the University of Minnesota as part of the USDA NRSP-8 Swine Subcommittee collaborative agreement and were partially subsidized by USDA-NRSP-8 funds provided through the Pig Genome Coordinator. We thank Sajeev Batra of Qiagen–Operon for producing and providing BLAST analyses data prior to distribution. We thank Kim Glenn and Dr. Steven M. Lonergan, for help with tissue sample collection, and Dr. Eileen Thacker for providing animals. We thank Oksana Kohutyuk for help with the GeneCluster 2 software. This project was supported by grants from the ISU Center for Integrated Animal Genomics to C.K.T. and D.N. and USDA-NRI-2004-35205-14202 to C.K.T., J.K.L., and D.N.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2005.08.001.

References

[1] M. Schena, D. Shalon, R.W. Davis, P.O. Brown, Quantitative monitoring of gene expression patterns with a complementary DNA microarray, Science 270 (1995) 467–470.
[2] A.I. Su, et al., Large-scale analysis of the human and mouse transcriptomes, Proc. Natl. Acad. Sci. USA 99 (2002) 4465–4470.
[3] E. Tumbleson, L.B. Schook, Advances in Swine in Biomedical Research, Plenum Press, New York, 1996.
[4] M.M. Swindle, A.C. Smith, Information Resources on Swine in Biomedical Research 1990–2000, in: C.P. Smith (Ed.), AWIC Resource Series, vol. 11, USDAM, ARS, AWIC, Beltsville, MD, 2000 (http://www.nal.usda.gov/awic/pubs/swine).
[5] D.J. Turner, P.B. Noble, M.P. Lucas, H.W. Mitchell, Decreased airway narrowing and smooth muscle contraction in hyperresponsive pigs, J. Appl. Physiol. 93 (2002) 1296–1300.
[6] C. Watremez, et al., An improved porcine model of stable methacholine-induced bronchospasm, Intens. Care Med. 29 (2003) 119–125.
[7] I.C. McMillen, B.S. Muhlhausler, J.A. Duffield, B.S. Yuen, Prenatal programming of postnatal obesity: fetal nutrition and the regulation of leptin synthesis and secretion before birth, Proc. Nutr. Soc. 63 (2004) 405–412.
[8] J. Pirenne, M. Kawai, Tolerogenic protocols for intestinal transplantation, Transplant. Immunol. 13 (2004) 131–137.
[9] P. Wolf, et al., The pig as a model in liver xenotransplantation, Vet. Res. 28 (1997) 217–222.
[10] A.D. Kirk, Crossing the bridge: large animal models in translational transplantation research, Immunol. Rev. 196 (2003) 176–196.
[11] S.H. Lee, et al., Transcriptional profiling using a novel cDNA array identifies differential gene expression during porcine embryo elongation, Mol. Reprod. Dev. 71 (2005) 129–139.
[12] S.H. Zhao, et al., Complementary DNA macroarray analyses of differential gene expression in porcine fetal and postnatal muscle, J. Anim. Sci. 81 (2003) 2179–2188.
[13] W. Ji, W. Zhou, K. Gregg, N. Yu, S. Davis, A method for cross-species gene expression analysis with high-density oligonucleotide arrays, Nucleic Acids Res. 32 (2004) e93.
[14] D.E. Moody, Z. Zou, L. McIntyre, Cross-species hybridisation of pig RNA to human nylon microarrays, BMC Genom. 3 (2002) 27.
[15] C.L. Afonso, et al., African swine fever virus multigene family 360 and 530 genes affect host interferon response, J. Virol. 78 (2004) 1858–1864.

[16] Q. Bai, et al., Development of a porcine skeletal muscle cDNA microarray: analysis of differential transcript expression in phenotypically distinct muscles, BMC Genom. 4 (2003) 8.
[17] T.N. Ledger, P. Pinton, D. Bourges, P. Roumi, H. Salmon, I.P. Oswald, Development of a macroarray to specifically analyze immunological gene expression in swine, Clin. Diagn. Lab. Immunol. 11 (2004) 691–698.
[18] R.J. Moser, A. Reverter, C.A. Kerr, K.J. Beh, S.A. Lehnert, A mixed-model approach for the analysis of cDNA microarray gene expression data from extreme-performing pigs after infection with Actinobacillus pleuropneumoniae, J. Anim. Sci. 82 (2004) 1261–1271.
[19] S.H. Zhao, J. Recknor, D. Nettleton, M.F. Rothschild, C.K. Tuggle, Current transcript profiling experience with the QiagenOperon porcine long oligonucleotide set (porcine AROS 1.0), Plant Anim. Genome Proc. W160 (2004) 45.
[20] J.M. Bland, D.G. Altman, Multiple significance tests—The Bonferroni method, Br. Med. J. 310 (1995) 170.
[21] L.L. Hsiao, et al., A compendium of gene expression in normal human tissues, Physiol. Genom. 7 (2001) 97–104.
[22] A.I. Su, et al., A gene atlas of the mouse and human protein-encoding transcriptomes, Proc. Natl. Acad. Sci. USA 101 (2004) 6062–6067.
[23] Hosack, et al., Identifying biological themes within a list of genes with EASE, Genome Biol. 4 (2003) R70.
[24] S. Wagner, N.J. Lynch, W. Walter, W.J. Schwaeble, M. Loos, Differential expression of the murine mannose-binding lectins A and C in lymphoid and nonlymphoid organs and tissues, J. Immunol. 170 (2003) 1462–1465.
[25] K. Kadota, et al., Detection of genes with tissue-specific expression patterns using Akaike's information criterion procedure, Physiol. Genom. 12 (2003) 251–259.
[26] R. Miki, et al., Delineating developmental and metabolic pathways in vivo by expression profiling using the RIKEN set of 18,816 full-length enriched mouse cDNA arrays, Proc. Natl. Acad. Sci. USA 98 (2001) 2199–2204.
[27] A. Saito-Hisaminato, T. Katagiri, S. Kakiuchi, T. Nakamura, T. Tsunoda, Y. Nakamura, Genome-wide profiling of gene expression in 29 normal human tissues with a cDNA microarray, DNA Res. 9 (2002) 35–45.
[28] H.D. Dawson, et al., Localized multigene expression patterns support an evolving Th1/Th2-like paradigm in response to infections with Toxoplasma gondii and Ascaris suum, Infect. Immun. 73 (2005) 1116–1128.
[29] H.D. Dawson, et al., Identification of key immune mediators regulating T helper 1 responses in swine, Vet. Immunol. Immunopathol. 100 (2004) 105–111.
[30] M.W. Karaman, et al., Comparative analysis of gene-expression patterns in human and African great ape cultured fibroblasts, Genome Res. 13 (2003) 1619–1630.
[31] H.Y. Wang, et al., Assessing unmodified 70-mer oligonucleotide probe performance on glass-slide microarrays, Genome Biol. 4 (2003) R5.
[32] Y.H. Yang, et al., Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation, Nucleic Acids Res. 30 (2002) e15.
[33] J.D. Storey, R. Tibshirani, Statistical significance for genome wide studies, Proc. Natl. Acad. Sci. USA 100 (2003) 9440–9445.
[34] M. Reich, K. Ohm, M. Angelo, O. Tamayo, J.P. Mesirov, GeneCluster 2.0: an advanced toolset for bioarray analysis, Bioinformatics 20 (2004) 1797–1798.
[35] M. Mohan, S. Ryder, P.L. Claypool, R.D. Geisert, J.R. Malayer, Analysis of gene expression in the bovine blastocyst produced in vitro using suppression–subtractive hybridization, Biol. Reprod. 67 (2002) 447–453.