Biochem Genet (2009) 47:117 125
DOI 10.1007/s10528-008-9211-5

Received: 15 Februar 2008 / Accepted: 15 August 2008 / Published online: 24 Januar 2009

T o non-normali ed cDNA libraries of uteri from Danish Landrace and Chinese Erhualian pigs ere constructed, and 13,756 e pressed sequence tags (ESTs) ere randoml sequenced. The ESTs ere clustered b Phrap soft are, and 6,139 distinct tentative consensus sequences ere produced, including 2,730 contigs and 3,409 singlets. Using Blast tools, these 6,139 candidate genes ere compared to the nr and nt databases; 5,210 of them ere assigned putative functions, hereas 929 potentiall represent ne genes. Highl e pressed genes appear to be associated ith basic energ metabolism, transferase activit , locali ation, cellular ph siological process, protein binding, and nucleic acid binding. Antileukoproteinase as the most highl e pressed gene, corresponding to endometrial differentiation and conceptus or fetal development.

Porcine uterus · Porcine ESTs · Gene e pression · Pig reproduction

H. Chai · D.-L. Yu
College of Life Sciences, Zhejiang Universit , Hang hou 310058, People's Republic of China

B. Zhang · S.-N. Hu
Beijing Institute of Genomics, Chinese Academ of Sciences, Beijing 100029,
People's Republic of China

Y. Fu (✉)
College of Animal Sciences, Zhejiang Universit , Hang hou 310029, People's Republic of China
e-mail: fu an@ ju.edu.cn

Y. Fu · S.-N. Hu
James D. Watson Institute of Genome Sciences, Zhejiang Universit , Hang hou 310008,
People's Republic of China

The reproductive process is central to pig production ef cienc , and the ovar and

interrupted on the linker locus. The ESTs that ere longer than 100 bp ere retained for later anal sis. All high-qualit and clean ESTs ere assembled b Phrap soft are, ith 40 minmatch and 0.95 repeat stringenc . Contigs and singlets ere called clusters. All clusters ere compared to the nonredundant nucleic acid (nt) and protein (nr) database provided b GenBank ith Blast tools. The best hit for each quer as used for function assignment and subsequentl manuall checked. Function categor as performed ith the GO database. All clusters ere also compared ith the human EST database for homologous sequences b Blast.

Signi cant Differentiation Statistic Test

The ESTs from porcine uterus cDNA libraries ere divided into t o groups, one from Danish Landrace and the other from Chinese Erhualian. The Web tool IDEG6 (http://telethon.bio.unipd.it/bioinfo/IDEG6_form/) as used to detect differentiall e pressed gene categories ith $P < 0.05$.

Overvie of cDNA Libraries and Clustering

In order to get an overvie of porcine genes in the uterus, t o nonnormali ed cDNA libraries ere constructed from different breeds (Danish Landrace and Chinese Erhualian). In total, 13,756 cDNA clones ere randoml selected (6,905 from Danish Landrace and 6,851 from Chinese Erhualian) and partiall sequenced from cDNA 5′ ends to generate ESTs (Table 1). The initial EST sequences ere

EST sequences (Table 2). More than half of the largest contigs had consensus sequences that ere homologous to genes involved in protein s nthesis (initiation factors, elongation factors, and ribosomal proteins). There as a tissue-speci c

gene, uteroferrin-associated basic protein-2 (UABP-2, NM_213845), hich is also abundantl e pressed.

As revealed b the cDNA frequenc , antileukoproteinase (ALP) as the most abundantl e pressed gene in porcine uterus. Other highl e pressed genes ere NADH deh drogenase (NP_008644), secreted phosphoprotein-I (SPPI, NM_214023), and elongation factor 1α (EF-1 alpha, NM_001097418).

The gene e pression pro les of the t o breeds ere different. The genes that ere more highl e pressed in Chinese Erhualian pigs ere c tochrome c o idase (COX) subunit I, 3-beta-h dro steroid deh drogenase/delta-5-delta-4 isomerase (3 beta-HSD), and c tochrome P450 11A1; in Danish Landrace the highl e pressed genes ere ribosomal protein S8 (RPS8), UABP-2, 60S acidic ribosomal protein P0 (RPP0), ribosomal protein L9 (RPL9), 60S acidic ribosomal protein P2 (RPP2), 60S ribosomal protein L6 (RPL6), ribosomal protein S23 (RPS23), and ribosomal protein L21 (RPL21).

The cDNAs ere classi ed according to the GO inde ith categories for cellular component, molecular function, and biological process. There ere 83 contigs (432 individual cDNA clones) ith GO cellular component annotations. Each contig contained more than si cDNA clones, and the numbers ere similar in the t o pig breeds. Based on the number of cDNA clones, the majorit of cellular mRNA-encoded component as ribonucleoprotein comple (Table 3). There ere 602 contigs (2,370 individual cDNA clones) clustered into the group involved in molecular function. The genes in this group e pressed higher in Chinese Erhualian than in Danish Landrace, and ere associated ith nucleic acid binding, nucleotide binding, protein binding, and h drolase activit . The consensus sequences for most contigs ere homologous to genes hose products ere involved in transferase activit (catal tic activit). There ere 727 contigs (2,833 individual cDNA clones) ith GO biological process annotations. More than half of the genes ere involved in cellular metabolism. The genes e pressed higher in Chinese Erhualian ere associated ith cellular ph siological processes and loca ation.

' ' ' '

A primar object of EST sequencing is gene identi cation (Jiang et al. 2003). One method to identif the factors that control ovarian function is to characte e the genes that are e pressed in the uterus (Caetano et al. 2003). The random sampling strategies resulted in highl e pressed genes represented b man EST sequences. The frequenc of cDNA ithin each tissue could be determined as each clone as sequenced from its original librar (Zhang et al. 2004). We anal ed 10,879 high-qualit ESTs generated from t o nonnorma ed porcine uterus cDNA libraries. The sequences clustered into 2,730 contigs, and the contig sequences ere blasted against the nr or nt databases in Genbank. The genes ere associated ith common cell functions, such as energ metabolism, protein s nthesis, signal transduction, cell communication, transport, development, and cell-c cle regulation. Genes associated ith uterus-speci c functions, such as *UABP-2*, ere also identi ed (Table 2).

Number of cDNAs within categories of the GO index.

GO index.	Total cDNAs	cDNAs from Danish Landrace	cDNAs from Chinese Erhualian
Cellular component			
Protein complex. respiratory chain complex. I	20	11	9
Protein complex. ribonucleoprotein complex.	40	18	22
Protein complex. transcription factor complex.	6	2	4
virion, viral capsid	17	11	6
Molecular function			
Binding	4	1	3
Binding, ion binding	31	17	14
Binding, lipid binding	81	41	40
Binding, nucleic acid binding	88	23	65
Binding, nucleotide binding	14	3	11
Binding, pattern binding	1	1	0
Binding, peptide binding	1	1	0
Binding, protein binding	92	35	57
Binding, ribonucleoprotein binding	1	0	1
Binding, selenium binding	5	2	3
Binding, steroid binding	4	1	3
Binding, vitamin binding	2	1	1
Catal tic activit	15	7	8
Catal tic activit , h drolase activit	33	13	20
Catal tic activit , isomerase activit	1	1	0
Catal tic activit , ligase activit	1	1	0
catal tic activit , l ase activit	5	2	3
Catal tic activit , o idoreductase activit	25	11	14
Catal tic activit , small protein conjugating e me activit	4	1	3
Catal tic activit , transferase activit	146	90	56
E me regulator activit , GTPase regulator activit	4	2	2
E me regulator activit , e me activator activit	10	4	6
E me regulator activit , e me inhibitor activit	3	1	2
E me regulator activit , kinase regulator activit	6	2	4
E me regulator activit , ornithine decarbo.lase regulator activit	1	0	1
Obsolete molecular function, FK506-sensitive peptid l-prol l cis-trans isomerase	1	1	0
Obsolete molecular function, Rho small monomeric GTPase activit	10	3	7
Obsolete molecular function, ba3-t pe COX	1	0	1
Obsolete molecular function, barbed-end actin capping/severing activit	2	1	1
Obsolete molecular function, cell adhesion molecule activit	10	2	8

continued

GO inde	Total cDNAs	cDNAs from Danish Landrace	cDNAs from Chinese Erhualian
Biological process			
Obsolete biological process, pero idase reaction	3	1	2
Ph siological process	3	0	3
Ph siological process, cellular ph siological process	112	43	69
Ph siological process, homeostasis	24	11	13
Ph siological process, local ation	127	41	86
ph siological process, metabolism	372	189	183
Ph siological process, organismal ph siological process	5	2	3
Ph siological process, response to stimulus	47	21	26
Regulation of biological process, regulation of development	2	1	1
Regulation of biological process, regulation of e me activit	3	1	2
Regulation of biological process, regulation of ph siological process	29	13	16

As revealed b the cDNA frequenc , ALP cDNA as the most abundant cDNA (Table 2). ALP is a ph siologic inhibitor of granuloc tic serine proteases. Other highl e pressed genes in porcine uterus ere NADH deh drogenase and SPPI. NADH deh drogenase, kno n as the NADH:Ubiquinone o idoreductase, is comple I of the mitochondrial electron transfer chain, and it catal es the transfer of electrons from NADH to coen me Q (Malath et al. 1990). It is ell kno n that SPPI is a highl phosphor lated form that has been associated ith cell transformation (Roberts and Be er 1988). The fourth highest e pressed gene is EF-1 alpha, an essential component of the eukar otic translational apparatus, hich is a GTP-binding protein that catal es the binding of aminoac l-transfer RNAs to the ribosome (Li et al. 2002). Other genes involved in protein s nthesis (including ribosomal proteins) ere also highl e pressed. We found 12 ribosomal protein genes for all 80 components of the ribosome among 6,139 clusters, hich indicates that e have found signi cant e pression information in the porcine uterus.

The gene e pression pro les ere different for the t o porcine species. The genes ith higher e pression in Chinese Erhualian than in Danish Landrace ere COX subunit I, 3 beta-HSD, and c tochrome P450 11A1. COX is one of a superfamil of proteins that act as the terminal en mes of respirator chains. The t o main classes are COXs and quinol o idases. Mitochondrial COX and its bacterial homologs catal e electron transfer and proton translocation reactions across membranes (Mammalian Gene Collection Program Team 2002). Three-beta-HSD catal es the o idative conversion of delta 5-3-beta-h dro steroids to the delta 4-3-keto con guration, and is therefore essential for the bios nthesis of all classes of hormonal steroids, namel , progesterone, glucocorticoids, mineralocorticoids, androgens, and estrogens (Adams et al. 1995). This ma be associated ith the larger litter si e of the Chinese Erhualian, compared ith the Danish Landrace. The function of the mammalian P450 s stem concerns its role as the rate-limiting

enzyme in the synthesis of all steroid hormones and many prostaglandins and leucotrienes. As such, the P450s play a central role in mineral balance, sugar regulation, reproduction, water balance, digestion of lipids, vascular tone, pain, and inflammation (Fahrenkrug et al. 2002).

In summary, our study provides a catalog of 2,730 contigs derived from 10,879 cDNA sequences obtained from porcine uterus tissues. For most contigs, the frequency of the sequenced genes was too low to study the gene expression reliably across tissues. Almost a quarter of the EST clusters did not have any match with nt or nr databases. Why are there so many anonymous ESTs? Cirera et al. (2000) reviewed various possible reasons, but here we propose two main explanations. First, the pig genome project is ongoing, and many genes expressed in the uterus still have not been identified. Second, a number of ESTs sequenced from uterus cDNA libraries are too short to be identified, and they may represent untranslated regions of the gene; 3′ UTR sequences vary more than the coding regions between organisms. Some of the ESTs may represent transcripts that have diverged to the extent that they are not recognized as orthologs; others may be inaccurate sequence data (Cirera et al. 2000). Thus, sequencing cDNA from the mammalian uterus is a good strategy for novel gene identification.

At present, the strategies of gene prediction include two basic approaches (Wang et al. 2003; Rogic et al. 2001). One is the ab initial computational prediction using statistic information, and the other is the integrated method of computational and sequence similarity search among species. The latter relies largely on cDNA resources and homologous comparison among relatively close organisms. Thus, these ESTs from the porcine uterus should be a useful resource for the pig genome sequencing project and its annotation.

This study was supported by the Sino-Danish Pig Genome Project.

Adams MD, Kerlavage AR, Fleischmann RD, Fuldner RA, Bult CJ, Lee NH, Kirkness EF, Weinstock KG, Gocayne JD, White O, Sutton G, Blake JA, Brandon RC, Chiu M-W, Clayton RA, Cline RT, Cotton MD, Hughes JE, Fine LD, Fitzgerald LM, Fitzhugh WM, Fritchman JL, Geoghagen NSM, Glodek A, Gnehm CL, Hanna MC, Hedblom E, Hinkle PS Jr, Kelley JM, Klimek KM, Kelley JC, Liu L-I, Marmaros SM, Merrick JM, Moreno-Palanques RF, McDonald LA, Nguyen DT, Pellegrino SM, Phillips CA, Ryder SE, Scott JL, Saudek DM, Shirley R, Small KV, Spriggs TA, Utterback TR, Weidman JF, Li Y, Barthlow R, Bednarik DP, Cao L, Cepeda MA, Coleman TA, Collins E-J, Dimke D, Feng P, Ferrie A, Fischer C, Hastings GA, He W-W, Hu J-S, Huddleston KA, Greene JM, Gruber J, Hudson P, Kim A, Kozak DL, Kunsch C, Ji H, Li H, Meissner PS, Olsen H, Raymond L, Wei Y-F, Wing J, Xu C, Yu G-L, Ruben SM, Dillon PJ, Fannon MR, Rosen CA, Haseltine WA, Fields C, Fraser CM, Venter JC (1995) Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. Nature 377(Suppl.):3 16

Caetano AR, Johnson RK, Pomp D (2003) Generation and sequence characterization of a normalized cDNA library from swine ovarian follicles. Mamm Genome 14:65 70

Cirera S, Winter AK, Fredholm M (2000) Why do we still find anonymous ESTs? Mamm Genome 11:689 693

Fahrenkrug SC, Smith TPL, Freking BA, Cho J, White J, Vallet J, Wise T, Rohrer G, Pertea G, Sultana R, Quackenbush J, Keele JW (2002) Porcine gene discovery by normalized cDNA-library sequencing and EST cluster assembly. Mamm Genome 13:475 478

Jiang Z, Zhang M, Wasem VD, Michal JJ, Zhang H, Wright RW (2003) Census of genes e pressed in porcine embr os and reproductive tissues b mining an e pressed sequence tag database based on human genes. Biol Reprod 69:1177 1182

Li N, Zhao ZH, Liu ZL, Zhao XB, Lian ZX, Wu CX (2002) Anal sis of e pressed sequence tags from porcine liver organ. Sci Agric Sinica 35(12):1525 1528

Malath PV, Imaka a K, Simmen RC, Roberts RM (1990) Molecular cloning of the uteroferrin-associated protein, a major progesterone-induced serpin secreted b the porcine uterus, and the e pression of its mRNA during pregnanc . Mol Endocrinol 4(3):428 440

Mammalian Gene Collection (MGC) Program Team (2002) Generation and initial anal sis of more than 15,000 full-length human and mouse cDNA sequences. PNAS 99(26):16899 16903

Roberts RM, Bā er FW (1988) The functions of uterine secretions. J Reprod Fertil 82(2):875 892

Rogic S, Mack orth AK, Ouellette FBF (2001) Evaluation of gene- nding programs on mammalian sequences. Genome Res 11:817 832

Wang J, Li ST, Zhang Y, Zheng HK, Xu Z, Ye J, Yu J, Wong GK-S (2003) Vertebrate gene predictions and the problem of large genes. Nat Rev Genet 4:741 749

Zhang B, Jin W, Zeng YW, Su ZX, Hu SN, Yu J (2004) EST-based anal sis of gene e pression in the porcine brain. Genomics Proteomics Bioinformatics 2:237 244