Mol. Cells *28*, 529-536, December 31, 2009
DOI/10.1007/s10059-009-0147-3

Molecules and Cells

Full-Length Enriched cDNA Library Construction from Tissues Related to Energy Metabolism in Pigs

Kyung-Tai Lee[1,7], Mi-Jeong Byun[1,7], Dajeong Lim[1,2,7], Kyung-Soo Kang[2], Nam-Soon Kim[3], Jung-Hwa Oh[4], Chung-Soo Chung[5], Hae-Suk Park[1], Younhee Shin[6], and Tae-Hun Kim[1,*]

Genome sequencing of the pig is being accelerated because of its importance as an evolutionary and biomedical model animal as well as a major livestock animal. However, information on expressed porcine genes is insufficient to allow annotation and use of the genomic information. A series of expressed sequence tags of 5′ ends of five full-length enriched cDNA libraries (SUSFLECKs) were functionally characterized. SUSFLECKs were constructed from porcine abdominal fat, induced fat cells, loin muscle, liver, and pituitary gland, and were composed of non-normalized and normalized libraries. A total of 55,658 ESTs that were sequenced once from the 5′ ends of clones were produced and assembled into 17,684 unique sequences with 7,736 contigs and 9,948 singletons. In Gene Ontology analysis, two significant biological process leaf nodes were found: gluconeogenesis and translation elongation. In functional domain analysis based on the Pfam database, the beta transducin repeat domain of WD40 protein was the most frequently occurring domain. Twelve genes, including SLC25A6, EEF1G, EEF1A1, COX1, ACTA1, SLA, and ANXA2, were significantly more abundant in fat tissues than in loin muscle, liver, and pituitary gland in the SUSFLECKs. These characteristics of SUSFLECKs determined by EST analysis can provide important insight to discover the functional pathways in gene networks and to expand our understanding of energy metabolism in the pig.

INTRODUCTION

The pig is an important model animal for evolutionary and biomedical research because it has physiological similarities to humans, and it is also considered one of the most important livestock animals for the meat industry around the world (Fang et al., 2005; Lunney et al., 2007). The pig transcriptome has been analyzed by many groups to answer biomedical, agricultural, and fundamental biological questions. The fat traits of pigs, such as intramuscular fat (IMF) content and backfat thickness (BFT), are among the most interesting traits because it they constitute major parameters in determining pork quality (Huff-Lonergan et al., 2002; van Wijk et al., 2005). The fat traits that influence pork quality may provide key information for tackling the problem of obesity in humans. Adipocytes are not only passive storage spaces, but also play active roles in the regulation of nutrition, energy homeostasis, and immune response, and in releasing many hormones and other signal transduction materials (Trayhurn et al., 2005).

Expressed sequence tag (EST) analysis is one of the most effective and widely used methods for studying gene expression (Adams et al., 1991). Therefore, the development of ESTs can provide powerful tools for structural and functional genomics (Stapleton et al., 2002). A cDNA library is a pool of transcripts that represent the profile of gene expression for particular tissues or cells. cDNA libraries are highly beneficial for gene discovery. An understanding of the function of a given tissue can be gained by analysis of genes that are abundantly expressed in that tissue (Chen et al., 2006). In pigs, approximately 1.7 million ESTs have been deposited in public databases such as dbEST and Trace Archive since the first EST project (Tuggle et al., 1994) and first large-scale EST project (Wintero et al., 1996). Mikawa et al. (2004) characterized 298 EST clones from pig backfat tissue. Although the amounts of porcine transcriptomic data have increased markedly, only a few research groups have reported ESTs derived from full-length enriched cDNA libraries. ESTs from the 5′ ends of full-length enriched cDNA libraries are more efficient for identifying protein coding information and functional implications of cDNA. A full-length cDNA library from porcine olfactory bulb was reported previously (Fujisaki et al., 2004). The Pig EST Data Explorer (PEDE) was developed to characterize a collection of porcine full-length cDNA sequences from the thymus, spleen, uterus, lung, liver, ovarian tissues, peripheral blood mononuclear cells, adrenal gland, alveolar macrophages, intestine, mesenteric lymph nodes, trachea, testis, and skin (Uenishi et al., 2004; 2007). In adipose tissues, Chen et al. (2006) constructed a full-length enriched cDNA library from subcutaneous

[1]Division of Animal Genomics and Bioinformatics, National Institute of Animal Science, Rural Development Administration, Suwon 441-707, Korea, [2]Department of Agricultural Biotechnology, Seoul National University, Seoul 151-742, Korea, [3]Laboratory of Human Genomics, Genome Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon 305-806, Korea, [4]Toxicogenomics Team, Korea Institute of Toxicology, Daejeon 305-343, Korea, [5]Department of Animal Science, Chungbuk National University, Cheongju 361-763, Korea, [6]Bioninformatics Division; Insilicogen, Inc., Suwon 441-813, Korea, [7]These authors contributed equally to this work.
*Correspondence: kth6160@rda.go.kr

Received June 16, 2009; revised September 2, 2009; accepted September 17, 2009; published online November 19, 2009

Keywords: fat, full-length cDNA, gene ontology, pig

adipose tissue and reported genes expressed abundantly in pig adipose tissue from analysis of 2,880 clones. In addition, we reported a full-length enriched cDNA library from porcine backfat tissue (Kim et al., 2006).

In the present study, we constructed five full-length enriched cDNA libraries from abdominal fat, induced fat cells, loin muscle, liver, and pituitary gland, which are related to fat metabolism in the pig. Large-scale ESTs were generated from five libraries, and their functional characteristics, including ESTs from a backfat cDNA library constructed previously, were analyzed to elucidate expression patterns among tissues and cells.

MATERIALS AND METHODS

Sample preparation and total RNA extraction

Four Landrace × Large White crossbred pigs were slaughtered at each of four ages, 1, 7, 12, 18, and 24 weeks, to maximize discovery of genes expressed in each tissue. The abdominal subcutaneous fat, loin muscle, liver, and pituitary gland were excised immediately after slaughter, snap-frozen in liquid nitrogen, and stored at -70°C until RNA extraction. The preadipocytes (stroma-vascular cells) were isolated from the backfat of ten male and female newborn pigs (Landrace × Yorkshire × Duroc), as described previously (Moon and Chung, 2004; Suryawan et al., 1997). The isolated cells were seeded and treated with insulin, transferrin, and hydrocortisone to induce differentiation one day after seeding (day 0). RNAs were prepared from the cells on days 1, 4, and 10 representing proliferation stage, early differentiation stage, and late differentiation stage, respectively. The pooled RNAs were used to construct a library. Total RNA was extracted from the pooled samples from four Landrace pigs at each age using an RNeasy Midi Kit (Qiagen, USA) in accordance with the manufacturer's instructions, as reported previously (Kim et al., 2006).

Full-length enriched cDNA library construction

The full-length enriched cDNA libraries were constructed using the oligo-capping method as described previously (Kato et al., 1994; Maruyama et al., 1994; Oh et al., 2003; Suzuki et al., 2001). Briefly, aliquots of 100 μg of total RNA were treated with bacterial alkaline phosphatase (TaKaRa, Japan) and then with 100 units of tobacco acid pyrophosphatase (Wako, Japan). The pretreated total RNA was ligated with 0.4 μg of 5′-oligoribonucleotide (5′-AGC AUC GAG UCG GCC UUG UUG GCC UAC UGG-3′). After completing the oligo-capping reactions, mRNA was isolated using an Oligotex Mini Kit (Qiagen). Synthesis of first-strand cDNA from the purified mRNA and cDNA amplification were performed as described previously (Maruyama et al., 1994). The amplified PCR products were then digested with *Sfi*I, and cDNAs longer than 1.3 kb were ligated into *Dra*II-digested pCNS-D2 (Oh et al., 2004) in an orientation-defined manner. The ligated cDNA was then transformed into *Escherichia coli* Top 10F′ (Invitrogen, USA) by electroporation (GenePulser II; BioRad, USA). The constructed cDNA libraries were normalized as described previously (Soares et al., 1994).

Plasmid isolation and cDNA sequencing

Colonies were picked at random, inoculated into individual wells of 96-well plates containing 500 μl of TB media, and incubated at 37°C for 18 h. Plasmid DNAs were extracted using a Montage Plasmid Miniprep 96 Kit (Millipore, USA) in accordance with the manufacturer's instructions. The cDNA inserts were sequenced once from the 5′ ends of clones using a BigDye Terminator Sequencing Kit ver 3.1 (Applied Biosystems, USA) and a 3730 DNA Analyzer (Applied Biosystems).

Characterization of the full-length enriched cDNA libraries

Base-calling of the sequencing reads was performed using Phred with an error probability cutoff of 0.01, and vector sequences were trimmed with Cross_match. Reads shorter than 200 bp were discarded with Seqclean. Then, the valid reads were assembled with CAP3 (Huang and Madan, 1999). The EST sequences were compared with human and mouse consensus coding DNA sequence (CCDS) to estimate the fullness of the SUSFLECK inserts (http://www.ncbi.nlm.nih.gov/projects/CCDS). The coding DNA sequences (CDSs) were downloaded from the NCBI ftp site (ftp://ftp.ncbi.nlm.nih.gov/pub/CCDS/, human build 36.2, mouse build 37.1). In addition, CDSs with high similarity with our EST sequences were selected using BLAST with an e-value cutoff of E-100. ESTs with sequences longer than the corresponding CDS and ATG translation initiation codon in the same frame with ORF of CDS upstream of the aligned regions with high similarity were considered to include full-length CDSs.

Sequence data analysis and EST clustering

The porcine EST trace data were base-called using Phred (Ewing and Green, 1998) and were vector-clipped by the Cross_match program (Gordon et al., 1998) with vector sequences. Vector-screened EST sequences were filtered according to repetitive sequence and low-complexity regions using RepeatMasker (http://ftp.genome.washington.edu/RM/RepeatMasker.html). The ESTs were clustered and assembled using CAP3. Transcripts of less than 200 bp were filtered from our dataset using SeqClean.

Gene Ontology annotation and functional analysis

To annotate porcine ESTs with Gene Ontology, a sequence similarity search was performed against the tentative consensus (TC) sequences of the *Sus scrofa* Gene Index (SsGI release 12.0) and UniGene cluster (Build #27) using BLASTN (Altschul et al., 1997) with cutoff values of 95% identity, 60% coverage, and an e-value < 1E-5 for GO identification. GO categorization of the ESTs was then performed with the GO profile of the Gene Ontology Consortium (http://www.geneontology.org). EST sequences that did not satisfy the above conditions were placed in the putative novel transcript group. For GO annotation of these novel transcripts, a BLASTX search was performed against the non-redundant protein sequence database downloaded from ftp://ftp.ncbi.nih.gov/blast/db/ (29 April 2007). A list of ESTs that share functional annotation was retrieved for each identified ontology term. To examine whether a functional bias was present in the ESTs in this study, Pearson's χ^2 test was used to determine statistical significance. We examined whether the ratio of genes associated with a particular GO term was the same between the gene list of interest and the total gene set (Zhong et al., 2004). Significance tests were performed for the 2nd level of GO terms to leaf terms by applying Bonferroni's correction (Bonferroni, 1936) to correct the multiple test problems. The hierarchical classification of GO terms in which the first level consists of Molecular Function, Biological Process, and Cellular Component terms was used. A significance level of 0.05 was used to reject the null hypothesis and identified only the significant leaf nodes with the algorithm described previously (Kim et al., 2007). Briefly, the algorithm searches all child nodes to determine whether any of its child nodes show significance. Another functional profiling analysis was conducted by comparing protein domains of pig ESTs and sequence profiles representing Pfam domains (Bateman et al., 2004) using the hmmpfam program. A total of 13,539 transcript sequences were translated in all six frames using ESTScan.

Table 1. Information of SUSFLECK library from each tissue

Library	Abdominal fat	Loin muscle	Backfat	Fat cell	Liver	Pituitary gland	Total
Non-normalized	5100	5791	13268	5168	7124	6274	42725
Matched to CCDS*	3235	2937	8454	3503	3574	4441	26144
Full CDS**	2318	568	5719	2703	2631	2703	16642
Fullness***	72%	19%	68%	77%	74%	61%	64%
Normalized	5121	5415	2075	5221	4692	5851	28375
Matched to CCDS	2539	942	1043	2785	1871	2579	11759
Full CDS	1209	250	497	1251	1001	1423	5631
Fullness	48%	27%	48%	45%	54%	55%	48%
Total ESTs	10221	11206	15343	10389	11816	12125	71100

*Matched to CCDS: The number of ESTs matched to consensus coding DNA sequences of human and mouse (BLAST e-value <1E-100)
**Full CDS: The number of ESTs containing the putative translation initiation codon (ATG) among the ESTs matched to CCDS
***Fullness: The proportion of full CDS within the ESTs matched to CCDS

Domain classes were also assigned according to Pfam to GO mapping (Harris et al., 2004) provided by the InterPro database.

Investigation of genes abundant in fat tissues

The redundancy of each assembled contig in each non-normalized library was counted to investigate abundant transcripts. Transcripts composed only of fat origin ESTs of SUSFLECK libraries were selected, and BLAST alignment was performed to exclude the ESTs detected in other tissues with identity of more than 95% and matched lengths of 100 bp or longer. GO analysis of the selected transcripts was performed using TCs of six organisms (human, dog, cattle, pig, mouse, and rat) from BLAST results as described above, except in this case, an e-value of 1E-5 was used as the cutoff, and χ^2 test of independence was performed. Bonferroni's correction was applied to correct the multiple test problems.

RESULTS AND DISCUSSION

Characterization of *Sus scrofa* full-length enriched cDNA libraries (SUSFLECK)

A total of 71,100 high-quality EST sequences (Phred quality score > 20 and at least 200 bp) were generated from both normalized and non-normalized libraries (Table 1). With the exception of EST sequences of backfat libraries reported previously (Kim et al., 2006), all 55,658 sequences were deposited in the dbEST division of GenBank (accession numbers: FD587621-FD643263, FD698977-FD698990). High-quality ESTs were assembled into 17,684 unique sequences with 7,736 contigs that included two or more ESTs and 9,948 singletons. The average number of ESTs per contig was approximately 8.5 for all libraries. We defined fullness as the proportion of cDNA clones containing the putative translational initiation codon (ATG) among the annotated cDNA clones by comparison with the human and mouse consensus CDS (CCDS) in NCBI. CDS sequences with high similarity were chosen for open reading frame and putative initiation site (BLAST e-value < 1E-100). The ESTs containing longer upstream sequences relative to the start site of the aligned CDS and also containing an ATG translation initiation codon in the same frame as the open reading frame of the CCDS were considered to have full-length CDSs. Of the 71,100 ESTs, 37,903 (53%) matched CCDS sequences, and 60% of these were predicted to contain a putative ATG translational initiation site (Table 1). The estimated average cDNA insert size was 1.7 kb.

Gene Ontology annotation and functional domain assignment

To classify transcripts by putative function, enrichment bias in the GO categories was manifested in both the number of transcripts in this study and in The Institute for Genome Research (TIGR) gene index of *Sus scrofa* (SsGI). In Fig. 1, the results are shown as three separate graphs, each representing the 2nd level of GO terms according to the GO Consortium (Ashburner et al., 2000). Further analyses were conducted to find unique functions using Fisher's exact test for GO terms from the 2nd level. The results indicated 32 significant leaf nodes in Biological Process, Molecular Function, and Cellular Component (Table 2). Approximately 90% (12,015) of the total transcript sequences were annotated with their functional domain based on the Pfam database (Bateman et al., 2004). The ten categories of the most frequently occurring domains including WD40 motif (Lee et al., 2006b) are listed in Table 3. A total of 12,014 transcripts were mapped to the functional assignment based on the Gene Ontology annotation of Pfam domains provided by the InterPro database (Mulder et al., 2005). Of these, 5,115 transcripts (38%) were found to have functional domains.

Investigation of genes abundant in fat tissues

These SUSFLECK libraries included cDNA libraries from three fat tissues, *i.e.*, backfat, abdominal fat, and cultured fat cells induced from preadipocytes in pigs. The genes in this EST data set showing fat-specific expression were investigated because of the importance of the fat traits of pigs. A total of 1,242 transcripts, with 263 contigs and 979 singletons from only fat tissues and cultured fat cells, were used in BLAST similarity searches (e-value < 1E-5) that involved 55,663 high-fidelity virtual transcripts called Tentative Consensus (TC) sequences of human, dog, cattle, pig, mouse, and rat (Quackenbush et al., 2000). Of the 2,711,770 TCs used for comparison, 55,987 were annotated with the Gene Ontology (GO) terms, and 418 transcripts from SUSFLECKs of fat tissues corresponded to the annotated TCs. The transcripts that had significantly different frequencies in SUSFLECKs of fat tissues as determined using Pearson's χ^2 test of difference between the observed and expected number of transcripts on the basis of GO terms from the 2nd level to all leaf nodes were investigated (Fig. 2). The different frequencies in SUSFLECKs of fat tissues were specifically shown in six GO terms. Fat tissue libraries were significantly overrepresented in the terms "extracellular matrix" (Cellular Component), "multicellular organismal process", "biological adhesion" (Biological Process), "molecular transducer activity",

Table 2. Significant GO terms from Fisher's exact test

GO number	Description	GO type	Subset	Total set	p-value
GO:0044456	Synapse part	Cellular component	238/16462	765/254210	0.00197
GO:0045202	Synapse	Cellular component	338/16462	1035/254210	0.0026
GO:0031975	Envelope	Cellular component	730/16462	2614/254210	0.00621
GO:0031974	Membrane\-enclosed lumen	Cellular component	1044/16462	2946/254210	0.00756
GO:0044421	Extracellular region part	Cellular component	1835/16462	4782/254210	0.01291
GO:0005576	Extracellular region	Cellular component	2009/16462	5609/254210	0.01493
GO:0032991	Macromolecular complex	Cellular component	2260/16462	7001/254210	0.01825
GO:0030188	Chaperone regulator activity	Molecular function	7/16462	11/254210	0.00012
GO:0015457	Auxiliary transport protein activity	Molecular function	42/16462	126/254210	0.0003
GO:0016209	Antioxidant activity	Molecular function	40/16462	132/254210	0.0003
GO:0009055	Electron carrier activity	Molecular function	103/16462	260/254210	0.0007
GO:0045182	Translation regulator activity	Molecular function	99/16462	272/254210	0.00071
GO:0030234	Enzyme regulator activity	Molecular function	396/16462	977/254210	0.0026
GO:0008369	Obsolete molecular function	Molecular function	444/16462	1046/254210	0.00264
GO:0005198	Structural molecule activity	Molecular function	559/16462	2491/254210	0.00559
GO:0005215	Transporter activity	Molecular function	1059/16462	3063/254210	0.00783
GO:0042056	Chemoattractant activity	Molecular function	4/16462	9/254210	0.00785
GO:0030528	Transcription regulator activity	Molecular function	1076/16462	3224/254210	0.00819
GO:0060089	Molecular transducer activity	Molecular function	1198/16462	3852/254210	0.00973
GO:0031386	Protein tag	Molecular function	5/16462	23/254210	0.03209
GO:0001906	Cell killing	Biological process	24/16462	49/254210	0.00014
GO:0016032	Viral reproduction	Biological process	45/16462	139/254210	0.00033
GO:0043473	Pigmentation	Biological process	67/16462	214/254210	0.00053
GO:0048511	Rhythmic process	Biological process	152/16462	402/254210	0.00108
GO:0008371	Obsolete biological process	Biological process	339/16462	1137/254210	0.00264
GO:0051704	Multi\-organism process	Biological process	503/16462	1229/254210	0.00276
GO:0040011	Locomotion	Biological process	526/16462	1728/254210	0.00387
GO:0022610	Biological adhesion	Biological process	559/16462	1819/254210	0.00414
GO:0022414	Reproductive process	Biological process	842/16462	2447/254210	0.0061
GO:0002376	Immune system process	Biological process	832/16462	2638/254210	0.00648
GO:0040007	Growth	Biological process	1006/16462	2703/254210	0.00697
GO:0000003	Reproduction	Biological process	1064/16462	3014/254210	0.00774
GO:0048519	Negative regulation of biological process	Biological process	1480/16462	4449/254210	0.01151
GO:0048518	Positive regulation of biological process	Biological process	1887/16462	5286/254210	0.01403
GO:0051234	Establishment of localization	Biological process	2031/16462	5838/254210	0.01544
GO:0051179	Localization	Biological process	2284/16462	6713/254210	0.01771
GO:0050896	Response to stimulus	Biological process	2526/16462	7194/254210	0.01917

Subset: Number of SUSFLECK genes in the particular GO term/total number of SUSFLECK genes
Total set: Number of porcine genes in the particular GO term/total number of porcine genes

and "catalytic activity" (Molecular Function). Moreover, "organelle" of the Cellular Component category and "gene expression" of the Biological Process category were underrepresented. These findings were more informative than were the results on fat reported by another group (Gorodkin et al., 2007). This group analyzed the porcine transcriptome in fat with 6,783 ESTs and found that fat was overrepresented in "extracellular matrix", which was also overrepresented in this study.

The most abundant transcripts in each non-normalized SUSFLECK library are shown in Table 4. The relative frequency

of sequences for a transcript can represent the level of expression of that transcript for non-normalized libraries (Audic and Claverie, 1997; Stekel et al., 2000; Tuggle et al., 2003). The most redundant transcripts, such as mitochondrial solute carrier family 25 member 6 (SLC25A6), translation elongation factor 1, swine leukocyte antigen (SLA), and vimentin, have been found in porcine backfat tissue (Kim et al., 2006). SLC25A6 is one of three isoforms of the ADP/ATP carrier protein (AAC) located in the inner mitochondrial membrane. It is expressed at lower levels in the brain, lung, kidney, liver, pancreas, heart, skeletal muscle, spleen, and thymus in humans, but is very abundant in proliferating cells (Palmieri, 2004). Moreover, it was highly expressed at the blastocyst stage during early embryo development in pigs (Lee et al., 2006a). In this study, SLC25A6 was shown to be markedly up-regulated in fat tissues. This suggests that high-level expression of SLC25A6 in porcine fat tissues resulted from marked cell proliferation and high energy metabolism in the adipose tissue. Therefore, the level of SLC25A6 transcript in fat cells induced from preadipocytes was double those in backfat and abdominal fat. In addition, cytochrome c Oxidase subunit I (COX1), actin alpha 1 (ACTA1), and swine leukocyte antigen (SLA) showed higher levels of expression in backfat than in abdominal fat or cultured fat cells. COX1 converts arachidonic acid to prostanoids that modulate immune and inflammatory responses (Gilroy et al., 1999; Tilley et al., 2001). Although ACTA1 and vimentin make up the cytoskele-

Table 3. Ten categories of the most frequently occurring domains

Domain	Description	Frequency
zf-C2H2	Zinc finger, C2H2 type	208
WD40	WD domain, G-beta repeat	142
Hormone_1	Somatotropin hormone family	114
Mito_carr	Mitochondrial carrier protein	104
Ank	Ankyrin repeat	96
MHC_I	Class I Histocompatibility antigen, domains	92
V-set	Immunoglobulin V-set domain	92
Fib_beta	Fibrinogen beta chain N terminal	60
RRM_1	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	56
Ras	Ras family	54

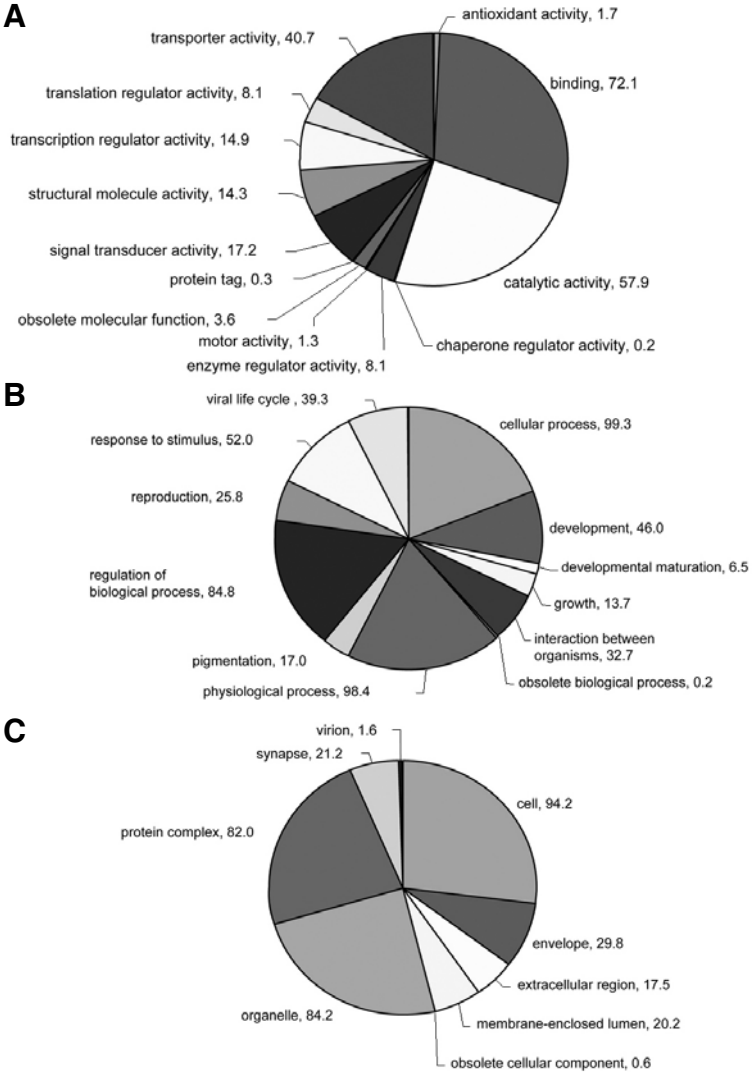

Fig. 1. Gene Ontology annotation of the unique transcripts in SUSFLECK libraries. (A), (B), and (C) indicate the distributions of Molecular Function, Biological Process, and Cellular Component, respectively.

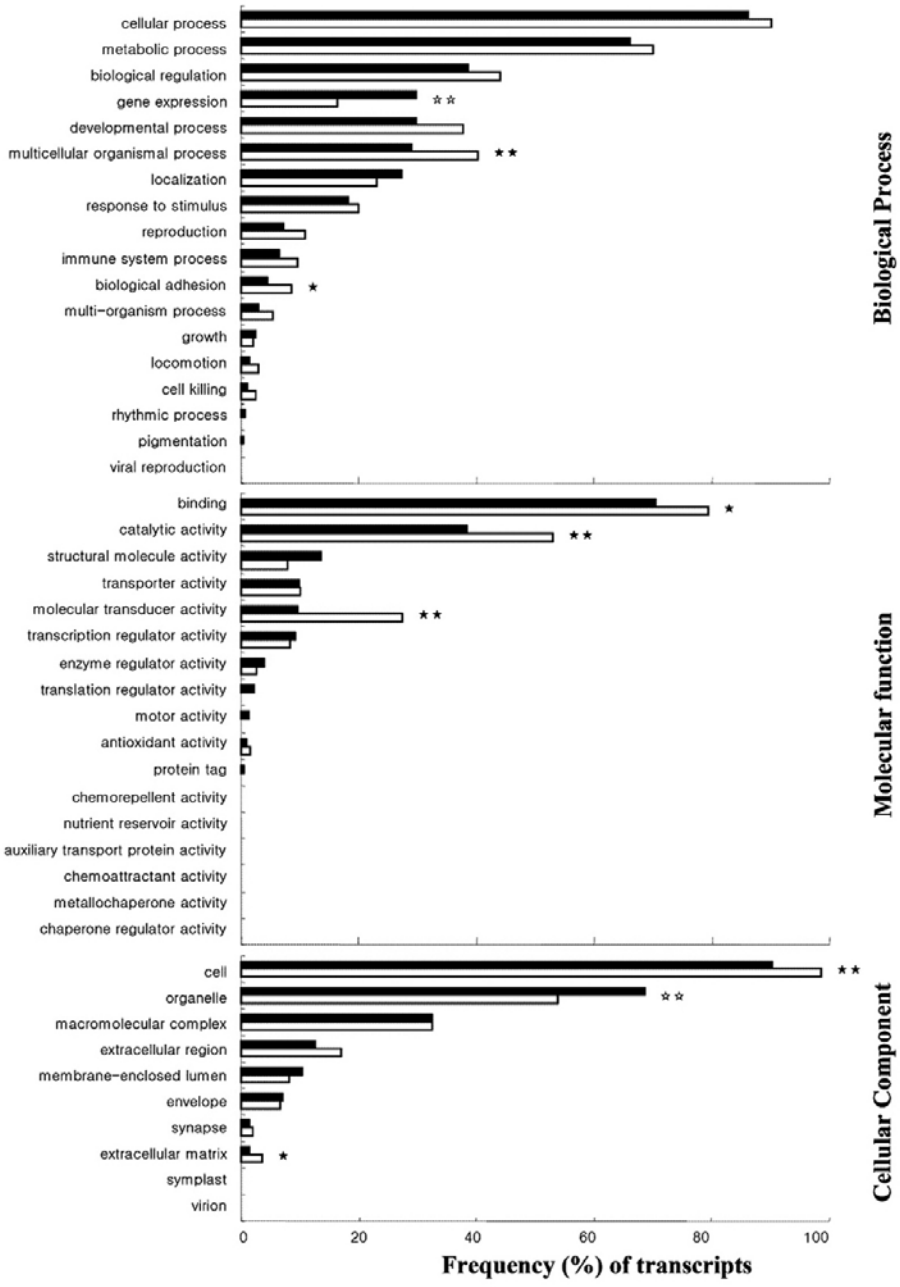

Fig. 2. Gene Ontology distribution of transcripts present exclusively in fat tissue libraries. Black bars represent the levels of expected frequency for the annotated TCs of multiple species. White bars show the observed transcripts levels of fat-related SUSFLECK libraries, which were matched to the annotated TCs of multiple species. To investigate the differences between the observed and expected number of transcripts on the basis of GO terms from the 2nd level to all leaf nodes, Pearson's χ^2 test was used. Black and white stars indicate that the levels of the transcripts of fat-related SUSFLECK libraries were significantly high or low, respectively, in comparison with the annotated TCs of multiple species (Biological Process ★: 0.003; ★★: 0.0006; Molecular Function ★: 0.003, ★★: 0.0006; Cellular Component ★: 0.005, ★★: 0.001).

ton, they showed different transcript levels. Vimentin showed similar high levels of expression in fat tissues, but the level of expression of ACTA1 was high only in backfat. Myo1c, a non-muscle myosin, binds to actin filaments, which control the movement of intracellular GLUT4-containing vesicles to the plasma membrane (Bose et al., 2002). Eukaryotic translation elongation factor 1 gamma (EEF1G) and alpha 1 (EEF1A1), which are essential for protein synthesis, showed tissue-specific expression. The levels of EEF1G expression were relatively low in the liver and pituitary gland, whereas EEF1A1 expression was low in the loin muscle (longissimus dorsi). These observations suggested that these redundancies of EEF1G and EEF1A1 should be involved in the synthesis of enzymes related to metabolism in the adipose tissue. Although it was reported that ribosomal protein L4 (RPL4) would be a good reference gene for highly abundant transcripts (Nygard et al., 2007), the frequency of RPL4 variant was very high in SUSFLECKs of the fat tissues. Thus, the RPL4 gene should be reconfirmed as a reference gene. Fructose-bisphosphate aldolase C (ALDOC) and enolase (ENO1) are glycolytic enzymes frequently expressed in fat tissues. It was reported that ENO1 would serve as a receptor of plasminogen in adipocyte differentiation (Wang et al., 2004). Annexin A2 (ANXA2) is a member of a widely distributed, phospholipid-binding, calcium-regulated, peripheral membrane protein family known as the annexins (Hajjar and Krishnan, 1999). The level of ANXA2 is decreased in small adipocytes, but vimentin was abundant in both large and small adipocytes of fat-specific insulin receptor knock-out mice (FIRKO) (Bluher et al., 2004). These observations suggest that transcripts expressed at high levels in the fat tissues, such as SLC25A6, COX1, ALDOC,

Table 4. Differently expressed genes among non-normalized SUSFLECKs

Gene	Gene description	BF	AF	FC	LM	LV	PG
SLC25A6	Mitochondrial solute carrier family 25 member 6	497	364	821	23	15	55
EEF1G	Eukaryotic translation elongation factor 1 gamma	250	194	302	103	15	63
EEF1A1	Eukaryotic translation elongation factor 1 alpha 1	206	620	541	39	123	283
ACTA1	Actin alpha 1	233	66	0	18	0	0
COX1	Cytochrome c oxidase subunit I	204	5	79	0	0	0
SLA 2	Swine leukocyte antigen 2/2	199	46	5	41	3	2
RPL4	Ribosomal protein L4 variant	165	97	243	0	22	45
ALDOC	Fructose-bisphosphate aldolase C	83	179	2	2	0	0
ANXA2	Annexin A2	81	113	229	0	0	3
PCPE	Procollagen C-proteinase enhancer	125	126	145	0	1	5
VIM	Vimentin isoform 1	119	128	112	0	0	2
ENO	enolase	89	88	168	0	3	6
EDG1	Endothelial differentiation, sphingolipid G protein-coupled receptor	34	16	2	828	0	0
FGB	Fibrinogen beta chain	1	0	0	0	667	5
GH	Growth hormone	0	0	0	0	27	1449

BF, back fat; AF, abdominal subcutaneous fat; FC, cultured fat cell; LM, loin muscle; LV, liver; PG, pituitary gland
The number of transcripts per 10,000 was calculated for comparison of expression level.

ANXA2, ENO1, and EEF1, may be necessary for cytoskeleton and protein synthesis. Further studies are required to elucidate the functional dynamics of these genes in adipocytes. The results of GO analysis and gene redundancy in full-length enriched cDNA libraries may represent direct or indirect evidence with regard to gene networks in tissues. These EST sequences with fullness of approximately 64% contributed to the international cDNA sequencing effort and will provide invaluable information for annotation of the porcine genome.

Note: Supplementary information is available on the Molecules and Cells website (www.molcells.org).

ACKNOWLEDGMENTS

This work was supported by a grant from the National Institute of Animal Science (Research title: Gene expression profiling and functional analysis from fat tissues in pig) (RIMS No. 20051390668000001).

REFERENCES

Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., et al. (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. Science *252*, 1651-1656.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389-3402.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. *25*, 25-29.

Audic, S., and Claverie, J.M. (1997). The significance of digital gene expression profiles. Genome Res. *7*, 986-995.

Ballard, F.J., Hanson, R.W., and Kronfeld, D.S. (1969). Gluconeogenesis and lipogenesis in tissue from ruminant and nonruminant animals. Fed. Proc. *28*, 218-231.

Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L., et al. (2004). The Pfam protein families database. Nucleic Acids Res. *32*, D138-141.

Bluher, M., Wilson-Fritch, L., Leszyk, J., Laustsen, P.G., Corvera, S., and Kahn, C.R. (2004). Role of insulin action and cell size on protein expression patterns in adipocytes. J. Biol. Chem. *279*, 31902-31909.

Bonferroni, C.E. (1937). Teoria statistica delle classi e calcolo delle probabilit?, Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze *8*, 3-62.

Bose, A., Guilherme, A., Robida, S.I., Nicoloro, S.M., Zhou, Q.L., Jiang, Z.Y., Pomerleau, D.P., and Czech, M.P. (2002). Glucose transporter recycling in response to insulin is facilitated by myosin Myo1c. Nature *420*, 821-824.

Chen, C.H., Lin, E.C., Cheng, W.T., Sun, H.S., Mersmann, H.J., and Ding, S.T. (2006). Abundantly expressed genes in pig adipose tissue: an expressed sequence tag approach. J. Anim. Sci. *84*, 2673-2683.

Ewing, B., and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. *8*, 186-194.

Fang, M., Hu, X., Jiang, T., Braunschweig, M., Hu, L., Du, Z., Feng, J., Zhang, Q., Wu, C., and Li, N. (2005). The phylogeny of Chinese indigenous pig breeds inferred from microsatellite markers. Anim. Genet. *36*, 7-13.

Fujisaki, S., Sugiyama, A., Eguchi, T., Watanabe, Y., Hiraiwa, H., Honma, D., Saito, T., and Yasue, H. (2004). Analysis of a full-length cDNA library constructed from swine olfactory bulb for elucidation of expressed genes and their transcription initiation sites. J. Vet. Med. Sci. *66*, 15-23.

Gilroy, D.W., Colville-Nash, P.R., Willis, D., Chivers, J., Paul-Clark, M.J., and Willoughby, D.A. (1999). Inducible cyclooxygenase may have anti-inflammatory properties. Nat. Med. *5*, 698-701.

Gordon, D., Abajian, C., and Green, P. (1998). Consed: a graphical tool for sequence finishing. Genome Res. *8*, 195-202.

Gorodkin, J., Cirera, S., Hedegaard, J., Gilchrist, M.J., Panitz, F., Jorgensen, C., Scheibye-Knudsen, K., Arvin, T., Lumholdt, S., Sawera, M., et al. (2007). Porcine transcriptome analysis based on 97 non-normalized cDNA libraries and assembly of 1,021,891 expressed sequence tags. Genome Biol. *8*, R45.

Hajjar, K.A., and Krishnan, S. (1999). Annexin II: a mediator of the plasmin/plasminogen activator system. Trends Cardiovasc. Med. *9*, 128-138.

Harris, M.A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C., et al. (2004). The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. *32*, D258-261.

Huang, X., and Madan, A. (1999). CAP3: A DNA sequence assem-

bly program. Genome Res. *9*, 868-877.
Huff-Lonergan, E., Baas, T.J., Malek, M., Dekkers, J.C., Prusa, K., and Rothschild, M.F. (2002). Correlations among selected pork quality traits. J. Anim. Sci. *80*, 617-627.
Humphray, S.J., Scott, C.E., Clark, R., Marron, B., Bender, C., Camm, N., Davis, J., Jenks, A., Noon, A., Patel, M., et al. (2007). A high utility integrated map of the pig genome. Genome Biol. *8*, R139.
Kato, S., Sekine, S., Oh, S.W., Kim, N.S., Umezawa, Y., Abe, N., Yokoyama-Kobayashi, M., and Aoki, T. (1994). Construction of a human full-length cDNA bank. Gene *150*, 243-250.
Kim, T.H., Kim, N.S., Lim, D., Lee, K.T., Oh, J.H., Park, H.S., Jang, G.W., Kim, H.Y., Jeon, M., Choi, B.H., et al. (2006). Generation and analysis of large-scale expressed sequence tags (ESTs) from a full-length enriched cDNA library of porcine backfat tissue. BMC Genomics *7*, 36.
Kim, H., Park, T.S., Lee, W.K., Moon, S., Kim, J.N., Shin, J.H., Jung, J.G., Lee, S.D., Park, S.H., Park, K.J., et al. (2007). MPSS profiling of embryonic gonad and primordial germ cells in chicken. Physiol. Genomics *29*, 253-259.
Lee, H.Y., Cui, X.S., Lee, K.A., and Kim, N.H. (2006a). Annealing control primer system identifies differentially expressed genes in blastocyst-stage porcine parthenotes. Zygote *14*, 71-80.
Lee, M.H., Lee, S.H., Kim, H., Jin, J.B., Kim, D.H., and Hwang, I. (2006b). A WD40 repeat protein, Arabidopsis Sec13 homolog 1, may play a role in vacuolar trafficking by controlling the membrane association of AtDRP2A. Mol. Cells *22*, 210-219.
Lunney, J.K. (2007). Advances in swine biomedical model genomics. Int. J. Biol. Sci. *3*, 179-184.
Maruyama, K., and Sugano, S. (1994). Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. Gene *138*, 171-174.
Mikawa, A., Suzuki, H., Suzuki, K., Toki, D., Uenishi, H., Awata, T., and Hamasima, N. (2004). Characterization of 298 ESTs from porcine back fat tissue and their assignment to the SSRH radiation hybrid map. Mamm. Genome *15*, 315-322.
Moon, H.S., and Chung, C.S. (2004). Effect of isomers of conjugated linoleic acid on porcine preadipocyte differentiation. J. Anim. Sci. Technol. *46*, 967-974.
Mulder, N.J., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bradley, P., Bork, P., Bucher, P., Cerutti, L., et al. (2005). InterPro, progress and status in 2005. Nucleic Acids Res *33*, D201-205.
Nafikov, R.A., and Beitz, D.C. (2007). Carbohydrate and lipid metabolism in farm animals. J. Nutr. *137*, 702-705.
Nygard, A.B., Jorgensen, C.B., Cirera, S., and Fredholm, M. (2007). Selection of reference genes for gene expression studies in pig tissues using SYBR green qPCR. BMC Mol. Biol. *8*, 67.
Oh, J.H., Kim, Y.S., and Kim, N.S. (2003). An improved method for constructing a full-length enriched cDNA library using small amounts of total RNA as a starting material. Exp. Mol. Med. *35*, 586-590.
Oh, J.H., Sohn, H.Y., Kim, J.M., Kim, Y.S., and Kim, N.S. (2004). Construction of multi-purpose vectors, pCNS and pCNS-D2, are suitable for collection and functional study of large-scale cDNAs. Plasmid *51*, 217-226.
Palmieri, F. (2004). The mitochondrial transporter family (SLC25): physiological and pathological implications. Pflugers Arch. *447*, 689-709.
Quackenbush, J., Liang, F., Holt, I., Pertea, G., and Upton, J. (2000). The TIGR gene indices: reconstruction and representation of expressed gene sequences. Nucleic Acids Res. *28*, 141-145.
Soares, M.B., Bonaldo, M.F., Jelene, P., Su, L., Lawton, L., and Efstratiadis, A. (1994). Construction and characterization of a normalized cDNA library. Proc. Natl. Acad. Sci. USA *91*, 9228-9232.
Stapleton, M., Carlson, J., Brokstein, P., Yu, C., Champe, M., George, R., Guarin, H., Kronmiller, B., Pacleb, J., Park, S., et al. (2002). A Drosophila full-length cDNA resource. Genome Biol. *3*, RESEARCH0080.
Stekel, D.J., Git, Y., and Falciani, F. (2000). The comparison of gene expression from multiple cDNA libraries. Genome Res. *10*, 2055-2061.
Suryawan, A., Swanson, L.V., and Hu, C.Y. (1997). Insulin and hydrocortisone, but not triiodothyronine, are required for the differentiation of pig preadipocytes in primary culture. J. Anim. Sci. *75*, 105-111.
Suzuki, Y., and Sugano, S. (2001). Construction of full-length-enriched cDNA libraries. The oligo-capping method. Methods Mol. Biol. *175*, 143-153.
Tilley, S.L., Coffman, T.M., and Koller, B.H. (2001). Mixed messages: modulation of inflammation and immune responses by prostaglandins and thromboxanes. J. Clin. Invest. *108*, 15-23.
Trayhurn, P., and Wood, I.S. (2005). Signalling role of adipose tissue: adipokines and inflammation in obesity. Biochem. Soc. Trans. *33*, 1078-1081.
Tuggle, C.K., and Schmitz, C.B. (1994). Cloning and characterization of pig muscle cDNAs by an expressed sequence tag approach. Anim. Biotechnol. *5*, 1-13.
Tuggle, C.K., Green, J.A., Fitzsimmons, C., Woods, R., Prather, R.S., Malchenko, S., Soares, B.M., Kucaba, T., Crouch, K., Smith, C., et al. (2003). EST-based gene discovery in pig: virtual expression patterns and comparative mapping to human. Mamm. Genome *14*, 565-579.
Uenishi, H., Eguchi, T., Suzuki, K., Sawazaki, T., Toki, D., Shinkai, H., Okumura, N., Hamasima, N., and Awata, T. (2004). PEDE (Pig EST Data Explorer): construction of a database for ESTs derived from porcine full-length cDNA libraries. Nucleic Acids Res. *32*, D484-488.
Uenishi, H., Eguchi-Ogawa, T., Shinkai, H., Okumura, N., Suzuki, K., Toki, D., Hamasima, N., and Awata, T. (2007). PEDE (Pig EST Data Explorer) has been expanded into Pig Expression Data Explorer, including 10 147 porcine full-length cDNA sequences. Nucleic Acids Res. *35*, D650-653.
van Wijk, H.J., Arts, D.J., Matthews, J.O., Webster, M., Ducro, B.J., and Knol, E.F. (2005). Genetic parameters for carcass composition and pork quality estimated in a commercial production chain. J. Anim. Sci. *83*, 324-333.
Wang, P., Mariman, E., Keijer, J., Bouwman, F., Noben, J.P., Robben, J., and Renes, J. (2004). Profiling of the secreted proteins during 3T3-L1 adipocyte differentiation leads to the identification of novel adipokines. Cell Mol. Life Sci. *61*, 2405-2417.
Wernersson, R., Schierup, M.H., Jorgensen, F.G., Gorodkin, J., Panitz, F., Staerfeldt, H.H., Christensen, O.F., Mailund, T., Hornshoj, H., Klein, A., et al. (2005). Pigs in sequence space: a 0.66X coverage pig genome survey based on shotgun sequencing. BMC Genomics *6*, 70.
Wintero, A.K., Fredholm, M., and Davies, W. (1996). Evaluation and characterization of a porcine small intestine cDNA library: analysis of 839 clones. Mamm. Genome *7*, 509-517.
Zhong, S., Storch, K.F., Lipan, O., Kao, M.C., Weitz, C.J., and Wong, W.H. (2004). GoSurfer: a graphical interactive tool for comparative analysis of large gene sets in gene ontology space. Appl. Bioinformatics *3*, 261-264.