

Identification of 1,531 cSNPs from Full-length Enriched cDNA Libraries of the Korean Native Pig Using *in Silico* Analysis

Younshin Oh, Dinh Truong Nguyen, Kwangha Park, Vijaya R. Dirisala, Hojun Choi and Chankyu Park*

Department of Animal Biotechnology, Konkuk University, Seoul 143-701, Korea

Abstract

Sequences from the clones of full-length enriched cDNA libraries serve as valuable resources for functional genomics related studies, genome annotation and SNP discovery. We analyzed 7,392 high-quality chromatograms (Phred value ≥ 30) obtained from sequencing the 5' ends of clones derived from full-length enriched cDNA libraries of Korean native pigs including brainstem, liver, cerebellum, neocortex and spleen libraries. In addition, 50,000 EST sequence trace files obtained from GenBank were combined with our sequences to identify cSNPs *in silico*. The process generated 11,324 contigs, of which 2,895 contigs contained at least one SNP and among them 610 contigs had a minimum of one sequence from Korean native pigs. Of 610 contigs, we randomly selected 262 contigs and performed *in silico* analysis for the identification of cSNPs. From the results, we identified 1,531 putative coding single nucleotide polymorphisms (cSNPs) and the SNP detection frequency was one SNP per 465 bp. A large-scale sequencing result of clones from full-length enriched cDNA libraries and identified cSNPs will serve as a useful resource to functional genomics related projects such as a pig HapMap project in the near future.

Keywords: cSNP, full-length cDNA library, functional genomics, Korean native pigs, pigs

Introduction

The pig, one of the first animals domesticated over 7000 years ago, is the major meat supply resource consumed worldwide (Rothschild, 2003). Further more, the porcine has served as sensitive and comprehensive tools to address biomedical, agricultural and fundamental biological

questions (Tuggle *et al.*, 2007). To better comprehend the physiological complexity of the pig genome, a functional gene analysis needs to be accomplished.

Single nucleotide polymorphisms (SNPs) are the most frequent form of DNA variations in the genome of organisms and useful for association studies, gene mapping and phylogenetic studies (Brookes, 1999; Dimmic *et al.*, 2005; Kim, *et al.*, 2007; Kim, *et al.*, 2008; Komar, 2007). They provide genomic landmarks of past events shaping individuals, breeds and species (Lee *et al.*, 2006a). Owing to high abundance and stability of SNPs in the genome, thousands of potentially informative SNP markers can be discovered and used to develop high-density SNP maps (Panitz *et al.*, 2007; Zimdahl *et al.*, 2004). These maps are an essential resource for the identification of genes that are responsible for variations in complex traits or quantitative trait loci (QTL; Glazier *et al.*, 2002; Rothschild *et al.*, 2007).

SNPs are classically identified by direct sequencing of amplicons of candidate genes from a set of individuals that represent the diversity in the interested populations. The draw backs of this approach are laborious and expensive (Useche *et al.*, 2001). A fast and cost-efficient approach to identify a multitude of novel SNPs involving in mining sequence data from public repositories needed to be developed. The improvement of sequence trace data in public repositories at the DNA chromatogram level was required (Panitz *et al.*, 2007).

An alternative and cheaper method of SNP identification exploits the redundancy of gene sequences generated by expressed sequence tags (ESTs; Gu *et al.*, 1998). - In this scenario, each SNP would also be associated with an expressed gene. EST information has been used for the detection of SNPs in mammalian genomes by many groups (Buetow *et al.*, 1999; Fitzsimmons *et al.*, 2004; Garg *et al.*, 1999; Guryev *et al.*, 2004; Hawken *et al.*, 2004; Kim *et al.*, 2003; Lee *et al.*, 2006a; Lee *et al.*, 2006b; Picoult-Newberg, *et al.*, 1999).

Many conventional cDNA libraries have been constructed with porcine material. As of May 2009, 1,532,429 ESTs were available in the GenBank porcine EST database and 167,740 SNPs were achievable in the GenBank pig SNP database (Table 1). Several researchers have reported mining cSNPs from porcine EST sequences (Dirisala *et al.*, 2005, 2007; Fahrrenkrug *et al.*, 2002; Grapes *et al.*, 2006; Kollers *et al.*, 2005; Panitz *et al.*, 2007; Park *et al.*, 2009; Uenishi *et al.*,

*Corresponding author: E-mail chankyu@konkuk.ac.kr
Tel +82-2-450-3697, Fax +82-2-457-8488
Accepted 9 June 2008

Table 1. The current status of reported SNPs from 6 animal species based on the dbSNP database at NCBI

Species	dbSNP build	No. of submissions	No. of RefSNP ¹ clusters
<i>Bos taurus</i>	128	2,485,282	2,457,983
<i>Canis familiaris</i>	126	3,527,071	3,301,322
<i>Equus caballus</i>	128	1,163,580	1,163,580
<i>Homo sapiens</i>	130	79,485,538	17,804,034
<i>Mus musculus</i>	128	18,667,322	14,380,528
<i>Sus scrofa</i>	128	167,740 ²	8,424

Data source: http://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary.cgi

¹Reference SNP.

²Our new 1,529 cSNPs were not included.

2004). Recently, full-length enriched cDNA libraries have also been constructed (Chen *et al.*, 2006; Dirisala *et al.*, 2005; Fujisaki *et al.*, 2004; Kim *et al.*, 2006; Uenishi *et al.*, 2007; Uenishi *et al.*, 2004). Although the number of ESTs has increased significantly, less progress has been made in analyzing this large volume of emerging data, and a multitude of markers is necessary for the identification of economically important traits in pigs (Kollers *et al.*, 2005).

A large-scale cSNP mining from EST sequences has been reported by the Sino-Danish pig genome sequencing initiative (Panitz *et al.*, 2007), these cSNPs were analyzed using cDNA sequences constructed by conventional methods. However, sequences from full-length enriched cDNA libraries serve as more valuable resources for functional genomics studies than sequences from conventional cDNA libraries as they extend further towards the 5' end and represent a critical tool for improving the quality of many genomic annotation parameters (Das *et al.*, 2001; Haas *et al.*, 2002). In spite of that, data from the analysis of full-length enriched cDNA libraries constructed using pig tissues and cells are still limited (Chen *et al.*, 2006; Dirisala *et al.*, 2005, 2007; Fujisaki *et al.*, 2004; Kim *et al.*, 2006; Uenishi *et al.*, 2004, 2007).

In this report, we have evaluated the process of cSNP discovery using the five full-length enriched cDNA libraries of Korean native pigs from cerebellum, liver, brainstem, neocortex and spleen libraries as a diverse genetic resource. In the result, 1,531 new cSNPs from 262 contigs were identified from the *in silico* analysis of the sequences from these libraries and the trace data files from the public database.

Methods

Tissue collection

The Korean native pig is a medium-size pig with black coat color originating in the Korean peninsula (Kim *et al.*, 2005). Neocortex, brainstem, liver, cerebellum, and spleen tissues were dissected from Korean native pigs that were maintained at the Jeju livestock experiment station in Jeju, Korea. The tissues were snap frozen in liquid nitrogen and stored at -80°C until use.

Construction of full-length enriched cDNA libraries

The neocortex, cerebellum and spleen libraries were constructed using a SMART kit (Clontech Laboratories, Mountain View, CA, USA). The brainstem and liver library were constructed by the modified oligo-capping and modified vector-capping method (Dirisala *et al.*, 2005, 2007), respectively. Plasmid DNA was prepared in the 96-well plate format using multi-well filter plates (Pall Corporation, Greenvale, NY, USA).

Sequencing of full-length enriched cDNA libraries

Sequencing of cloned cDNA inserts was performed using an ABI PRISM Bigdye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions, and resultant sequences were analyzed on ABI 3700 automated sequencers.

SNP identification

The 7,392 EST sequences with high-quality chromatograms were transferred into the Linux workstation. The method for bioinformatic analysis of full-length cDNA clones were performed as previously described (Park *et al.*, 2009). Briefly, EST sequence trace files were base called using Phred (Ewing & Green, 1998a; Ewing *et al.*, 1998b) with the options trim_alt and trim_phd. Sequences assessed for having a Phred quality score higher than 30 in 20 bp windows were subjected to Cross_match (P. Green, Unpublished) for vector trimming. RepeatMasker was used for removing repeat sequences (<http://repeatmasker.genome.washington.edu/>; A.F.A. Smit and P. Green, Unpublished). The processed sequences were assembled into contigs by Phrap (<http://bozeman.mbt.washington.edu/phrap.docs/phrap.html>), integrated with Polyphred (<http://droog.mbt.washington.edu/PolyPred.html>) for finding SNP candi-

date sites and viewed with the Consed program (<http://bozeman.mbt.washington.edu/consed/consed.html>) using the polymorphism option (Gordon *et al.*, 1998). The polymorphic sites were identified, and the results were confirmed in individual chromatograms. The new cSNPs were determined by comparing our data with NCBI BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>) and the previously published paper (Park *et al.*, 2009).

Results

Characteristics of cerebellum and liver libraries

Library titers for the cerebellum and liver libraries were 5.0×10^5 pfu and 6.5×10^4 cfu, respectively. The average cDNA insert size estimated by restriction analysis of 20 clones picked up randomly from each library was 1.7 kb and 1.1 kb for the cerebellum and liver libraries, respectively. The analysis of other characteristics of libraries was performed as previously described (Dirisala *et al.*, 2005, 2007).

A total of the 7,392 clones from five full-length enriched libraries (4,416 from cerebellum, 288 from liver, 1,056 from brainstem, 672 from spleen, 960 from neocortex) randomly picked into 96 well plates was sequenced from the only 5' end direction. Although it is debatable to set the criteria for full-length clones, we used the presence of translation initiation codon 'ATG' in the cDNA sequences as the definition. According to our definition, the percentage of full-length clones was estimated to be 50% and 75% for the cerebellum and the liver libraries, respectively (data not shown). The percentage of full length clones for neocortex, brain stem and spleen was 40%, 80% and 60%, respectively (Dirisala *et al.*, 2005, 2007).

Since these libraries were specifically constructed for generating full-length clones, 65% of clones had longer 5' end sequences than currently available EST sequences in the GenBank EST database. To further eval-

uate the characteristics of five libraries, 610 contigs containing at least one sequence of the Korean native pig was selected and analyzed (Table 2). The result showed that 98% of sequences were matched with sequences of either pigs or other species with an E value $< 10^{-100}$.

Identification of putative csnp's using a bioinformatics approach

We analyzed cDNA clones from five libraries where the cerebellar and liver libraries were constructed in this study and the neocortex, spleen and brainstem libraries were from our previous study (Dirisala *et al.*, 2005, 2007) by 5' end sequencing, and a total of 7,392 high-quality chromatograms (Phred value ≥ 30) were obtained from sequencing the 5' ends of clones. Furthermore, 50,000 chromatograms of porcine EST sequences were retrieved from the GenBank trace file archive, combined with the sequencing results from the cDNA libraries and analyzed for the presence of cSNPs.

The process yielded a total of 11,324 contigs, and 2,895 out of 11,324 contigs had cSNPs (data not shown). As we were interested in identifying cSNPs using Korean native pig sequences, only 610 contigs which had overlapping regions contributed by at least one sequence each from GenBank and Korean native pig were chosen for further analysis. The remaining contigs assembled only with Genbank sequences were excluded from our analysis. Of 610 contigs, only a part of contigs (262) were randomly selected for the discovery of cSNPs in this study due to the limitation from the manual confirmation of generated data through bioinformatic analysis. The analysis yielded the discovery of 1,533 cSNPs including 2 previously reported among the contigs containing Korean native pig sequences.

We analyzed the contribution in number of cDNA sequences for each contig used in cSNP discovery (Table 3). A total of 59 contigs were consisted of both one Korean native pig and one GenBank sequences which is the most abundant type. Out of 610 contigs, 182 were from the group with one Korean native pig sequence and at least one GenBank sequence. One hundred three contigs contained two Korean native pig sequences and at least one public EST sequence.

Table 2. The functional characteristics of contigs used for SNP discovery using bioinformatic analysis from cerebellar, liver, neocortex, brainstem and spleen cDNA libraries

Blast match	No. of clones (%)
Non redundant (nr)	506 (83.0)
Pig	208 (41.1)
Mitochondrial transcripts	13 (2.6)
Genomic sequence	22 (4.3)
Other species	263 (52.0)
EST	93 (15.2)
Non	11 (1.8)
Total	610 (100)

Discussion

Our primary goal for this study was to discover new cSNPs for pigs and to test the efficiency of cSNP identification by mixing cDNA sequences of full-length enriched cDNA libraries constructed from Korean native pigs and GenBank sequences. The application of our

Table 3. The pattern of contribution in number of sequences for 610 contigs used for cSNP discovery from Korean native pigs and GenBank sequences

No. of KNP ¹ sequences	NO. of GenBank sequences												Total
	0	1	2	3	4	5	6	7	8	9	10	Others	
1	-	59 ²	36 ²	26 ²	15 ²	17	17	6	1	5	-	-	182
2	25 ²	21 ²	17 ²	9 ²	8	12	3	4	4	-	-	-	103
3	9 ²	14 ²	7 ²	6	2	5	1	1	-	-	-	-	45
4	11 ²	10 ²	4	2	2	2	2	-	-	-	-	-	33
5	3 ²	6	3	2	2	3	-	-	-	-	-	-	19
6	4	3	4	3	3	-	-	-	-	-	-	-	17
7	2	0	0	2	-	-	-	-	-	-	-	-	4
8	3	1	0	-	-	-	-	-	-	-	-	-	4
9	3	0	-	-	-	-	-	-	-	-	-	-	3
10	2	-	-	-	-	-	-	-	-	-	-	-	2
Others	-	-	-	-	-	-	-	-	-	-	-	198 ³	198
Total	62	114	71	50	32	39	23	11	5	5	-	198	610

¹The KNP is Korean native pig.

²Contigs used for the discovery of 1,529 cSNPs by *in silico* analysis.

³198 contigs were consisted of more than 10 sequences per contig.

study using the cDNA sequences of Korean native pigs as a source of genetic diversity in the porcine genome can provide additional resources for the identification of cSNPs. The genetic relationship between Korean native pigs and other pig breeds was expected to be rather distant (Porter, 1993). Thus, we predicted that there would be a higher possibility of identifying genetic differences between Korean native pigs and other breeds than between the commonly used breeds alone. The use of full-length enriched cDNA libraries also allowed to detect SNPs in the promoter regions of genes (Dirisala *et al.*, 2005, 2007; Park *et al.*, 2009)

The observed differences in SNP detection frequency arising from these reports might reflect genetic diversity within species and/or differences in application methods. In humans, the SNP detection frequency is reported to be about 1 SNP per 700 bp across the whole genome (The International SNP Map Working Group, 2001), and in some regions it is as high as 1 SNP per 300 bp (Wang *et al.*, 1998). In chickens, the frequency of putative SNPs was variously reported to be 1 SNP per 2,119 bp (Kim *et al.*, 2003) to 1 SNP per 1,900 bp (Fitzsimmons *et al.*, 2004). In Korean native cattle, the SNP detection frequency of 1 SNP per 787 bp has been reported (Lee *et al.*, 2006b). In our previous study, in which a small number of Korean native pig sequences and a relatively high stringency were used in selecting the candidate sequences, an *in silico* method yielded 1 SNP per 614 bp (Park *et al.*, 2009). The results are very similar to our results in this study.

The removal of false SNPs from a large cSNP dataset

generated from a large scale cDNA sequencing project is a difficult process since the most reliable method for the evaluation of SNPs is experimental confirmation based on animals genotyping against genetically diverse populations. One possible idea to solve this problem is to exploit the redundancy of cSNPs with existing EST sequences. If we can develop an algorithm to select the SNPs from polymorphisms existing in multiple times in the public database and evaluate some statistical values for the probability, it will greatly help to utilize this type of *in silico* approach for SNP discovery.

Acknowledgements

This work was supported by grants from the Technology Development Program for Agriculture and Forestry, Ministry for Agriculture, Forestry and Fisheries and the BioGreen21 program (20070401-034-029-010), Rural Development Administration, Republic of Korea.

References

- Barker, G., Batley, J., Sullivan, H.O., Edwards, K.J., and Edwards, D. (2002). Redundancy based detection of sequence polymorphisms in expressed sequence tag data using autoSNP. *Bioinformatics* 19, 421-422.
- Brookes, A.J. (1999). The essence of SNPs. *Gene* 234, 177-186.
- Buetow, K.H., Edmonson, M.N., and Cassidy, A.B. (1999). Reliable identification of large number of candidate SNPs from public EST data. *Nat. Genet.* 21, 323-325.
- Chen, C.H., Lin, E.C., Cheng, W.T.K., 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.
- Dimmic, M.W., Sunyaev, S., and Bustamante, C. (2005). Inferring SNP function using evolutionary, structural and computational methods. *Pac. Symp. Biocomput.* 10, 382-384.
- Dirisala, V.R., Kim, J., Park, K., Kim, N., Lee, K.T., Oh, S.J., Oh, J.H., Kim, N.S., Um, S.J., Lee, H.T., Kim, K.I., and Park, C. (2005). cSNP mining from full-length enriched cDNA libraries of the Korean native pig. *Kor. J. Genet.* 27, 329-335.
- Dirisala, V.R., Kim, J., Park, K., Lee, H.T., and Park, C. (2007). Discovery of cSNPs in Pig Using Full-length Enriched cDNA Libraries of the Lorean Native Pig as a source of Genetic Diversity. *BBE*, 12, 424-432
- Ewing, B., and Green, P. (1998a). Base calling of automated sequencing tracers using phred. II. Error probabilities. *Genome Res.* 8, 186-194.
- Ewing, B., Hillier, L., Wendl, M., and Green, P. (1998b). Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Res.* 8, 175-185.
- Fahrenkrug, S.C., Freking, B.A., Smith, T.P.L., Rohrer, G.A., and Keele, J.W. (2002). Single nucleotide polymorphism (SNP) discovery in porcine expressed genes. *Anim. Genet.* 33, 186-195.
- Fitzsimmons, C.J., Savolainen, P., Amini, B., Hjalms, G., Lunderberg, J., and Andersson, L. (2004). Detection of sequence polymorphisms in red junglefowl and white leg-horn ESTs. *Anim. Genet.* 35, 391-396.
- 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.
- Garg, K., Green, P., and Nickerson, D.A. (1999). Identification of candidate coding region single nucleotide polymorphisms in 165 human genes using assembled expressed sequence tags. *Genome Res.* 9, 1087-1092.
- Glazier A.M., Nadeau J.H., and Aitman, T.J. (2002). Finding genes that underlie complex traits. *Science* 298, 2345-2349.
- Gordon, D., Abajian, C., and Green, P. (1998). Consed: a graphical tool for sequence finishing. *Genome Res.* 8, 195-202.
- Grapes, L., Rudd, S., Fernando, R.L., Megy, K., Rocha, D., and Rothschild, M.F. (2006). Prospecting for pig single nucleotide polymorphisms in the human genome: have we struck gold? *J. Anim. Breed. Genet.* 123, 145-151.
- Gu, Z., Hillier, L., and Kwok, P.Y. (1998). Single-nucleotide polymorphism hunting in cyberspace. *Hum. Mutat.* 12, 221-225.
- Guryev, V., Berezikov, E., Malik, R., Plasterk, R.H., and Cuppen, E. (2004). Single nucleotide polymorphisms associated with rat expressed sequences. *Genome Res.* 14, 1438-1443.
- Hawken, R.J., Barris, W.C., McWilliam, S.M., and Dalrymple, B.P. (2004). An interactive bovine *in silico* SNP database (IBISS). *Mamm. Genome* 15, 819-827.
- Kim, H., Schmidt, C.J., Decker, K.S., and Emará, M.G. (2003). A double-screening method to identify reliable candidate non-synonymous SNPs from chicken EST data. *Anim. Genet.* 34, 249-254.
- Kim, J.H., Yim, S.H., Jeong, Y.B., Jung, S.H., Xu, H.D., Shin, S.H., and Chung, Y.J. (2008). Comparison of Normalization Methods for Defining Copy Number Variation Using Whole-genome SNP Genotyping Data. *G&I* 6, 231-234.
- Kim, T.H., Kim, K.S., Choi, B.H., Yoon, D.H., Jang, G.W., Lee, K.T., Chung, H.Y., Lee, H.Y., Park, H.S., and Lee, J.W. (2005). Genetic structure of pig breeds from Korea and China using microsatellite loci analysis. *J. Anim. Sci.* 83, 2255-2263.
- 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., Lee, H.Y., Chung, H.Y., and Kim, H. (2006). Generation and analysis of large-scale expressed sequence tags (ESTs) from a full-length enriched cDNA library of porcine back-fat tissue. *BMC Genomics* 7, 36.
- Kim, Y.H., and Kim, H. (2007). Application of Random Forests to Association Studies Using Mitochondrial Single Nucleotide Polymorphisms. *G&I* 5, 168-173.
- Kollers, S., Megy, K., and Rocha, D. (2005). Analysis of public single nucleotide polymorphisms in commercial pig populations. *Anim. Genet.* 36, 426-431.
- Komar, A.A. (2007). SNPs, silent but not invisible. *Science* 315, 466-467.
- Lee, M.A., Keane, O.M., Glass, B.C., Manley, T.R., Cullen, N.G., Dodds, K.G., McCulloh, A.F., Morris, C.A., Schreiber, M., Warren, J., Zadissa, A., Wilson T., and McEwan, J.C. (2006a). Establishment of a pipeline to analyse non-synonymous SNPs in *Bos taurus*. *BMC Genomics* 7, 298.
- Lee, S.H., Park, E.W., Cho, Y.M., Lee, J.W., Kim, H.Y., Lee, J.H., Oh, S.J., Cheon, I.C., and Yoon, D.H. (2006b). Confirming single nucleotide polymorphisms from expressed sequence tag datasets derived from three cattle cDNA libraries. *J. Biochem. Mol. Biol.* 39, 183-188.
- Panitz, F., Stengaard, H., Hornshøj, H., Gorodkin, J., Hedegaard, J., Cirera, S., Thomsen, B., Madsen, L.B., Høj, A., Vingborg, R.K., Zahn, B., Wang, X., Wang, X., Wernersson, R., Jørgensen, C.B., Scheibye-Knudsen, K., Arvin, T., Lumholdt, S., Sawera, M., Green, T., Nielsen, B.J., Havgaard, J.H., Brunak, S., Fredholm, M., Bendixen, C. (2007). SNP mining porcine ESTs with MAVIANT, a novel tool for SNP evaluation and annotation. *Bioinformatics* 23, i387-i391.
- Park, K., Dirisala, V.R., Oh, Y., Choi, H., Lee, K.T., Kim, J.H., Lee, H.T., Seo, K.H., and Park, C. (2009). Reporting 678 putative cSNPs from full-length enriched cDNA sequences of the Korean native pig. *J. Anim. Breed Genet.* 126(2), 127-133.
- Picoult-Newberg, L., Idekar, T.E., Pohl, M.G., Taylor, S.L., Donaldson, M.A., Nickerson, D.A., and Boyce-Jacino, M. (1999). Mining SNPs from EST databases. *Genome Res.* 9, 167-174.
- Porter, V. (1993) Pigs, A Handbook to the Breeds of the

- World, Helm information Ltd., UK.
- Rothschild, M.F. (2003). Advances in pig genomics and functional gene discovery. *Comp. Funct. Genom.* 4, 266-270.
- Rothschild, M.F., Hu, Z.L., and Jiang, Z. (2007). Advances in QTL Mapping in pigs. *Int. J. Biol.* 3, 192-197
- Sambrook, J., Fritsch, E., and Maniatis, T. (1989). *Molecular cloning: A laboratory manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA.
- The International SNP Map Working Group. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409, 928-933.
- Tuggle, C.K., Wang, Y., and Couture, O. (2007). Advances in Swine Transcriptomics. *Int. J. Biol. Sci.* 3, 132-152
- Uenishi, H., Eguchi, T., Suzuki, K., Sawazaki, T., Toki, D., Shinkai, H., Okumura, N., Hamasina, N., and Awata, T. (2004). PEDE (Pig EST Data Explorer): Construction of a database for ESTs derived from porcine full-length cDNA libraries. *Nucl. Acids Res.* 32, 484-488.
- 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 10147 porcine full-length cDNA sequences. *Nucleic Acids Res.* 35, D650-D653.
- Useche, F.J., Gao, G., Harafey, M., and Rafalski, A. (2001). High-throughput identification, database storage and analysis of SNPs in EST sequences. *Genome Inform. Ser. Workshop Genome Inform.* 12, 194-203.
- Wang, D.G., Fan, J.B., Siao, C.J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M.S., Shen, N., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T.J., and Lander, E.S. (1998). Large-scale identification, mapping, genotyping of single nucleotide polymorphisms in the human genome. *Science* 280, 1077-1082.
- Zimdahl, H., Nyakatura, G., Brandt, P., Schulz, H., Hummel, O., Fatmann, B., Brett, D., Droege, M., Monti, J., Lee, Y.A., Sun, Y., Zhao, S., Winter, E.E., Pontig, C.P., Chen, Y., Kasprzyk, A., Birney, E., Ganten, D., and Hubner, N. (2004). A SNP map of rat genome generated from cDNA sequences. *Science* 303, 807.

Table S1. The list of 1,529 new cSNPs identified from *in silico* analysis

Acc. No. of blast match	Description	e-value	Sequence variation
EU650787	hydroxyacyl-Coenzyme A dehydrogenase/3-ketoacyl-Coenzyme A thiolase/enoyl-Coenzyme A hydratase (trifunctional protein), alpha subunit	0	AAAGAAAATGGG (G/A) CTGGTTGACC
AY996813	RAB1A, member RAS oncogene family	0	CA (A/G) ATATGGGACACAGCAGGCCA (A/G) GAA CAATCACCT (C/A) CAG (T/C) TACTA(C/T) AGAGG
BC009018	enolase 2, gamma neuronal	0	CCGTGGGAATC (C/T) CACAGTGG
XM_001925729	similar to Gamma-enolase	0	TGGTATCTATGA (A/G) GCCCTGGAGCTA GTGGACCACATC (G/A) ACACCACCATCG CTTTGCCCCCAA (C/T) ATCCTGGAGAATA
XR_027670	similar to lipocalin-interacting membrane receptor	0	TCCTGCCC (C/T) TTCCATCA CAACTGTCCCTC(G/A) TCTTCCTCA
BC103155	stromal cell-derived factor 2	0	GGGGGAAGAC (T/C) GCCACAGTGT
BC112833	phosphatidylserine synthase 1	0	GCA (C/T) CTGGTTGGGCATGGTCGTCTG (T/C) CG
XM_864974	similar to cell adhesion molecule JCAM	8e-122	GAAAGGCACGG (T/G) GGTCTGAGCGTCA
EW070413	rduo412b_c14.y1 duo Sus scrofa cDNA 5', mRNA sequence	5e-111	CGGGAATTGA (C/T) GGGGATAGGG
XM_001924706	similar to asparagine-linked glycosylation 14 homolog	0	TGAAC (T/C) CAATCGAGCTGATC (A/G) GAAC CCGGGAGGTTCA (A/G) CAGTCCTGGCCGTCC
BC151565	intraflagellar transport 57 homolog	0	TATTATCTGAGCTT (C/T) GATCATTCCGGAGAAC
AK225203	suppressor of variegation 4-20 homolog 1	0	CATAATAACTGTAG (G/A) AAGTGACAGTGTACTGA
XM_001113664	similar to single-stranded DNA-binding protein isoform a	2e-138	GCCGCCCCGAGCCCCGTGCT (A/T) GG (A/C) AACA CCC (A/C) (G/A) (G/A) (A/C) GATGG (C/G) AT C (A/G) G (T/G) AGG (T/C) CC (T/C) AT (C/A) C C (A/G) GGGTCTTTTCAGCCTTTTATGTCACC (T/G) C
XM_001167299	spermatogenesis associated 2	8e-112	ATGCCACCTGGAG (T/C) GGCGAGCTAGAGGCC
XM_908972	triple functional domain (PTPRF interacting)	6e-126	ACTCGCTGTC (T/C) CGAGACACTGCAGCGTCC GCAGCACTGTGGCA (A/G) AAACCACGCTGGTG
XM_001928320	similar to Brain and acute leukemia cytoplasmic protein	0	CCACACCAC (G/A) TCTCTCTCCTCCACCTTACC GAATC (G/A) TAGCATCACAGTCA (C/T) GGGCAAT GTGTTTCTCCAGCTTTCC (A/T) CTTCTTTCTTGTG TTTCTTGTGCTGAGC (G/A) TTTTTCTCTCCATTCA AATCAAAGACC (G/A) GCCACTGCTCACAAGAGGGA TTAGCCTCTCCTTGACACA (T/C) CCAGGACTGTGTCA GCATCCTGGGAAAGA (G/A) AGTGTTCCTTCCG
BC114879	GIMAP4 GTPase, IMAP family member 4	1e-136	
XR_028356	similar to zinc finger protein 650	0	GCCCAGT (A/G) GTGGTGATCTTCATG
BC142261	transmembrane protein 223	2e-164	TGCCGCCTTA (G/A) CCCGCCCCCGG
XM_580320	STEAP family member 3	3e-100	GAGCGCATT (T/C) GGTCTTTAGAGAA GGGCCCTGA (T/C) TGAGCGGCAGCTGTG
XM_001914786	autism susceptibility candidate 2	0	TCCGAAACGGTAA (T/C) GGAGGACGTGAG
BC112491	glutathione S-transferase mu 3	0	ATGAAGAGAAAC (A/G) GTACATCTGCGGG
XM_001496172	ARP6 actin-related protein 6 homolog	0	GACCCCTTCTGGACTC (C/T) TTTATATCCTTCT
XR_045452	similar to Copine VI	0	GATGGCATGGG (T/C) GCCTGAGCCCCAGCC C (G/A) CTGGG (A/G) GCCTCG (C/A) GAGT (A/G) G CTGCCACGGCCT (T/C) CT (A/C) GACCGAGAC CC (T/C) CA (T/C) CCCTG (C/T) GTGCT (C/G) CT CAAGTCCACTC (T/C) GATGAGCAGTGGGTGGA GA (T/G) CGCAC (G/A) GAGGTGCTGC (C/T) CC GCCC (T/C) GTCTTCTCCCG (G/C) GT (G/A) CTGGC CCTGGAGTA (T/C) TTCTT (T/C) GAGGAGAAGCA GCAGTTCCA (C/T) GTGTT (C/T) GACGCTGAGGA CCAGGAAGGCGG (T/C) GAGCTTAACTGAAGCA
BC010357	arginine/serine-rich coiled-coil 1	0	

Table S1. Continued

EF635914	COP9 constitutive photomorphogenic homolog subunit 4	0	GATCTGGCGGGCAAGT (G/A) TCGTCAGATCCTAGAAA
BC146134	chromosome 21 open reading frame 51 ortholog	4e-122	GGCCCGGAGCT (C/-) (A/-) (G/-) GACAGTGAGACCTGGC TCAAGATATACC (G/A) AAGGAAAAGATTGGAAGCAA
EW172950	rfbs3932b_p13.y1 fbs Sus scrofa cDNA 5', mRNA sequence	0	TGAGCTCGTGAGA (C/T) GAGCTTTCCCCACT CAGAAGGAAGTGAA (T/C) AGAAATTGATAACA TCCCCTACCTTCC (G/A) TGTGCATCTCTGCC
XM_001929242	similar to Uncharacterized protein C9orf78	0	AGAGGTACAGAA (T/C) TTGAGAAAAGAGGCCCA
AY610017	Sus scrofa clone Clu_5617_scr.msk.p1.Contig3, mRNA sequence	1e-117	TTCACTC (C/T) TGCAC (T/A) GTG ATTGCCACCAA (T/C) ATTCTCCTGGT TGA (T/C) GA (G/A) AT (C/T) ATG (A/C) GAGCT GG (A/T) ATGTCTTCT (C/T) TAAAAGG (T/G) TGAA
AF380118	selenoprotein W, 1	0	GCCGTTTCGAGTCGTCTAT (C/T) GTGGCGCTTGAGGCTA
BC114078	glutaminyl-tRNA synthase (glutamine-hydrolyzing)-like 1	0	GCGCAGCGGCCG (C/T) GTCAGCGTTCCACC CCAAGCTATGGCT (C/T) AGTGTCTCGTCATG
XM_844642	ataxin 2-like	0	ATCGGCAGC (G/C) CC (A/C) GAGGACAAAGCAC ACAGCTGTTGTGGGCTC (T/C) ACTTGTGATGTA GCTGTGCACCGGAAAGC (C/A) TCTGAGCCA
XM_584414	superoxide dismutase 1, soluble	0	ACGAACACTTGGGA (G/A) GATGTGTATAATTT
NM_004447	epidermal growth factor receptor pathway substrate 8	7e-111	ATCCTGTGAAAGA (T/C) TCTAGGTAGAGAAA
AK235024	regulator of G-protein signaling 5	0	TATTGACCACTTCC (G/A) AAGGAGATCACAA
XM_001141083	F-box protein 5	4e-180	AGGCTGTGGATTTGA (C/T) TATTGTACAAGGT
XM_001100585	schwannomin interacting protein 1	0	GGCACTTGAAGC (C/T) TCTTTGATGATGGC
AK239800	programmed cell death 10	0	TGC (G/A) GACAATCAAGGATATAGCC (G/A) GTG
AK231106	similar to vacuolar protein sorting 26	0	TTCAATGACAAGAGTA (G/A) TACTCACGAATT
AK238862	similar to DEAD (Asp-Glu-Ala-Asp) box polypeptide 55	0	CTGAAGGTTGGTAC (C/T) TTTATCCTTTGTCT
XR_045734	similar to peptidylprolyl isomerase (cyclophilin) -like 1	0	TCCCAAGACCGCCC (T/C) GTGGACGATGTG
AK238476	RAB32, member RAS oncogene family	0	GTGTGAACCTTCC (T/G) GGGGCTGTGCCTT AAAATGACAAGG (A/G) TGGCACCCAGGA
EF416571	SPARC-like 1	0	CTTAAGTTCCTT (C/T) CAGCAGGCTGAAACG
NM_138962	musashi homolog 2	0	GTCACAA (G/A) AA (C/A) AAA (G/A) AAAAT
XM_858153	protein kinase, cAMP-dependent, regulatory, type II, alpha	9e-124	TGGCATATTCTTTG (A/G) GTAGCAGAGAGCC
XM_001927219	similar to single-strand selective monofunctional uracil DNA glycosylase	0	GCAATCTCACCCC (T/C) GCCGAAGTCCCGGC
BC102483	abhydrolase domain containing 1	0	CTCTTCCCTGAACCC (A/C) CAGAATGACAGCT
BC103326	OCIA domain containing 2	0	GAAGAGAATCGAAGACAA (A/G) GTCA (C/T) CATG CT (G/T) CAGT (A/G) TC (T/C) AC (T/A) (G/C) AT G (G/A) AAACCAAGA (C/A) AAAG (C/G) TCCCAT GCCACCAC (T/C) (T/A) AGCAAGCAGAGTCTGTTGTT AAA (G/A) CTGCACATCCA (T/C) AGA (G/T) CAGAG GAT (A/T) ATCCGGGAATGTCAAGA (G/A) GAAAGTTT TT (A/G) TCACCCAAGG (G/A) CTGGTCCA (C/T) CA TG GCAGC (T/G) AACCCGAG (G/A) TTTGGATCATTG
BC109995	tetraspanin 31	2e-168	AGACGGTCCCGG (C/G) GACCCTCCCCC
XR_045530	similar to topoisomerase (DNA) II beta 180kDa	0	GTTTTATTTACC (C/T) GTTTCAAGAGGAATGAT TGTGAATTTGGCTCA (A/G) AACTTTGTGGGAA
BC103122	SH3 domain binding glutamic acid-rich protein like	1e-168	ATTCTGAAAATTTT (C/T) AAGAAGACAAGAAAT
XM_522913	KIAA1737	0	TGGTTTAGATA (T/C) GACCA (A/G) TTTAGAAA TTTGGCAGAATATG (C/T) TTGGCAGCAAAGCT
AK235840	paraoxonase 2	0	TAAAGGAATTGAAG (T/C) CGGCTCTGAAGATA
J03489	dihydroliipoamide dehydrogenase	0	AAACACTTGGTGGGA (G/A) CATGCTTGAATGTT

Table S1. Continued

AK232420	alpha-1-antichymotrypsin 2	0	GCAAGAACAGGACGGTGA (G/A) GGTGCCCATG ATG (G/A) GCATT (A/C) G (G/A) G (C/G) CCT G (A/G) (C/A) (C/A) (G/A) (T/C) (G/A) CC (A/T) TA ATCCTCCCTGACGA (C/T) GGCAGG ATGG (A/C) GG (A/C) CGTGGAG CAAG (T/C) TGCTCCC (T/G) GAG
BC148062	dynein, light chain, Tctex-type 3	0	AAATTTTCAGGCAT (C/T) ACCTGTATGTTACT
XM_001499678	similar to Beta-2-glycoprotein I apolipoprotein H	0	TTGAAATCATCCTA (C/T) GCGCCAGGGGAGG
AK094092	defects in morphology 1 homolog	5e-155	ACTGACCTCTGTA (T/C) TCAGAACTGGTGCG
BC141993	ribosomal protein L26	0	GGT (T/G) GT (A/C) CGAGG (G/A) CACTA CAA (G/A) GG (G/T) CAGCA (G/A) ATTGG AAGTAGTCCAGGT (T/G) TACAGGAAGAAAT TCTACAT (T/C) GA (A/G) CG (A/G) GTGCAGC TAATGG (C/T) ACAACTGTCCA (T/C) GTGGGCA (C/G) AGCAAGGT (G/A) GTTATCAC (T/C) AG (A/G) A (G/A) GA (C/T) CG (C/G) AAAAA (G/A) AT (C/T) C TTGA (A/C) CG (T/C) AAAGCCAAATCTCG (C/A) CAA GT (A/T) GGAAA (G/A) GAAAA (G/A) GG (A/C) AAAT AGAAACCATTGAGAA (G/A) ATGCAGGAGTGAATATAC ATGGACATTAAGATA (C/T) TGAATAGAGAAGA
FD632396	susfleck_FC_N_14_E02 SUSFLECK Fat Cell Normalized Sus scrofa cDNA clone 14_E02, mRNA sequence	0	ATGGACATTAAGATA (C/T) TGAATAGAGAAGA
BT020945	BCS1-like	0	ACTCAGCACCTCAG (C/T) GTCGAGACTTCGTA
AK239939	cAMP-regulated phosphoprotein	0	AAGCGAGGTATCC (C/T) CATCTGGGACAAAAG CTGGTGACCACAT (C/T) CCCACTCCACAGGAC ATTAAGAGCTG (G/A) ACTGCATGAATCTTCT
XM_001918249	hypothetical LOC100056802	0	CCCACGCGTCGCC (T/C) AGGACCCGCTCG CAGCACTACCTCC (T/G) CCGGACTACCGCC
U49178	protein kinase C substrate 80K-H	0	AAGAAGGCCCGAG (G/A) GGAGAAGCAGAAA
XM_001914994	hypothetical LOC100073052	0	CCTGGAAGGCCAG (T/G) ACGCGTTCTTCCTG
XM_001926357	similar to isocitrate dehydrogenase	0	CTGGCAGTTCATCA (G/A) GGAGAAGCTCATC CTGCCCCACGTGG (G/A) TGTCCAGCTCAAGT
XM_001927037	similar to Rraga protein	0	GGGCCAGTCAGC (T/G) AGCCCCGCGACTC GGGCAACCTGGT (T/C) TGAACCTGTGGGAC GTTTCCGAACATCC (G/A) TCTGGGATGAAACG
BC149559	protein (peptidylprolyl cis/trans isomerase) NIMA-interacting, 4	0	CACCGAAAGGAAAA (G/A) GCGCATCGGGAAAA AAAAAGCCTCAGGG (T/C) CCCAAAGGTGGTGG ATTTCAAGAAGCAGC (G/A) TTGCCTTGCCAT
BC022524	fibroblast growth factor 12	2e-146	TGTC (G/T) GCAGATTTATGGATAT (C/T) ATCA
XM_001928575	similar to Nck-associated protein 1	0	CATACATTATCTC (C/T) ATGTTTTTGATATTTG TGTAACACTGTAT (T/A) ACATTAATAAATGTTT
XM_607424	zinc finger, CCHC domain containing 11	0	CT (G/T) TGGTTGGGACT (T/G) CTTCCG (T/C) TTC TA (C/T) AC (A/T) GA (G/A) GA (A/G) TTTGA TTT (T/C) AAAGA (A/G) (C/T) ATGT (T/A) ATTAG CAT (C/T) (A/C) GGCAGAAAA (G/A) CTG (C/T) T (T/G) ACAACTTTT (A/G) AGAA (A/G) CAGTG GAC (C/T) TC (A/C) AA (G/A) T (A/G) (C/T) AT TG (T/C) (T/G) ATTGAAGATCC (A/T) TTTGAT AA (T/C) CTTGG (A/T) GCTGGA (T/G) TT (A/T) CT GTGCCACCATGCT (G/A) GCGGCAAGACTCGTG
XM_001927393	similar to growth hormone inducible transmembrane protein	0	GTGCCACCATGCT (G/A) GCGGCAAGACTCGTG
XM_516815	transmembrane 4 L six family member 1	2e-159	GACTGCTGTGGCTGCTG (C/T) GGCCATGAAA
DQ104414	small muscle protein, X-linked	0	GGAGATCCAGAACAT (C/T) AAAAGTGAAGTGA
AK233760	similar to tetratricopeptide repeat domain 5	0	GAAGCTGGAGCCCAA (G/A) CTGGTGGGAG TCAATGGTGCTTCG (C/T) CAGCTCGGACT

Table S1. Continued

XM_001927571	similar to PC4 and SFRS1-interacting protein	0	CAAACCTACTAGTAA (C/A) TTACTTCTTTGC
XR_045568	similar to Uncharacterized protein KIAA1107	0	GCTGA (G/T) AGCTTAATA (T/C) TAGAAAA TTAACTCTTTTTTC (A/G) GATATATCCAGAGA
AB000425	neurolysin (metallopeptidase M3 family)	0	CTGTAAGAAGTG (T/C) GGCAAGCACCAAC
XM_001141202	catenin (cadherin-associated protein), delta 1	0	CTTTTGGGGACCT (T/A) TAACTTTTTTTTTT
XM_001929509	similar to Mitogen-activated protein kinase 1	0	GTCCCTGGCAGCCC (G/T) GTGTGTGTAC
XM_001500465	similar to template acyivating factor-I alpha	0	AAAGGGTGGGTTTGA (A/G) ATGTTTGGGTGGGT
XR_045658	similar to mitochondrial ribosomal protein S28	0	GAGTGTTTTTCTTCTC (T/C) CGGCCCTGTAGAGG AAATGTGGAGTCCTT (T/C) GCATCTATGCTGAGA
BC142418	NECAP endocytosis associated 1	3e-175	GT (C/T) CT (G/C) TGTGT (G/C) AAGCC (C/T) G A (C/G) GTCAGCGTCTACCG (G/C) AT (T/C) CCG CC (C/G) CGGGCC (T/A) CCAACCG (C/T) GGTTA CAGGGC (A/T) TC (T/G) GA (C/G) TGG (A/C) A (A/G) (T/C) T (A/G) GA (T/C) CA (A/G) CC AT (G/C) ATGGA (C/G) (A/T) GG (T/C) C G (C/G) CT (C/G) CG (A/G) ATCACT (T/G) C (C/A) AAAGGG (A/C) AGGTGGC (G/C) T A (T/C) ATCAA (A/G) CTCGAGGA (T/C) A (A/G) (A/G) (G/A) (T/C) (T/C) TCAGGGGAG C (A/C) CC (A/T) GT (A/G) GA (A/T) CA (A/G) T (A/T) TCCTGG (T/C) A (T/C) (C/A) GC (T/G) GT GGAGA (C/G) (A/T) GTG (G/A) C (A/G) GATTC CAGC (C/A) G (C/G) TACTT (T/C) GT (A/T) AT CCG (G/C) AT (C/T) (C/G) A (G/A) GA (T/C) G G(T/A) A (C/A) TGGGCG (C/A) (A/C) G (T/G) G C (T/A) TT (C/T) ATTGG (C/A) AT (C/T) GGC TTC (A/G) (C/G) (A/G) GA (T/C) CGAGGT GA (T/C) GCCTT (C/T) GACTT (T/C) AATG T (C/T) (T/G) C (T/C) TT (A/G) CA (A/G) GA CCA (C/T) TTCAA (G/A) TGGGT (A/G) AAGC AG (G/C) A (A/G) T (C/G) TGAGATTT (C/G) CA
XM_534074	similar to small acidic protein	0	CGATCTTGA (C/T) CTAGCAATTGGGA
BC116061	breast cancer metastasis suppressor 1	0	GGCAGCTCGA (G/A) GACCTGGGACTC
AK234663	RAB3A, member RAS oncogene family	0	ACTACATGTT (T/C) AAG (A/C) TCCTCAT CAT (C/G) GG (T/C) AA (A/C) AGCAGT GT (G/T) GGCAA (G/A) AC (A/T) TCCTT TATGC (A/C) GA (C/T) GAC (T/A) CCTT CAC (G/C) CC (T/A) GCCTT (T/C) GTCA CAC (T/C) GTGGG (A/C) ATCGACTTTAA TCG (G/C) (A/C) A (T/A) GA (T/G) AAG CGT (A/G) T (C/G) AAGCT (C/T) CAGA T (C/T) TGGGACAC (A/G) GC (G/T) GG AC (G/C) ACAGCCTATTACCG (G/T) GG CAT (C/T) CT (C/G) ATGTA (T/C) GACA ATGAAGA (G/A) TCCTTCAATGCAGTGCA
XM_001929333	similar to LOC781710 protein	0	ACAAGCTGGC (T/A) GG (G/A) (A/C) AG AAC (A/C) A (G/T) GGCTTCC (C/G) TCT CTC (C/T) ACGGTGC (A/G) (G/A) TCTCA C (G/C) AAGGGCATCTGGATGTGGTGTGCC (T/C) C ACCCC (A/T) (G/C) (G/C) AATCCAGACCACA (C/T) C CTGGT (T/C) CT (G/T) CT (T/G) GACACCGAGGGCC TGGG (G/T) GATGT (A/G) GAGAAGGG (A/C) GA C (A/T) (A/C) CCAGAA (T/C) GACTC (C/G) TG

Table S1. Continued

			G (C/A) TCTTTGCCCTGGC (A/C) GT (T/G) C T (C/T) CTGAGCAGC (A/G) CC (C/T) T (T/C) A T (A/C) TACAACAG (T/C) (G/A) TGGG (G/C) ACC AGCAGGCC (A/C) TGGA (C/G) CA (A/G) CTGC ACTA (T/C) GTGAC (A/T) GAACTGACAAAACCTCA GAAAAATGATGAGAA (G/A) TCTGCAAAAACAGAT CCGCTCGCTGCCA (A/G) ACTCCAATCCTGTCCG
XR_011082	hypothetical protein LOC702894	0	
AK240550	Sus scrofa mRNA, clone:UTR010096H04, expressed in uterus	4e-147	
XM_001087923	NHL repeat containing 3	0	GTTGG (A/T) CGGGTGTGGGT (G/A) GCTGACC
AJ697885	cytochrome P450, family 2, subfamily E, polypeptide 1	0	C (C/A) AC (C/A) AGCACCACCCTGAGATA (C/T) G
XM_001498463	similar to lecithin-cholesterol acyltransferase	0	TCCATCAAGCCCAT (A/G) CTAGTCTTGGCCTCA
BC134783	carbonic anhydrase VIII	0	TTCTGATCAAGAAC (G/A) GCATACAGCTTCAAAA GAATGTGAAGGAT (C/T) AGTAAAAATCCTCTCCC
XM_518045	similar to C5orf3 protein	1e-150	GTTTGTGTTTGGCA (G/A) ATATTCTGATTCTCCAG GTGATGGGGTGGT (A/G) GTGTTTGAATACTGTTG
NM_183381	ring finger protein 13	0	CTGCCTGTAGAAGC (A/C) GACGTTTTAGCATATA
XM_001109916	similar to Zinc finger CCCH-type domain-containing protein 6	0	TGGTGCAAATATGTA (C/T) ATAACGTACTGTA TATTCGCCCCCTTT (C/T) GACTCCAGATCT
BI340566	365876 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence	0	TGGTCACTATAGCCC (A/T) CTAATAAATATTCCA TGAGATCCTAGCTAA (G/T) CAGAAAACCATTGG AGCCGAGAAGAGATTGACA (G/A) TGGTCTTGT TGTTAC (G/A) TGAG (C/T) GACA (C/T) GATT AAAATCTG (C/T) GTGTA(T/C) ATTAATGATATA CGCCTGCATTGG (G/A) GAGAAGCTAGATGAG TGTGTGGGCCAT (C/T) GGTACCGGCAAGACT CCCGCATCATTTA (C/T) GGGGGCTCTGTGACT
AK234489	hypothetical protein LOC100157582	0	
XM_001915087	hypothetical LOC100057720	0	T (A/G) TACCC (C/A) (G/A) ACCTC (T/A) C (C/G) C G (G/A) CT (C/G) T (C/G) (C/G) GACCC (A/C) CA GGT (C/G) TGG (G/A) T (A/G) GATGC (C/T) G G (G/C) AC (G/C) CAGATCTT (T/C) TTCTC T (A/T) CGCCATCTGCCT (G/C) GG (T/G) T G (T/C) CTGAC (T/G) GC (T/C) CTGGG (A/C) AG CTA (T/C) AACAA (T/G) TA (T/C) (A/C) ACAACAA CATC (A/G) (T/C) (G/C) CTCTG (T/C) T (G/T) C CT (G/C) AA (C/T) AGCGGCACCAGCTTCGTGGCCG GGTTTGCCATCTTCTC (G/T) (G/A) TCCTGGGCTTC ATG (G/T) C (G/T) (T/C) A (C/G) GAGCAGGGGG CAT (T/C) (G/T) C (C/T) GAGGTGGC (A/C) GAGT GCACAGAGAACTTCT (T/C) ATGAATCAAAAAGGGGT
BC112821	family with sequence similarity 107, member B	0	
BC123755	RAB8A, member RAS oncogene family	0	AAGTTCATGGAGACC (A/G) GCGCGAAGGCCAAC TCAAAATCACACCGG (G/A) CCAGCAGAAGAGGA GTGAGCCCTCACTC (G/A) GCCGGGGCCCTCC CAATATTAATCCCGG (T/C) GTGCGGAAAATGGA
XM_001928952	similar to semaphorin 6D	0	CATGCCAAACGTGTA (G/A) CAATTATGCCAAAAG
BC134734	H3 histone, family 3A	0	ATGGA (A/G) TATTACGAGAAGAAGGAAAAG AACCAGGTACCA (G/A) GTGCTGTTGGATGG
AK231872	similar to Vacuolar proton pump subunit E 1	0	
BC148980	hypothetical LOC530597	9e-177	GCCCAGGGCCTCA (G/A) GTTGGGGCTGGAGGA
BC103354	enolase 1, (alpha)	0	AAGGCTGGCGCTG (C/T) TGAGAAGGGGGTGCC CAGAGTATGACTCG (A/C) GGGTGAAGGCCATGG CATTGTCTGCCAA (C/T) GAACCACAAGGAGAAG GGGTGCTGATGTAC (G/T) GGCCCCAGGTACAG
XM_001491791	similar to 26S protease regulatory subunit 6A	0	CACCGTTCCTGTCC (G/A) TGACCTGGGCTCGAC TCAGAGAGGTCCTG (G/A) AGAGGGCCAATGTGG AAAACATGAGCAAG (A/G) CCCCTCACTTGGTTCC
XM_001928345	similar to Acetyl-Coenzyme A acetyltransferase 2	0	

Table S1. Continued

AK233032	similar to transmembrane 4 superfamily member 8	0	CGAGCAGGGCCGC (G/A) GCCGACCGCCGGCGC CTGTAGGGGCTTG (T/C) TCTTCGTTATTGGGCTG CAGCATTGAGAAAGT (G/A) TATAAGACCTACAATG AAGCTACAAGAAATCA (C/T) GATGCATGTTATCTG
BC151302	NDRG family member 3	0	CTTAAATCTGAAAAGC (G/A) TCATTGGGGTTGGA ACATTATTTGGCTCA (C/T) CACTTTGGGCAGGA
XM_001929069	pyruvate kinase, muscle	0	ACCTTGAGGCTGAGG (T/C) AGCCGATTCCCGCA GACTCGCCACCCATC (G/A) CGGCCCGGAACACC AGTCTGGAATGAAACG (C/T) GGCTCGTTTGAACCT ATGAGTACCATGCAG (G/A) GACCATCAAGAACGT GGGCAGCGGCACTGC (T/C) GAGGTGGAGCTCAA GAAGTGGAGCAATAC (T/C) GCAGAGAGAGAGAG
EU282339	ATPase, H ⁺ transporting, lysosomal 13kDa, V1 subunit G2	0	
BX674424	BX674424 Sus Scrofa library (scac) Sus scrofa cDNA clone scac0042.j.05 5prim, mRNA sequence	0	TGAGAAGACATTCTTC (C/T) CCAAATCTGTTTTCT GCAGAAGTAGAAGAC (G/A) CACTAGAGTCTGACC
XM_001925157	similar to SH3-domain GRB2-like 2	0	GGTCTCAGGTTCTTTA (C/T) TTCAATAGAAGAGAA
BC122645	oligodendrocyte myelin glycoprotein	0	AATCTGTCTTATAACCA (T/C) TTTACTGATCTGCAT
BC109483	solute carrier family 4, sodium bicarbonate transporter, member 10	0	AGATGATCCATCTGTG (G/A) TCAATATTTCTGATGA
AJ958193	AJ958193 KN404_2 Sus scrofa cDNA clone C0007757o19 5', mRNA sequence	3e-145	TGATTATTTGGGACCA (G/T) AATTAAGTTAGGTTTT
BC118366	metallothionein 1E	8e-117	CTGCTGTGCCTGATGT (T/C) GGGGAGCACCTGCC CAACCTGCGTTTTTACA (T/C) GACCCTGACTCGAT ACCCTGTTTGTGATCG (C/T) GCCAGTTCTCGGACC GGAGCCCAATTTATGG (C/T) AACTTATTTACTGGAT GAATGAGTCTAGAAA (C/G) CAGACTTATCTAGC TCTTACAGTTGGGTTG (C/T) TATCTTAGGTCAG
NM_001123139	myelin oligodendrocyte glycoprotein	0	
BC126658	phosphotyrosine interaction domain containing 1	3e-169	ACCTCCAGTCTTTGCC (T/A) ACTCGTGCCTATCT TCCATTGAGTCTTCT (G/A) TGAACGAATCTGCTC CCCTTGGCGTTTTCT (C/T) CCCTAGTCTGTCCC
XM_532923	dpy-30 homolog	0	
BC103460	OCIA domain containing 1	0	
XM_001928641	similar to ubiquitin related modifier 1 homolog	0	
AY566232	soluble epoxide hydrolase	0	CCGTCGAGTCCGAGG (C/A) GGAGGCCGCCATGG TCGGGAATACGCCCTG (C/T) CTGAGAATTTCTCCGT GACAAGGCGCTTTCAG (C/A) GAAAAAGATCAACT ACCCCATGCTGCAGGC (T/C) GCTCTCACTCTCAA GG (G/C) AAGGCGCCTCCTACTGGACCAGTGAATC G (C/T) TCCTGCAGCTGCGCAGGCTCCTG (C/T) A TGCTGCTCCTGCTGCCCC (A/G) CGGGCTGTGCCA (A/G) GTGTGCCCAGG (C/G) CTGCATCTGCAAAG AAGTGCAGCTGCTGTGCCTGATGTC (A/G) GGGAG ACCCTG (T/C) CCCAGAAAT (T/C) AATAGAGCAAC CTGCATTTTA (T/A) TATTTTT (T/C) AATACAAGCTG TGAATATGTGAGTTGTAATAAA (A/T) GTTGTCGATTT
M29515	metallothionein	0	
BC118450	fumarate hydratase	0	ATCCAGGCCAACAC (T/C) GAAAGGATCAACAAGCTG
BI337595	361073 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence	0	A (A/G) GACTC (T/C) (T/C) TGGACAT (G/A) G A (A/G) GT (G/C) AAGCAGAACTTCAT (T/C) G A (C/T) CCCCT (T/C) CAGAA (T/C) C T (T/G) (C/T) (A/G) TGACAAGGA (T/C) CT GA (G/A) GGA (A/G) AT (T/C) CAGCA (T/C) C A (T/C) CT (A/G) AAGAAG (T/C) TGGA (G/A) G G (T/C) CG (A/C) CGCCT (A/G) GA (T/C) TTTG TTA (T/C) AAGAA (A/G) AA (A/G) CG (A/G) C A (A/G) GGCAAGAT (T/C) CCCGA (T/C) G A (A/G) GAG (T/C) T (A/G) CGTCA (A/G) G C (G/C) (C/A) TGGAGAA (A/G) TTTGA (T/G) GA GTC (T/C) AAAGAAATTGCTGAGTCCAGCATGTTCAA

Table S1. Continued

XM_001498803	hypothetical protein LOC100053147	0	ACGGTCATCAGAGGC (C/T) CTCAGGCCGAGCTCAA GTGCCCCGTGTGTC (C/T) TTTGGAATTTGAGGAGGA CATTCTGCCCTGGCTGAGCAAG (G/A) CAAACTCCTG CCCTGTGCCG (T/C) CATGAGCTGCCACAGATGATG
U95969	metallothionein-III	1e-179	GCCCTGCCGAGTGTG (G/A) GAAATGCGCCAAGGAT GTGCCCCCTCCGTGC (A/G) AAATGTGTGTAATAGTGC
AY864608	PRA1 family protein-like protein	0	GTGGTGGCTGCCATG (C/A) TGATTTCCGTCGTGGGG CGGTCTGAAGAGGAC (A/G) CCCATGGCATTGTCCT
XM_001926122	similar to Acylphosphatase-1	0	TTATGAAGTTTTTGA (T/A) AGGTGCAAGGAGTGTTT CGCAAGTACTC (G/A) GGCTGAGGGTAAAAAGCTT
XM_001926433	similar to UPF0463 protein C6orf35 homolog	0	GCTGGGACTGCGGA (A/G) TCGTCTCTCGATTG TTAGGAGTTCACAGT (A/G) TGAAAGACTTTTCGAA
EU650784	cardiac muscle ATP synthase H+ transporting mitochondrial F1 complex alpha subunit 1	0	AGTTCGAGGCTGC (T/G) GCTGCAGAGGAGTCA CGGC (C/T) GTGGCCCGCGC (T/C) CTCCTCGG TTTGTGCTGCAAGAAA (T/C) CTCATGCCTCTAA GACCTTGAAGA (A/G) AC (T/C) GGGCGTGTTTTA GT (A/G) TACACGGGCTGAGAAATGT (C/T) CA
XM_851307	SNRPN upstream reading frame	0	TCATGAC (C/T) GTGGG (A/C) AAGAGCAGCA AGCACAT (T/C) GACTA (T/C) AG (A/G) ATG AG (A/G) TG (T/C) ATCCTGCAAGA (T/C) GG CG (G/T) ATCTTCATTGGCACCTTTAA (G/A) GC A (C/T) TTGATCCTCTGTGA (T/C) TGTGA (T/C) G A (A/G) TTCAG (G/A) AAGATCAA (G/A) CCAAAG AC (G/T) CAAA (G/A) CA (G/A) (C/G) C (A/G) G AG (C/A) G (T/A) GAAGA (A/G) AAGCG (G/A) G T (T/C) (T/C) T (G/T) GGTCTGGT (T/C) T (G/A) C T (G/T) CG (T/A) GGGGA (A/G) AA (T/C) (T/C) T GGT (T/C) TC (C/A) ATGAC (T/G) GT (G/A) G A (G/A) GGACCACC (C/T) CCCAAAGATACCGG
DQ499447	tRNA-yW synthesizing protein 3 homolog	0	ACCCTAGCTCTCTC (G/T) TCGTATGGATCGGAG GAACAGTTTTTTCAC (C/T) ACCAGTTCCTGCGCT CCTTGTGTGAAAGA (T/C) GATGTGATTGTAGCC AGGAAAGACTATGT (T/A) GGCTGTCCGGAGCA
EW606688	rsto213_n2.y1 sto Sus scrofa cDNA 5', mRNA sequence	8e-155	ATTCAGTAATTTGT (T/A) CAATATTTAAAAATT
AK236522	cytochrome b5 type A	0	GAGGTTCCGGC (G/T) TTACGAAATGGCCGAAC G (G/A) GAAGAAGT (C/T) (T/C) T (A/G) A (G/T) GG A (A/G) GCTGG (A/T) (G/A) G (T/A) GATGC (T/A) A CTGAAAA (T/C) TTTGA (G/A) GATGTTGGACA AC (A/T) GATGC (T/C) (C/A) GAGAG (T/A) T G (T/C) CCAA (A/G) (A/C) (C/A) GTTCATCAT TGG (G/T) GA (G/T) (C/G) TGCATCCGAATGA
XM_001925711	similar to Protein fam69a	0	TGTTGGAAGCTGGAT (T/A) ATCTATGTGCAATA
BC105398	CWC15 spliceosome-associated protein homolog	0	GAGAGAGCTGTGC (G/A) AGAGAAAAAATAG TTGTAATGATACGCT (A/G) CGATCTGAATTTCA ACAGTTTTATGTGCTT (A/T) ATTAAGGCTGTAA
XM_001504320	similar to phosphatidylinositol transfer protein	0	GTTCAAGTGGTGGGG (T/C) CTGCAGAACAAG
EW310740	rp1do0124_e24.y1 Ido Sus scrofa cDNA 5', mRNA sequence	0	AGAGGGCTGTCCCC (G/C) CCCTCTGATGAGGA
CR860378	succinate-CoA ligase, ADP-forming, beta subunit	0	CTGCAGACCGCCG (T/C) CAGGTTCTGGGAAG ATGAATA (T/C) ATGAGTA (C/T) GGAATTATTG
AK231804	Sus scrofa mRNA, clone:LNG010025F10, expressed in lung	0	TAGCATCCATGAGGA (C/T) GCGGGTTCCATCCC CAGAGGGTTAAGG (C/A) TCTGG (T/C) GTTG TGAGCT (G/T) (T/C) GG (T/C) G (T/A) AG TC (G/A) CA (G/C) (A/T) (T/C) GTGGCTTG ATCC (T/C) GCATTGCTG (T/C) G (G/C) CTG TGG (T/C) (G/A) TAGG (T/C) CAG (T/C) (G/A) G

Table S1. Continued

			CT (A/G) CAGCTCT (T/G)/ ATTG (A/G)/ ACCCCTAG CCTGGGAAC (C/T) TCCATATGCCTCAAGTGTGGCCCTAA
AK293500	Nik related kinase	0	CGCATTAAAATTGCT (C/T) TTCGCCTCTGGATCC
XM_001925969	similar to microtubule-associated protein 1A	0	GTGGTCATCTGCTGTG (G/A) ACCGGGTGTGTAAC GGTAAAAAGCCACT (G/A)/ CCTGACATCCCAAC ATGTGTGACTCCCCA (A/G) CATTCCACTATGCCA
XM_001929570	heat shock 90kD protein 1, beta	0	TGGTGTGTTGTGATCC (C/T) TATCCTTTAAGATGCC
AF268463	voltage-dependent anion channel 5, pseudogene	0	TGTGA (C/T) CT (T/A) GG (C/A) AA (A/G) GCTG CCA (G/A) (A/G) GAT (A/G) T (T/C) TTCAACAAA GGAT (T/A) TGG (T/G) TTTGG (G/C) (T/A) TGG T (G/C) AAA (C/A) T (G/A) GAT (G/C) TGA (A/G) AA C (A/C) AAGTCATG (C/T) AGTGG (C/A) GT G (A/-) (T/-) (G/-) GAATT(C/T) TC (A/T) AC (A/T) T C (T/G) GGTCATGCT (A/T) A (T/C) AC (A/G) GA (C/T) A C (T/A) GG (T/G) AAAG (T/C) (T/A) (A/T) C (T/A) G G (G/C) A (C/A) C (T/C) TGGAGACCAAATA (T/C) AAGA T (G/C) TGTGA (G/C) TATGG (T/A) CTGAC (T/C) T T (C/T) AC (A/C) (G/C) AAAAGTGAACAC (T/A) G A (T/C) AA (C/T) ACTCT (G/T) GG (A/G) ACAGAAAT C (G/T) CTTTG (A/G) AGA (C/A) (C/T) A (G/A) (A/G) T T (T/G) GCT (C/G) AAGG (T/G) TTGAAACTGAC (A/T) (T/C) TTGATACCA (C/T) (C/A) TTT (T/G) (C/T) ACCAAA AA (A/G) AGTGG (T/G) AAA (A/T) T (C/G) AAG (T/G) C (T/C) TC (T/C) TA (C/T) AA (G/A) (A/C) G (G/A) G A (G/A) TGT (G/T) T (A/C) A (A/G) (C/T) (C/A) TTG G (T/C) (T/A) GT (G/A) ATGTTGA (C/T) (T/A) T (T/A) GATTTTGCTGGACC (T/A) (G/A) C (T/C) ATC (C/T) ATG G(T/C) T (C/G) (A/G) GC (T/C) G (C/T) (C/G) (T/C) T (T/G) G (G/C) (T/C) (T/C) (A/T) TGA (G/A) GGCTGGC TTGCTGG (G/C) TA (C/T) CAGATGA (C/G) (C/T) TTTGAC A (G/C) (T/A) GCCAAATC (A/C) AA (G/A) CTG (A/T) CA TACACGCCAGGTTATTGGGG (T/C) GCCTGCAGTATCCTCTAC TGGGACTTCCAATTTTCCTTT (G/A) TTTGCACTGAGGATTTAG
XM_001929579	similar to SNAP-associated protein	0	
XM_602219	collagen, type XXVII, alpha 1	3e-149	TTTTTAAGAAAAAAAAAAAA (C/A) CCTTATTTAAGATTCTGAAG ACAGACTTTGAAGAAACAT (T/A) TGGACGTGGGGTGTGATCC
BC153246	prune homolog 2	2e-165	ACCAAGGGGTCATTGGGG (T/C) AAACGTATTGTTTTCCCAAAG A (C/T) CTTCTAAGACATGCTCTGACCTATGTGGAGAA (G/A) T GTTGAAGCCAAAACCCTC (C/T) GGCAGCTGAGTAATGTGGTTC
AK235825	similar to Mitochondrial import inner membrane translocase subunit Tim23	0	
AK232705	Sus scrofa mRNA, clone:LVRM10010B01, expressed in liver	1e-160	GCG (T/G) CGTGATGAATGGAAGCTCGCTCGCAGTGATTTTT
BC102976	NADH dehydrogenase (ubiquinone) 1 alpha ubcomplex, 11, 14.7kDa	9e-169	CATGGCCAAGACGCTTCT (T/C) CACAAGTACTCGGATATCC
AY613916	chromogranin B	0	TTTCAAATGCCCTGTGCA (A/G) GTCCAATGCTCCTCCCAT
NM_014342	mitochondrial carrier homolog 2	0	ATCCCCCTCTTATTCCCC (G/A) ATATATACTTCTTGATAG
XM_001503183	similar to CGI-110 protein	0	GGCGGGGTCTTTTTGTTG (G/A) TGTCCTATATCTCACTCGGT
NM_021992	thymosin beta 15a	5e-160	CTCCCAAGAGTGAATTTCA (A/G) CATTGCTTGAGTGTCTTGA CATCTTCTGATGTCTTCTCA (T/C) CTGTAATGCCTGGCTAAG ACC (T/A) TCAGTGGTGT (A/C) TAGATCCAGC (C/T) GAC AAGAGTGAAGGGATAATG (G/A) CCTGGGAGCTAGACCGGCT
XM_001504143	phosphatidylinositol glycan anchor biosynthesis, class S	0	
XM_001929662	similar to MLLT11 protein	0	TCCTTTCTTTTCTGGAGGA (C/T) GCCCATTCCAGAACTGGAT
XM_001928514	similar to antioxidant protein	0	AATTCGCCTTTAGCACCA (T/G) TTCTTCCCACCATCCCCTGCC AAATTATTGCTTTCAAGT (G/A) CAAAGCCAATGAATTTTCATGA CCCCTATCCCCACACCC (T/C) AGGAGCTGGGGACTTCTTACCT
BC102595	proteasome (prosome, macropain) 26S subunit, ATPase, 4	0	

Table S1. Continued

XM_001929267	similar to ubiquinol-cytochrome c reductase binding protein	0	AGAACCTTTATAACGAC (G/A) GGGTGTTCGCATTAAGAGAG TGGACAAAATATGAGGA (A/G) GATAAATTCTACCTTGAACCA
XM_001144002	FGFR1 oncogene partner 2	1e-148	CAGTTTTTATTTTTAACT (T/A) ATTCTATTGATTTGGGGTAA
BC118277	tetraspanin 7	0	TTTCATGGAGACCAACA (G/T) GGGGATCATCGCGGAGTGGC ACAACGATGGAGTAAGA (G/A) ACCTGTTTCCAAAAGGCATGG
S73478	ribonuclease, RNase A family, 4	0	AGCGCTTCAACAC (T/C) TT (C/T) TTCATGAAGACATTTGGAA
XM_001928222	similar to Myelin-associated oligodendrocyte basic protein	0	GAAAGTAGCTAAGGAGG (A/G) CCCCAGACTCTCCAAGA CTGAACGTCAGCC (G/A) CG (A/C) CCCCAGCCAGAAGCCC
XM_001493074	similar to DM-20	0	CAGTA (C/T) GTCATCTA (C/T) GG (C/A) A (T/C) (C/T) G C (G/C) TC (C/T) TTCTTCTC (T/C) T (C/T) TA (C/T) G GG (A/G) (T/C) C (A/C) TCCTG (T/C) TGGC (A/T) GAGG GCTTCTACACCACC (A/G) GCGCAGTCAGGCAGATCTT
XM_001497146	adaptor-related protein complex 2, mu 1 subunit	0	CAGG (C/T) GAGCAGGGGGCGGGCGGACA (T/C) CTTG TTTGTGCTCATATATGAG (C/T) TGCTTGTAGATCCTGGAC GGCAGATTGGCTGGCGG (G/C) GGGAAAGGCATCAAGTATCG CTCCCGCGGTTTCGGAC (T/G) GACGCTGGCACAGCTCCTG
XM_001103332	sepiapterin reductase	0	AGTGTATTTTGTAGTATGTAATAAAAA (C/T) TTTAAAAGT
CR860856	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex,1, 7.5kDa	2e-126	
BI346028	375023 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence	0	TGCGATCACAACCTGAAC (G/A) TAGATAGCAAAACACTTA
NM_001113234	anaphase promoting complex subunit 13	3e-168	ACATTGGCTTATTTTGT (G/A) GCAGCAAAATGCAATGTG
AW787203	120964 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence	0	TCATTGACCAATTGACAA (T/C) GTGGTTAACTATCAAGA
XM_853711	H3 histone, family 3B (H3.3B)	0	CGTTCGTTGGAGTTGCT (A/T) CTTGCTTTGGTTCCGGCGA TTGGTGAGGGAGATCG (T/C) CCAGGATTTCAAACCTGAC CGTGGTATCAGTATCC (G/A) ACCATTGATAAGGCATTAT
BC102690	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 5,16kDa	0	
BC150010	SEC11 homolog A	0	GATTGTGACGATCCTC (G/A) TGAATGACTATCCTAAATTTA
AK161097	FUS interacting protein (serine-arginine rich) 1	0	AATACGTCTCTGTTTCGT (A/G) AGGAACGTGGCCGACGATA
BC102721	NAD(P)H dehydrogenase, quinone 2	0	AAACCTCGCGCAGCCTT (C/T) CCGTGCGCCCTCAGGCG GCGCCCTCTCCAACCCC (A/G) GCTTCTTCAATTACGGGGTG TACAC (T/A) TCTGTCTGCTGGTATTAGCAGC (T/C) TGAG TGAGCCTCTGTGTTCCAA (G/A) GGCTGCGTCTGCTGCTTGA
CJ025998	CJ025998 full-length enriched swine cDNA library, adult trachea Sus scrofa cDNA clone TCH01B050048 5', mRNA sequence	0	
BC102779	septin 6	0	GCT (G/C) CGG (C/T) CC (C/A) A (G/T) ACCTA (C/T) GAC T (C/T) CAAGAGAGCAACGTG (C/G) (A/G) (G/C) CT (C/G) A CT (G/C) ACCATCGT (G/T) (G/A) (A/G) C (G/A) C (C/A) GT (T/A) GA (G/C) AG (T/C) TACA (G/A) GCCCATCGT (C/G) G A (C/A) T (A/T) CATCGACGC (G/C) CA (G/A) TTTGA (A/G) (A/G) (A/C) CTA (C/T) CTGCAGGAGGAGCTGAAGATCCG C (C/A) G (C/G) (T/G) (C/T) (G/A) CT (C/G) (T/C) (T/A) C (G/A) (A/C) CTACCACGAC (A/T) C (G/C) (A/C) G (G/C) AT CCACG (T/C) CTGC (C/T) T (C/G) TACTTCATCACGCCACCG TCTGGCTCTGCGAGTACT (G/A) TCTCAAGTACATGAAATATGA GGAGCC (G/A) TTCGTCTTCTACATCTTGAC (T/C) GAGGTGGAC AAGCCGCTGTCTGACCTG (G/A) GCAAGCTCAGCTACCCGACG CCCCTGACTGGCGCAGAC (C/T) ACAAGAGGCCCCCGCAGTA CGGCAACCCGGGCCAGT (A/G) CACTCGCAGGCCTCCCTCC AGCCCATGGATCCAGACGC (A/G) TTGGGATGTTGGCTTTGCTC AGAGTCTGCTGGGCCTTTG (G/A) AAAGGGATGTCTCCTCCAT GCCCTCTCCCGCGGTGCC (G/A) GGGAGTTTGTGGAACGGGA AGTTGCCCATTAACCTCA (G/A) TTCCAGTCTTGATTCTCATC GGTGTTCACTGTGAGGGGC (C/A) GTGGCTGTGAGCCTTCTAGT
XM_001500749	hypothetical protein LOC100064910	0	
BC148053	solute carrier family 25, member 38	0	
AK236031	similar to MRPL43 protein	0	

Table S1. Continued

BC126591	FYN oncogene related to SRC, FGR, YES	0	ATGATCCACTGCTGG (A/C) (A/C) AAAGGACCC (C/G) GAAG A (G/A) (A/C) GGCCCAC (T/C) TTCGAGTAC (T/C) TGCAGG (G/C) CTTCTGGAAG (C/A) CTACTTCACCGCGAC (A/C) GA CCCAGTA (T/C) CA (A/G) CCCGG (T/A) GA (A/G) AACCTGT
D37917	cathepsin L	0	ATCCTGTTG (C/G) TAT (T/A) GATGC (G/A) (G/A) GCCA T (T/G) (C/A) ATC (T/C) TTCCAGTT (C/T) TA (T/C) A (A/G) (G/C) (T/A) CAGGCATTATT (A/T) TGA (T/A) CC A (G/A) A (C/G) TGCAGCAGC (A/G) AAGA (C/T) CTGGA G (T/C) GTT (T/C) TGGTGGT (T/C) GGCTATGGC (T/G) T GGA (A/G) C (T/A) (G/C) A (T/C) TCAAATAATAA (A/G) T (T/A) TTGG (A/C) TTGTCAAGAACAG (T/C) TGGGGTAATAC
XM_525719	abhydrolase domain containing 1	0	TTCCAGGGGAGCAGCCC (C/T) TCCAGGAAGCCGATGTGGCC
XM_001498796	similar to phosphatidylinositol glycan, class F	0	CTTCTCAATATTGGAAAC (G/A) CACTTGACATGGTTGTGCATC AGTAGTGAATCAAATAC (A/G) TCATCTAAAAGAAGTTCATT
D90073	poly (ADP-ribose) polymerase 1	0	TCACTG (G/A) ATCTCCTTAAAGGATTTACATGAACA
NM_001040457	rhomboid domain containing 2	0	CCGTCCGCGCTCGGATG (A/C) GGCGGGCCCTGGTGTTTGGC
BC102096	eukaryotic translation initiation factor 3, subunit H	0	CAGCTCCACCACCGGC (G/A) CTGCTGGGAAAGGCAAAGGC ACTTCTGGATTCTCAGTT (C/T) AGTTACCAGCATGCGATTGAA ATAAAACTGCCAAGG (G/A) TCTCTCTCACTAAAGGCATAC
AY911314	splicing factor 3b, subunit 5, 10kDa	5e-160	CTTCTCTCTGCTGCGGC (C/T) TCTGACCTGCGAGCAAAGGA
AY166682	transmembrane BAX inhibitor motif containing 6	0	AATTCTCTGCTAGGCCTG (G/A) GAATCTGAGTTACATCTCTTG
XM_001928357	similar to RAB25	0	TTTGTCTTCAA (A/G) GTGGTGCT (C/G) ATCGG (G/C) G A (T/G) TC (G/A) GG (A/C) GTGGGAAGA (G/C) CA A (C/T) CTGCT (G/A) TCCCG (C/A) TTCAC (C/T) CG AA (C/T) GAGTTCA (A/G) CC (T/A) (G/C) GA (G/C) A GC (A/C) (A/G) (G/C) A (G/C) CACCATCGG (C/G) G T (G/T) GAGTTT (G/T) CCACCCGCA (G/C) (C/G) (A/G) T C (C/A) (A/T) G (G/C) TGG (A/G) C (G/A) (G/C) C (A/G) (A/C) (G/C) (A/G) C (C/T) (A/G) TCAAGGC (A/T) C AGATCTGGGACAC (C/G) GC (C/T) GGCC (A/T) GGAG CG (C/A) TACCG (G/A) GCCATCACCTC (T/G) GC (A/G) TA TA (C/T) CGTGGTGCAAGTGGG (T/G) GC (G/C) CT (G/A) CT GGTGT (A/T) TGA (T/C) (A/C) T (T/A) (G/A) CCAAGCAC C (T/A) GACATACGAGAACGTGGAGCGCTGGCTGAAG
XM_001926473	hypothetical protein LOC100156815	0	GGCCTGGAAGCAAATGTC (G/C) TGGTTCTACTACCAGTACCT
AY609829	similar to 19 kDa subunit of NADH:ubiquinone oxidoreductase complex (complex I)	0	GTCGACCCACGCG (G/T) CC (C/G) CTCGGCGGTCCGGAAGG GGCATC (T/A) CCTCGGTCCGGGCTACTGCCA (T/A) CATGC AAGCCCAACAAGGAGTTCA (C/T) GCTCTGCCGGTGGGAAGA
BF441917	258379 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence	0	AAGAAC (C/T) TGACATAGT (G/A) TCCA (C/T) (A/G) (G/A) G GA (A/T) GC (C/A) GGTT (C/T) (A/G) ATCCCT (G/T) G (C/T) (A/C) T (C/T) (A/G) CTCAG (T/G) GG (A/G) TTAAGGA
BP436711	BP436711 full-length enriched swine cDNA library, adult lung Sus scrofa cDNA clone LNG010069D10 5', mRNA sequence	6e-101	AACTTGGAGTTC (T/C) CTTGTGG (T/C) (G/A) CA (A/G) C AGGTT (A/G) AGGAT (T/C) (T/C) (G/A) GCATTGTC (A/C) C TGC (C/T) (A/G) TG (A/G) (C/T) G (A/C) (G/A) (A/G) G TT (T/C) (G/A) ATCCTTGGC (C/T) (C/T) AGAACTTCCACA
BT021491	ribosomal L1 domain containing 1	0	AAGCCA (G/A) GCTCCGCCTTTGGGCG (T/C) TTTGACT
BC109537	alanyl-tRNA synthetase domain containing 1	0	CACCGAGAGAGCGCTGAC (A/G) GCCCTGCTGAAGTGTGCA
XM_001917680	similar to pyruvate dehydrogenase kinase 2 subunit p45	0	ACCCTCGGTGCAGCTGGT (G/A) CAGAGCTGGTATGTCCAGA ACCCTGAGCCAGTTCACC (A/G) ACGCCCTGGTCACCATCCG
BC119909	IMP4, U3 small nucleolar ribonucleoprotein, homolog	0	GAAGGACAAGGTTCCGGCG (C/T) GCGCTGAAGAGAACC GC GAGTTACGCAGAGAGGCT (T/C) TGGCCTTACAGGGATCCCT CGGCACGAGGTCGGGGC (A/G) CTGGTGCAGCCTGCAAAG
BC116064	immediate early response 3 interacting protein 1	0	ATTAATCTCAACTAATGAA (C/T) CTGATCGATCTGTAAGAA CAGCAGAAGTTACTACTTCA (A/G) TCTTTGTTGGAATATACAT
XM_600126	metastasis suppressor 1-like	3e-30	GCTGAAATCCCAGCTG (A/G) (A/G) TGTCTCTTTCCCTAAA

Table S1. Continued

EF416570	P311 protein	0	TGAATGTTGTTCTCTAGAC (C/T) GGATTGACTTGTTCCTTG AGTAAGTGCCACCTCTGAG (G/A) GCAGACACTCGTCTCAGG TTGGGGAAAAAAGGGCAA (C/A) CTGGTTTGGGTTTTTTTGT
XM_001929537	ubiquitin-conjugating enzyme E2L 3	0	CACGTTTAAAACGAAGATCT (G/A) TCACCCCAACATCGATGA
XM_001168089	contactin 1	0	AACTAAATCTTTTGCCAAAT (G/A) CATGCTTGCCTTTATTTT
XM_001503420	similar to visinin-like protein	0	AGAGCGGCCCTTGGCTCCT (T/C) GGTGTGGTCCCCACACAA
XM_533961	cold inducible RNA binding protein	0	ACTCAGTGGTGACCATGG (C/G) ATCAGACGAGGGCAAACCTTT ATGGCCATGAACGGGAAG (C/T) CTGTGGATGGCGGCAGAT GCCGGAGTCAGGGTGGC (A/G) GCTATGGTGACCGGAGCTCA CCTGCCACCCTGAGTGCC (G/A) CATTTCGAGGCCACCCCAT
XM_001927751	similar to kidney predominant protein NCU-G1	0	GACTGATTTGTAGTCTGCC (G/C) AATCCCTGCTGCACCTGAGG
XM_001495383	similar to Protein ARMET precursor	0	TTAAATATTAAGAAGACC (G/A) TCTCTCTATTAAATTACAGT
XM_614056	cullin 3	0	A (G/A) AGG (T/C) CT (G/A) AAGAATGT (G/A) TTT (A/G) A (T/C) GA (G/A) GC (T/A) AT (C/A) (C/T) T (A/G) GC TGCCCT (C/G) GAGCCTCC (G/A) GAACCGAAGA (G/A) GA
NM_001791	cell division cycle 42	0	CCGGTGTGTTTGCTCCTCG (C/T) CCATCTACTGGTCCCCACAA GCTGTTTCATCGTATTACACC (G/T) GAGGAGGCCAAGTACAAG CTAACCTGGGAAGAATTGG (C/T) GTGATCACCAACAGGGAGA
AK233807	ribosomal protein S4	0	GGCCATGACAGATGCTGC (C/T) GCTCTCAACATCCTTGCCCT TATAAG (G/A) TGGTTTCCTCTGTGATGAAAA (C/T) GCCTGA GAGGAGCG (C/G) C (C/A) GGATGA (G/C) CACGG (A/C) TACA TC (G/T) C (G/C) CGCGAGTTCACCGCCGCTACCGC (T/C) TG CCGCCTGGTGTC (A/G) ACCCGGCGGCCGTGACGTCCGCTCTG
AK234517	glucosidase, beta; acid	0	CCTGCACC (A/C) G (G/C) TTCAC (G/C) GACGAGTA (T/C) C AGCT (T/C) T (T/A) CGAGGAG (C/A) TTGG (A/C) AAGGGGG C (A/T) TTCTC (A/T) GTGGTG (A/C) GAA G (A/C) TGTATGA A (A/G) (A/C) TC (C/T) C (T/C) AC (T/C) GG (A/C) CA (A/C) G A (A/G) TA (T/C) GC (T/A) GCCAA (A/G) ATTATCAACACCAA (A/G) AAGCT (T/C) TCTGC (T/C) AG (G/A) GATCATCA (A/G) A (A/G) CTGGA (A/G) AGGGA (A/G) GCT (A/C) GAATCTGCCG CTT (T/C) TGAAGCAC (C/T) CAA (T/C) AT (T/C) GTGCGACT CTT (T/C) CATTACTTGGT (G/C) TT (T/C) GA (T/C) (T/C) TGG T (T/C) ACTGG (A/C) GG (T/G) GA (A/G) CT (G/C) TTTGAA GACAT (A/C) GTGGCGAGAGAATATTACAG (T/C) GA (A/G) G C (T/G) GA (T/C) GCCAGTCA (T/C) TGTATACAGCAGATTCTA GACCAACACCGACTCTA (C/T) CAGCAGATCAAGGCTGGAGCT
XM_001925668	similar to Neurocalcin-delta	0	TGGCCCGTTGGCAGAGC (C/T) ACGTGGATAACCTGATGTGCG CTGCCAGGAAGCCGCCA (C/T) TGTCGGCTACTGCGACGCCAA GTGGGGAGCCAACATAC (G/A) ACGTTGCTGTGCGCAGAGCTG GTGCCCATCAGACCTCG (C/T) TCAAACCCTCATGCATTTCGAG
BC126729	heat shock protein, alpha-crystallin-related, B6	0	CAAGTTGCTCCAGAAAA (G/A) CCTGTAAAGAAGCAAAGAC TGGACTGTCAATGTT (T/C) ATAACCTAAACCTGGGTACAT
U73504	calcium/calmodulin-dependent protein kinase II delta	0	GCCATGTCTGCCCTTTT (T/A) TTGTGGTGTCTAGCATTAACTT
BC122616	profilin 2	0	AAGCTGCACCAG (T/G) TGGCAA (T/G) (G/A) CAACAGTCTCAT CCCATGACGCA (T/C) GGCAACACCGGATTTCAG (T/C) GGCATT (A/C) TCCAGCTCTCCAGAGGTGAAAAGGCTATTGGGCA (G/A) G TTGGATGCATCT (G/A) CTCA (G/A) AC (T/C) ACTTCTCATGAA AACGA (T/C) TTGATTGGCTGCATAATCGGGCG (T/C) CAAGGC CAA (A/C) AT (C/T) AATGAGATCCG (T/C) CAGATGTC (T/C) G AAAATTGC (G/C) AA (C/T) CCAGTGAAGG (A/C) TCT (A/T) C (A/G) TAGGCAGGTTACCATCACTGGATCTGCTGCCAG (C/T) A TAG (C/T) CTGGC (T/C) CAATATCTAATCAATG (T/C) CAGGCT
AK236785	Sus scrofa mRNA, clone:PBL010004E06, expressed in periphral blood mononuclear cell	0	
BC134490	SUB1 homolog	0	
XM_857035	heterogeneous nuclear ribonucleoprotein A3	0	
XM_001928960	similar to Astrocytic phosphoprotein PEA-15	0	
AK238034	poly(rC) binding protein 2	0	

Table S1. Continued

BC153235	chimerin (chimaerin) 1	0	GAAGA (C/T) AT (A/C) AACAT (C/T) ATCACTGG (A/T) G C (C/G) CT (T/G) AAAGTACTTTCAG (A/G) GA (C/T) T TACC (C/A) ATTCC (T/A) (G/C) TCATCAC (C/G) TATG A (T/C) (A/G) CCTA (T/C) (T/C) C (T/G) AA (A/G) T T (T/C) ATAGA (A/G) (G/T) C (A/T) GC (A/C) AAAATC TAAGACAGCTATGTCTCA (A/G) CCTGACTCTTGACATTAA
AY609871	Sus scrofa clone Clu_36963_scr.msk.p1.Contig1, mRNA sequence	0	
BQ597566	MI-P-A2-afc-a-12-1-UM,s1 MI-P-A2 Sus scrofa cDNA clone MI-P-A2-afc-a-12-1-UM 3', mRNA sequence	0	GGAGTTCCTC (G/A) TCGTGGCT (C/A) AGTGG (T/A) TAAT GAA (C/T) (A/C) (C/T) GACTGGCATCCATGAGGACACAGG CCTGGC (C/A) TC (G/A) CT (C/T) AGTGGGTTAA (G/A) G ATCTGG (T/C) GTT (G/A) C (C/A) GTGA (G/A) CTGTGGT GTAGG (T/G) TG (T/C) AGATG (C/T) GGCTCA (G/A) ACC GC (G/A) T (T/C) G (T/C) TGTG (C/G) CTGTGG (C/T) G CCAGCGG (T/C) T (A/G) CAGCTCCAATTCA (A/G) CCCCT GCCTGAGAACTCCATATGC (T/C) GC (T/A) GGTGCAGCC TCTCCATCTCCGGCTGGGTG (C/T) TGGTGTCTCCACGCT GCCACCGACTACTGGAA (G/A) GTGTCCACCATCGACGGC
XM_001925602	similar to claudin 10	0	
BE013772	125231 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence	4e-179	TTTTTATCAG (G/A) GTCTTCACGG TTGTAGGG (G/A) AGACGTTGTACT CTTCTTTTC (G/A) ATTGTGGCTCT GTCGCGGA (T/-) (T/-) TGTAGTGTAG TAACTTCTCAAATC (T/C) CTT GCTTCCTTTTT (T/A) GTCTCAATGG GTCCAGGTT (T/C) TGAATCCACTCA TCCTCTT (C/T) CTGTGATGATTCCATCTGTTT
CN157594	946310 MARC 4PIG Sus scrofa cDNA 5', mRNA sequence	0	
EW089596	recc2802c_h7.y1 ecc Sus scrofa cDNA 5', mRNA sequence	0	AGATAAATCACAGA (G/A) CGTGCCGTT
AW347194	30566 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence	5e-134	CCGCCTCTTC (T/C) ACTCTTGGTGAC GTTTTGTGTCT (C/T) GATCCCCGTGG AAGAGCCCCG (T/C) GCACACACAGC CCAGCTGTAGA (G/A) AAGAATATATC
AK235753	Sus scrofa mRNA, clone:OVRM10111F12, expressed in ovary	0	
EV903443	rcbl0_004449.y1 cbl Sus scrofa cDNA 5', mRNA sequence	0	AAT (A/G) ATCAAATGCCACTGTCA (T/C) TGC
CN158340	947125 MARC 4PIG Sus scrofa cDNA 5', mRNA sequence	0	TCATTGGAAACCT (T/C) GTCTAACAACAGAAAC
BG895945	359644 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence	1e-179	GCT (G/A) TGGC (T/C) GTGG (T/C) A (T/C) AGG CTC (T/C) GATTCAAC (A/C) (A/C) C (T/A) AGCC CCTGGGAGTTTC (T/C) ATATGCCACAGGTACAACCC
BW982861	BW982861 full-length enriched swine cDNA library, adult intestine Sus scrofa cDNA clone ITT010077F03 5', mRNA sequence	0	AGGA (T/G) CCTCCTCAACTCTGTCT (G/A) (G/T) GA TG (C/T) TGGGTCAGC (A/G) TGTTCCGCA (A/G) TTT GGTGAGAA (G/A) CTGGTCCGAAGGCGGCCACTGGAG CCCAG (C/G) GACGAGTCTGAGATTTCCACACCTCATC T (A/C) AACA (C/G) C (C/T) T (C/T) (A/G) A (T/C) C TG (C/G) T (T/C) TTTGGTGTGAA (T/C) A (T/G) AAC CT (A/T) AGG (C/G) T (C/T) GGT (A/G) T (A/G) TA CTGGT (G/T) C (C/A) (A/G) TGGCCGTG (C/T) ATG TAGCTGGACCTG (C/G) (T/G) ATCATC (A/G) TCTC C (T/C) TCTTG (C/G) T (C/G) G (T/C) (T/C) G (G/C) C TT (A/C) T (T/C) T (G/T) C (C/T) (G/A) TGTT G (T/G) (G/C) TG (C/G) (A/G) (A/T) T (C/G) T (G/A) C T (T/A) (T/C) GC (T/A) GAGTTTGG (G/T) GCCCGGG T (A/C) (T/C) CC (T/C) G (C/T) T (G/C) (T/A) GG G (T/C) GTATCTCTACAGCTATGT (G/C) AC (A/T) (A/G) T GGG (A/T) (C/G) AACT (G/C) T (A/G) (T/G) GCCTTC A (T/C) CACTGGCTGGACCCTCATACTGCATTACGCGT

Table S1. Continued

BF191147	237886 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence	0	CCTATATTCCATTC (G/A) TAGTTGAACCATGCGAGTACC
DR066496	rp11895 Normalized porcine adipose tissues cDNA library Sus scrofa cDNA, mRNA sequence	4e-170	C (T/A) TCCAGTGTGCTGTGGCTGTGGC (A/G) TAGG (T/C) C AGCA (C/G) CTGCAGCTCC (G/A) A (G/T) T (T/C) G A (T/C) CCCT (G/A) GCC (C/T) GGGAACCTCCATATGCTGC
XM_001493250	kelch repeat and BTB (POZ) domain containing 7	0	CT (G/T) TA (C/T) GC (C/T) (C/G) T (C/G) AA GCAAGCGCATG (T/C) TCTGCT (T/A) (T/C) GAT CCCAGCCA (C/T) AACATGTGGCTGAA (G/C) TG TGC (G/T) TCTCTAAACGCAGTGACTTCCAGGAG
XM_001929045	hypothetical LOC100155420	0	AAATCTAGTTTTTGA (T/C) TTGTTTTGTCTTGT
XM_585907	carbonic anhydrase XIV	1e-162	GTCTGTGTGAATAGC (A/G) AAGGAAGGAAGAGC TCGCCAGCAACCCAG (G/T) CAACCCCCCTCCC
EW646381	rthg04_l18.y1 thg Sus scrofa cDNA 5', mRNA sequence	0	ACAATCATTTTTTTTTC (T/C) TTCATTAGCCATTAT
BE234333	141171 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence	0	CCACA (G/A) AACATCTACTG (T/A) AACTGGCA GA (A/T) C (T/C) TAA (G/A) (C/T) CTC (A/C) AA
BQ597486	MI-P-A2-ae-y-h-04-1-UM,s1 MI-P-A2 Sus scrofa cDNA clone MI-P-A2-ae-y-h-04-1-UM 3', mRNA sequence	0	CACACAGCATGAA (G/C) TGATCAAGAAAAGCCACA CACAACCTGTTTA (G/C) ACTGATGAAGCTTCTAGGA GAACTGTGTC (A/G) TGAAAACCAAGTGAGGAAGAAG
BC120117	NDRG family member 4	0	AGGG (G/A) AACCG (C/G) CC (A/C) G (C/T) CATCC TGAC (G/A) T (G/T) GG (T/C) (C/A) T (C/G) AACCC ACAAG (C/A) (T/C) (G/C) TGCT (T/A) CA A (T/C) (A/C) CC (T/C) TCTTCAACT (T/C) (T/C) GAG
XM_001926775	similar to glycoprotein M6A	0	GAAAATGAAAGTGC (T/C) CCTCATCAATGATAT
BG610286	324863 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence	0	GATCAGGTGACAAG (C/T) TAACTGGCTACCAAC
AY216477	glutamate-ammonia ligase	2e-169	GGTGGGCCTTCCA (C/T) TCCTTAGAAGGTTAGA
EW345807	rplun0127_a19.y1 lun Sus scrofa cDNA 5', mRNA sequence	6e-104	GATTAAGGATGATG (G/A) TGTTTTGGGGAGAG
CA778985	MPL384_11_K21 MPL Sus scrofa cDNA clone pSPORT1 5', mRNA sequence	0	TTTAGGAAGCA (T/C) TTGATTAAC ATTTAGTAGTTA (C/A) AGAAATAG
BG894744	355473 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence	0	CTGTCCTTGCTCAGT (A/G) GGTTAAAGGATCTGG AGCTGCAGC (G/A) TAGGTTGCAG(C/A) TGTAGC
BE234314	141142 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence	0	GGAAATGCTGGTAG (T/C) CGTCAGTATGTCCATT
BC146251	synaptotagmin XI	1e-153	ATTGCCCTCACCGT (C/A) ATCCTTCTCCGATGTG
BC109983	nuclear transport factor 2	0	GAGGCCGGGTGAC (A/G) CTCCAGAATGGGAGAC AAGGAC (T/A) GCAATTCAGGG (G/A) AAAACTG GTGGGAGTGGGCGC (G/A) GTGCGCTGCTGCTAC
XM_001254701	calponin 3, acidic	0	ATCAGAAGCTAACAC (C/T) ACAACCCGTGGACAA
XM_001928448	similar to Mitogen-activated protein-binding protein-interacting protein	0	GGGATCAGGTTGTGA (G/A) AGCCCAGGGTTAGGA GCCACAGGCA (C/T) GCTGCGTCCCAAGGCTTTGA CTAAGCCAAGC (T/C) AACACTGGAGGTGTCCAGA GCACCCTGC (C/T) GCTGAATAACGAGGGATCTCT GCATGTATGCC (T/A) AGGAGACTGTTGGCTTCGG
AB436774	MHC class I antigen 2	0	G (A/G) CCCTGGCCCTGACCGGGACCCAGGCGGGT (T/C) C CCGCTTC (C/T) TCA (C/T) (C/T) GTCGG (G/C) TACGTG T (C/T) CGACAGCGACGCC (G/A) AAT (C/G) C (G/A) A (G/A) G (G/A) (A/T) GGAGCCGCGGGCGC (A/C) GTGGA TA (A/C) AGCAGGAGGGGC (C/A) GGAGTATTGGGATCGGGA GGA (A/C) ACCGCACAGACTT (A/T) CCGAG (T/G) G (G/A) (G/A) CCTG (A/C) (A/G) (G/C) AC (C/A) G CTC (G/T) CGGCTACTACAACCAGAGCGAGGCGGGTCTCACA (C/T) TCCAGA (G/C) (A/C) ATGTACGGCTGC (T/G) A (T/G) TGGG (A/G) CCAGACGGGCTCCTCCTCC (G/A) C ACA (T/C) GGCGGCTCAGATC (T/A) CCAAGCGCAAGTGGGA

Table S1. Continued

GGCGGCC (G/A) ATG (A/T) GGCGGAGCAG (C/G) AGAG GAGCTACCTGCAGGGCC (T/G) GTGTGTGGAGGGGCTCCAGAA GTCACCTTGAGGTGCTGG (A/G) CCCTGGGCTTCTACCCTAAGG A (C/G) CTGGCAGCGGGAGGGCCAGGACCAGAGCCA (A/G) G GACCTCCAGAAGTGGGC (G/A) GCCCTGGTGGTGCCCTCCTGG
