FIRST SURVEY AND FUNCTIONAL ANNOTATION OF PROHORMONE AND CONVERTASE GENES IN THE PIG

BY

KENNETH I. PORTER

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Animal Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Adviser:

Professor Sandra Rodriguez-Zas

ABSTRACT

Neuropeptides are small molecules that originate as a result of cleavage of prohormones by convertases as well as other posttranslational modifications of the prohormones. Neuropeptides are important for their roles in intercellular communication as well as regulation of many processes including growth, development, immune defense, and reproduction. The use of sequence homology to annotate genes that code for neuropeptides in a species that have been sequenced recently, such as the pig, is not ideal because of the variability of posttranslational modification of prohormones, from which neuropeptides are derived. So far only 40 prohormones have been confirmed empirically in the pig, as a result of this incomplete knowledge, this study was undertaken to detect prohormones and convertases utilizing the recent release of the pig genome coupled with a proven bioinformatics pipeline. Using 35 gene expression microarray experiments, prohormone and convertase genes that were identified were functionally annotated. As a result, 95 unique prohormone genes, 2 alternative calcitonin-related sequences, 7 prohormone convertases and 1 cleavage facilitator were identified in the pig genome 10.2 assembly and trace archives. The genome search performed identified 12 prohormone genes that were previously unreported in the UniGene, UniProt, and Gene databases. These genes are Intermedin (ADM2), Cortistatin (CORT), Insulin-like 5 (INSL5), Orexigenic neuropeptide QRFP (OX26), Prokineticin 2 (PROK2), Prolactin-releasing peptide (PRRP), neuropeptide S (NPS), Parathyroid hormone 2 (TIP39), Urocortin (UCN1), Urocortin 2 (UCN2), Urocortin 3 (UCN3) and Urotensin 2 related peptide (UTS2B). A UniProt record was found for Neuropeptide S, the result from our search, however, returned a more reliable result, as the existing entry was identical to the rabbit NPS. Furthermore, the gene expression experiments analyzed were partitioned into seven separate categories with the categories of experiments

containing the most differentially expressed prohormone and convertase genes being immune response tissues, the embryo and placenta, and brain and central nervous system. In the immune category, the general trend was for prohormone genes to be under-expressed when the pigs were immune challenged relative to the un-challenged control pigs. There were two trends in the embryo and placenta category, one was for prohormone genes to be over-expressed in younger relative to older embryos while the other trend was for prohormone genes to be over-expressed in pregnant sows relative to their non-pregnant counterparts. The tendency in the brain and central nervous system category was for prohormone genes to be over-expressed in neural tissue relative to non-neural tissue. The cleavages of the prohormones were also predicted using a variety of models; the human cleavage prediction models provided the highest performance with a correct classification rate of 92% which suggests that prohormone processing in pigs is similar to that of humans. These models also found no evidence for cleavage of CRSP2, UCN2, UCN3, TOR2X, TKN4, or IAPP to produce active peptides in the pig. The genomic and functional characterization presented supports the use of the pig as an animal model that is effective in furthering the understanding of prohormones, convertases, and neuropeptides in biomedical and agricultural research.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1: LITERATURE REVIEW

1.1 The Pig as a Model for Biomedical and Agricultural Research

Research in the fields of agriculture and biomedicine has used the pig since before 1966, when Bustad and McClellan published a review of recent research. This review showed that the pig was useful in areas such as genetic, dental, skeletal, and cardiovascular research and that projects had already started before that time to develop smaller pigs more tailored for biomedical research. Aside from the adult pigs, piglets are also advantageous in terms of research into early human development (Book and Bustad, 1974). These previous findings were confirmed in recent studies. In a recent review, Butler et al. (2009) showed the development of the piglet as a model for the immune system and enumerated the characteristics that make it the optimal model. Among the positive characteristics were the large litters produced by pigs, which can be raised apart from the mothers to facilitate separate experimental and control groups, and ease of study of fetal development offered by the long gestation period. Usefulness of the pig skeletal system has also been shown in a recent study that sought to determine the differences of using bypassed or non-bypassed cortical windows and their contribution to the risk of periprosthetic fracture (Wilson et al., 2011). The authors used pig femurs in place of human femurs because the mechanical properties of both are comparable. However, the morphology of the femurs is different, with the pig femur having a shorter length. In another study, which was concerned with the polycystic kidney disease 1 (*PKD1*) gene, a *PKD1* homolog from the pig was used to discover that, through cloning and characterization of the homolog sequence, that the pig gene is highly similar to the human gene (He et al., 2011). This gene was selected because mutations in it cause the majority of cases of autosomal dominant polycystic kidney disease (ADPKD) and

the pig was used because the mouse has shortcomings including a short lifespan that is not ideal for the study of ADPKD because the disease has a chronic nature. The pig is useful in neurological studies because of the large brain it possesses, and is the species used most frequently in organ xenotransplantation (Sorensen et al., 2011 and Ekser et al., 2011).

The use of pigs in agriculture also warrants extensive research, including the impact of pig farming on the environment (Basset-Mens and van der Werf, 2011). Of further importance are transgenic pigs that could produce less phosphorus in their waste, healthier meat, and more live piglets (Niemann and Kues, 2007). These techniques of increasing production of individual animals are important in contributing to the feasibility of overall production increase which is important to meet the demand for pork. According to the National Pork Producers Council (NPPC), in November of 2011 the pork industry in the United States was on a course to set a new record for exports (http://nppc.org/uploadedfiles/CPR_OCT_NOV_11.pdf). The demand for pork in the United States in 2012 is expected to increase because of beef and poultry supplies being smaller, this in conjunction with the projection that increase in production will be a result of more pigs produced per litter as opposed to herd expansion (http://www.porknetwork.com/pork-news/latest/Swine-herd-expansion-modest-despite-return-to-profitability-137686823.html). These projections underscore the importance of furthering our understanding of the pig at the genomic level in order to optimize production and reduce negative environmental impact. The pig is both an important model for biomedical research as well as a significant portion of the human diet worldwide. These qualities make the pig an ideal candidate for further investigation.

1.2 Pig Genome

The Swine Genome Sequencing Consortium, which was formed in late 2003, released the Sscrofa9 genome sequence which was derived from BAC clones in 2009 (Archibald et al.,

2010). Sscrofa10, a bacterial artificial chromosome (BAC) and whole-genome shotgun (WGS) derived assembly, was released in 2011. BACs are a bacterial cloning system based on the *E. coli* F plasmid which has a low likelihood for recombination between DNA fragments it carries as a result of the low copy number of the F plasmid (Shizuya et al., 1992). The bacterial DNA carrying F factors can be useful for cloning large fragments of DNA (Shizuya et al., 1992). Whole-genome shotgun sequencing is a method of determining nucleotide sequences in DNA that uses DNA polymerase as well as inhibitors to terminate synthesized chains (Sanger et al., 1977). DNA fragments of defined lengths are used to acquire reads which are assembled into contigs using overlapping sections of the sequence as a reference (Kaiser et al., 2003). The process was later modified by sequencing cloned inserts from both ends to get paired-end reads (Edwards et al., 1990). This version contained redundancy and mapping errors which were corrected in Sscrofa10.2 which was released in early 2012 (http://www.ncbi.nlm.nih.gov/projects /genome/guide/pig/, http://www.ncbi.nlm.nih.gov/projects/mapview/map_search.cgi?taxid =9823, Archibald et al., 2010 and Schook et al., 2005). The pig genome is diploid with 18 pairs of autosomes and 2 sex chromosomes for a total of 38 chromosomes, which is comparable to the human chromosome number. The pig genome is about 3 billion base-pairs (or letters of genetic code) and is estimated to encompass approximately 24,000 protein-coding genes, a number similar to the human (http://www.sanger.ac.uk/about/press/2006/060116.html). Among these genes, several hundred code for proteins and peptides that are involved in inter and intra cellular signaling.

1.3 Prohormones, Convertases and Neuropeptides

Neuropeptides

The messenger molecules that are involved in signaling between cells and control a variety of regulatory functions are called neuropeptides (Hummon et al., 2006). Neuropeptides result from the process of cleavage via enzyme of larger prohormone precursor molecules and are generally between 3 and 40 amino acids long (Hook et al., 2008). Neuropeptides include both peptide neurotransmitters and peptide hormones, although some neuropeptides function as both. The term neuropeptides stems from the direct effects that these molecules have on neurons. However, neuropeptides act on non-neural organs and tissues in addition to neurons (Strand, 1999). As more is learned about neuropeptides, the better their functions are understood; the insight that neuropeptides affect non-neural systems lead to awareness of one of the most important functions of neuropeptides, integration of the bodily systems with the function of the brain (Strand, 1999). This integration is demonstrated by the involvement of neuropeptides in the regulation of reproduction; food and water intake; cardiovascular, gastrointestinal, and respiratory control; water and salt metabolism; temperature control; growth; and memory (Strand, 1999). An example is vasoactive intestinal peptide (VIP) reported to be isolated in the 1970s. This neuropeptide was deemed unlike other peptides because it lacked glycine and proline residues (Said and Mutt, 1970). Early classification of neuropeptides relied mostly on where these molecules were found, as a result the new peptide, which was isolated in the intestine of a pig, was named VIP. Many neuropeptides have multiple roles; VIP is implicated in gastrointestinal secretion and vasodilation, as well as neuronal survival in the central nervous system (Strand, 1999). More recently, VIP has been shown to have anti-inflammatory properties as well (Deldago et al., 2004; Szema et al., 2011). Another neuropeptide, NPY, is very

important in the gastrointestinal tract and in obesity. NPY also enhances memory processes in mice, is involved in regulation of the cardiovascular system, and has stimulatory effects on the neuroendocrine system (Strand, 1999).

Prohormone Convertases

Prohormone convertases are precursor processing endopeptidases that cleave the active segments from prohormones at single or paired basic residues (Strand, 1999). Prohormone convertases are a family of proteases that are similar to subtilisin, a bacterial protease, based on strong evolutionary conservation demonstrated by extensive homology in the N-terminal sections (Strand, 1999). All prohormone convertases have a p-domain, an approximate 150 residue long region located downstream of the catalytic segment that is necessary for the production of active convertases; a catalytic segment; a prosegment; and a signal peptide (Zhou et al., 1998; Strand, 1999).

Prohormones

Neuropeptides are derived from the actions of prohormone convertases on prohormones. A prohormone is inert following the removal of the signal peptide, the remaining prohormone can contain one or more copies of the neuropeptide which, in order to become active, would require processing. Following separation from the signal peptide, the prohormone moves into the Golgi apparatus and further are sorted into granules from the *trans*-Golgi network. It is within these granules that the proteolysis begins (Strand, 1999). Prohormones have a variety of roles and are diverse in their size. However, the size of the prohormone does not determine the size of the derived neuropeptide. The enkephalins are an example, they are derived from proenkephalin which is a large prohormone and the resulting neuropeptides are only 5 amino acids in length (Strand, 1999).

1.4 The Central Dogma

The presence of any active neuropeptide in sufficient amount to affect inter-cellular signaling depends on the presence of corresponding gene in the genome, the appropriate nucleotide sequence (polymorphisms or sequence variants can alter the transcription or translation), transcription into mRNA and subsequent translation into a prohormone sequence that results in neuropeptides after cleavage and post-translational modification. This pipeline follows the ubiquitous central dogma of molecular biology. The protein is only an end product, as the protein to protein, protein to RNA, and protein to DNA transfers were thought to be very unlikely (Crick, 1970). This concept makes necessary the examination of both genes and proteins. The identification of genes and proteins in the genome and understanding of their role requires the consideration of multiple genomic and proteomic repositories and use of complementary bioinformatics tools.

1.5 Repositories of Sequence Information

UniProt. The UniProt database (Universal Protein Resource, http://www.uniprot.org/) is the result of collaboration between the EBI (European Bioinformatics Institute), the Swiss Institute of Bioinformatics, and the PIR (Protein Information Resource). UniProt is a centralized source of information on protein sequences, concentrating data that would otherwise be need to be found in a variety of large sources (The UniProt Consortium, 2010). The database primarily significant to this study, UniProtKB (The UniProt Knowledgebase) is composed of two distinct sections UniProtKB/Swiss-Prot and UniProtKB/TrEMBL. The Swiss-Prot section is made up of manually annotated records, while the TrEMBL section utilizes computer-assisted annotation. Two databases are required because of the volume of data that is generated. The manual annotation process demands significant time and labor commitments, and cannot be expected to

contend with all the available data. Therefore, a large portion of the new protein sequences need to be annotated automatically by computer in order to be publically available quickly (UniProt 2012; http://www.uniprot.org/faq/7). This process follows rules created using systems such as UniRule to annotate the translated sequences from EMBL, GenBank, and DDBJ (The UniProt Consortium 2010). Automatically annotated records are later selected to be manually annotated with priority given to model organisms (Uniprot 2012, http://www.uniprot.org/program/). The process of manual annotation follows the subsequent steps: sequence curation, sequence analysis, literature curation, family-based curation, evidence attribution, and quality assurance and integration of completed entries (UniProt 2012, http://www.uniprot.org/faq/45). The manual annotations include information on function, associated diseases, tissue specificity, structure, isoforms, interactions, and posttranslational modifications (O'Donovan and Apweiler, 2011). In this study, both sections of UniProtKB (release 2011_11) were considered.

Gene. The NCBI Gene database, also called Entrez Gene, is specific to gene information (Maglott et al., 2010). This database focuses primarily on genomes that are sequenced in their entirety that also have ongoing research to provide additional data (NCBI 2011). The data in the Gene database is the result of the Reference Sequence Projet (RefSeq) and uses unique identifiers (GeneID) for the loci and genes for specific organisms (Maglott et al., 2010). As of November 2010, the Gene database contained information on 6.7 million genes that represents in excess of 6,700 organisms (Sayers et al., 2010). The information that is consolidated includes chromosomal localization, associated markers and phenotypes, nomenclature, protein interactions and reports of pathways (Maglott et al., 2010).

UniGene. The purpose of UniGene is to produce an organized view of the transcriptome. This organization was necessary because the high redundancy of transcribed sequences makes using the data difficult (The NCBI Handbook, Ch. 21, http://www.ncbi.nlm.nih.gov/books/NBK2108 3/). The UniGene (http://www.ncbi.nlm.nih.gov/unigene, build #41) database contains mRNA sequences of genes that are well known as well as expressed sequence tag (EST) sequences. UniGene organizes this information into non-redundant clusters that are gene-oriented (Wheeler et al., 2003). This is to say that each cluster represents a unique gene, known or putative, in the organism the sequences are from (NCBI 2012, http://www.ncbi.nlm.nih.gov/books/NBK21106/ def-item/app187/). The procedures that UniGene is organized by are: removing foreign sequences, mitochondrial sequences, and rRNA from clusters; soft-masking repetitive sequences, allowing those sequences to be used in a sequence alignment although not to start a sequence alignment; and requiring clusters to be anchored at the 3' end to help prevent multiple clusters being identified for one gene (The NCBI Handbook, Ch. 21). The clusters also link to information on protein similarities, as well as tissues where the gene is expressed.

GenBank. The database GenBank (http://www.ncbi.nlm.nih.gov/genbank/) contains all public DNA sequences and is one third of the International Nucleotide Sequence Database Collaboration, which also includes DDBJ (the DNA DataBank of Japan) and ENA (the European Nucleotide Archive) (Benson et al., 2010). GenBank is built from sequence data submitted by individual authors as well as EST (Expressed Sequence Tags) and other high-throughput data (Benson et al., 2010). GenBank is organized into twenty divisions, eleven taxonomic divisions, seven high-throughput divisions, one PAT division with records from patent offices and one WGS division that constitute all sequences from whole genome shotgun projects (Benson et al., 2010).

Genome. The Genome database contains the most recent public build of the pig genome (Sscrofa10.2) along with the sequence and map data for over 1000 other genomes (NCBI 2011, http://www.ncbi.nlm.nih.gov/books/NBK3837/). In late 2011 the Genome database was restructured to provide simpler navigation. While the database previously contained separate records for strains and genome assemblies of organisms, the records are now consolidated by organism with all assemblies found in a central location. Also included in the changes is that the database has expanded to include INSDC genomes as well as RefSeq genomes (NCBI 2011, http://www.ncbi.nlm.nih.gov/About/news/17Nov2011.html). As of 2011, the Genome database contains complete genomes of 37 higher eukaryote species, and data obtained from more than 1,100 genome sequencing projects. The Map Viewer at NCBI can be used for a variety of applications within this database including displaying maps from multiple organisms or assemblies simultaneously in a single view (Sayers et al., 2012)

Traces. The Trace archive is a database that was established following the Human Genome Sequencing Project, as a result, traces of human origin account for only 12% of the database. As of 2012, the Trace Archive has data from over 10,000 species including EST libraries, as well as shotgun and BAC clone projects (Sayers et al., 2012)

EST. Expressed sequence tags (ESTs) are small sections of cDNA that correspond to mRNA (Adams et al., 1991). These cDNAs are produced by reverse transcribing mRNAs that are isolated on the basis of their 3' poly-A tails (Parkinson and Blaxter, 2009). After these cDNAs are cloned into a vector, single sequencing reactions are performed on some of the clones. Low quality sequencing information and vector sequence contamination are removed using bioinformatics pipelines downstream of the sequencing reaction, and the resulting sequence is catalogued in the dbEST (http://www.ncbi.nlm.nih.gov/nucest) database (Parkinson and Blaxter,

2009). The use of ESTs is beneficial both as an alternative process and complementary process to the more expensive procedure of full genome sequencing. As of 2009, there were more than 45 million ESTs generated from in excess of 1400 different species of Eukaryote (Parkinson and Blaxter, 2009). The number of ESTs per cluster in different cDNA libraries should correlate to gene expression levels provided the libraries have not been normalized. In non-normalized libraries, differential expression of genes between the libraries can be identified via statistical analysis, which is a useful alternative to approaches based on microarrays (Parkinson and Blaxter, 2009). The dbEST database is then clustered using BLAST programs for the UniGene database.

1.6 Tools to Identify and Characterize Sequences

BLAST. The Basic Local Alignment Search Tool (BLAST) is a group of programs that calculates the statistical significance of a match of a sequence of a protein or a nucleotide to its respective sequence database (NCBI 2012 http://blast.ncbi.nlm.nih.gov/Blast.cgi). Using maximal segment pair (MSP) score, which gauges local similarity of two sequences, and a heuristic approach, BLAST is able to find a relationship even if similarity occurs only in isolated sections of the sequence (Altschul et al., 1990). BLAST is similar to FASTA in that these two tools use a heuristic algorithm and search for words common to the sequences being used. However, BLAST filters the search by finding only those words that are most significant (Mount, 2007). The comparison of two sequences for homology allows extrapolation of function from the known sequence to the unknown sequence. Homology is assumed only when the statistical significance, expressed as an E-value, exceeds a predetermined threshold.

ClustalW. Clustal W is a multiple sequence alignment tool that can be used for multiple sequence or profile alignments (Thompson et al., 1994). The basic method for alignment used in

Clustal W consists of three steps: initially, a distance matrix that shows the divergence of sequence pairs is calculated and, all sequence pairs are separately aligned; then the distance matrix is used to calculate a guide tree; finally the branching order in the guide tree is used to progressively align all sequences (Thompson et al., 1994). Clustal W can be accessed online at the European Bioinformatics Institute (http://www.ebi.ac.uk/Tools/msa/clustalw2/).

Wise2. Wise2 is a tool used to parse gene sequence information using that GeneWise algorithm first detailed by Birney et al. in 2004. The GeneWise model was made to integrate a gene prediction model and a protein homology model. GeneWise allows comparison of one protein sequence to genomic DNA and accounts for sequencing errors and the known statistical properties of gene structures such as the poly(A) signal AAUAAA, the G+U rich site to recognize the poly(A) site at the beginning of the gene or the 3' splice acceptor site that is always AG (Zhang, 2002; Birney et al., 2004). The GeneWise algorithm can be represented in two pair-HMMs, the function of pair-HMMs is to convert one sequence of letters into another sequence. An example is the pair-HMMs S and T, where S maps letters from an alphabet A to letters in alphabet B, and T maps letters from alphabet B to letters in alphabet C. This example is present in GeneWise with A being the genomic sequence, B being the predicted protein sequence, and C being the homologous protein sequence used to guide gene prediction (Birney et al., 2004). The method the GeneWise uses to handle insertion and deletion errors is deliberately naïve, because correctly handling the errors is difficult and expensive from a calculation standpoint (Birney et al., 2004). The GeneWise algorithm was developed by combining two highly accurate hidden Markov models (HMMs).Wise2 can be used to predict the gene, gene structure and associated protein on a region in the genome that could be identified using BLAST.

1.7 Prediction of Cleavage Sites in Prohormones

Known-motif Model. The known-motif model, which is detailed in Southey et al., (2006a), outlines five basic cleavage motifs. These cleavage motifs are x-x-K-K, x-x-K-R, x-x-R-R, R-x-x-K, and R-x-x-R where K, R, and x represent Lysine, Arginine, and any amino acid, respectively. These motifs are found prior to cleavage sites, such that the cleavage would occur immediately after the last R or K. The probability of cleavage is approximately 0.88 and 0.997 when there were one and multiple motifs present, respectively (Southey et al., 2006). Multiple motifs would be present simultaneously and not in sequence, such an instance could occur where x-x-K-K and R-x-x-K would combine to become R-x-K-K and give a 0.997 probability of cleavage.

Logistic and Neural Network Models. The use of a logistic regression model to predict cleavage sites was first proposed by Hummon et al. (2003) and applied to an Aplysia dataset. Amare et al. (2006) showed a logistic regression model that was more sensitive and for mammalian precursors. As explained in Tegge et al. (2008) the probability of cleavage is denoted π_i which applies to the *i*th window, while the regression coefficient β_j corresponds to the *j*th model term (j=1 to p, amino acid or amino acid property at peri-cleavage locations) and x_{ij} represents the presence or absence of that *j*th model term in the *i*th window. This is represented in the following binary logistic regression model:

$$log\left[\frac{\pi_i}{1-\pi_i}\right] = \sum_{j=1}^{p} \beta_j x_{ij}$$

The human cleavage model is based on the human genome and the cleavage information is based on a large number of precursors and neuropeptides that have been confirmed (Tegge et al., 2008). This level of information is also true of the mouse and the rat which are included in the

mammalian model along with the human, cow, and pig information. The use of models that include experimentally derived cleavage information helps avoid pitfalls of using sequence homology and motif finding in determining cleavage sites (Tegge et al., 2008).

Neuropred. Neuropred is a neuropeptide cleavage site prediction tool available on the internet at http://neuroproteomics.scs.uiuc.edu/neuropred.html (Southey et al., 2006b). Neuropred uses known motif, and logistic regression models in order to predict cleavages and can also estimate the mass of the predicted peptides as well as calculate the accuracy of the model (Southey et al., 2006b).

1.8 Microarray Gene Expression Experiments

A gene expression microarray is a platform in which a series of gene-specific probes are affixed to a substrate such as glass or plastic. This platform is used to analyze the expression of thousands of genes simultaneously. The probes are segments of single stranded DNA which will hybridize to their complimentary DNA or RNA if that complimentary sequence is present in the sample added to the microarray (Trachtenberg et al., 2012). The probes can be PCR products, oligos from 25 to 85 base pairs, or gene fragments (Trachtenberg et al., 2012). Microarrays can use one or two dyes, these fluorescent labels are used to indicate the locations where the DNA/RNA from the sample has hybridized to the probe on the array. Methods that use one dye consider transcripts from a single sample whereas two dye methods are used to examine differential expression of genes on one microarray by assigning one dye to the experimental group and the other dye to the control group. Results derived from the one dye methods tend to be more consistent among replicates than the two dye methods (Trachtenberg et al., 2012; Kuo et al., 2006).

Affymetrix. The Affymetrix in situ synthesized oligonucleotide arrays are widely used microarray platforms available for a wide range of species. Affymetrix arrays are one dye platforms which have superior performance compared to two dye platforms as well as more consistent results compared to arrays produced in-house at academic institutions (Liu et al., 2012). There are several approaches to normalize data from microarray experiments including: RMA, GCRMA, quantile, and log2. The log2 transformation normalizes microarray data that has a tendency to be skewed. As the name implies, the log2 transformation involves taking the log2 of all values which results in a relatively normal distribution of data. The log2 method has several drawbacks including the fact that it is a one-size-fits-all solution that ignores potential noise differences associated with separate runs (Lin et al., 2008). Quantile normalization is used to make the distribution of data on the individual arrays identical. The process involves standardizing data, arranging measurements on each array from lowest to highest, calculating a new distribution using the average of all arrays for each feature, and replacing all original measurements with newly calculated averages (Bolstad et al., 2003). RMA (robust multichip analysis) is an additive model based on quantile normalized and background corrected PM s (perfect matches). PMs are oligos that are exactly complementary to genes, while MMs (mismatches) are identical to PMs but have transversions in the middle. This model can be represented as:

$$T(PM_{ij}) = e_i + \alpha_j + \varepsilon_{ij}$$

In this model *T* represents the transformation to normalize, log, and background correct PM intensities; $e_i$represents the log2 value on the arrays; α_j represents log affinity effects; and ε_{ij} represents the errors (Irizarry et al., 2003). To estimate e_i, median polish, a robust linear regression, is used. The GCRMA (GeneChip Robust Multichip average) is a model very similar

to and an improvement on RMA. The difference between the models is that GCRMA considers the sequence specificity of the oligos in the background correction step (Binder et al., 2010). These normalizations are available the software suite Beehive (http://stagbeetle.animal.uiuc.edu/Beehive/).

1.9 Normalization and Analysis of Microarray Experiments

Preprocessing, Normalization and Analysis

Beehive. Beehive (http://stagbeetle.animal.uiuc.edu/Beehive/) is a set of tools for management, statistical analysis, and query and interpretation of microarray experiment data. Beehive can be used to preprocess and normalize the data gathered from microarray gene expression experiments as well as analyze them to determine which probes displayed differential expression. The normalized gene expression data can be analyzed using ANOVA (analysis of variance) models. ANOVA can be used to identify differential gene expression between levels of one (one-way ANOVA) or multiple factors such as tissue and age (multifactor ANOVA).ANOVA is useful in considering the possible multiple sources of variation in microarray experiments (Churchill, 2004). The multifactor ANOVA model can be written as:

$$y_{ijk} = \mu + a_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

in which y_{ijk} is the response of the kth subject to the ith level of factor α and the jth level of factor β; μ is the overall effect, a_i is the effect of the ith level of factor α; β_j is the effect of the jth level of factor ß, $(\alpha\beta)_{ij}$ is the interaction of factors α and ß; and ε_{ijk} is the error.

Gene Expression Omnibus. The Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) is a repository for gene expression data that was established in 2000. GEO is the largest all public resource of gene expression data holding more than 50,000 samples from over 100 organisms contributed by nearly 1,500 laboratories (Barrett and

Edgar, 2006). The Affymetrix in situ oligonucleotide platforms was the most commonly used platform across multiple species including pig (GEO GPL3533)

Current State of Pig Neuropeptide Research and Objectives

The pig is a biomedical model and neuropeptides have a major role in health, behavior and traits of agricultural importance. Lines of research on neuropeptide that have been studied in pigs include defining the distribution of NPS and the corresponding receptor (NPSR), in which NPS was found to be highly expressed in the central nervous system and brain (Yao et al., 2009). A study to determine how ACTH responds differently to hemorrhage when several different anesthetics are used found that ACTH concentration increased when propofol or pentobarbitone were used but not when alphaxalone-alphadolone was used (Ruane-O'Hora et al., 2011). Examining the impact that both severe sepsis and hemofitration has on VIP concentrations in plasma and tissues, Kuncova et al. (2011) found that sepsis caused concentrations of VIP in the mesenteric artery and plasma to increase while simultaneously decreasing in the coronary artery. Kuncova et al. also determined that hemofiltration decreased VIP concentration in the mesenteric artery. A study in 2010 explored the effect that LEPR (leptin receptor) genotype has on expression of NPY, CART, and LEPR in the hypothalamus. Ovilo et al. (2010) determined that NPY was over-expressed in the LEPR TT genotype in the Iberian-Duroc backcross, and that CART was under-expressed in the Iberian-Landrace backcross, the T allele having a negative additive effect. Robich et al. (2010) examined the role NPY has in angiogenesis during chronic myocardial ischemia. The result was that exogenous NPY improved myocardial function and up-regulated pro-angiogenic receptors, increasing angiogenesis and arteriogenesis. Hausman et al. (2008) investigated the gene expression of NPY as well as other secreted factors in adipose

tissue. In this experiment, in middle subcutaneous adipose tissue expression of NPY was decreased significantly with age. Some understanding of the role of neuropeptides in the pig has been uncovered. However, few prohormones that produce neuropeptides following the cleavage by prohormone convertases have been confirmed experimentally in the pig. This situation can be resolved by integrating information from the recently released pig genome sequence with gene expression experiments and bioinformatics tools. The objectives of this study were: a) to identify prohormone and prohormone convertase genes in the pig genome, b) to functionally annotate the identified genes using gene expression microarray data, c) to predict the cleavage sites of prohormones and determining the resulting neuropeptides.

REFERENCES

1. Bustad LK, McClellan RO: **Swine in biomedical research.** Science 1966, **152**:1526-1530.
2. Book SA, Bustad LK: **The fetal and neonatal pig in biomedical research.** J Anim Sci 1974, **38**:997-1002.
3. Butler JE, Lager KM, Splichal I, Francis D, Kacskovics I, Sinkora M, Wertz N, Sun J, Zhao Y, Brown WR, DeWald R, Dierks S, Muyldermans S, Lunney JK, McCray PB, Rogers CS, Welsh MJ, Navarro P, Klobasa F, Habe F, Ramsoondar J: **The piglet as a model for B cell and immune system development.** Vet Immunol Immunopathol 2009, **128**:147-170.
4. Wilson LJ, Richards CJ, Irvine D, Tillie A, Crawford RW: **Risk of periprosthetic femur fracture after anterior cortical bone windowing: a mechanical analysis of short versus long cemented stems in pigs.** Acta Orthop 2011, **82**:674-678.
5. He J, Wang Q, Ye J, Hu X, Li N: **Identification of porcine polycystic kidney disease 1 (PKD1) gene: molecular cloning, expression profile, and implication in disease model.** Gene 2011, **490**:37-46.
6. Sorensen JC, Nielsen MS, Rosendal F, Deding D, Ettrup KS, Jensen KN, Jorgensen RL, Glud AN, Meier K, Fitting LM, Moller A, Alstrup AK, Ostergaard L, Bjarkam CR: **Development of neuromodulation treatments in a large animal model--do neurosurgeons dream of electric pigs?** Prog Brain Res 2011, **194**:97-103.
7. Ekser B, Gridelli B, Veroux M, Cooper DK: **Clinical pig liver xenotransplantation: how far do we have to go?** Xenotransplantation 2011, **18**:158-167.

8. Basset-Mens C, van der Werf H: **Scenario-based environmental assessment of farming systems: the case of pig production in France.** Agriculture, Ecosystems & Environment 2011, **105**:127-144.
9. Niemann H, Kues WA: **Transgenic farm animals: an update.** Reprod Fertil Dev 2007, **19**:762-770.
10. **The National Pork Producers Council** [http://nppc.org/uploadedfiles/CPR_OCT_NOV_11.pdf]
11. **Pork Network** [http://www.porknetwork.com/pork-news/latest/Swine-herd-expansion-modest-despite-return-to-profitability-137686823.html]
12. Archibald AL, Bolund L, Churcher C, Fredholm M, Groenen MA, Harlizius B, Lee KT, Milan D, Rogers J, Rothschild MF, Uenishi H, Wang J, Schook LB, Swine Genome Sequencing Consortium: **Pig genome sequence--analysis and publication strategy.** BMC Genomics 2010, **11**:438.
13. Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M: **Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector.** Proc Natl Acad Sci U S A 1992, **89**:8794-8797.
14. Sanger F, Nicklen S, Coulson AR: **DNA sequencing with chain-terminating inhibitors.** Proc Natl Acad Sci U S A 1977, **74**:5463-5467.
15. Kaiser O, Bartels D, Bekel T, Goesmann A, Kespohl S, Puhler A, Meyer F: **Whole genome shotgun sequencing guided by bioinformatics pipelines--an optimized approach for an established technique.** J Biotechnol 2003, **106**:121-133.

16. Edwards A, Voss H, Rice P, Civitello A, Stegemann J, Schwager C, Zimmermann J, Erfle H, Caskey CT, Ansorge W: **Automated DNA sequencing of the human HPRT locus.** Genomics 1990, **6**:593-608.
17. **Pig Genome Resources** [http://www.ncbi.nlm.nih.gov/projects/genome/guide/pig/]
18. **NCBI Map Viewer** [http://www.ncbi.nlm.nih.gov/projects/mapview/map_search.cgi?taxid=9823]
19. Schook LB, Beever JE, Rogers J, Humphray S, Archibald A, Chardon P, Milan D, Rohrer G, Eversole K: **Swine Genome Sequencing Consortium (SGSC): a strategic roadmap for sequencing the pig genome.** Comp Funct Genomics 2005, **6**:251-255.
20. **Wellcome Trust Sanger Institute** [http://www.sanger.ac.uk/about/press/2006/060116.html]
21. Hummon AB, Amare A, Sweedler JV: **Discovering new invertebrate neuropeptides using mass spectrometry.** Mass Spectrom Rev 2006, **25**:77-98.
22. Hook V, Funkelstein L, Lu D, Bark S, Wegrzyn J, Hwang SR: **Proteases for processing proneuropeptides into peptide neurotransmitters and hormones.** Annu Rev Pharmacol Toxicol 2008, **48**:393-423.
23. Strand FL: *Neuropeptides.* Massachusetts Institute of Technology, Cambridge, Massachusetts; 1999.
24. Said SI, Mutt V: **Polypeptide with broad biological activity: isolation from small intestine.** Science 1970, **169**:1217-1218.
25. Delgado M, Pozo D, Ganea D: **The significance of vasoactive intestinal peptide in immunomodulation.** Pharmacol Rev 2004, **56**:249-290.

26. Szema AM, Hamidi SA, Koller A, Martin DW: **Vasoactive Intestinal Peptide Knockout (VIP KO) Mouse Model of Sulfite-Sensitive Asthma: Up-regulation of Novel Lung Carbonyl Reductase.** BMC Immunol 2011, **12**:66.

27. Zhou A, Martin S, Lipkind G, LaMendola J, Steiner DF: **Regulatory roles of the P domain of the subtilisin-like prohormone convertases.** J Biol Chem 1998, **273**:11107-11114.

28. Crick F: **Central dogma of molecular biology.** Nature 1970, **227**:561-563.

29. **UniProt** [http://www.uniprot.org/]

30. UniProt Consortium: **The Universal Protein Resource (UniProt) in 2010**. Nucleic Acids Res 2010, **38**:D142-8.

31. **UniProt** [http://www.uniprot.org/faq/7]

32. **Uniprot** [http://www.uniprot.org/program/].

33. **UniProt** [http://www.uniprot.org/faq/45]

34. O'Donovan C, Apweiler R: **A guide to UniProt for protein scientists.** Methods Mol Biol 2011, **694**:25-35.

35. Maglott D, Ostell J, Pruitt KD, Tatusova T: **Entrez Gene: gene-centered information at NCBI.** Nucleic Acids Res 2011, **39**:D52-7.

36. Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Federhen S, Feolo M, Fingerman IM, Geer LY, Helmberg W, Kapustin Y, Landsman D, Lipman DJ, Lu Z, Madden TL, Madej T, Maglott DR, Marchler-Bauer A, Miller V, Mizrachi I, Ostell J, Panchenko A, Phan L, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Shumway M, Sirotkin K, Slotta D, Souvorov A, Starchenko G, Tatusova TA, Wagner L, Wang Y, Wilbur WJ, Yaschenko E, Ye J:

Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 2011, **39**:D38-51.

37. Pontius J, Wagner L, Schuler G: The NCBI Handbook, Ch. 21, [http://www.ncbi.nlm.nih.gov/books/NBK21083/]
38. **UniGene** [http://www.ncbi.nlm.nih.gov/unigene]
39. Wheeler DL, Church DM, Federhen S, Lash AE, Madden TL, Pontius JU, Schuler GD, Schriml LM, Sequeira E, Tatusova TA, Wagner L: **Database resources of the National Center for Biotechnology.** Nucleic Acids Res 2003, **31**:28-33
40. **UniGene Cluster** [http://www.ncbi.nlm.nih.gov/books/NBK21106/def-item/app187/]
41. **GenBank** [http://www.ncbi.nlm.nih.gov/genbank/]
42. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW: **GenBank.** Nucleic Acids Res 2011, **39**:D32-7.
43. **Entrez Help** [http://www.ncbi.nlm.nih.gov/books/NBK3837/]
44. **NCBI** [http://www.ncbi.nlm.nih.gov/About/news/17Nov2011.html]
45. Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, Chetvernin V, Church DM, Dicuccio M, Federhen S, Feolo M, Fingerman IM, Geer LY, Helmberg W, Kapustin Y, Krasnov S, Landsman D, Lipman DJ, Lu Z, Madden TL, Madej T, Maglott DR, Marchler-Bauer A, Miller V, Karsch-Mizrachi I, Ostell J, Panchenko A, Phan L, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Shumway M, Sirotkin K, Slotta D, Souvorov A, Starchenko G, Tatusova TA, Wagner L, Wang Y, Wilbur WJ, Yaschenko E, Ye J: **Database resources of the National Center for Biotechnology Information.** Nucleic Acids Res 2012, **40**:D13-25.

46. Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF: **Complementary DNA sequencing: expressed sequence tags and human genome project.** Science 1991, **252**:1651-1656.

47. Parkinson J, Blaxter M: **Expressed sequence tags: an overview.** Methods Mol Biol 2009, **533**:1-12.

48. **EST** [http://www.ncbi.nlm.nih.gov/nucest]

49. **NCBI BLAST** [http://blast.ncbi.nlm.nih.gov/Blast.cgi]

50. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** J Mol Biol 1990, **215**:403-410.

51. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** Nucleic Acids Res 1994, **22**:4673-4680.

52. **Clustal W- Multiple Sequence Alignment** [http://www.ebi.ac.uk/Tools/msa/clustalw2/]

53. Birney E, Clamp M, Durbin R: **GeneWise and Genomewise.** Genome Res 2004, **14**:988-995.

54. Zhang MQ: **Computational prediction of eukaryotic protein-coding genes.** Nat Rev Genet 2002, **3**:698-709.

55. Southey BR, Amare A, Zimmerman TA, Rodriguez-Zas SL, Sweedler JV: **NeuroPred: a tool to predict cleavage sites in neuropeptide precursors and provide the masses of the resulting peptides.** Nucleic Acids Res 2006, **34**:W267-72.

56. Southey BR, Rodriguez-Zas SL, Sweedler JV: **Prediction of neuropeptide prohormone cleavages with application to RFamides.** Peptides 2006, **27**:1087-1098.

57. Hummon AB, Hummon NP, Corbin RW, Li L, Vilim FS, Weiss KR, Sweedler JV: **From precursor to final peptides: a statistical sequence-based approach to predicting prohormone processing.** J Proteome Res 2003, **2**:650-656.

58. Amare A, Hummon AB, Southey BR, Zimmerman TA, Rodriguez-Zas SL, Sweedler JV: **Bridging neuropeptidomics and genomics with bioinformatics: Prediction of mammalian neuropeptide prohormone processing.** J Proteome Res 2006, **5**:1162-1167.

59. Tegge AN, Southey BR, Sweedler JV, Rodriguez-Zas SL: **Comparative analysis of neuropeptide cleavage sites in human, mouse, rat, and cattle.** Mamm Genome 2008, **19**:106-120.

60. **Neuropred** [http://neuroproteomics.scs.uiuc.edu/neuropred.html]

61. Trachtenberg A, Robert J, Abdalla A, Fraser A, He S, Lacy J, Rivas-Morello C, Truong A, Hardiman G, Ohno-Machado L, Liu F, Hovig E, Kuo W: **A Primer on the Current State of Microarray Technologies.** In *Next Generation Microarray Bioinformatics. Volume 802.* Edited by Wang J. New York: Springer; 2012: 3-17.

62. Kuo WP, Liu F, Trimarchi J, Punzo C, Lombardi M, Sarang J, Whipple ME, Maysuria M, Serikawa K, Lee SY, McCrann D, Kang J, Shearstone JR, Burke J, Park DJ, Wang X, Rector TL, Ricciardi-Castagnoli P, Perrin S, Choi S, Bumgarner R, Kim JH, Short GF,3rd, Freeman MW, Seed B, Jensen R, Church GM, Hovig E, Cepko CL, Park P, Ohno-Machado L, Jenssen TK: **A sequence-oriented comparison of gene expression measurements across different hybridization-based technologies.** Nat Biotechnol 2006, **24**:832-840.

63. Liu F, Kuo WP, Jenssen T, Hovig E: **Performance Comparison of Multiple Microarray Platforms for Gene Expression Profiling.** In *Next Generation Microarray Bioinformatics. Volume 802.* Edited by Wang J. New York: Springer; 2012:141-155.

64. Lin SM, Du P, Huber W, Kibbe WA: **Model-based variance-stabilizing transformation for Illumina microarray data.** Nucleic Acids Res 2008, **36**:e11.

65. Bolstad BM, Irizarry RA, Astrand M, Speed TP: **A comparison of normalization methods for high density oligonucleotide array data based on variance and bias.** Bioinformatics 2003, **19**:185-193.

66. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP: **Summaries of Affymetrix GeneChip probe level data.** Nucleic Acids Res 2003, **31**:e15.

67. Binder H, Preibisch S, Berger H: **Calibration of microarray gene-expression data.** Methods Mol Biol 2010, **576**:375-407.

68. **Beehive** [http://stagbeetle.animal.uiuc.edu/Beehive/]

69. Churchill GA: **Using ANOVA to analyze microarray data.** BioTechniques 2004, **37**:173-5, 177.

70. **GEO** [http://www.ncbi.nlm.nih.gov/geo/]

71. Barrett T, Edgar R: **Gene expression omnibus: microarray data storage, submission, retrieval, and analysis.** Methods Enzymol 2006, **411**:352-369.

72. Yao Y, Lin X, Su J, Yang G, Hou Y, Lei Z: **Cloning and distribution of neuropeptide S and its receptor in the pig.** Neuropeptides 2009, **43**:465-481.

73. Ruane-O'Hora T, Hall WJ, Markos F: **The effect of alphaxalone-alphadolone, propofol, and pentobarbitone anaesthesia on the beta-endorphin and ACTH response to haemorrhage in the pig.** Can J Physiol Pharmacol 2011, **89**:521-526.

74. Kuncova J, Chvojka J, Sykora R, Sviglerova J, Stengl M, Nalos L, Krouzecky A, Matejovic M: **Tissue concentrations of vasoactive intestinal peptide are affected by peritonitis-induced sepsis and hemofiltration in pigs.** Physiol Res 2011, **60**:531-540.

75. Ovilo C, Fernandez A, Fernandez AI, Folch JM, Varona L, Benitez R, Nunez Y, Rodriguez C, Silio L: **Hypothalamic expression of porcine leptin receptor (LEPR), neuropeptide Y (NPY), and cocaine- and amphetamine-regulated transcript (CART) genes is influenced by LEPR genotype.** Mamm Genome 2010, **21**:583-591.

76. Robich MP, Matyal R, Chu LM, Feng J, Xu SH, Laham RJ, Hess PE, Bianchi C, Sellke FW: **Effects of neuropeptide Y on collateral development in a swine model of chronic myocardial ischemia.** J Mol Cell Cardiol 2010, **49**:1022-1030.

CHAPTER 2: FIRST SURVEY AND FUNCTIONAL ANNOTATION OF PROHORMONE AND CONVERTASE GENES IN THE PIG

2.1 Introduction

In addition to its importance in livestock production, the pig is a well-established biomedical model to study human health due to the genomic, anatomic and physiologic similarities with humans. A wide variety of human health traits including cancer, reproductive health, drug metabolism, wound healing, and cardiovascular disease have been successfully studied using the pig (Book and Bustad, 1974; Anzenbacherova et al., 2003; Sullivan et al., 2001; Kurahashi et al., 2004). Underlying these and other important traits such as growth and development, stress, memory and substance abuse are neuropeptides, a class of cell-cell signaling peptides that have paracrine, endocrine, and autocrine effects (Fricker, 2005; Hook, 2008). Research in selected pig neuropeptides has offered insights into cell transplantation, nervous system diseases, and injury (Schwartz et al., 2005). For example, Yang et al. (2010) reported on the inhibitory effect of *neuromedin S* (*NMS*) on luteinizing hormone secretion which is mediated via melanocyte-stimulating hormone (MSH) neurons in the hypothalamus-pituitary axis of ovariectomized pigs. Kaminski et al. (2010) concluded that orexin A and B, hypothalamic peptides involved in the control of food intake, sleep patterns, autonomic and neuroendocrine systems, may also affect reproductive functions through the influence on the hypothalamic-pituitary-ovarian axis in pigs.

The identification of neuropeptides is a difficult process because neuropeptides are derived from larger prohormone proteins as a result of complex processing. The conversion of the large prohormone to one or multiple smaller neuropeptides involves cleavage by one or many prohormone convertases and additional post-translational modifications such as amidation and

glycosilation (Fricker et al., 2005). This complex processing of prohormones into neuropeptides challenges the identification of neuropeptide genes across genomes solely based on sequence homology to better understood species (Fricker, 2005; Hook , 2008).

Only 40 prohormone and two prohormone convertase genes have been empirically confirmed in the pig compared to approximately 100 genes identified in human, rat, mouse, cow and rhesus monkey (Southey et al., 2008; Southey et al., 2009; Tegge et al., 2008). This partial knowledge of the prohormone, prohormone convertase and associated neuropeptides in the pig is a critical shortcoming, especially considering the important role of pig in biomedical and agricultural research (Book and Bustad, 1974). In addition, few gene expression studies have discussed the expression profile of prohormone or prohormone convertase genes. Ross et al. (2007) found that estradiol treatment at day 9 of gestation was associated with changes in the expression of the prohormone *neuromedin* (*NMB*) in the endometrium of glits. Hausman et al., (2008) concluded that the expression of *neuropeptide Y* (*NPY*) was down-regulated with age in gilts ranging from 90 to 210 days old.

Understanding the role of neuropeptides in human and livestock traits using the pig as biomedical model requires a more comprehensive knowledge of the neuropeptide complement in the recently released pig genome (SScrofa 10.2v18, http://pre.ensembl.org/Sus_scrofa/Location/Genome). This understanding includes the identification of prohormone and prohormone convertase genes, prediction of cleavage sites in prohormones that may result in potentially bioactive neuropeptides, and characterization of gene expression and protein abundance across conditions to gain insights into the role of neuropeptides. A complete survey of the prohormone and prohormone convertase genes in the pig is critical not only to support the interpretation of gene expression experiments and the

effectiveness of tandem mass spectrometry studies to identify neuropeptides (Li and Sweedler, 2008). Functional annotation of these genes can be obtained by the analysis of the large number of gene expression experiments already available (e.g. Ross et al., 2009; Chen et al., 2009). To address the lack of a comprehensive understanding of the prohormone and prohormone convertase genes in the pig, we present the first comprehensive survey and functional annotation of these genes. An all-inclusive catalogue of prohormone and prohormone convertase genes known in other species was used to search complementary pig genome databases. These genes were then characterized by analyzing a large number of gene expression experiments across a wide range of conditions. The potential cleavage sites of prohormones that can result in bioactive neuropeptides were predicted and compared to the cleavages based on known neuropeptide sequences.

2.2 Methods

Search for Pig Prohormone and Convertase Genes

A registry of approximately 100 candidate mammalian prohormone and convertase genes was built from public proteomic and genomic databases (including NCBI Gene –release date September 2011, UniGene – release date April 13 2011, and UniProt –release 2011_11 November 16, 2011) and a literature review (Amare et al, 2006; Southey et al., 2008; Southey et al., 2009; Delfino et al., 2010).

Candidate genes were searched for in the pig genome 10.2 assembly using the bioinformatics pipeline developed by Southey et al. (2009). The protein sequence of each candidate gene in the registry was searched on the pig genome assembly using the NCBI BLAST (version 2.18) with default parameters settings (E-value < 10 and BLOSUM62 scoring matrix) and filtering disabled.

In addition, sequences not used in the pig genome assembly (including unassigned genomic regions, whole genome shotgun sequencing and trace archives) were searched when there was no suitable BLAST match to a candidate gene or when the alignment to the genome assembly suggested a missing genomic region. This strategy allowed the annotation of genomic regions that were partly or not included in the assembly.

The BLAST matches were examined based on the alignment score and E-value to identify the most likely matches and genomic location of the corresponding prohormone. The identified pig genomic region that encompassed the BLAST match was further extended approximately 500 base pairs to the 5' and 3' ends of the match. Matches were also screened for alignments to multiple homologous prohormone genes that could indicate gene duplication events in the pig genome. The gene parsing tool Wise2 was used to predict the protein sequence within the genome regions detected with BLAST (http://www.ebi.ac.uk/Wise2, Birney et al., 2004). In this study, Wise2 compared the target protein (preference was given to pig protein sequences, followed by human, cattle and other mammals) to the pig genomic DNA sequence identified by BLAST to infer the gene structure based on a model that includes introns and frameshift errors. Each predicted gene was compared to the UniProt (http://www.uniprot.org/) and NCBI Gene (http://www.ncbi.nlm.nih.gov/gene/)databases to assess the accuracy of the prediction based on previously reported pig genes. To further confirm the Wise2 predictions, the protein sequence predicted from the gene model was also compared to the corresponding published mammalian sequences using the multiple sequence alignment tool Clustalw (www.ebi.ac.uk/Tools/msa/clustalw2/). The predicted sequences were also searched against the pig EST database to confirm the presence of the predicted protein sequence. The EST database was also used to complete the protein sequence when the genome coverage was incomplete.

Functional Annotation of the Pig Prohormone and Convertase Genes

A review of the pig microarray gene expression experiments available in the NCBI GEO database (http://www.ncbi.nlm.nih.gov/geo/) indicated that the Affymetrix Porcine Genome Array GPL3533 was the most commonly used platform (http://www.affymetrix.com/support/technical/byproduct.affx?product=porcine). The UniGene database was searched for sequences that represent prohormone and prohormone convertase genes. This information was used to identify the probes representing prohormone and prohormone convertase genes in the Affymetrix Porcine Genome platform.

Thirty-five experiments were identified in GEO that used The Affymetrix Porcine Genome platform. Selected experiments had a minimum of 6 microarrays and a maximum of 80 microarrays. Table 1 summarizes the main features of these experiments. The wide range of selected microarray experiments (Table 1) available supported a comprehensive characterization of the association of prohormone and associated neuropeptide and convertase genes with various biological processes.

The experiments were grouped into 7 classes: primary immune-response tissues, embryo and placenta, brain and central nervous system, reproduction, muscle, fat, and gut. Within each class, the number of experiments including GEO series identification and tissue identification whenever multiple tissues were studied was immune was: 6 (GSE7313, GSE7314, GSE11787, GSE17492, GSE14758-mediastinal lymph nodes, and GSE14790), embryo and placenta: 5 (GSE18467, GSE18641, GSE18343, GSE11853, and GSE12705), brain and nervous system: 5 (GSE16855, GSE12604, GSE14739-hypothalamus, GSE14739-thyroid, and GSE14739-adenohypophsis), reproduction: 2 (GSE11590, and GSE14739-gonads), muscle: 7 (GSE18653,

GSE19275, GSE8974, GSE14643, GSE15211, GSE21096, and GSE16348-skeletal muscle), fat: 8 (GSE17309, GSE14373, GSE14739-fat, GSE9333, GSE18359-fat, GSE18359-liver, GSE13528-fat and GSE13528-liver) and gut: 2 (GSE14357 and GSE15256). When experiments encompassed multiple tissues (GSE14739, GSE18359, GSE13528), the samples corresponding to each tissue were analyzed separately to facilitate the interpretation of results.

The gene expression data was pre-processed and normalized using the Affy R package (http://www.biostat.jhsph.edu/~ririzarr/affy/). Steps included the log-2 transformation and GC-robust multichip average normalization of the gene expression measurements. All probes in the platform were analyzed using ANOVA to identify those that exhibited differential expression across the conditions studied. The false discovery rate (Benjamini and Hochberg, 1995) approach was used to adjust the statistical significance of the differential expression and account for multiple testing across all probes. The normalization, one or two-way ANOVA and multiple test adjustment of the results were done using Beehive (http://stagbeetle.animal.uiuc.edu/Beehive). Due to the multiple probes analyzed, a minimum false discovery rate, multiple-test adjusted P-value < 0.05 threshold (corresponds to an approximate unadjusted P-value < 0.005) and a minimum fold-change equal to 1.2 were used to identify differentially expressed prohormone and prohormone convertase genes.

Prediction of Cleavage Sites

The location of the cleavage in pig prohormone proteins that would result in potentially active neuropeptides was predicted using NeuroPred

(http://neuroproteomics.scs.uiuc.edu/neuropred.html; Southey et al., 2006a). Complete prohormone sequences from UniProt were used to predict cleavage in preference to the predicted sequences. In limited cases, EST sequences were combined with the genomic data and published partial sequences to predict the complete prohormone sequence. For example, for *Chromogranin-A* (*CMGA*), three Glutamic acids were missing in the genome-based predictions that were present in the corresponding Uniprot fragment sequence (P04404) and EST sequence EW261315 permitted the prediction of the complete pig CMGA protein sequence. The location of the potential cleavage sites in the pig prohormones were inferred by homology to human data.

Complementary cleavage prediction models trained on confirmed cleavages from mammalian sequences (Amare et al. 2006; Southey et al., 2006b; Tegge et al., 2008) were used to predict cleavages in the pig prohormone sequences. These models included the known motif model that searches for sites with specific combinations of basic amino acid associated to reported cleavages in other species (Southey et al., 2006a), mammalian logistic regression (Amare et al., 2006), and human logistic regression and artificial neural network models based on amino acids only or amino acids combined with the physicochemical properties of amino acids (Tegge et al., 2008).

Known or predicted cleavage sites on all 97 prohormone sequences were used to assess the performance of the models to predict cleavage. The “observed” cleavage sites known or inferred from homology to other species were compared to the cleavage sites predicted by the models. The counts of the true positives (number of correctly predicted cleaved sites), true negatives (number of correctly predicted non-cleaved sites), false positives (the number of incorrectly predicted cleaved sites) and false negatives (number of incorrectly predicted non-cleaved sites) or functions of the counts were used to assess the model performance. These measurements were

used to compute the correct classification rate (number of correctly predicted sites divided by the total number of all sites), sensitivity (number of true positives divided by the total number of cleaved sites), specificity (number of true negatives divided by the total number of non-cleaved sites), positive predictive power (number of true positives divided by the total number of sites predicted to be cleaved), negative predictive power (number of true negatives divided by the total number of sites predicted to not be cleaved), Mathew's correlation coefficient between observed and predicted cleavage. The area under the receiver operator characteristic or ROC curve relating sensitivity and 1 - specificity (Southey et al., 2006a) was also calculated where area values lower than 0.7 indicate poor model performance.

2.3 Results and Discussion

Pig Prohormone Genes

A comprehensive catalogue of 95 potential pig prohormone genes, 7 prohormone convertase genes and one prohormone convertase facilitator gene (7B2) were identified in the pig genome. Table 2 lists the genes and the corresponding BLAST matches on the pig Genome, UniProt, and UniGene databases. Our search enabled the detection of 11 previously unreported (i.e. without empirical confirmation) prohormone genes in pig and complete sequences when only partially or incomplete sequences have been previously reported. Newly identified genes included *intermedin* (*ADM2*), *cortistatin* (*CORT*), *insulin-like 5* (*INSL5*), *orexigenic neuropeptide QRFP* (*OX26*), *prokineticin 2* (*PROK2*), *prolactin-releasing peptide* (*PRRP*), *parathyroid hormone 2* (*TIP39*), *urocortin* (*UCN1*), *urocortin 2* (*UCN2*), *urocortin 3* (*UCN3*), and *urotensin II-related peptide* (*UTS2B*). Our search also identified two different calcitonin protein entries in public

databases that are isoforms of other calcitonin genes. Additional information on the comprehensive catalogue of genes is available in supplementary material Table S1. The predicted prohormone sequences and corresponding prohormone cleavage predictions are available at http://neuroproteomics.scs.uiuc.edu/neuropred.html.

The genome predicted prohormone genes had 63 UniProt entries (Table 3) and 74 NCBI Gene entries on the database releases studied. All UniProt entries were present in Gene database with three exceptions: *gastrin-releasing peptide* (*GRP*), *neuromedin-U* (*NMU*), and *prothyroliberin* (*TRH*). For these three cases, "-like" versions of these prohormone genes were present in the Gene database. Of the prohormones available in UniProt, 48 sequences had evidence at the protein level, 14 sequences had evidence at the transcript level, and 4 were inferred by homology. With 14 exceptions, all the prohormone genes present in the Gene database had a corresponding record in UniGene. The exceptions are *ADM2*, *apelin* (*APEL*), *CORT*, *endothelin-2* (*EDN2*), *progonadoliberin-2* (*GON2*), *INSL5*, *OX26*, *PRRP*, *TIP39*, *prothyroliberin* (*TRH*), *UCN1*, *UCN2*, *UCN3*, and *UTS2B*. The absence of UniGene entries stems from the lack of ESTs information for these genes. However, Saida et al. (2004) has reported the nucleic and protein sequence of *EDN2* have been reported.

Table 3 compares the evidence of the 95 unique prohormone genes and two alternative sequences (*Preprocalcitonin gene-related peptide* and *Calcitonin-2*) to previously published UniProt evidence. Of our 79 complete genome identifications, 41 and 14 sequences have been previously reported in UniProt as complete and fragment sequences, respectively. Additionally 17 complete genome identifications had other experimental support including, ESTs, UniGene clusters or reported sequences that have not been assigned to a UniGene cluster and 10 predictions had no external validation (Tables 2 and 3). In addition, 4 incomplete predictions had

complete UniProt sequences, 3 incomplete predictions had partial UniProt sequence, 4 incomplete predictions has experimental support but no UniProt entry, and 1 incomplete prediction (*EDN2*) has been reported but is not in present in public databases. A partial match to *UCN1* was found in the trace archives but was not found in the current assembly and lacks any external validation. *Progonadoliberin-2* (*GON2*) was not matched in the current pig genome assembly but was matched in earlier revisions and trace archives and has no current experimental evidence outside homology to other mammalian species. The apparent lack of *UCN1* and *GON2* in the assembly and fragment evidence of other prohormones is most likely due to poor coverage of the genomic regions where these prohormone genes are located.

At least four calcitonin (*CALC*s) genes, also known as or calcitonin receptor-stimulating peptide genes (*CRSP*s), were identified with two genes exhibiting alternative splicing. The genome permitted the assignment of the UniProt *pre-procalcitonin* (*A6P7L6*) and *preprocalcitonin gene-related peptide* (*A6P7L7*) entries to the same gene that also produces the UniProt calcitonin (*CALC*, P01259) and calcitonin gene-related peptide (*CALCA*, P30880) peptides, respectively. This gene corresponds to *CALCA* gene found in other mammalian species.

Our gene prediction pipeline also found that the separate Q766Y6 and A0A761 UniProt entries are alternatively spliced variants of the same *calcitonin receptor-stimulating peptide 3* (*CRSP3*) gene as initially reported by Rezaeian et al. (2008). A single genome match was identified for *calcitonin receptor-stimulating peptide 2* (*CRSP2*, Q766Y7). While the *calcitonin receptor-stimulating peptide 1* (*CRSP1*) gene has been reported, the actual gene maybe inaccurately assembled in the 10.2 genome release because the region appears to contain a small duplication

leading to two starting locations. Further supporting this argument, a conserved 19 amino acid region in all CRSP-related protein sequences matched to an additional 5th genome site which was part of a discontinued NCBI Gene entry (Gene ID 100624618). There is insufficient information to conclude whether there is a separate coding gene involved or an assembly-related problem.

Neuropeptide S (*NPS*) is a potential 12th prohormone gene discovered by our bioinformatics strategy. Although UniProt has a pig *NPS* entry, the pig *NPS* protein and nucleic sequences are 100% and 99% identical, respectively, to the rabbit sequence (EU978456). The similarity between the UniProt pig and rabbit sequences was also evident in the phylogenetic relationships among *NPS* sequences reported by Yao et al. (2009). These findings call into question the present pig *NPS* entry in UniProt. Our approach identified two matches on different chromosomes for the *NPS* gene. However, the complete identity of the matched sequence across chromosome implied that this was an assembly error rather than a duplication event. The predicted protein sequence from these matches differed from the UniProt partial pig sequence (B5M997). Our pig genome sequence was more similar to the bovine than the partial UniProt pig sequence as expected from the observed homology of other mammalian prohormones.

Prohormone Genes Previously Unreported in Pig

Our genome search identified 11 prohormone genes that do not have empirical confirmation in the UniProt, UniGene or Gene databases (Table 2). These genes are *ADM2*, *CORT*, *INSL5*, *OX26*, *PROK2*, *PRRP*, *NPS*, *TIP39*, *UCN1*, *UCN2*, *UCN3*, and *UTS2B*. Of these genes, *NPS* has a questionable UniProt entry. Only inferred sequences are available for *ADM2* and *CORT* in

UniProt and the current pig *NPS* entry is identical to the rabbit *NPS*. There is evidence for mammalian homologs of all these genes in the UniProt database. The protocol followed to identify these genes included a high percentage of identities and similarities with a minimum percentage of mismatches and gaps and conservation of the region encompassing the potential neuropeptide. *Intermedin* or *adrenomedullin 2* (*ADM2*) is part of the CGRP/calcitonin family of peptides and has effects similar to those of *adrenomedullin* (*ADML*). In humans, *intermedin* causes hypotension when given peripherally and augments blood pressure and causes sympathetic activation when given to the central nervous system (Hong et al., 2011). This neuropeptide induces prolactin release, has anti-diuretic and natriuretic properties and reduces food intake. The amino acid sequences of *CORT* and *somatostatin* are highly similar and both reduce neuronal activity. In addition, *CORT* has unique roles such as induction of slow-wave sleep, reduction of locomotor activity, and activation of cation selective currents not responsive to *somatostatin* (Spier and deLecea, 2000). Although the function of *INSL5* is still being determined, high expression in the colon, as well as in the brain and hypothalamus, indicates roles in gut contractility and neuroendocrine signaling (Huagaard-Jonsson et al., 2009). Likewise, the function of *OX26* is still being elucidated, although studies in chicken confirm the orexigenic, appetite stimulating activity of this neuropeptide (Ukena et al., 2010). Takayanagi and Onaka (2010) demonstrated that *PRRP* plays a role in control of energy metabolism and stress response. Prokineticins are involved in tumorigenesis process (prostate, testicles, neuroblastoma, colon, and pancreas) acting as a growth factor for cancer cells, an angiogenic and a chemotactic factor for pro-inflammatory neutrophils (Monnier and Samson, 2010). *NPS* has anxiolytic-like effects (stress reduction) and can induce arousal and wakefulness (Reinscheid, 2008). *TIP39* and its receptor form a neuromodulator system and the anatomical distribution

indicates a role in limbic, endocrine, viscerosensory, and auditory functions. This system has been postulated as potential drug target in anxiety, depression and chronic pain management (Dobolyi et al., 2010). Urocortins and their receptors has been found in the central nervous, digestive, reproductive, cardiovascular, immune and endocrine systems, suggesting a variety of roles including cardiovascular activity and cell survival (Venkatasubramanian et al., 2010). *UTSB2* is a paralog of *urotensin 2* that exerts similar biological effects including relaxation of muscles and reduction of blood pressure (Vaudry, 2008).

Pig Prohormone Convertase Genes

The sequence of 7 prohormone convertase genes and the *7B2* facilitator gene were identified in the pig genome (Table 2). Our approach identified the complete sequence of 6 prohormone convertase genes that were previously unreported or not based on empirical evidence (*furin* or *PCSK3*, *PCSK4*, *PCSK5*, *PCSK6*, *PCSK7*, and *PCSK9*). The UniProt and Gene databases only had supporting evidence for *PCSK1*, *PCSK2*, and *7B2*. Only transcript evidence supports the *PCSK1* and *PCSK2* entries in UniProt, meanwhile protein evidence is available for *7B2* also known as *secretogranin V* (*SCG5*). Dai et al. (1995) isolated *PCSK1* from the ovary cDNA library of a pregnant sow and Renegar et al. (2000) detected *PCSK1* in the corpus luteum and brain of pregnant sows. Also, mRNA from *PCSK1* and *PCSK2* has been identified in the pituitary neurointermediate lobes of pigs (Seidah et al., 1992). Among the prohormone convertases, *furin*, *PCSK4*, *PCSK5*, *PCKS6* and *PCSK7* do not have UniGene entries. The present catalogue enhances the currently limited work on pig prohormone convertases.

Functional Characterization of the Pig Prohormone and Prohormone Convertase Genes

Analysis of the large number of microarray gene expression experiments enabled the first comprehensive characterization of the role of prohormone and prohormone convertase genes in biological processes in the pig. The results from these analyses augmented the understanding of the role of these genes on reproduction, health, growth, and other traits of importance to biomedical research and agricultural production.

The query of Affymetrix Porcine Genome Array (GPL3533) identified 77 probes representing 56 prohormone and 3 prohormone convertase genes. Table 4 lists the total number of differentially expressed probes (P-value < 0.005) within the seven experimental classes considered. Supplementary materials Table S2 presents the detailed distribution of the differential expression level of each probe and experiment. A discussion of the findings for the three groups with highest number of differentially expressed probes (immune-related, embryo and placenta, and brain and central nervous system) is presented below. Although neuropeptides expressed in the brain and the immune system interact with circulating cytokines to support two-way communications between the brain and immune system (Jessop, 2008), we describe the profiles of prohormones in immune-related tissues separately from the brain and central nervous system tissues to facilitate the interpretation of results.

Immune-related Profiling. Several studies have demonstrated that prohormone genes play an important role in pig immune response (Pampusch et al., 2000). This was evidenced by the high number of differentially expressed prohormone and prohormone convertase genes (24 genes) among experiments that evaluated immune-response in blood, spleen, and lymph nodes (Table 4). Differentially expressed genes were: *Adrenomedullin* (*ADML*), *Augurin* (*AUGN*),

Cholecystokinin (*CCKN*), *CRSP3*, *Endothelin-1* (*EDN1*), *Galanin* (*GALA*), *Galanin-like peptide* (*GALP*), *Progonadoliberin-1* (*GON1*), *Insulin-like growth factor I* (*IGF1*), *Insulin-like growth factor II* (*IGF2*), *Neuromedin-B* (*NMB*), *Neuromedin-U-25* (*MNU*), *Neuropeptide Y* (*NPY*), *Platelet-derived growth factor subunit A* (*PDGFA*), *Proenkephalin-A* (*PENK*), *Prorelaxin 1* (*REL1*), *Secretogranin-1* (*SCG1*), *Secretogranin-2* (*SCG2*), *Secretogranin-3* (*SCG3*), *Somatostatin* (*SMS*), *Vascular endothelial growth factor C* (*VEGFC*), *Vascular endothelial growth factor D* (*VEGFD*), *PCSK1* and *PCSK7*.

In general, prohormone genes were under-expressed in pigs under immune challenge relative to the un-challenged controls. *AUGN* was differentially expressed in two experiments; GSE7313 (Wang et al. 2007) that profiled lymph nodes and GSE14790 (Tomás et al. 2010) that profiled blood. In GSE14790, 7 day-old pigs were inoculated with porcine circovirus type 2 (PCV2), a virus that is widely spread across pig farms, and gene expression was profiled at 0, 7, 14, 21 and 29 dpi. *AUGN* was over expressed in un-inoculated pigs at 29 dpi relative to 7 dpi, regardless of inoculation and relative 21 dpi inoculated pigs (P-value $< 2.5 \times 10^{-4}$). Both contrasts indicate that the expression of *AUGN* increases with age and this trend is slower in pigs infected with PCV2. In GSE7313, the gene expression of seven week old piglets inoculated with *Salmonella Typhimurium* was profiled at 8 hours post inoculation (hpi), 24 hpi, 48 hpi, and 21 days post inoculation (dpi). AUGN was over expressed at 21 dpi relative to 24 hpi and 48 hpi (P-values $< 6.8 \times 10^{-5}$ and 2.7×10^{-6}, respectively). Consistent with the differential expression in relation to immune-response observed in this study, *AUGN* is a putative tumor suppressor gene and is down-regulated in many cancers (Gonzalez et al., 2011).

IGF2, a member of the insulin family and is involved in development and growth, was differentially expressed across immune-related experiments. *IGF2* was represented by 12 probes

in the microarray platform and six probes were differentially expressed across experiments. Five probes (Ssc.9365.1.S1_at, Ssc.9365.2.S1_a_at, Ssc.9365.5.A1_at, Ssc.9365.5.S1_at, Ssc.9365.6.S1_x_at) and one probe (Ssc.9365.3.S1_a_at) were differentially expressed in experiments GSE14790 (Tomás et al., 2010) and GSE7314 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE7314), respectively. In GSE14790, *IGF2* was under-expressed in non-inoculated piglets at 7 dpi relative to inoculated pigs at various days (P-value < 1.7×10^{-5}, fold change = 0.71). In experiment GSE7314, *IGF2* was over-expressed in pigs inoculated with *Salmonella choleraesuis* at 21 dpi relative to non-inoculated pigs (P-value < 8.1×10^{-4}). These results are consistent with reports that *IGF2* is down-regulated in pigs immune-challenged with lypopolysaccharide (Solinhac et al. 2011).

SCG1, *SCG2*, *SCG3*, members of the secretogranin family, exhibited differential expression among immune-challenge experiments consistent with the known association of these genes with cell activation, cytotoxicity and microbial defense (Radek et al., 2008). Probes on all three SCGs exhibited differential expression on two immune-related experiments. *SCG1* and *SCG2* both had one probe differentially expressed in GSE14790 (probes Ssc.15718.1.A1_at, and Ssc.13645.1.A1_at, respectively) while *SCG3* had one probe (Ssc.6770.1.A1_at) differentially expressed in GSE11787 (Chen et al., 2009). In GSE11787 *SCG3* was under-expressed in inoculated pigs relative to controls (P-value < 1.2×10^{-3}, fold change = 0.33). Our results are consistent with the lack of synthesis of endogenous granins in rat PC12 cells infected with recombinant vaccinia viruses (Krömer et al., 1998). In GSE14790, *SCG1* and *SCG2* were under-expressed in pigs inoculated with PCV2 relative to un-inoculated control pigs (P-value < 1.8×10^{-4}, fold change = 0.88 and P-value < 1.5×10^{-5}, fold change = 0.93, respectively).

Two members of the vascular endothelial growth factor family, *vascular endothelial growth factor C* (*VEGFC*, probe Ssc.12790.1.A1_at) and *vascular endothelial growth factor D* (*VEGFD*, probe Ssc.29289.1.A1_at) were under-expressed in PCV2 inoculated pigs relative to control pigs (P-value < 1.8×10^{-5}, fold change = 0.62) and also under-expressed at early stages (7 dpi) relative to later stages (19 and 29 dpi) in GSE14790. In agreement with our findings, a loss of endothelial growth factor transcription and increase in pro-inflammatory indicators were reported in the endometrial lymphocytes of pigs at sites of fetal arrest (Tayade et al., 2007).

NPY (probe Ssc.15981.1.A1_at) was under-expressed in PCV2-inoculated pigs relative to control pigs (P-value < 6.6×10^{-4}) and within infection level, *NPY* was under-expressed at earlier stages relative to 29 dpi in GSE14790. Consistent with our findings, the levels of *NPY* mRNA decreased in the blood of rats treated with vinblastine, an anti-cancer drug known to decrease the number of white blood cells of the immune system involved in defense (Ericsson et al., 1991). Similarly, *NPY* was found to decrease in cattle infected with Bovine Spongiform Encephalopathy (Almeida et al., 2011).

ADML was differentially expressed in GSE14758-D and GSE7314. In GSE14758-D (Tomas et al., 2010), *ADML* was under-expressed in the mediastinal lymph nodes of PCV2-infected pigs relative to control pigs at 29 dpi (P-value < 1.7×10^{-3}, fold change = 0.6). Whereas, in GSE7314 *ADML* was over-expressed at 48 dpi in the blood of pigs inoculated with *Salmonella choleraesuis* relative to controls (P-value 4.8×10^{-3}). The latter result is consistent with the up-regulation of *ADML* gene expression and increases in systemic circulatory concentrations of *ADML* in response to the onset and progression of trauma, infection, and sepsis (Elasser and Kahl, 2002). The apparent inconsistency between both experiments may be associated with the

differential effects that *ADML* has on cellular metabolism, immune function, endocrine function, and cardiovascular function.

Of the three prohormone convertases available in the microarray platform, *PCSK1* (probe Ssc.141.1.S1_at) and *PCSK7* (probe Ssc.5628.1.S1_at) were significantly differentially expressed (P-value < 1.3×10^{-3}) and *PCSK2* (probe Ssc.109.1.S1_at) was marginally significantly differentially expressed (P-value < 6.5×10^{-3}) in GSE14790. *PCSK1* was under-expressed in PCV2-inoculated pigs already at 7 dpi relative to 29 dpi, regardless of inoculation at the later stage (P-value < 5.7×10^{-5}). Likewise, *PCSK7* is under-expressed in PCV2-inoculated pigs relative to controls already at 7 dpi (P-value < 4.2×10^{-4}) and, within controls, *PCSK7* was under-expressed at early stages (7 dpi, 21 dpi) relative to 29 dpi (P-value < 4.1×10^{-4}). These results are in agreement with similar findings that *furin*, another prohormone convertase, was dysregulated in the immune cells of advanced human atherosclerotic plaques (Turpeinen et al., 2011) and imply that prohormone convertase, like prohormone genes, are down regulated under immune challenges.

Embryo and Placenta Profiling. In GSE18641 (Østrup et al., 2010), *IGF2* (Ssc.9365.2.S1_a_at) was over-expressed in pregnant sows relative to non-pregnant sows (P-value < 2.7 x 10-3, fold change 1.23). In GSE12705 (Ross et al., 2009), *IGF2* (Ssc.9365.4.S1_a_at, and Ssc.9365.5.S1_at) was over-expressed in earlier stage (day 11 spherical and day 11 and 12 tubular) relative to later stage (day 12 and 14 filamentous) conceptuses (P-value < 2.1×10^{-4}). This *IGF2* profile is supported by Pantaleon et al. (2003) that showed that *IGF2* is needed in order for mouse embryos to progress from early stages to blastocyst stages. Gupta et al. (2007, 2008) reported that the expression of the embryo survival related gene *IGF2* increased with the

addition of nonessential amino acids or phytohemaglutinin in pig embryos and blastocysts, respectively.

PENK (probes Ssc.11281.1.A1_at, Ssc.11281.2.S1_at) was over-expressed in tubular and spherical conceptuses relative to filamentous conceptuses (P-Value < 2.6×10^{-6}) in experiment GSE12705 (Ross et al., 2009). This is consistent with results that found *PENK* mRNA to increase linearly during gestation in the hippocampus of pigs (Pittius et al., 1987) *PTHR* was under-expressed in tubular and spherical relative to filamentous conceptuses (P-Value < 6.4×10^{-7}, fold change = 0.02) in experiment GSE12705. This finding is supported by reports that *PTHR* is present in higher concentrations in fetal pigs than in sows (Abbas et al., 1994). *VEGFC* is a representative of the vascular endothelial growth factor family of prohormones that have an important role in the survival and mitogenesis of endothelial cells and lymphaniogenesis and angiogenesis of embryos (Anisimov et al., 2009). *VEGFC* (was over-expressed in pregnant sows relative to non-pregnant sows (P-Value < 7.8×10^{-4}) in experiment GSE18641 (Østrup et al., 2010). This finding is supported by a study in the chicken, demonstrating that the chorioallantoic membrane (analogous to the placenta in mammals) contained growth of embryonic microvessels stimulated by *VEGFC* (Cao et al., 1998). Our profile is also supported by the finding that in mice embryos, *VEGFC* is required for successful lymphatic vasculature development and lymphatic endothelial cell migration (Kakkainen et al., 2003).

Brain and Central Nervous System. Eleven differentially expressed prohormone genes were identified in experiments concerning the hypothalamus, thyroid, and olfactory bulb (neuroblasts). These genes are *Adrenomedullin-5* (*ADM5*), *ADML*, *C-type natriuretic peptide* (*ANFC*), *Cocaine and amphetamine regulated transcript* protein *CART*, *IGF1*, *IGF2*, *NPY*, *Platelet-derived growth factor subunit A* (*PDGFA*), *Prodynorphin* (*PDYN*), *PTHR*, and *VEGFC*.

ADML was over-expressed in the immortalized porcine olfactory bulb neuroblasts relative to the non-neural epithelial cells (P-value < 2.2 x 10^{-6}, fold change > 10) in experiment GSE16855 (Uebing-Czipura et al., 2009). This result is supported by a previous study that found that *ADML* is important for regulation of proliferation and differentiation of neural stem/progenitor cells using the mouse olfactory bulb (Vergaro-Vera et al., 2010).

IGF1 was over-expressed in the neuroblasts relative to non-neural epithelial cells (average P-value < 5 x 10^{-7}, fold change > 10) in experiment GSE16855 (Uebing-Czipura et al., 2009). Our result is supported by a study in chickens showing that *IGF1* was expressed in the olfactory bulb (Mathonnet et al., 2001). Also, *IGF2*, (probe Ssc.9365.6.S1_x_at) was consistently over-expressed in the hypothalamus of male Iberian pigs relative to all other seven breed-gender combinations (on average, P-value < 2.3 x 10^{-4}, fold change = 2.42) in experiment GSE14739-H (Perez-Enciso et al., 2009 and Yang et al., 2011). *NPY* was over-expressed (P-value < 8.1 x 10-4, fold change = 7.94) in neuroblasts relative to non-neuronal cells in GSE16855 (Uebing-Czipura et al., 2009). This result is consistent with reports that the olfactory bulb exhibit high levels of immunoreactive *NPY* in the brain of pigs (Busch-Sørensen et al., 1989) and that *NPY* may inhibit excitatory neurotransmission in the rat olfactory bulb (Blakemore et al., 2006). *VEGFC* was over-expressed in neuroblasts relative to non-neuronal cells (P-value < 1.5 x 10^{-9}, fold change > 10) in experiment GSE16855. This result agrees with a 30% increase in dividing neuroblasts in olfactory bulb in culture stimulated with *VEGFC* compared to controls reported by Le Bras et al. (2006). *PTHLH* was under-expressed in neuroblasts relative to non-neuronal cells (P-value < 2.6 x 10^{-4}, fold change = 0.20) in GSE16855. This finding is consistent with reports that *PTHLH* may be a negative regulator in the differentiation of chondrocytes (Zenmyo et al., 2000). *PDGFA* was over-expressed in neuroblasts relative to non-neuronal cells (P-value < 1.2 x

10^{-4}) in experiment GSE16855. Related to this result, Fressinaud et al. (1991) reported that platelet-derived growth factors increase the glutamine synthetase activity in astrocytes in the brain.

Prediction of Cleavage Sites in Pig Prohormones

All 97 prohormone sequences were used to predict cleavage and confirm the prediction against known or predicted cleavage sites. These sequences were inferred to have 228 cleavage sites that resulting in a 14.6% prevalence rate. Most sites were cleaved at an Arginine (R) such that the most frequently cleaved motifs were xxKR (71%), RxxR (34%) and xxRR (41%), where x denotes any amino acid and K denotes Lysine. There were 5% (38) C-terminal single R sites that were cleaved without a basic amino acid in the second and fourth positions preceding the cleavage site (P2 or P4 locations, respectively).

The performance of the cleavage prediction models is presented in Table 5. The correct classification rate ranged from 82% to 92% indicating that a large proportion of the sites were accurately predicted across all models. The Human cleavage prediction models had the best performance for most of the statistics followed by the Mammalian model. The Known Motif model provided the highest number of true positive predictions but also the highest number of false positive predictions. The Known Motif model provided the highest sensitivity, 77%, indicating more than three quarters of the cleaved sites were correctly predicted as cleaved. However this model also provided the highest number of false positive predictions. Consequently the Known Motif positive predictive power was 35% indicating that, on average, only 35% of sites predicted to be cleaved are expected to be true cleavage sites.

The Human models provided the highest number of true negatives resulting in the best model performance compared to the Known Motif and Mammalian models. The Human artificial neural network models had approximately 60% positive predictive power indicating that most sites predicted as cleaved are expected to be true positives. Although the Human logistic models had lower sensitivity than their artificial neural network counterparts, the differences with the Human artificial neural network model were only 4 cleaved and 11 non-cleaved sites. The high performance of the Human models suggests that the cleavage of prohormones that result in potential biologically active neuropeptides in the pig is similar to humans. Noteworthy is that the Mammalian model was trained on 51 mammalian prohormones that included 8 pig prohormones. This model provided slightly more true positive predictions and a higher sensitivity than the Human logistic model. However, the Mammalian model had noticeably more false positive predictions than the Human logistic model resulting in lower performance in the other accuracy measures.

The comparison of results across models also provides information on the accuracy of the cleavage assignment, prediction accuracy and potential for a gene to produce bioactive peptides. For ten prohormones, at least five of the models did not predict any cleavage site. However, it must be noted that four of the prohormones (*ANF*, *GHRL*, *IGF1* and *PDGFD*) are likely to have sites cleaved by proteases other than prohormone convertases. For example, *ANF* is cleaved by type II transmembrane serine protease Corin (Wu, 2007).

Genes with no predicted cleavage or assigned cleavage that differ from other species can be used to identify proteins are not cleaved to form smaller peptides. There is no evidence for cleavage of *UCN2* and *UCN3* to produce mature peptides in mammals (Fekete and Zorrilla, 2007). All models failed to predict two cleavage sites in *TOR2X*. The first site, an N-terminal dibasic 'RK',

is known to be rarely cleaved across species (Southey et al. 2006a). The second site is a cleavage found in humans that forms alpha- and beta-salusin but this site may not be cleaved in the pig since the pig sequence, like the bovine sequence, only has a single basic site instead of the human dibasic 'RR' site. Similarly for *TKN4*, the genomic prediction and supporting EST data indicates a change from an R in other species to a Glycine amino acid in the pig sequence that may prevent the formation of the 'Hemokinin' peptide.

The pig *CRSP2* protein sequence lacks the 'KR' and a C-terminal cleavage site that are cleaved in human *CALC* and *CALCB* genes to produce '*Calcitonin gene-related peptide 1*' and '*Calcitonin gene-related peptide 2*'. Therefore it is unlikely that this protein would provide these calcitonin peptides. The assigned cleavages in the *RES18* protein are necessary to provide a potential 'Triskadecapeptide' peptide reported by Bloomquist et al. (1994). This potential peptide has flanking dibasic cleavage sites in the mouse and rat but this peptide has not been experimentally confirmed. However, the corresponding region in human, bovine and pig sequences are monobasic and lack common PC cleavage motifs suggesting that these species probably cannot form this peptide.

Examination of the potential cleavage sites in *IAPP* indicated that a mutation from R to Q in the N-terminal cleavage sites necessary to produce the Islet amyloid polypeptide. Examination of the corresponding ESTs indicated that two swine ESTs (AJ649149andAJ649469) were 100% identical to the rat genome and consequently invalid sequences. Two other ESTs (EW569366, BF712755) matched the region that supported the genomic prediction. The predicted protein sequence including the potential cleavages sites of the expected *IAPP* was less than 80% identical to other mammals sequences compared to typically over 85% identity between the human and most other mammalian sequences. Potter et al. (2010) questioned the capability of

IAPP to form amyloids after examining the functionality of a synthesized pig sequence based on the BF712755 EST sequence. Our prohormone sequence and cleavage prediction results also strongly suggest that the pig is unlikely to be able to form *IAPP*. This reflects the importance of proteomic studies involving cleavage to first determine that a species can produce a peptide.

2.4 Conclusion

The pig is a fundamental biomedical and agricultural research species. Results from the first genome-wide study of pig prohormone and prohormone convertase genes, functional annotation and prediction of prohormone cleavage are presented. This study was enabled by the availability of the pig genome sequence and of 35 gene expression experiments that evaluated a wide range of conditions in pigs. These results offer more insights into the role of neuropeptides on biological processes such as reproduction, development, growth, and health and support targeted empirical confirmation. The bioinformatics strategy used in this study can be used to identify prohormones or other sets of genes in species with similar sequence resources. Confirmatory insight into the pig prohormones can be expected from proteomic mass spectrometry studies.

Combining complementary bioinformatic resources, 95 prohormone genes, 7 prohormone convertases and one cleavage facilitator were discovered in the pig genome and raw sequence repositories. We uncovered 11 prohormone genes that have not been previously reported and one potentially incorrectly reported. The high performance of the models used to predict cleavage in the pig prohormones suggests that the prohormone cleavage in pigs is similar to humans. The analysis of 35 gene expression experiments identified various neuropeptide genes differentially expressed in immune-related tissues, embryo and placenta and the central nervous system

including *AUGN*, *IGF2*, the family of *SCG*s, *NPY*, *ADM* and *ADML*, *PENK*, *PTHR*, and *VEGFC*. Experiments are required to confirm that the pig does not produce the bioactive neuropeptides *UCN2*, *UCN3*, *TOR2X*, *TKN4*, *IAPP*, and *CRSP2* as suggested by the cleavage prediction models.

ACKNOWLEDGMENTS

The project described was supported by Award Numbers R21DA027548 and P30DA018310 from the National Institute on Drug Abuse (NIDA), Award Number 1R03CA143975 from the National Cancer Institute (NCI), and by Award Number ILLU-538-311 from the United States Department of Agriculture (USDA), Cooperative State Research, Education and Extension Service.

REFERENCES

1. Book SA, Bustad LK: **The fetal and neonatal pig in biomedical research.** J Anim Sci 1974, **38**: 997-1002
2. Anzenbacherova E, Anzenbacher P, Svoboda Z, Ulrichova J, Kvetina J, Zoulova J, Perlik F, Martinkova J: **Minipig as a model for drug metabolism in man: comparison of in vitro and in vivo metabolism of propafenone.** Biomed Pap Fac Univ Palacky Olomouc Czech Repub 2003, **147**:155-159.
3. Sullivan TP, Eaglstein WH, Davis SC, Mertz P: **The pig as a model for human would healing.** Wound Repair Regen 2001, **9**:66-76
4. Kurahashi M, Miyake H, Takagi T, Tashiro S: **Changes of lymphatic flow in case of pancreatic duct obstruction in the pig—as a model of pancreatic cancer**. J Med Invest 2004, **51**: 70-75.
5. Fricker LD: **Neuropeptide-processing enzymes: applications for drug discovery.** AAPS J 2005, **7**: E449-455.
6. Hook V, Funkelstein L, Lu D, Bark S, Wegrzyn J, Hwang SR: **Proteases for processing proneuropeptides into peptide neurotransmitters and hormones.** Annu Rev Pharmacol Toxicol 2008, **48**: 393-423.
7. Schwartz PH, Nethercott H, Kirov II, Ziaeian B, Young MJ, Klassen H: **Expression of neurodevelopmental markers by cultured porcine neural precursor cells.** Stem Cells 2005, **23**:1286-1294.
8. Yang G, Su J, Yao Y, Lei Z, Zhang G, Li X: **The regulatory mechanism of neuromedin S on luteinizing hormone in pigs.** Anim Reprod Sci 2010, **122**:367-374.

9. Kaminski T, Smolinska N, Nitkiewicz A, Przala J: **Expression of orexin receptors 1 (OX1R) and 2 (OX2R) in the porcine hypothalamus during the oestrous cycle.** J Physiol Pharmacol 2010, **61**:363-371.

10. Southey BR, Sweedler JV, Rodriguez-Zas SL: **A python analytical pipeline to identify prohormone precursors and predict prohormone cleavage sites.** Front Neuroinform 2008, **2**:7.

11. Southey BR, Rodriguez-Zas SL, Sweedler JV: **Characterization of the prohormone complement in cattle using genomic libraries and cleavage prediction approaches.** BMC Genomics 2009, **10**:228.

12. Tegge AN, Southey BR, Sweedler JV, Rodriguez-Zas SL: **Comparative analysis of neuropeptide cleavage sites in human, mouse, rat, and cattle.** Mamm Genome 2008, **19**:106-120.

13. Ross JW, Ashworth MD, White FJ, Johnson GA, Ayoubi PJ, DeSilva U, Whitworth KM, Prather RS, Geisert RD: **Premature estrogen exposure alters endometrial gene expression to disrupt pregnancy in the pig.** Endocrinology 2007, **148**: 4761-4773.

14. Hausman GJ, Barb CR, Dean RG: **Patterns of gene expression in pig adipose tissue: insulin-like growth factor system proteins, neuropeptide Y (NPY), NPY receptors, neurotrophic factors and other secreted factors.** Domest Anim Endocrinol 2008, **35**:24-34.

15. **Pig Genome Release Sscrofa10.2v18**

[http://pre.ensembl.org/Sus_scrofa/Location/Genome]

16. Li L, Sweedler JV: **Peptides in the brain: mass spectrometry-based measurement approaches and challenges.** Annu Rev Anal Chem (Palo Alto Calif) 2008, **1**:451-483.

17. Ross JW, Ashworth MD, Stein DR, Couture OP, Tuggle CK, Geisert RD: **Identification of differential gene expression during porcine conceptus rapid trophoblastic elongation and attachment to uterine luminal epithelium**. Physiol Genomics 2009, **36**: 140-148.

18. Chen H, Li C, Fang M, Zhu M, Li X, Zhou R, Li K, Zhao S: **Understanding Haemophilus parasuis infection in porcine spleen through a transcriptomics approach.** BMC Genomics 2009, **10**: 64.

19. Amare A, Hummon AB, Southey BR, Zimmerman TA, Rodriguez-Zas SL, Sweedler JV: **Bridging neuropeptidomics and genomics with bioinformatics: Prediction of mammalian neuropeptide prohormone processing.** J Proteome Res 2006, **5**:1162-1167.

20. Delfino KR, Southey BR, Sweedler JV, Rodriguez-Zas SL: **Genome-wide census and expression profiling of chicken neuropeptide and prohormone convertase genes.** Neuropeptides 2010, **44**:31-44.

21. **Wise2 – Intelligent algorithms for DNA searches** [http://www.ebi.ac.uk/Tools/Wise2/]

22. Birney E, Clamp M, Durbin R: **GeneWise and Genomewise.** Genome Res 2004, **14**:988-995.

23. **UniProt** [http://www.uniprot.org/]

24. **The NCBI Gene Database** [http://www.ncbi.nlm.nih.gov/gene/]

25. **ClustalW2 – Multiple Sequence Alignment** [http://www.ebi.ac.uk/Tools/msa/clustalw2/]

26. **The NCBI GEO Database** [http://www.ncbi.nlm.nih.gov/geo/]

27. **The Affymetrix Porcine Genome Array** [http://www.affymetrix.com/support/technical/byproduct.affx?product=porcine]

28. Wang Y, Qu L, Uthe JJ, Bearson SM, Kuhar D, Lunney JK, Couture OP, Nettleton D, Dekkers JC, Tuggle CK: **Global transcriptional response of porcine mesenteric lymph nodes to Salmonella enterica serovar Typhimurium.** Genomics 2007, **90**:72-84.

29. Paczkowski M, Yuan Y, Fleming-Waddell J, Bidwell CA, Spurlock D, Krisher RL: **Alterations in the transcriptome of porcine oocytes derived from prepubertal and cyclic females is associated with developmental potential.** J Anim Sci 2011, **89**:3561-3571.

30. Zhou QY, Fang MD, Huang TH, Li CC, Yu M, Zhao SH: **Detection of differentially expressed genes between Erhualian and Large White placentas on day 75 and 90 of gestation.** BMC Genomics 2009, **10**:337.

31. Majo F, Rochat A, Nicolas M, Jaoude GA, Barrandon Y: **Oligopotent stem cells are distributed throughout the mammalian ocular surface.** Nature 2008, **456**:250-254.

32. Lkhagvadorj S, Qu L, Cai W, Couture OP, Barb CR, Hausman GJ, Nettleton D, Anderson LL, Dekkers JC, Tuggle CK: **Microarray gene expression profiles of fasting induced changes in liver and adipose tissues of pigs expressing the melanocortin-4 receptor D298N variant.** Physiol Genomics 2009, **38**:98-111.

33. Jameel MN, Li Q, Mansoor A, Qiang X, Sarver A, Wang X, Swingen C, Zhang J: **Long-term functional improvement and gene expression changes after bone marrow-derived multipotent progenitor cell transplantation in myocardial infarction.** Am J Physiol Heart Circ Physiol 2010, **298**:H1348-56.

34. Perez-Enciso M, Ferraz AL, Ojeda A, Lopez-Bejar M: **Impact of breed and sex on porcine endocrine transcriptome: a bayesian biometrical analysis.** BMC Genomics 2009, **10**:89.

35. Yang B, Bassols A, Saco Y, Perez-Enciso M: **Association between plasma metabolites and gene expression profiles in five porcine endocrine tissues.** Genet Sel Evol 2011, **43**:28.

36. Tomas A, Fernandes LT, Sanchez A, Segales J: **Time course differential gene expression in response to porcine circovirus type 2 subclinical infection.** Vet Res 2010, **41**:12.

37. Mulder IE, Schmidt B, Stokes CR, Lewis M, Bailey M, Aminov RI, Prosser JI, Gill BP, Pluske JR, Mayer CD, Musk CC, Kelly D: **Environmentally-acquired bacteria influence microbial diversity and natural innate immune responses at gut surfaces.** BMC Biol 2009, **7**:79.

38. Banduseela VC, Ochala J, Chen YW, Goransson H, Norman H, Radell P, Eriksson LI, Hoffman EP, Larsson L: **Gene expression and muscle fiber function in a porcine ICU model.** Physiol Genomics 2009, **39**:141-159.

39. Galindo RC, Munoz PM, de Miguel MJ, Marin CM, Labairu J, Revilla M, Blasco JM, Gortazar C, de la Fuente J: **Gene expression changes in spleens of the wildlife reservoir species, Eurasian wild boar (Sus scrofa), naturally infected with Brucella suis biovar 2.** J Genet Genomics 2010, **37**:725-736.

40. Uebing-Czipura AU, Dawson HD, Rutherford MS, Scherba G: **Transcriptome profile and cytogenetic analysis of immortalized neuronally restricted progenitor cells derived from the porcine olfactory bulb.** Anim Biotechnol 2009, **20**:186-215.

41. Lkhagvadorj S, Qu L, Cai W, Couture OP, Barb CR, Hausman GJ, Nettleton D, Anderson LL, Dekkers JC, Tuggle CK: **Gene expression profiling of the short-term adaptive response to acute caloric restriction in liver and adipose tissues of pigs differing in feed efficiency.** Am J Physiol Regul Integr Comp Physiol 2010, **298**:R494-507.

42. Ostrup E, Bauersachs S, Blum H, Wolf E, Hyttel P: **Differential endometrial gene expression in pregnant and nonpregnant sows.** Biol Reprod 2010, **83**:277-285.

43. Kim SS, Kim JR, Moon JK, Choi BH, Kim TH, Kim KS, Kim JJ, Lee CK: **Transcriptional alteration of p53 related processes as a key factor for skeletal muscle characteristics in Sus scrofa.** Mol Cells 2009, **28**:565-573.

44. Canovas A, Quintanilla R, Amills M, Pena RN: **Muscle transcriptomic profiles in pigs with divergent phenotypes for fatness traits.** BMC Genomics 2010, **11**:372.

45. Depre C, Park JY, Shen YT, Zhao X, Qiu H, Yan L, Tian B, Vatner SF, Vatner DE: **Molecular mechanisms mediating preconditioning following chronic ischemia differ from those in classical second window.** Am J Physiol Heart Circ Physiol 2010, **299**:H752-62.

46. **affy: Methods for Affymetrix Oligonucleotide Arrays** [http://www.biostat.jhsph.edu/~ririzarr/affy/]

47. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** Journal of the Royal Statistical Society, Series B (Methodological) 1995, **57**:289-300.

48**.** **Beehive** [http://stagbeetle.animal.uiuc.edu/Beehive]

49. **Neuropred** [http://neuroproteomics.scs.uiuc.edu/neuropred.html]

50. Southey BR, Amare A, Zimmerman TA, Rodriguez-Zas SL, Sweedler JV: **NeuroPred: a tool to predict cleavage sites in neuropeptide precursors and provide the masses of the resulting peptides.** Nucleic Acids Res 2006, **34**:W267-72.

51. Southey BR, Rodriguez-Zas SL, Sweedler JV: **Prediction of neuropeptide prohormone cleavages with application to RFamides.** Peptides 2006, **27**:1087-1098.

52. Saida K, Uchide T, Zhang S, Adur J, Kawano Y, Ogiso M, Oka S, Takizawa S: **Molecular cloning and sequence analysis of the porcine precursor of endothelin-2.** J Cardiovasc Pharmacol 2004, **44 Suppl 1**:S244-7.

53. Rezaeian AH, Katafuchi T, Yoshizawa M, Hiraiwa N, Saito T, Nishibori M, Hamano K, Minamino N, Yasue H: **Genomic organization, expression and evolution of porcine CRSP1, 2, and 3.** Cytogenet Genome Res 2008, **121**:41-49.

54. Yao Y, Lin X, Su J, Yang G, Hou Y, Lei Z: **Cloning and distribution of neuropeptide S and its receptor in the pig.** Neuropeptides 2009, **43**:465-481

55. Hong Y, Hay DL, Quirion R, Poyner DR: **The pharmacology of Adrenomedullin 2/Intermedin.** Br J Pharmacol 2011.

56. Spier AD, de Lecea L: **Cortistatin: a member of the somatostatin neuropeptide family with distinct physiological functions.** Brain Res Brain Res Rev 2000, **33**:228-241.

57. Haugaard-Jonsson LM, Hossain MA, Daly NL, Craik DJ, Wade JD, Rosengren KJ: **Structure of human insulin-like peptide 5 and characterization of conserved hydrogen bonds and electrostatic interactions within the relaxin framework.** Biochem J 2009, **419**:619-627.

58. Ukena K, Tachibana T, Iwakoshi-Ukena E, Saito Y, Minakata H, Kawaguchi R, Osugi T, Tobari Y, Leprince J, Vaudry H, Tsutsui K: **Identification, localization, and function of a novel avian hypothalamic neuropeptide, 26RFa, and its cognate receptor, G protein-coupled receptor-103.** Endocrinology 2010, **151**:2255-2264.

59. Takayanagi Y, Onaka T: **Roles of prolactin-releasing peptide and RFamide related peptides in the control of stress and food intake.** FEBS J 2010, **277**:4998-5005.

60. Monnier J, Samson M: **Prokineticins in angiogenesis and cancer.** Cancer Lett 2010, **296**:144-149.

61. Reinscheid RK: **Neuropeptide S: anatomy, pharmacology, genetics and physiological functions.** Results Probl Cell Differ 2008, **46**:145-158.

62. Dobolyi A, Palkovits M, Usdin TB: **The TIP39-PTH2 receptor system: unique peptidergic cell groups in the brainstem and their interactions with central regulatory mechanisms.** Prog Neurobiol 2010, **90**:29-59.

63. Venkatasubramanian S, Newby DE, Lang NN: **Urocortins in heart failure.** Biochem Pharmacol 2010, **80**:289-296.

64. Vaudry H: **Biologically active peptides urotensin II (UII) and urotensin II-related peptide (URP), and to their cognate receptor (UT). Editorial.** Peptides 2008, **29**:647-648.

65. Dai G, Smeekens SP, Steiner DF, McMurtry JP, Kwok SC: **Characterization of multiple prohormone convertase PC1/3 transcripts in porcine ovary.** Biochim Biophys Acta 1995, **1264**:1-6.

66. Renegar RH, Owens CR, Dai G, McMurtry JP, Kwok SC: **Expression and localization of prohormone convertase 1/3 (SPC3) in porcine ovary.** Mol Reprod Dev 2000, **57**:361-365.

67. Seidah NG, Fournier H, Boileau G, Benjannet S, Rondeau N, Chretien M: **The cDNA structure of the porcine pro-hormone convertase PC2 and the comparative processing by PC1 and PC2 of the N-terminal glycopeptide segment of porcine POMC.** FEBS Lett 1992, **310**:235-239.

68. Jessop DS: **Neuropeptides in the Immune System: Mediators of Stress and Inflammation.** In *Handbook of Neurochemistry and Molecular Neurobiology. 3rd Edition.* Edited by Lajtha A. New York: Springer; 2008:19-35.

69. Pampusch MS, Serie JR, Osinski MA, Seybold VS, Murtaugh MP, Brown DR: **Expression of nociceptin/OFQ receptor and prepro-nociceptin/OFQ in lymphoid tissues.** Peptides 2000, **21**:1865-1870.

70. Gonzalez AM, Podvin S, Lin SY, Miller MC, Botfield H, Leadbeater WE, Roberton A, Dang X, Knowling SE, Cardenas-Galindo E, Donahue JE, Stopa EG, Johanson CE, Coimbra R, Eliceiri BP, Baird A: **Ecrg4 expression and its product augurin in the choroid plexus: impact on fetal brain development, cerebrospinal fluid homeostasis and neuroprogenitor cell response to CNS injury.** Fluids Barriers CNS 2011, **8**:6.

71. **GSE7314** [http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE7314]

72. Solinhac R, Mompart F, Martin P, Robelin D, Pinton P, Iannuccelli E, Lahbib-Mansais Y, Oswald IP, Yerle-Bouissou M: **Transcriptomic and nuclear architecture of immune cells after LPS activation.** Chromosoma 2011, **120**:501-520.

73. Radek KA, Lopez-Garcia B, Hupe M, Niesman IR, Elias PM, Taupenot L, Mahata SK, O'Connor DT, Gallo RL: **The neuroendocrine peptide catestatin is a cutaneous antimicrobial and induced in the skin after injury.** J Invest Dermatol 2008, **128**:1525-1534.

74. Kromer A, Glombik MM, Huttner WB, Gerdes HH: **Essential role of the disulfide-bonded loop of chromogranin B for sorting to secretory granules is revealed by expression of a deletion mutant in the absence of endogenous granin synthesis.** J Cell Biol 1998, **140**:1331-1346.

75. Tayade C, Fang Y, Croy BA: **A review of gene expression in porcine endometrial lymphocytes, endothelium and trophoblast during pregnancy success and failure.** J Reprod Dev 2007, **53**:455-463.

76. Ericsson A, Hemsen A, Lundberg JM, Persson H: **Detection of neuropeptide Y-like immunoreactivity and messenger RNA in rat platelets: the effects of vinblastine, reserpine, and dexamethasone on NPY expression in blood cells.** Exp Cell Res 1991, **192**:604-611.

77. Almeida LM, Basu U, Williams JL, Moore SS, Guan LL: **Microarray analysis in caudal medulla of cattle orally challenged with bovine spongiform encephalopathy.** Genet Mol Res 2011, **10**.

78. Elsasser TH, Kahl S: **Adrenomedullin has multiple roles in disease stress: development and remission of the inflammatory response.** Microsc Res Tech 2002, **57**:120-129.

79. Turpeinen H, Raitoharju E, Oksanen A, Oksala N, Levula M, Lyytikainen LP, Jarvinen O, Creemers JW, Kahonen M, Laaksonen R, Pelto-Huikko M, Lehtimaki T, Pesu M: **Proprotein convertases in human atherosclerotic plaques: The overexpression of FURIN and its substrate cytokines BAFF and APRIL.** Atherosclerosis 2011, **219**:799-806.

80. Pantaleon M, Jericho H, Rabnott G, Kaye PL: **The role of insulin-like growth factor II and its receptor in mouse preimplantation development.** Reprod Fertil Dev 2003, **15**:37-45.

81. Gupta MK, Uhm SJ, Lee SH, Lee HT: **Role of nonessential amino acids on porcine embryos produced by parthenogenesis or somatic cell nuclear transfer.** Mol Reprod Dev 2008, **75**:588-597.

82. Gupta MK, Uhm SJ, Han DW, Lee HT: **Embryo quality and production efficiency of porcine parthenotes is improved by phytohemagglutinin.** Mol Reprod Dev 2007, **74**:435-444.

83. Pittius CW, Ellendorff F, Hollt V, Parvizi N: **Ontogenetic development of proenkephalin A and proenkephalin B messenger RNA in fetal pigs.** Exp Brain Res 1987, **69**:208-212.

84. Abbas SK, Ratcliffe WA, Moniz C, Dixit M, Caple IW, Silver M, Fowden A, Care AD: **The role of parathyroid hormone-related protein in calcium homeostasis in the fetal pig.** Exp Physiol 1994, **79**:527-536.

85. Anisimov A, Alitalo A, Korpisalo P, Soronen J, Kaijalainen S, Leppanen VM, Jeltsch M, Yla-Herttuala S, Alitalo K: **Activated forms of VEGF-C and VEGF-D provide improved vascular function in skeletal muscle.** Circ Res 2009, **104**:1302-1312.

86. Cao Y, Linden P, Farnebo J, Cao R, Eriksson A, Kumar V, Qi JH, Claesson-Welsh L, Alitalo K: **Vascular endothelial growth factor C induces angiogenesis in vivo.** Proc Natl Acad Sci U S A 1998, **95**:14389-14394.

87. Karkkainen MJ, Haiko P, Sainio K, Partanen J, Taipale J, Petrova TV, Jeltsch M, Jackson DG, Talikka M, Rauvala H, Betsholtz C, Alitalo K: **Vascular endothelial growth factor C is required for sprouting of the first lymphatic vessels from embryonic veins.** Nat Immunol 2004, **5**:74-80.

88. Vergano-Vera E, Fernandez AP, Hurtado-Chong A, Vicario-Abejon C, Martinez A: **Lack of adrenomedullin affects growth and differentiation of adult neural stem/progenitor cells.** Cell Tissue Res 2010, **340**:1-11.

89. Mathonnet M, Comte I, Lalloue F, Ayer-Le Lievre C: **Insulin-like growth factor I induced survival of axotomized olfactory neurons in the chick.** Neurosci Lett 2001, **308**:67-70.

90. Busch-Sorensen M, Sheikh SP, O'Hare M, Tortora O, Schwartz TW, Gammeltoft S: **Regional distribution of neuropeptide Y and its receptor in the porcine central nervous system.** J Neurochem 1989, **52**:1545-1552.

91. Blakemore LJ, Levenson CW, Trombley PQ: **Neuropeptide Y modulates excitatory synaptic transmission in the olfactory bulb.** Neuroscience 2006, **138**:663-674.

92. Le Bras B, Barallobre MJ, Homman-Ludiye J, Ny A, Wyns S, Tammela T, Haiko P, Karkkainen MJ, Yuan L, Muriel MP, Chatzopoulou E, Breant C, Zalc B, Carmeliet P, Alitalo K, Eichmann A, Thomas JL: **VEGF-C is a trophic factor for neural progenitors in the vertebrate embryonic brain.** Nat Neurosci 2006, **9**:340-348.

93. Zenmyo M, Komiya S, Hamada T, Hiraoka K, Suzuki R, Inoue A: **P21 and Parathyroid Hormone-Related Peptide in the Growth Plate.** Calcif Tissue Int 2000, **67**:378-381.

94. Fressinaud C, Weinrauder H, Delaunoy JP, Tholey G, Labourdette G, Sarlieve LL: **Glutamine synthetase expression in rat oligodendrocytes in culture: regulation by hormones and growth factors.** J Cell Physiol 1991, **149**:459-468.

95. Wu Q: **The serine protease corin in cardiovascular biology and disease.** Front Biosci 2007, **12**:4179-4190.

96. Fekete EM, Zorrilla EP: **Physiology, pharmacology, and therapeutic relevance of urocortins in mammals: ancient CRF paralogs.** Front Neuroendocrinol 2007, **28**:1-27.

97. Bloomquist BT, Darlington DN, Mueller GP, Mains RE, Eipper BA: **Regulated endocrine-specific protein-18: a short-lived novel glucocorticoid-regulated endocrine protein.** Endocrinology 1994, **135**:2714-2722.

98. Potter KJ, Abedini A, Marek P, Klimek AM, Butterworth S, Driscoll M, Baker R, Nilsson MR, Warnock GL, Oberholzer J, Bertera S, Trucco M, Korbutt GS, Fraser PE,

Raleigh DP, Verchere CB: **Islet amyloid deposition limits the viability of human islet grafts but not porcine islet grafts.** Proc Natl Acad Sci U S A 2010, **107**:4305-4310.

TABLES

Table 1. Main features of the 35 microarray experiments analyzed to functionally annotate prohormone and prohormone convertase genes

Experiment[a]	Tissue	Class	Gender	Age	Comparison(s)	Reference
GSE7313	Mesenteric Lymph Nodes	Immune	NA[b]	7 weeks	Non-Infected, 8 hpi, 24 hpi, 48 hpi, 21 dpi	Wang et al. (2007)
GSE7314	Mesenteric Lymph Nodes	Immune	NA	7 weeks	Non-Infected, 8 hpi, 24 hpi, 48 hpi, 21 dpi	GEO
GSE8974	Endothelial Cells	Muscle	Female	1 week	Cultured, Regenerated	GEO
GSE9333	Adipose Tissue	Fat	NA	6-7 months	Korean Native Pig, Yorkshire	GEO
GSE11590	Oocytes	Reproduction	Female	42-44 hours	Sow, Gilt	Paczkowski et al. (2011)
GSE11787	Spleen	Immune	M/F[c]	30 days	Haemophilusparasuis infected, non-infected	Chen et al. (2009)
GSE11853	Placenta	Embryo and Placenta	Female	75 or 90 days after gestation	Erhualian, Large White, day 75, day 90	Zhou et al. (2009)
GSE12604	Corneal/Conjuctival clone	Brain and Nervous System	NA	NA	Corneal, Conjuctival clone	Majo et al. (2008)

Table 1. (cont.)

Experiment[a]	Tissue	Class	Gender	Age	Comparison(s)	Reference
GSE12705	Embryo	Embryo and Placenta	Female	11, 12 or 14 days	Spherical, tubular, filamentous	Ross et al. (2009)
GSE13457	Jejunum	Gut	NA	21-35 days	day 21, 24, 28,35, animal, vegetable protein diet	GEO
GSE13528-F	Fat	Fat	Female	6 months	fed ad lubitum, fasted 3 days, homozygous D298, homozygous N298	Lkhagvadorj et al. (2009)
GSE13528-L	Liver	Fat	Female	6 months	fed ad lubitum, fasted 3 days, homozygous D298, homozygous N298	Lkhagvadorj et al. (2009)
GSE14373	Kidney	Fat	NA	NA	naïve, transplanted, non-transplanted, 4 or 24 hours, CO treated or untreated	GEO
GSE14643	Heart	Muscle	Female	3 months	stem cell injection, saline injection	Jameel et al. (2010)
GSE14739-B	Fat	Fat	M/F	80, 83, 87, 89 days	Male, Female, Large White, Iberian, Duroc, Youli	Perez-Enciso et al. (2009), Yang et al. (2011)

Table 1. (cont.)

Experiment[a]	Tissue	Class	Gender	Age	Comparison(s)	Reference
GSE14739-G	Gonads	Reproduction	M/F	80, 83, 87, 89 days	Male, Female, Large White, Iberian, Duroc, Youli	Perez-Enciso et al. (2009), Yang et al. (2011)
GSE14739-H	Hypothalamus	Brain and Nervous System	M/F	80, 83, 87, 89 days	Male, Female, Large White, Iberian, Duroc, Youli	Perez-Enciso et al. (2009), Yang et al. (2011)
GSE14739-T	Thyroid Gland	Brain and Nervous System	M/F	80, 83, 87, 89 days	Male, Female, Large White, Iberian, Duroc, Youli	Perez-Enciso et al. (2009), Yang et al. (2011)
GSE14739-A	Adenohypophsis	Brain and Nervous System	M/F	80, 83, 87, 89 days	Male, Female, Large White, Iberian, Duroc, Youli	Perez-Enciso et al. (2009), Yang et al. (2011)
GSE14758	Mediastinal Lymph nodes	Immune	NA	7 days	Inoculated with PCV2, uninoculated, 1, 2, 5, 8, or 29 dpi	Tomas et al. (2010)
GSE14790	Blood	Immune	NA	7-36 days	Inoculated with PCV2, uninoculated, 0, 7, 14, 21 or 29 dpi	GEO
GSE15211	Aortic Valve	Muscle	Female	Adult	normotensive, hypertensive	GEO

Table 1. (cont.)

Experiment[a]	Tissue	Class	Gender	Age	Comparison(s)	Reference
GSE15256	Ileum	Gut	NA	5, 28 and 56 days	isolation/antibiotics/indoors, with sow indoors, outdoors, 5, 28 or 56 days	Mulder et al. (2009)
GSE16348	Skeletal Muscle	Muscle	Female	NA	sepsis, corticosteroids, NBMA/CS/sepsis, control	Banduseela et al. (2009)
GSE16855	Neuroblasts/ epithelial cells	Brain and Nervous System	Female	neuroblasts from female newborn	neuroblasts, epithelial cells	Uebing-Czipura et al. (2009)
GSE17309	Liver	Fat	M/F	211 days	Male, Female, high, low feeding level	GEO
GSE17492	Spleen cells	Immune	Male	Adult	Brucella suis infected, uninfected	Galindo et al. (2010)
GSE18343	Endometrium	Embryo and Placenta	Female	day 13 and 15 of psuedopregnancy	estrogen treated, non estrogen treated, day 13, day 15	GEO
GSE18359-A	Adipose Tissue	Fat	Female	6 Months	Low, High RFI, ad lib, calorie restricted	Lkhagvadorj et al. (2010)
GSE18359-L	Liver	Fat	Female	6 Months	Low, High RFI, ad lib, calorie restricted,	Lkhagvadorj et al. (2010)
GSE18467	Cortical	Embryo and Placenta	NA	60 or 80 days	day 60, day 80	GEO

Table 1. (cont.)

Experiment[a]	Tissue	Class	Gender	Age	Comparison(s)	Reference
GSE18641	Endometrium	Embryo and Placenta	Female	day 14 after insemination	Non-pregnant, pregnant, Yorkshire, Landrace	Ostrup et al. (2010)
GSE18653	Longissimus Dorsi	Muscle	Male	6 or 7 months	Korean Native Pig, Yorkshire, 6 or 7 months	Kim et al. (2009)
GSE19275	Gluteus Medius Muscle	Muscle	Male	190 days	High, low carcass/plasma/muscle fat content	Canovas et al. (2010)
GSE21096	Heart	Muscle	Female	NA	control, RCO, SWOP, RCS	Depre et al. (2010)

[a] Experiment ID, Gene Expression Omnibus GEO Series identifier.

[b] NA, not available.

[c] M/F, male and female.

Table 2. Prohormone and convertase genes identified across three pig genome resources.

Symbol	Gene Name	Genome Sequence[1]	UniProt ID	UniGene Cluster ID
Prohormone Genes				
ADM2	*Intermedin*	Complete		
ADM5	*Adrenomedullin-5*	Complete	A5LHG2	Ssc.26627
ADML	*Adrenomedullin*	Complete	P53366	Ssc.314
ANF	*Atrial natriuretic factor*	Complete	P24259	Ssc.16245
ANFB	*Natriuretic peptide B*	Complete	P07634	Ssc.629
ANFC	*C-type natriuretic peptide*	Complete	P18104	Ssc.23867
APEL	*Apelin*	Complete		
AUGN	*Augurin*	Complete		Ssc.22487
CALC[2]	*Calcitonin/calcitonin gene-related peptide 1*	Complete	A6P7L6	Ssc.14052
	Preprocalcitonin gene-related peptide	Complete	A6P7L7	Ssc.56129
CART	*Cocaine- and amphetamine-regulated transcript protein*	Complete	Q307W6	Ssc.15900
CCKN	*Cholecystokinin*	Complete	P01356	Ssc.717
CMGA	*Chromogranin-A*	Complete	P04404	Ssc.4653
COLI	*Pro-opiomelanocortin*	Complete	P01192	Ssc.14556
CORT	*Cortistatin*	Complete		
CRF	*Corticoliberin*	Complete	P06296	Ssc.69887
CRSP1	*Calcitonin receptor-stimulating peptide 1*	Complete	Q862B1	Ssc.3741
CRSP2	*Calcitonin receptor-stimulating peptide 2*	Complete	Q766Y7	Ssc.18558
CRSP3[2]	*Calcitonin receptor-stimulating peptide 3*	Complete	Q766Y6	Ssc.17879
	Calcitonin-2	Complete	A0A761	
EDN1	*Endothelin-1*	Complete	P09558	Ssc.9364
EDN2	*Endothelin-2*	Fragment		
EDN3	*Endothelin-3*	Complete	A5A752	Ssc.31972
GALA	*Galanin*	Complete	P07480	Ssc.713
GALP	*Galanin-like peptide*	Complete	Q9TT95	Ssc.4875
GAST	*Gastrin*	Complete	P01351	Ssc.644
GHRL	*Obestatin*	Complete	Q9GKY5	Ssc.440
GIP	*Gastric inhibitory polypeptide*	Complete	P01281	Ssc.38713

Table 2. (cont.)

Symbol	Gene Name	Genome Sequence[1]	UniProt ID	UniGene Cluster ID
Prohormone Genes				
GLUC	*Glucagon*	Complete	P01274	Ssc.17225
GON1	*Progonadoliberin-1*	Complete	P49921	Ssc.16310
GON2	*Progonadoliberin-2*	Not Present		
GRP	*Gastrin-releasing peptide*	Fragment	P63153	Ssc.13923
HEPC	*Hepcidin*	Complete	Q8MJ80	Ssc.376
IAPP	*Islet amyloid polypeptide*	Complete	Q29119	Ssc.8324
IGF1	*Insulin-like growth factor 1*	Complete	P16545	Ssc.16231
IGF2	*Insulin-like growth factor 2*	Fragment	P23695	Ssc.9365
INS	*Insulin*	Fragment	P01315	Ssc.583
INSL3	*Insulin-like 3*	Complete	P51461	Ssc.11990
INSL5	*Insulin-like 5*	Complete		
INSL6	*Insulin-like 6*	Complete		Ssc.46919
KISS1	*Metastasis-suppressor KiSS-1*	Complete	B5M447	Ssc.73565
MCH	*Pro-melanin-concentrating hormone*	Complete		Ssc.3287
MOTI	*Motilin*	Complete	P01307	Ssc.714
NEU1	*Oxytocin*	Complete	P01177	Ssc.15668
NEU2	*Neurophysin-II*	Complete	P01183	Ssc.4210
NEUT	*Neurotensin*	Complete		Ssc.38680
NMB	*Neuromedin-B*	Fragment	B0LUW4	Ssc.2083
NMS	*Neuromedin-S*	Complete	C3UZJ1	Ssc.12508
NMU	*Neuromedin-U*	Complete	P34964	Ssc.12508
NPB	*Neuropeptide B*	Complete		Ssc.82498
NPFF	*Neuropeptide FF*	Complete		Ssc.44958
NPS	*Neuropeptide S*	Complete		Ssc.73596
NPW	*Neuropeptide W*	Complete	Q8MI35	Ssc.15796
NPY	*Neuropeptide Y*	Complete	P01304	Ssc.15981
OREX	*Orexin*	Complete	O77668	Ssc.15983
OSTN	*Osteocrin (Musclin)*	Complete		Ssc.5148
OX26	*Orexigenic neuropeptide QRFP*	Complete		
PACA	*Pituitary adenylatecyclase-activating polypeptide*	Complete	P41535	Ssc.27598
PAHO	*Pancreatic polypeptide*	Fragment	P01300	Ssc.456

Table 2. (cont.)

Symbol	Gene Name	Genome Sequence[1]	UniProt ID	UniGene Cluster ID
Prohormone Genes				
PCSK1N	*Proproteinconvertasesubtilisin/kexin type 1 inhibitor*	Fragment		Ssc.17429[3]
PDGFA	*Platelet-derived growth factor alpha polypeptide*	Complete		Ssc.6173
PDGFB	*Platelet-derived growth factor beta polypeptide*	Fragment	P20034	Ssc.54182
PDGFD	*Platelet-derived growth factor D*	Complete		Ssc.49835
PDYN	*Proenkephalin-B*	Complete	P01214	Ssc.121
PENK	*Proenkephalin*	Fragment		Ssc.11281
PNOC	*Prepronociceptin*	Complete	P55791	Ssc.15910
PROK2	*Prokineticin 2*	Fragment		
PRRP	*Prolactin-releasing peptide*	Fragment		
PTHR	*Parathyroid hormone-related peptide*	Complete	Q866H2	Ssc.9991
PTHY	*Parathyroid hormone*	Complete	P01269	Ssc.668
PYY	*Peptide YY*	Complete	P68005	Ssc.63650
REL1	*Pro-relaxin 1*	Complete	P01348	Ssc.162
REL3	*Relaxin 3*	Complete	Q8HY17	Ssc.42647
RES18	*Regulated endocrine-specific protein 18*	Complete		Ssc.49266
RFRP	*Neuropeptide VF precursor*	Complete	C4P9W1	Ssc.75350
SCG1	*Secretogranin-1*	Fragment	Q9GLG4	Ssc.15718
SCG2	*Secretogranin-2*	Complete	Q5FZP5	Ssc.13645
SCG3	*Secretogranin-3*	Complete		Ssc.6770
SECR	*Secretin*	Complete	P63298	Ssc.710
SLIB	*Somatoliberin*	Complete	P01287	Ssc.71374
SMS	*Somatostatin*	Complete	P01168	Ssc.19520
SPXN	*Spexin*	Complete		Ssc.57764
TIP39	*Parathyroid hormone 2*	Complete		
TKN1	*Tachykinin, precursor 1*	Complete		Ssc.18075
TKN4	*Tachykinin-4*	Complete		Ssc.23153
TKNK	*Tachykinin 3*	Complete	P67934	Ssc.19565
TOR2X	*Torsin family 2, member A*	Fragment		Ssc.67158
TRH	*Prothyroliberin*	Complete	P62968	
UCN1	*Urocortin*	Traces		

Table 2. (cont.)

Symbol	Gene Name	Genome Sequence[1]	UniProt ID	UniGene Cluster ID
Prohormone Genes				
UCN2	*Urocortin 2*	Complete		
UCN3	*Urocortin 3*	Complete		
UTS2	*Urotensin 2*	Complete	Q95J46	Ssc.437
UTS2B	*Urotensin II-related peptide*	Complete		
VEGFC	*Vascular endothelial growth factor C*	Fragment		Ssc.12790
VEGFD	*Vascular endothelial growth factor D*	Complete		Ssc.29289
VGF	*Neurosecretory protein VGF*	Fragment		Ssc.90772
VIP	*Vasoactive intestinal peptide*	Complete	E0Y441	Ssc.47759
Prohormone Convertase-related genes				
PCSK1	*Proprotein convertase subtilisin/kexin type 1 PC1/3*	Complete	Q28959	Ssc.92884
PCSK2	*Proprotein convertase subtilisin/kexin type 2*	Complete	Q03333	Ssc.109
PCSK3	*Furin*	Complete		Ssc.94009
PCSK4	*Proprotein convertase subtilisin/kexin type 4*	Complete		
PCSK5	*Proprotein convertase subtilisin/kexin type 5*	Complete		
PCSK6	*Proprotein convertase subtilisin/kexin type 6*	Complete		
PCSK7	*Proprotein convertase subtilisin/kexin type 7*	Complete		
PCSK9	*Proprotein convertase subtilisin/kexin type 9*	Complete		Ssc.84357[3]
7B2	*Neuroendocrine protein 7B2 (secretogranin V)*	Complete	P01165	Ssc.155

[1] Gene sequences detected in this study. Complete = the complete gene sequence was identified in the pig genome assembly 10.2; Traces = a fragment of the gene sequence was identified in the pig trace archives; Fragment = fragment gene sequence was identified in the pig genome assembly 10.2; Not Present = gene sequence is not detected in the 10.2 assembly but a match was detected in earlier gene assemblies.

[2] Alternative splicing resulted in multiple reported/published protein sequences.

[3] Not exact assignments (e.g. moderately similar or –like gene)

Table 3. Distribution of the prohormone gene predictions across resources.

Genome Prediction	**UniProt Status**	**Experimental Support** (EST, UniGene or sequence)	**No Experimental Support**
Complete	UniProt	50	1
	No UniProt	17	10
Fragment	UniProt	7	0
	No UniProt	7	1
Not found	No UniProt	0	2

Table 4. Differentially expressed prohormone and prohormone convertase genes (P-value < 0.005) across 35 microarray experiments by tissue class.

Name	Probe[a]	Imm.[b]	Emb.	CNS	Repro.	Musc.	Fat	Gut	Total
Prohormone Genes									
ADM5	Ssc.26627.1.A1_at	0	0	1	0	0	0	0	1
ADML	Ssc.314.1.S1_at	2	0	1	0	1	0	1	5
ANF	Ssc.16245.1.S1_at	0	0	0	0	1	0	0	1
ANFB	Ssc.629.1.S1_at	0	0	0	0	1	0	0	1
ANFC	Ssc.23867.1.A1_at	0	1	1	0	0	0	0	2
AUGN	Ssc.22487.1.S1_at	2	0	0	0	1	1	0	4
CART	Ssc.15900.1.S1_at	0	1	1	0	0	0	0	2
CCKN	Ssc.717.1.S1_at	1	0	0	1	1	0	0	3
CMGA	Ssc.4653.1.S1_at	0	0	0	0	0	1	1	2
COLI	Ssc.14556.1.S1_at	0	1	0	0	0	0	0	1
CRSP1	Ssc.3741.1.S1_at	0	0	0	0	0	0	0	0
CRSP2	Ssc.18558.1.S1_at	0	1	0	0	0	0	0	1
CRSP3	Ssc.17879.1.S1_at	1	0	0	0	0	0	0	1
EDN1	Ssc.9364.1.S1_at	2	0	0	0	0	1	0	3
GALA	Ssc.713.1.S1_at	1	1	0	0	0	0	1	3
GALP	Ssc.4875.1.S1_at	1	1	0	0	1	0	0	3
GAST	Ssc.644.1.S1_at	0	1	0	0	0	0	0	1
GHRL	Ssc.440.1.S1_at	0	0	0	0	0	0	0	0
GLUC	Ssc.17225.1.S1_at	0	1	0	1	0	0	1	3
GON1	Ssc.16310.1.S1_at	1	1	0	0	0	0	0	2
HEPC	Ssc.376.1.S1_at	0	0	0	0	0	0	0	0
IAPP	Ssc.8324.1.A1_at	0	1	0	0	0	0	0	1
IGF1	Ssc.16231.1.S1_a_at	1	0	1	0	0	0	0	2
	Ssc.16231.2.A1_a_at	0	0	0	0	0	0	0	0
	Ssc.16231.3.S1_a_at	0	0	1	0	0	0	0	1
IGF2	Ssc.9365.1.S1_at	1	0	0	0	0	0	0	1
	Ssc.9365.2.S1_a_at	1	1	0	0	0	1	0	3
	Ssc.9365.3.S1_a_at	1	0	0	0	0	0	0	1
	Ssc.9365.3.S1_x_at	0	0	0	0	0	0	0	0
	Ssc.9365.4.S1_a_at	0	1	0	0	0	0	0	1
	Ssc.9365.5.A1_at	1	0	0	0	0	0	0	1
	Ssc.9365.5.S1_at	1	1	0	0	0	0	0	2
	Ssc.9365.5.S1_a_at	0	0	0	0	0	1	0	1
	Ssc.9365.6.A1_a_at	0	0	0	0	0	0	0	0
	Ssc.9365.6.A1_x_at	0	0	0	0	0	0	0	0
	Ssc.9365.6.S1_x_at	1	0	1	0	0	0	0	2
	Ssc.9365.7.A1_x_at	0	0	0	0	0	0	0	0
INS	Ssc.583.1.S1_at	0	0	0	0	0	0	0	0
INSL3	Ssc.11990.1.S1_at	0	1	0	0	0	0	0	1

Table 4. (cont.)

Name	Probe[a]	Imm.[b]	Emb.	CNS	Repro.	Musc.	Fat	Gut	Total
Prohormone Genes									
MCH	Ssc.3287.1.S1_at	0	0	0	0	0	0	0	0
MOTI	Ssc.714.1.S1_at	0	0	0	0	0	0	0	0
NEU1	Ssc.15668.1.A1_at	0	0	0	0	0	0	0	0
NEU2	Ssc.4210.1.S1_at	0	0	0	0	1	0	0	1
NMB	Ssc.2083.1.A1_at	1	0	0	0	0	0	0	1
NMU	Ssc.12508.1.A1_at	1	0	0	0	0	0	0	1
NPW	Ssc.15796.1.S1_at	0	1	0	0	0	0	0	1
NPY	Ssc.15981.1.A1_at	1	1	0	0	1	0	0	3
	Ssc.15981.1.S1_at	0	2	1	0	0	0	0	3
OREX	Ssc.15983.1.S1_at	0	0	0	0	0	0	0	0
PACA	Ssc.27598.1.S1_at	0	1	0	0	0	0	0	1
PAHO	Ssc.456.1.S1_at	0	1	0	0	0	0	0	1
PCSK1N	Ssc.17429.1.S1_at	0	1	0	1	0	0	0	2
PDGFA	Ssc.6173.3.S1_a_at	1	0	1	0	0	0	1	3
PDYN	Ssc.121.1.S1_at	0	1	1	0	0	0	0	2
PENK	Ssc.11281.1.A1_at	0	1	0	1	0	1	1	4
	Ssc.11281.2.S1_at	1	1	0	0	0	0	0	2
PNOC	Ssc.15910.1.A1_at	0	0	0	0	0	0	0	0
	Ssc.15910.1.S1_at	0	0	0	0	0	0	0	0
PTHR	Ssc.9991.1.S1_at	0	1	1	2	0	0	0	4
PTHY	Ssc.668.1.S1_at	0	1	0	0	0	0	0	1
REL1	Ssc.162.1.S1_at	1	1	0	0	0	0	0	2
SCG1	Ssc.15718.1.A1_at	1	1	0	0	0	0	1	3
SCG2	Ssc.13645.1.A1_at	1	0	0	0	0	1	1	3
SCG3	Ssc.6770.1.A1_at	1	1	0	1	0	0	0	3
SECR	Ssc.710.1.S1_at	0	1	0	0	0	0	0	1
SMS	Ssc.19520.1.A1_at	1	1	0	0	1	0	0	3
TKN1	Ssc.18075.1.A1_at	0	0	0	0	0	0	0	0
	Ssc.18075.2.S1_at	0	1	0	0	0	0	0	1
TKN4	Ssc.23153.1.S1_at	0	0	0	0	0	0	0	0
TKNK	Ssc.19565.1.S1_at	0	0	0	0	0	0	0	0
	Ssc.19565.2.A1_at	0	0	0	0	0	0	0	0
UTS2	Ssc.437.1.S1_a_at	0	1	0	0	0	0	0	1
VEGFC	Ssc.12790.1.A1_at	1	1	1	0	1	0	1	5
VEGFD	Ssc.29289.1.A1_at	1	1	0	0	0	0	0	2
Prohormone Total		30	35	12	7	10	7	9	110
Prohormone Convertase									
PCSK1	Ssc.141.1.S1_at	1	1	0	0	0	0	1	3
PCSK2	Ssc.109.1.S1_at	0	0	0	0	0	0	0	0
PCSK7	Ssc.5628.1.S1_at	1	1	0	0	0	0	1	3
Prohormone Convertase Total		2	2	0	0	0	0	2	6

[a] Affymetrix microarray gene probe identifier

[b] Experiment classes: Imm: primary immune-response tissues, Emb: embryo and placenta, CNS: brain and central nervous system, Repro: reproduction, Musc: muscle, fat, and gut.

Table 5. Performance of various models to predict cleavage in pig prohormones.

		Mammalian	Human			
	Known	Logistic	Logistic Regression		Artificial Neural Network	
Model	**Motif**	**Regression**	**AA[2]**	**AA Prop.[3]**	**AA**	**AA Prop.**
True Positives[1]	181	165	160	158	164	167
True Negatives	1520	1640	1724	1670	1735	1747
False Positives	329	209	125	179	114	102
False Negatives	54	70	75	77	71	68
Correct Classification	0.8162	0.8661	0.904	0.8772	0.9112	0.9184
Sensitivity	0.7702	0.7021	0.6809	0.6723	0.6979	0.7106
Specificity	0.8221	0.887	0.9324	0.9032	0.9383	0.9448
Positive predictive power	0.3549	0.4412	0.5614	0.4688	0.5899	0.6208
Negative predictive pow	0.9657	0.9591	0.9583	0.9559	0.9607	0.9625
Correlation	0.4358	0.4856	0.5645	0.4944	0.5919	0.6184
AUC	0.8006	0.8470	0.8600	0.8186	0.8589	0.8802

[1] True positives: number of correctly predicted cleaved sites; True negatives: number of correctly predicted non-cleaved sites; False positives: number of incorrectly predicted cleaved sites; False negatives: number of incorrectly predicted non-cleaved sites; Correct classification rate: number of correctly predicted sites divided by the total number of sites; Sensitivity (one minus false positive rate): number of true positives divided by the total number of sites cleaved; Specificity (one minus false negative rate): number of true negatives divided by the total number of sites not cleaved; Positive predictive power: number of true positives divided by the total number of sites predicted to be cleaved; Negative predictive power: number of true negatives divided by the total number of sites predicted to not be cleaved; Correlation coefficient: Mathew's correlation coefficient between observed and predicted cleavage; and Area under the receiver operator characteristic or ROC curve relating sensitivity and 1-specificity.
[2] Models trained only on amino acids.
[3] Models trained with amino acids combined with the physicochemical properties of amino acids.

Supplementary Table S1. Comprehensive catalogue of prohormone and prohormone convertase genes

Symbol	Prohormone Name	Genome Sequence[1]	UniProt ID	UniGene ID	SwissProt Evidence	Gene Database ID	Ensembl DNA ID	Ensembl Protein ID	Ensembl Gene ID
ADM2	Intermedin	complete		None			ENSSSCT00000001052	ENSSSCP00000001030	ENSSSCG00000000962
ADM5	Adrenomedullin-5	complete	A5LHG2	Ssc.26627	Protein	100101476	ENSSSCT00000003545	ENSSSCP00000003462	ENSSSCG00000003190
ADML	Adrenomedullin	complete	P53366	Ssc.314	Protein	397195	ENSSSCT00000014648	ENSSSCP00000014253	ENSSSCG00000013408
ANF	Atrial natriuretic factor	complete	P24259	Ssc.16245	Protein	397496	ENSSSCT00000003808	ENSSSCP00000003721	ENSSSCG00000003430
ANFB	Natriuretic peptide B	complete	P07634	Ssc.629	Protein	396844	ENSSSCT00000003809	ENSSSCP00000003722	ENSSSCG00000003431
ANFC	C-type natriuretic peptide	complete	P18104	Ssc.23867	Protein	493772	nomatch	nomatch	nomatch
APEL	Apelin	complete		None			ENSSSCT00000001847	ENSSSCP00000001800	ENSSSCG00000001657
AUGN	Augurin (ECRG4)	complete		Ssc.22487			ENSSSCT00000008919	ENSSSCP00000008693	ENSSSCG00000008142
CALC	Calcitonin/calcitonin gene-related peptide 1	complete	A6P7L6	Ssc.14052	Protein	100125547			
CALCalt	Preprocalcitonin gene-related peptide	complete	A6P7L7	Ssc.56129	Protein	100124407			
CART	Cocaine- and amphetamine-regulated transcript protein	complete	Q307W6	Ssc.15900	Inferred from Homology	397252	ENSSSCT00000018481	ENSSSCP00000017981	ENSSSCG00000016972
CCKN	Cholecystokinin	complete	P01356	Ssc.717	Protein	397468	ENSSSCT00000012346	ENSSSCP00000012023	ENSSSCG00000011277
CMGA	Chromogranin-A	complete	P04404	Ssc.4653	Protein	397540	ENSSSCT00000002727	ENSSSCP00000002658	ENSSSCG00000002456

Supplementary Table S1. (cont.)

Symbol	Prohormone Name	Genome Sequence[1]	UniProt ID	UniGene ID	SwissProt Evidence	Gene Database ID	Ensembl DNA ID	Ensembl Protein ID	Ensembl Gene ID
COLI	Pro-opiomelanocortin	complete	P01192	Ssc.14556	Protein	396863	ENSSSCT00000011914	ENSSSCP00000011608	ENSSSCG00000010890
CORT	Cortistatin	complete		None			ENSSSCT00000003775	ENSSSCP00000003688	ENSSSCG00000003403
CRF	Corticoliberin	complete	P06296	Ssc.69887	Protein	100127468	ENSSSCT00000006811	ENSSSCP00000006626	ENSSSCG00000006215
CRSP1	CRSP1_PIG	complete	Q862B1	Ssc.3741	Protein	396563	ENSSSCT00000014627	ENSSSCP00000014232	ENSSSCG00000013390
CRSP2	CRSP2_PIG	complete	Q766Y7	Ssc.18558	Transcript	396574	ENSSSCT00000014623	ENSSSCP00000014228	ENSSSCG00000013387
CRSP3	CRSP3_PIG	complete	Q766Y6	Ssc.17879	Transcript	396573	ENSSSCT00000014621	ENSSSCP00000014226	ENSSSCG00000013386
CRSP3alt	Calcitonin-2	complete	A0A761						
EDN1	Endothelin-1	complete	P09558	Ssc.9364	Protein	396915	ENSSSCT00000001144	ENSSSCP00000001121	ENSSSCG00000001050
EDN2	Endothelin-2	complete		None			nomatch	nomatch	nomatch
EDN3	Endothelin-3	complete	A5A752	Ssc.31972	Transcript	100049663	ENSSSCT00000008255	ENSSSCP00000008036	ENSSSCG00000007527
GALA	Galanin	complete	P07480	Ssc.713	Protein	397465	ENSSSCT00000014084	ENSSSCP00000013698	ENSSSCG00000012883
GALP	Galanin-like peptide	complete	Q9TT95	Ssc.4875	Protein	396772	ENSSSCT00000003683	ENSSSCP00000003598	ENSSSCG00000003318
GAST	Gastrin	complete	P01351	Ssc.644	Protein	445524	ENSSSCT00000018972	ENSSSCP00000018467	ENSSSCG00000017426
GHRL	Obestatin	complete	Q9GKY5	Ssc.440	Inferred from Homology	396728	ENSSSCT00000012660	ENSSSCP00000012327	ENSSSCG00000011568

Supplementary Table S1. (cont.)

Symbol	Prohormone Name	Genome Sequence[1]	UniProt ID	UniGene ID	SwissProt Evidence	Gene Database ID	Ensembl DNA ID	Ensembl Protein ID	Ensembl Gene ID
GIP	Gastric inhibitory polypeptide	complete	P01281	Ssc.38713	Protein		nomatch	nomatch	Nomatch
GLUC	Glucagon	complete	P01274	Ssc.17225	Protein	397595	ENSSSCT00000017307	ENSSSCP00000016844	ENSSSCG00000015895
GON1	Progonadoliberin-1	complete	P49921	Ssc.16310	Protein	397516	ENSSSCT00000010581	ENSSSCP00000010305	ENSSSCG00000009651
GON2	Progonadoliberin-2	Not in 10.2v18	(F1S8B1)	None			ENSSSCT00000007832	ENSSSCP00000007621	ENSSSCG00000007159
GRP	Gastrin-releasing peptide	fragmentv18	P63153	Ssc.13923	Protein		ENSSSCT00000005421	ENSSSCP00000005286	ENSSSCG00000004913
HEPC	Hepcidin	complete	Q8MJ80	Ssc.376		397207	ENSSSCT00000003189	ENSSSCP00000003108	ENSSSCG00000002886
IAPP	Islet amyloid polypeptide	complete	Q29119	Ssc.8324			ENSSSCT00000000626	ENSSSCP00000000612	ENSSSCG00000000582
IGF1	Insulin-like growth factor 1	complete	P16545	Ssc.16231	Transcript	397491	ENSSSCT00000000936	ENSSSCP00000000916	ENSSSCG00000000857
IGF2	Insulin-like growth factor 2	fragmentv18	P23695	Ssc.9365	Protein	396916	nomatch	nomatch	Nomatch
INS	Insulin	fragmentv18	P01315	Ssc.583	Protein	397415	nomatch	nomatch	Nomatch
INSL3	Insulin-like 3	complete	P51461	Ssc.11990	Transcript	397024	ENSSSCT00000015173	ENSSSCP00000014767	ENSSSCG00000013887
INSL5	Insulin-like 5	complete		None			ENSSSCT00000005747	ENSSSCP00000005605	ENSSSCG00000005214
INSL6	Insulin-like 6	complete		Ssc.46919		100158105	ENSSSCT00000005747	ENSSSCP00000005605	ENSSSCG00000005214
KISS1	Metastasis-suppressor KiSS-1	complete	B5M447	Ssc.73565		100145896	ENSSSCT00000016651	ENSSSCP00000016205	ENSSSCG00000015280

Supplementary Table S1. (cont.)

Symbol	Prohormone Name	Genome Sequence[1]	UniProt ID	UniGene ID	SwissProt Evidence	Gene Database ID	Ensembl DNA ID	Ensembl Protein ID	Ensembl Gene ID
MCH	Pro-melanin-concentrating hormone	complete		Ssc.3287		396962	ENSSSCT00000000937	ENSSSCP00000000917	ENSSSCG00000000858
MOTI	Motilin	complete_10.2.f	P01307	Ssc.714	Protein	397466	ENSSSCT00000001694	ENSSSCP00000001650	ENSSSCG00000001521
NEU1	Oxytocin	complete	P01177	Ssc.15668	Protein	100152272	ENSSSCT00000007837	ENSSSCP00000007626	ENSSSCG00000007164
NEU2	Neurophysin-II	complete	P01183	Ssc.4210	Protein	396995	ENSSSCT00000007836	ENSSSCP00000007625	ENSSSCG00000007163
NEUT	Neurotensin	complete		Ssc.38680			ENSSSCT00000001018	ENSSSCP00000000996	ENSSSCG00000000932
NMB	Neuromedin-B	fragmentv18	B0LUW4	Ssc.2083	Protein	100141313	ENSSSCT00000019060	ENSSSCP00000018554	ENSSSCG00000017507
NMS	Neuromedin-S	complete_10.2.f	C3UZJ1	Ssc.12508		100294685	ENSSSCT00000008951	ENSSSCP00000008725	ENSSSCG00000008173
NMU	Neuromedin-U	complete_10.2.f	P34964	Ssc.12508	Protein		ENSSSCT00000009752	ENSSSCP00000009500	ENSSSCG00000008907
NPB	Neuropeptide B	complete		CA778466.1/ EW120651.2			ENSSSCT00000003809	ENSSSCP00000003722	ENSSSCG00000003431
NPFF	Neuropeptide FF	complete		Ssc.44958			ENSSSCT00000000298	ENSSSCP00000000293	ENSSSCG00000000277
NPS	Neuropeptide S	complete		Ssc.73596		100188981	ENSSSCT00000006864	ENSSSCP00000006678	ENSSSCG00000006265
NPW	Neuropeptide W	complete	Q8MI35	Ssc.15796	Protein	396680	ENSSSCT00000008806	ENSSSCP00000008582	ENSSSCG00000008037
NPY	Neuropeptide Y	complete	P01304	Ssc.15981	Protein	397304	ENSSSCT00000018199	ENSSSCP00000017708	ENSSSCG00000016718
OREX	Orexin	complete	O77668	Ssc.15983	Transcript	397305	ENSSSCT00000018952	ENSSSCP00000018447	ENSSSCG00000017410

Supplementary Table S1. (cont.)

Symbol	Prohormone Name	Genome Sequence[1]	UniProt ID	UniGene ID	SwissProt Evidence	Gene Database ID	Ensembl DNA ID	Ensembl Protein ID	Ensembl Gene ID
OSTN	Osteocrin (Musclin)	complete		Ssc.5148		100049691	ENSSSCT00000012927	ENSSSCP00000012586	ENSSSCG00000011815
OX26	Orexigenic neuropeptide QRFP	complete		None			ENSSSCT00000006274	ENSSSCP00000006112	ENSSSCG00000005705
PACA	Pituitary adenylate cyclase-activating polypeptide	complete	P41535	Ssc.27598	Protein	414283	ENSSSCT00000004097	ENSSSCP00000004005	ENSSSCG00000003698
PAHO	Pancreatic polypeptide	fragmentv18	P01300	Ssc.456	Protein	397272	ENSSSCT00000018907	ENSSSCP00000018403	ENSSSCG00000017368
PCSK1N	Proprotein convertase subtilisin/kexin type 1 inhibitor	fragmentv18		Ssc.17429			ENSSSCT00000018690	ENSSSCP00000018189	ENSSSCG00000017166
PDGFA	Platelet-derived growth factor alpha polypeptide	complete		Ssc.6173			ENSSSCT00000008273	ENSSSCP00000008054	ENSSSCG00000007541
PDGFB	Platelet-derived growth factor beta polypeptide	fragmentv18	P20034	Ssc.54182	Protein	100126843	nomatch	nomatch	nomatch
PDGFD	Platelet-derived growth factor D	complete		Ssc.49835			ENSSSCT00000016357	ENSSSCP00000015918	ENSSSCG00000014994
PDYN	Proenkephalin-B	complete	P01214	Ssc.121	Protein	445529	ENSSSCT00000007854	ENSSSCP00000007643	ENSSSCG00000007180
PENK	Proenkephalin	fragmentv18		Ssc.11281			ENSSSCT00000006841	ENSSSCP00000006655	ENSSSCG00000006243
PNOC	Prepronociceptin	complete	P55791	Ssc.15910	Protein	397257	ENSSSCT00000010609	ENSSSCP00000010333	ENSSSCG00000009675
PROK2	Prokineticin 2	fragmentv18		None			ENSSSCT00000007453	ENSSSCP00000007257	ENSSSCG00000006805

Supplementary Table S1. (cont.)

Symbol	Prohormone Name	Genome Sequence[1]	UniProt ID	UniGene ID	SwissProt Evidence	Gene Database ID	Ensembl DNA ID	Ensembl Protein ID	Ensembl Gene ID
PRRP	Prolactin-releasing peptide	fragmentv18		None			ENSSSCT00000000079	ENSSSCP00000000078	ENSSSCG00000000073
PTHR	Parathyroid hormone-related peptide	complete	Q866H2	Ssc.9991		396951	ENSSSCT00000000587	ENSSSCP00000000574	ENSSSCG00000000544
PTHY	Parathyroid hormone	complete	P01269	Ssc.668	Protein	399502	ENSSSCT00000014631	ENSSSCP00000014236	ENSSSCG00000013394
PYY	Peptide YY	complete	P68005	Ssc.63650	Protein	445018	ENSSSCT00000018902	ENSSSCP00000018398	ENSSSCG00000017363
REL1	Pro-relaxin 1	complete	P01348	Ssc.162		396891	ENSSSCT00000005749	ENSSSCP00000005607	ENSSSCG00000005216
REL3	Relaxin 3	complete	Q8HY17	Ssc.42647	Protein	503836	ENSSSCT00000015039	ENSSSCP00000014634	ENSSSCG00000013765
RES18	Regulated endocrine-specific protein 18	complete		Ssc.49266		100154377	ENSSSCT00000017660	ENSSSCP00000017182	ENSSSCG00000016219
RFRP	Neuropeptide VF precursor	complete_10.2.f	C4P9W1	Ssc.75350		100302024	ENSSSCT00000018193	ENSSSCP00000017702	ENSSSCG00000016712
SCG1	Secretogranin-1	fragmentv18	Q9GLG4	Ssc.15718	Protein	397154	ENSSSCT00000007714	ENSSSCP00000007507	ENSSSCG00000007045
SCG2	Secretogranin-2	complete	Q5FZP5	Ssc.13645	Transcript	497237	ENSSSCT00000017563	ENSSSCP00000017090	ENSSSCG00000016130
SCG3	Secretogranin-3	complete		Ssc.6770		100154760	ENSSSCT00000005113	ENSSSCP00000004989	ENSSSCG00000004630
SECR	Secretin	complete	P63298	Ssc.710	Protein	397464	ENSSSCT00000014047	ENSSSCP00000013662	ENSSSCG00000012852
SLIB	Somatoliberin	complete	P01287	Ssc.71374	Protein		ENSSSCT00000008023	ENSSSCP00000007809	ENSSSCG00000007332

Supplementary Table S1. (cont.)

Symbol	Prohormone Name	Genome Sequence[1]	UniProt ID	UniGene ID	SwissProt Evidence	Gene Database ID	Ensembl DNA ID	Ensembl Protein ID	Ensembl Gene ID
SMS	Somatostatin	complete	P01168	Ssc.19520	Protein	494469	ENSSSCT00000012920	ENSSSCP00000012579	ENSSSCG00000011808
SPXN	Spexin	complete		Ssc.57764		100155886	ENSSSCT00000000622	ENSSSCP00000000608	ENSSSCG00000000578
TIP39	Parathyroid hormone 2	complete		None			ENSSSCT00000003521	ENSSSCP00000003438	ENSSSCG00000003171
TKN1	Tachykinin, precursor 1	complete		Ssc.18075			ENSSSCT00000016711	ENSSSCP00000016265	ENSSSCG00000015338
TKN4	Tachykinin-4	complete		Ssc.23153			ENSSSCT00000019109	ENSSSCP00000018603	ENSSSCG00000017554
TKNK	Tachykinin 3	complete	P67934	Ssc.19565		492314	ENSSSCT00000000452	ENSSSCP00000000446	ENSSSCG00000000418
TOR2X	Torsin family 2, member A	fragment		FS688259/BW972021			ENSSSCT00000006179	ENSSSCP00000006020	ENSSSCG00000005619
TRH	Prothyroliberin	complete	P62968	None			ENSSSCT00000012693	ENSSSCP00000012360	ENSSSCG00000011596
UCN1	Urocortin	fragment_traces		None			ENSSSCT00000016739	ENSSSCP00000016293	ENSSSCG00000015364
UCN2	Urocortin 2	complete		None			ENSSSCT00000012431	ENSSSCP00000012107	ENSSSCG00000011354
UCN3	Urocortin 3	complete		None			ENSSSCT00000012195	ENSSSCP00000011882	ENSSSCG00000011143
UTS2	Urotensin 2	complete	Q95J46	Ssc.437	Transcript	397268	ENSSSCT00000005848	ENSSSCP00000005703	ENSSSCG00000005310
UTS2B	Urotensin Ii-related peptide	complete		None			ENSSSCT00000012928	ENSSSCP00000012587	ENSSSCG00000011816
VEGFC	Vascular endothelial growth factor C	fragmentv18		Ssc.12790			ENSSSCT00000017171	ENSSSCP00000016710	ENSSSCG00000015770

Supplementary Table S1. (cont.)

Symbol	Prohormone Name	Genome Sequence[1]	UniProt ID	UniGene ID	SwissProt Evidence	Gene Database ID	Ensembl DNA ID	Ensembl Protein ID	Ensembl Gene ID
VEGFD	Vascular endothelial growth factor D	complete		Ssc.29289		100155670	ENSSSCT00000013274	ENSSSCP00000012918	ENSSSCG00000012135
VGF	Neurosecretory protein VGF	fragmentv18		FS665444					
VIP	Vasoactive intestinal peptide	complete	E0Y441	Ssc.47759	Protein		ENSSSCT00000004508	ENSSSCP00000004406	ENSSSCG00000004079
	Convertase								
PCSK1	Proprotein convertase subtilisin/kexin type 1 PC1/3	complete	Q28959	Ssc.92884	Transcript	397103	ENSSSCT00000018690	ENSSSCP00000018189	ENSSSCG00000017166
PCSK2	Neuroendocrine convertase 2	complete	Q03333	Ssc.109	Transcript	445533	ENSSSCT00000007755	ENSSSCP00000007546	ENSSSCG00000007083
PCSK3	Furin	complete		Ssc.94009		100156882	ENSSSCT00000002036	ENSSSCP00000001987	ENSSSCG00000001817
PCSK4	Proprotein convertase subtilisin/kexin type 4	complete					ENSSSCT00000015478	ENSSSCP00000015068	ENSSSCG00000014169
PCSK5	Proprotein convertase subtilisin/kexin type 5	complete				100519237	ENSSSCT00000005810	ENSSSCP00000005666	ENSSSCG00000005275
PCSK6	Proprotein convertase subtilisin/kexin type 6	complete					ENSSSCT00000006359	ENSSSCP00000006197	ENSSSCG00000005783
PCSK6	Proprotein convertase subtilisin/kexin type 6 Isoform	complete							

Supplementary Table S1. (cont.)

Symbol	Convertase	Genome Sequence[1]	UniProt ID	UniGene ID	SwissProt Evidence	Gene Database ID	Ensembl DNA ID	Ensembl Protein ID	Ensembl Gene ID
PCSK7	Proprotein convertase subtilisin/kexin type 7	complete				100523009	ENSSSCT00000016438	ENSSSCP00000015996	ENSSSCG00000015072
7b2	Neuroendocrine proetin 7B2	complete	P01165	Ssc.155	Protein				

[1] Gene sequences detected in this study. Complete = the complete gene sequence was identified in the pig genome assembly 10.2; Traces = a fragment of the gene sequence was identified in the pig trace archives; Fragment = fragment gene sequence was identified in the pig genome assembly 10.2; Not Present = gene sequence is not detected in the 10.2 assembly but a match was detected in earlier gene assemblies.

Supplementary Table S2. Detailed distribution of differential expression by tissue group

Symbol	Affy Probe ID	Experiments Immune[1]					
		GSE7313[2]	GSE7314	GSE11787	GSE17492	GSE14758-D	GSE14790
ADM5	Ssc.26627.1.A1_at	5.40E-01	2.60E-02	8.00E-02	7.20E-02	8.10E-01	3.10E-01
ADML	Ssc.314.1.S1_at	1.90E-02	3.40E-04	2.40E-02	5.20E-01	1.70E-03	7.50E-02
ANF	Ssc.16245.1.S1_at	9.70E-01	5.00E-01	1.70E-01	5.90E-01	5.00E-01	2.30E-02
ANFB	Ssc.629.1.S1_at	5.20E-01	3.70E-01	6.70E-01	9.60E-02	9.40E-01	2.10E-01
ANFC	Ssc.23867.1.A1_at	9.20E-01	1.30E-01	3.40E-01	7.40E-02	5.10E-01	3.20E-01
AUGN	Ssc.22487.1.S1_at	4.10E-05	9.00E-03	3.90E-01	6.10E-01	1.90E-01	7.30E-04
CART	Ssc.15900.1.S1_at	2.00E-01	8.40E-02	7.70E-01	4.60E-02	7.10E-01	1.40E-02
CCKN	Ssc.717.1.S1_at	5.40E-01	6.00E-01	5.30E-01	5.00E-02	9.50E-01	1.20E-03
CMGA	Ssc.4653.1.S1_at	1.20E-01	7.30E-01	5.00E-01	2.60E-01	9.70E-01	9.50E-02
COLI	Ssc.14556.1.S1_at	5.90E-01	3.70E-01	6.00E-01	2.10E-01	6.50E-01	8.10E-02
CRSP1	Ssc.3741.1.S1_at	9.00E-01	1.60E-01	8.10E-01	1.50E-02	2.50E-01	1.80E-01
CRSP2	Ssc.18558.1.S1_at	8.10E-01	5.30E-01	6.70E-01	1.00E-01	5.10E-01	8.10E-03
CRSP3	Ssc.17879.1.S1_at	9.30E-01	1.50E-01	9.20E-01	6.30E-02	6.50E-01	2.30E-03
EDN1	Ssc.9364.1.S1_at	2.00E-03	5.20E-04	1.50E-01	1.20E-01	8.70E-02	6.60E-02
GALA	Ssc.713.1.S1_at	3.80E-01	1.10E-01	8.30E-01	2.00E-02	4.00E-01	2.50E-04
GALP	Ssc.4875.1.S1_at	1.10E-01	3.90E-01	1.90E-01	2.20E-01	5.20E-01	1.30E-03
GAST	Ssc.644.1.S1_at	5.80E-01	3.20E-01	6.30E-01	1.40E-01	2.40E-01	4.60E-01
GHRL	Ssc.440.1.S1_at	2.40E-01	1.70E-01	7.50E-01	1.40E-01	5.40E-01	5.80E-01
GLUC	Ssc.17225.1.S1_at	4.40E-01	3.30E-01	9.30E-01	1.90E-01	4.60E-01	2.30E-02
GON1	Ssc.16310.1.S1_at	9.70E-01	1.80E-01	9.10E-01	4.40E-02	6.10E-01	4.00E-05
HEPC	Ssc.376.1.S1_at	9.10E-01	4.80E-01	9.40E-01	2.60E-02	7.90E-01	1.20E-02
IAPP	Ssc.8324.1.A1_at	4.40E-01	4.20E-01	7.20E-01	2.10E-02	7.50E-01	1.80E-01
IGF1	Ssc.16231.1.S1_a_at	5.70E-01	4.30E-03	4.30E-01	1.10E-01	4.20E-01	1.50E-01
	Ssc.16231.2.A1_a_at	6.40E-01	1.20E-01	5.70E-01	2.20E-01	7.50E-01	1.20E-01
	Ssc.16231.3.S1_a_at	6.50E-01	8.10E-02	3.00E-01	1.40E-01	1.20E-01	1.00E-02

Supplementary Table S2. (cont.)

Symbol	Affy Probe ID	Immune[1]					
		GSE7313[2]	GSE7314	GSE11787	GSE17492	GSE14758-D	GSE14790
IGF2	Ssc.9365.1.S1_at	3.90E-01	2.40E-01	2.70E-01	2.60E-01	9.40E-01	2.70E-03
	Ssc.9365.2.S1_a_at	3.40E-02	5.50E-02	1.00E-01	5.50E-03	2.70E-01	9.60E-04
	Ssc.9365.3.S1_a_at	1.90E-01	3.10E-03	1.50E-01	6.40E-02	1.80E-01	6.90E-03
	Ssc.9365.3.S1_x_at	3.90E-02	9.60E-02	2.00E-01	3.70E-02	2.60E-01	3.50E-02
	Ssc.9365.4.S1_a_at	7.50E-01	1.50E-01	7.20E-01	5.00E-01	6.00E-01	5.10E-03
	Ssc.9365.5.A1_at	6.90E-01	1.50E-01	2.70E-01	1.10E-01	3.30E-01	1.70E-03
	Ssc.9365.5.S1_at	9.30E-01	9.60E-01	8.20E-01	2.80E-02	2.30E-01	7.10E-04
	Ssc.9365.5.S1_a_at	6.30E-01	4.10E-01	1.00E-01	9.30E-02	1.20E-01	9.00E-03
	Ssc.9365.6.A1_a_at	1.90E-01	3.30E-01	5.30E-01	8.10E-01	4.20E-01	4.50E-02
	Ssc.9365.6.A1_x_at	4.20E-01	6.40E-01	9.90E-01	6.40E-01	6.10E-01	6.60E-03
	Ssc.9365.6.S1_x_at	3.20E-02	4.20E-02	1.10E-01	2.70E-02	1.10E-01	2.90E-03
	Ssc.9365.7.A1_x_at	7.00E-01	4.50E-01	3.30E-01	5.10E-01	6.70E-01	2.00E-01
INS	Ssc.583.1.S1_at	8.50E-01	1.10E-01	7.70E-01	1.60E-02	6.20E-01	1.90E-02
INSL3	Ssc.11990.1.S1_at	3.10E-01	8.90E-01	4.40E-01	6.30E-02	5.10E-01	7.20E-02
MCH	Ssc.3287.1.S1_at	5.70E-02	1.90E-01	6.60E-01	2.60E-02	5.80E-01	4.80E-02
MOTI	Ssc.714.1.S1_at	8.20E-01	1.80E-01	5.40E-01	1.90E-01	7.80E-01	2.20E-01
NEU1	Ssc.15668.1.A1_at	5.60E-01	4.60E-02	5.40E-01	9.40E-02	2.80E-01	8.00E-03
NEU2	Ssc.4210.1.S1_at	6.50E-01	3.30E-02	4.80E-01	2.00E-01	3.30E-01	6.90E-02
NMB	Ssc.2083.1.A1_at	9.70E-01	3.80E-02	9.00E-01	2.20E-01	2.50E-01	3.20E-03
NMU	Ssc.12508.1.A1_at	9.10E-01	4.80E-01	5.80E-01	1.50E-02	8.50E-01	1.40E-03
NPW	Ssc.15796.1.S1_at	4.10E-01	2.80E-01	3.90E-01	3.20E-02	1.50E-01	6.20E-02
NPY	Ssc.15981.1.A1_at	3.30E-01	2.40E-01	4.10E-01	1.80E-02	6.20E-01	6.40E-04
	Ssc.15981.1.S1_at	5.40E-01	8.30E-01	1.50E-01	2.20E-02	3.40E-01	1.30E-01
OREX	Ssc.15983.1.S1_at	4.40E-01	6.70E-03	3.70E-01	1.10E-01	1.30E-01	4.10E-02
PACA	Ssc.27598.1.S1_at	5.10E-01	4.00E-01	9.20E-01	2.70E-02	6.50E-01	2.60E-01
PAHO	Ssc.456.1.S1_at	7.30E-01	1.50E-01	7.00E-01	3.00E-02	4.70E-01	9.40E-03

Supplementary Table S2. (cont.)

Symbol	Affy Probe ID	Immune[1]					
		GSE7313[2]	GSE7314	GSE11787	GSE17492	GSE14758-D	GSE14790
PCSK1	Ssc.17429.1.S1_at	2.00E-01	2.80E-01	3.10E-01	8.90E-01	1.40E-01	1.70E-02
PDGFA	Ssc.6173.3.S1_a_at	3.00E-01	5.20E-03	3.90E-01	2.20E-01	1.60E-01	1.30E-04
PDYN	Ssc.121.1.S1_at	5.30E-01	4.30E-01	5.30E-01	1.10E-01	4.90E-01	2.10E-01
PENK	Ssc.11281.1.A1_at	2.90E-02	1.20E-01	6.00E-01	1.60E-02	4.60E-01	5.80E-01
	Ssc.11281.2.S1_at	1.90E-01	4.30E-01	6.50E-01	9.50E-02	7.20E-01	2.60E-05
PNOC	Ssc.15910.1.A1_at	9.00E-01	9.00E-01	9.00E-01	2.50E-01	5.00E-01	1.00E-01
	Ssc.15910.1.S1_at	2.90E-01	5.10E-01	4.60E-01	8.10E-02	6.00E-01	1.50E-01
PTHR	Ssc.9991.1.S1_at	6.10E-02	4.20E-01	3.80E-01	8.70E-02	5.50E-01	3.40E-02
PTHY	Ssc.668.1.S1_at	3.90E-02	3.20E-01	9.00E-01	1.30E-02	7.40E-01	8.10E-03
REL1	Ssc.162.1.S1_at	6.50E-01	2.00E-02	8.90E-01	4.60E-02	9.80E-01	1.30E-03
SCG1	Ssc.15718.1.A1_at	9.50E-01	3.10E-02	9.60E-01	8.20E-02	8.00E-01	1.90E-03
SCG2	Ssc.13645.1.A1_at	8.40E-02	8.20E-03	6.90E-01	7.60E-02	3.90E-01	4.80E-05
SCG3	Ssc.6770.1.A1_at	6.70E-01	2.60E-01	1.20E-03	7.50E-01	3.50E-02	7.70E-02
SECR	Ssc.710.1.S1_at	3.90E-01	2.40E-02	1.00E+00	5.10E-02	4.40E-01	5.20E-02
SMS	Ssc.19520.1.A1_at	7.60E-01	1.60E-01	9.40E-01	2.00E-01	1.00E+00	7.40E-05
TKN1	Ssc.18075.1.A1_at	7.00E-01	7.10E-02	5.30E-01	1.50E-02	4.00E-01	8.00E-03
	Ssc.18075.2.S1_at	6.80E-01	8.40E-01	5.10E-01	1.00E-01	6.70E-01	6.30E-03
TKN4	Ssc.23153.1.S1_at	7.20E-01	4.80E-01	5.00E-01	7.40E-02	7.10E-01	6.90E-02
TKNK	Ssc.19565.1.S1_at	7.80E-01	5.10E-02	3.20E-01	2.10E-02	5.60E-02	1.90E-01
	Ssc.19565.2.A1_at	1.50E-01	1.90E-01	3.80E-01	5.40E-02	3.20E-02	5.30E-01
UTS2	Ssc.437.1.S1_a_at	9.30E-01	3.90E-01	4.10E-01	2.00E-02	8.60E-01	3.40E-02
VEGFC	Ssc.12790.1.A1_at	4.10E-01	2.30E-02	1.40E-02	7.10E-01	1.30E-01	1.20E-03
VEGFD	Ssc.29289.1.A1_at	2.70E-01	2.30E-01	7.10E-01	1.20E-01	4.00E-01	1.00E-04
Prohormone Convertase							
PCSK1	Ssc.141.1.S1_at	9.50E-01	1.70E-01	8.60E-01	1.50E-02	5.80E-01	3.40E-04

Supplementary Table S2. (cont.)

Symbol	Affy Probe ID	Immune[1]					
		GSE7313[2]	GSE7314	GSE11787	GSE17492	GSE14758-D	GSE14790
PCSK2	Ssc.109.1.S1_at	4.40E-01	2.70E-01	3.50E-01	2.50E-02	5.80E-01	6.50E-03
PCSK7	Ssc.5628.1.S1_at	3.80E-01	2.70E-01	7.30E-01	2.90E-02	7.40E-02	1.30E-03
Total P-value in Column <0.005		2	4	1	0	1	24
Totals by Group							32

Symbol	Affy Probe ID	Experiments				
		Embryo and Placenta				
		GSE18467	GSE18641	GSE18343	GSE11853	GSE12705
ADM5	Ssc.26627.1.A1_at	1.10E-02	1.50E-01	3.70E-01	7.00E-01	1.80E-01
ADML	Ssc.314.1.S1_at	8.50E-01	8.60E-02	7.20E-01	4.70E-02	1.20E-02
ANF	Ssc.16245.1.S1_at	1.80E-01	7.80E-01	2.70E-01	7.90E-01	7.40E-02
ANFB	Ssc.629.1.S1_at	4.00E-01	3.00E-01	5.40E-01	1.10E-01	3.00E-01
ANFC	Ssc.23867.1.A1_at	6.10E-01	4.70E-01	3.10E-01	5.90E-01	2.90E-05
AUGN	Ssc.22487.1.S1_at	9.00E-01	7.10E-01	6.50E-01	4.00E-01	1.50E-01
CART	Ssc.15900.1.S1_at	3.40E-01	3.90E-01	2.90E-01	3.40E-01	2.30E-03
CCKN	Ssc.717.1.S1_at	1.60E-01	9.80E-02	5.30E-01	5.30E-01	1.10E-01
CMGA	Ssc.4653.1.S1_at	1.40E-01	8.20E-01	4.10E-01	9.40E-01	2.00E-01
COLI	Ssc.14556.1.S1_at	2.40E-03	6.30E-01	9.20E-01	7.00E-01	1.20E-01
CRSP1	Ssc.3741.1.S1_at	4.40E-01	3.70E-01	7.70E-01	7.40E-01	7.10E-02
CRSP2	Ssc.18558.1.S1_at	2.20E-01	4.10E-01	8.80E-01	5.50E-01	3.30E-05
CRSP3	Ssc.17879.1.S1_at	1.10E-01	4.20E-01	8.50E-01	8.00E-01	1.20E-01

Supplementary Table S2. (cont.)

Symbol	Affy Probe ID	Embryo and Placenta				
		GSE18467	GSE18467	GSE18467	GSE18467	GSE18467
EDN1	Ssc.9364.1.S1_at	9.60E-01	1.20E-01	2.70E-01	1.50E-01	4.70E-01
GALA	Ssc.713.1.S1_at	1.30E-02	7.30E-01	7.50E-01	6.80E-01	1.10E-03
GALP	Ssc.4875.1.S1_at	2.90E-01	6.40E-01	4.60E-01	7.20E-01	2.10E-04
GAST	Ssc.644.1.S1_at	1.50E-01	2.60E-01	8.60E-01	5.90E-01	5.30E-04
GHRL	Ssc.440.1.S1_at	2.00E-01	6.70E-01	8.20E-01	7.30E-01	7.50E-01
GLUC	Ssc.17225.1.S1_at	2.60E-01	3.00E-01	6.10E-01	2.20E-01	5.30E-05
GON1	Ssc.16310.1.S1_at	2.20E-01	4.50E-01	7.80E-01	7.80E-01	2.10E-04
HEPC	Ssc.376.1.S1_at	1.50E-01	9.50E-01	6.00E-01	6.70E-01	2.30E-01
IAPP	Ssc.8324.1.A1_at	1.30E-01	9.00E-01	5.70E-01	7.90E-01	1.20E-03
IGF1	Ssc.16231.1.S1_a_at	2.40E-01	3.70E-01	1.60E-01	9.50E-02	1.10E-02
	Ssc.16231.2.A1_a_at	1.20E-01	3.70E-01	9.50E-01	7.80E-01	1.30E-02
	Ssc.16231.3.S1_a_at	5.80E-01	3.80E-01	9.50E-02	5.30E-01	3.60E-02
IGF2	Ssc.9365.1.S1_at	2.00E-01	8.50E-01	6.20E-01	5.00E-01	8.90E-03
	Ssc.9365.2.S1_a_at	8.60E-01	2.70E-03	7.10E-01	6.80E-01	2.60E-02
	Ssc.9365.3.S1_a_at	9.40E-01	1.50E-02	5.90E-01	3.90E-01	3.50E-01
	Ssc.9365.3.S1_x_at	7.20E-01	2.00E-02	7.50E-01	3.90E-01	1.90E-02
	Ssc.9365.4.S1_a_at	2.40E-01	1.00E+00	9.30E-01	5.80E-01	5.20E-04
	Ssc.9365.5.A1_at	8.50E-01	5.40E-01	8.00E-01	6.20E-01	1.90E-02
	Ssc.9365.5.S1_at	1.80E-01	5.20E-01	2.80E-01	7.10E-01	5.30E-04
	Ssc.9365.5.S1_a_at	8.70E-01	1.60E-02	2.10E-01	6.80E-01	3.00E-01
	Ssc.9365.6.A1_a_at	2.30E-01	7.00E-01	9.10E-01	3.40E-01	7.70E-02
	Ssc.9365.6.A1_x_at	3.40E-01	2.90E-01	3.00E-01	1.80E-01	6.20E-02
	Ssc.9365.6.S1_x_at	1.30E-01	3.40E-02	7.00E-01	4.80E-01	1.20E-01
	Ssc.9365.7.A1_x_at	1.90E-01	6.90E-01	5.30E-01	3.90E-01	2.60E-02
INS	Ssc.583.1.S1_at	2.20E-01	4.90E-01	8.90E-01	9.10E-01	4.40E-01
INSL3	Ssc.11990.1.S1_at	3.50E-01	6.10E-01	8.60E-01	8.80E-01	1.30E-03

Supplementary Table. S2. (cont.)

Symbol	Affy Probe ID	Embryo and Placenta				
		GSE18467	GSE18467	GSE18467	GSE18467	GSE18467
MCH	Ssc.3287.1.S1_at	3.60E-01	9.20E-01	8.40E-01	8.50E-01	2.50E-01
MOTI	Ssc.714.1.S1_at	3.90E-01	9.60E-01	6.20E-01	6.10E-01	9.10E-01
NEU1	Ssc.15668.1.A1_at	5.50E-01	3.50E-01	6.30E-01	7.80E-01	6.30E-02
NEU2	Ssc.4210.1.S1_at	5.30E-01	5.30E-01	6.70E-01	4.70E-01	6.20E-01
NMB	Ssc.2083.1.A1_at	2.10E-01	1.40E-02	7.70E-01	8.40E-01	1.40E-02
NMU	Ssc.12508.1.A1_at	3.00E-01	1.30E-02	3.90E-01	8.50E-01	1.30E-01
NPW	Ssc.15796.1.S1_at	3.00E-01	5.00E-01	7.60E-01	4.90E-01	3.30E-05
NPY	Ssc.15981.1.A1_at	7.50E-01	9.30E-01	7.60E-01	8.80E-01	7.30E-04
OREX	Ssc.15981.1.S1_at	7.20E-02	1.40E-03	5.70E-01	9.70E-01	3.50E-04
	Ssc.15983.1.S1_at	5.60E-01	2.20E-01	9.60E-01	7.50E-01	1.00E-01
PACA	Ssc.27598.1.S1_at	6.40E-01	5.80E-01	1.90E-01	5.30E-01	3.90E-04
PAHO	Ssc.456.1.S1_at	3.00E-01	4.20E-01	5.50E-01	8.20E-01	6.00E-05
PCSK1	Ssc.17429.1.S1_at	1.60E-01	4.90E-01	5.30E-01	4.20E-01	4.80E-03
PDGFA	Ssc.6173.3.S1_a_at	6.70E-02	4.50E-01	3.50E-01	3.50E-01	6.50E-03
PDYN	Ssc.121.1.S1_at	9.30E-01	9.90E-01	7.80E-01	6.20E-01	3.00E-03
PENK	Ssc.11281.1.A1_at	1.90E-01	5.40E-02	8.50E-02	7.90E-01	1.90E-04
PNOC	Ssc.11281.2.S1_at	2.40E-01	9.40E-02	7.90E-01	6.30E-01	7.70E-06
	Ssc.15910.1.A1_at	1.60E-01	4.00E-01	5.10E-01	5.40E-01	6.80E-01
	Ssc.15910.1.S1_at	5.70E-02	3.50E-01	8.90E-01	2.00E-01	1.20E-01
PTHR	Ssc.9991.1.S1_at	2.70E-01	7.60E-03	2.00E-01	1.50E-01	3.50E-07
PTHY	Ssc.668.1.S1_at	1.90E-01	5.10E-01	8.30E-01	6.80E-01	3.30E-04
REL1	Ssc.162.1.S1_at	3.40E-01	7.70E-01	4.50E-01	2.90E-01	4.50E-04
SCG1	Ssc.15718.1.A1_at	1.10E-01	8.10E-01	6.10E-01	6.60E-01	3.60E-05
SCG2	Ssc.13645.1.A1_at	3.20E-02	8.70E-01	8.30E-01	4.10E-01	1.70E-02
SCG3	Ssc.6770.1.A1_at	4.40E-01	5.10E-01	6.10E-01	7.30E-01	9.60E-05
SECR	Ssc.710.1.S1_at	5.90E-01	8.60E-01	8.20E-01	6.40E-01	3.50E-05

Supplementary Table. S2. (cont.)

Symbol	Affy Probe ID	Embryo and Placenta				
		GSE18467	GSE18467	GSE18467	GSE18467	GSE18467
SMS	Ssc.19520.1.A1_at	4.80E-01	3.90E-02	1.30E-01	8.60E-01	6.80E-04
TKN1	Ssc.18075.1.A1_at	6.00E-01	3.00E-01	3.50E-01	5.40E-01	1.50E-01
	Ssc.18075.2.S1_at	6.30E-01	4.30E-01	7.00E-01	7.20E-01	1.30E-04
TKN4	Ssc.23153.1.S1_at	2.00E-01	9.30E-01	1.00E+00	2.30E-01	5.60E-03
TKNK	Ssc.19565.1.S1_at	2.60E-01	6.10E-01	5.20E-01	5.30E-01	6.60E-01
	Ssc.19565.2.A1_at	2.20E-01	3.20E-01	5.30E-01	7.10E-01	3.20E-01
UTS2	Ssc.437.1.S1_a_at	1.50E-01	6.60E-01	4.90E-01	7.60E-01	7.80E-05
VEGFC	Ssc.12790.1.A1_at	1.90E-02	7.80E-04	9.00E-01	8.70E-01	6.40E-01
VEGFD	Ssc.29289.1.A1_at	7.40E-02	1.40E-01	3.10E-01	2.70E-01	3.60E-05
Prohormone Convertase						
PCSK1	Ssc.141.1.S1_at	2.70E-02	6.60E-01	5.10E-01	8.40E-01	1.40E-04
PCSK2	Ssc.109.1.S1_at	9.90E-02	4.20E-01	5.70E-01	8.30E-01	5.10E-01
PCSK7	Ssc.5628.1.S1_at	6.20E-01	2.40E-01	4.10E-01	7.30E-01	9.70E-04
Total P-value in Column <0.005		1	3	0	0	33
Totals by Group						37

Supplementary Table S2. (cont.)

Symbol	Affy Probe ID	Experiments						
		Brain and Central Nervous System					Reproduction	
		GSE16855	GSE12604	GSE14739-H	GSE14739-T	GSE14739-A	GSE11590	GSE14739-G
ADM5	Ssc.26627.1.A1_at	3.40E-01	5.30E-01	9.70E-01	1.90E-03	5.00E-01	3.90E-01	6.20E-02
ADML	Ssc.314.1.S1_at	2.20E-06	2.60E-01	5.20E-01	7.20E-01	8.70E-02	1.60E-01	5.20E-02
ANF	Ssc.16245.1.S1_at	3.50E-01	9.70E-01	9.30E-01	3.30E-01	5.40E-01	2.00E-01	8.60E-01
ANFB	Ssc.629.1.S1_at	1.90E-01	6.20E-01	7.30E-01	5.10E-01	1.00E-01	6.00E-02	6.60E-01
ANFC	Ssc.23867.1.A1_at	9.40E-02	4.20E-01	9.40E-04	3.20E-01	4.80E-01	1.60E-01	6.10E-01
AUGN	Ssc.22487.1.S1_at	3.50E-01	1.30E-01	4.70E-02	2.80E-01	9.00E-02	1.20E-01	7.90E-02
CART	Ssc.15900.1.S1_at	3.60E-01	7.60E-01	3.60E-01	2.40E-03	2.50E-01	1.50E-01	3.20E-01
CCKN	Ssc.717.1.S1_at	6.30E-01	1.60E-01	2.60E-01	4.10E-01	5.00E-01	1.80E-01	1.20E-07
CMGA	Ssc.4653.1.S1_at	4.10E-01	2.70E-01	5.00E-01	8.70E-01	5.30E-01	1.50E-01	4.70E-01
COLI	Ssc.14556.1.S1_at	8.70E-01	6.40E-01	9.10E-01	3.50E-01	2.90E-01	1.50E-01	7.40E-01
CRSP1	Ssc.3741.1.S1_at	7.00E-02	4.80E-01	7.70E-01	6.50E-01	6.10E-03	1.40E-01	6.60E-01
CRSP2	Ssc.18558.1.S1_at	3.20E-01	5.40E-01	6.00E-01	1.70E-01	5.10E-01	1.40E-01	7.10E-01
CRSP3	Ssc.17879.1.S1_at	8.50E-01	8.80E-01	4.00E-01	7.30E-01	3.30E-02	1.60E-01	2.20E-01
EDN1	Ssc.9364.1.S1_at	9.70E-01	7.90E-01	7.10E-01	6.60E-01	5.40E-02	1.60E-01	1.60E-02
GALA	Ssc.713.1.S1_at	8.80E-01	5.10E-01	8.90E-01	3.30E-01	4.10E-01	1.50E-01	6.10E-01
GALP	Ssc.4875.1.S1_at	8.10E-01	4.70E-01	6.20E-01	1.50E-01	6.70E-02	1.40E-01	1.70E-02
GAST	Ssc.644.1.S1_at	1.90E-01	3.40E-01	9.00E-01	3.10E-01	5.30E-01	1.50E-01	5.20E-01
GHRL	Ssc.440.1.S1_at	2.90E-01	6.40E-01	9.80E-01	3.60E-01	2.50E-01	1.60E-01	5.50E-01
GLUC	Ssc.17225.1.S1_at	6.80E-01	8.00E-01	7.00E-01	5.00E-01	3.60E-01	5.60E-04	6.70E-01
GON1	Ssc.16310.1.S1_at	8.40E-01	6.90E-01	4.10E-01	3.20E-01	2.20E-01	1.40E-01	7.50E-01
HEPC	Ssc.376.1.S1_at	5.50E-01	6.50E-01	8.20E-01	6.20E-01	1.50E-01	1.60E-01	8.40E-01
IAPP	Ssc.8324.1.A1_at	9.20E-01	9.70E-01	2.30E-01	3.50E-01	8.20E-02	1.20E-01	6.10E-01
IGF1	Ssc.16231.1.S1_a_at	4.70E-06	4.60E-01	7.60E-01	3.80E-01	1.60E-01	1.70E-01	8.60E-03
	Ssc.16231.2.A1_a_at	7.70E-01	7.30E-01	8.40E-01	1.30E-01	4.70E-01	1.40E-01	7.90E-01
	Ssc.16231.3.S1_a_at	2.40E-07	4.90E-01	9.70E-01	4.90E-01	5.40E-01	1.70E-01	7.20E-02
IGF2	Ssc.9365.1.S1_at	4.70E-01	4.30E-01	8.20E-01	1.50E-01	2.40E-01	7.80E-01	3.60E-01
	Ssc.9365.2.S1_a_at	2.90E-01	3.20E-01	8.70E-02	4.40E-01	8.50E-01	1.40E-01	1.40E-01

Supplementary Table S2. (cont.)

Symbol	Affy Probe ID	Brain and Central Nervous System					Reproduction	
		GSE16855	GSE12604	GSE14739-H	GSE14739-T	GSE14739-A	GSE11590	GSE14739-G
	Ssc.9365.3.S1_a_at	6.10E-02	5.70E-01	1.10E-02	6.30E-01	9.40E-01	1.70E-01	2.50E-01
	Ssc.9365.3.S1_x_at	7.70E-01	6.00E-01	5.10E-01	5.50E-01	6.60E-01	1.50E-01	3.50E-01
	Ssc.9365.4.S1_a_at	3.50E-01	4.70E-01	9.10E-01	2.00E-01	1.00E-01	1.60E-01	3.90E-01
	Ssc.9365.5.A1_at	1.30E-01	3.40E-01	9.10E-01	7.90E-01	4.70E-01	1.60E-01	4.70E-01
	Ssc.9365.5.S1_at	9.30E-01	4.80E-01	7.50E-01	5.20E-01	3.40E-01	1.50E-01	9.20E-01
	Ssc.9365.5.S1_a_at	6.40E-02	6.00E-01	9.60E-03	6.30E-01	6.20E-01	1.70E-01	3.70E-01
	Ssc.9365.6.A1_a_at	7.60E-01	6.50E-01	6.90E-01	6.10E-02	6.80E-02	1.60E-01	3.00E-01
	Ssc.9365.6.A1_x_at	2.70E-01	7.50E-01	4.60E-01	6.00E-02	3.00E-01	5.30E-01	3.20E-01
	Ssc.9365.6.S1_x_at	1.80E-02	3.40E-01	2.30E-03	5.70E-01	3.40E-01	1.40E-01	3.50E-01
	Ssc.9365.7.A1_x_at	3.80E-01	4.50E-01	8.50E-01	1.80E-01	1.60E-01	2.20E-01	3.20E-01
INS	Ssc.583.1.S1_at	6.00E-01	4.00E-01	9.10E-01	2.20E-01	4.20E-01	1.50E-01	6.80E-01
INSL3	Ssc.11990.1.S1_at	2.00E-01	5.40E-01	8.10E-01	3.80E-01	4.70E-01	1.50E-01	7.20E-02
MCH	Ssc.3287.1.S1_at	9.40E-01	2.10E-01	3.50E-01	3.30E-01	1.60E-01	1.60E-01	6.10E-01
MOTI	Ssc.714.1.S1_at	7.90E-01	6.00E-01	8.20E-01	6.70E-01	1.10E-01	1.60E-01	7.30E-01
NEU1	Ssc.15668.1.A1_at	3.90E-01	6.30E-01	1.30E-01	3.90E-01	9.80E-01	1.40E-01	1.80E-01
NEU2	Ssc.4210.1.S1_at	1.20E-01	6.60E-01	1.60E-01	3.00E-01	7.20E-01	1.60E-01	4.30E-01
NMB	Ssc.2083.1.A1_at	5.00E-01	5.20E-01	2.10E-01	2.10E-01	1.50E-02	1.50E-01	6.80E-01
NMU	Ssc.12508.1.A1_at	6.90E-01	3.90E-02	5.50E-01	3.40E-01	1.70E-01	1.30E-01	5.60E-01
NPW	Ssc.15796.1.S1_at	6.60E-01	4.30E-01	9.40E-01	4.10E-01	5.20E-01	1.60E-01	6.30E-01
NPY	Ssc.15981.1.A1_at	7.60E-01	9.50E-01	8.50E-01	5.10E-01	1.00E-01	1.60E-01	5.50E-01
	Ssc.15981.1.S1_at	8.10E-04	3.20E-01	9.80E-01	7.20E-01	2.30E-01	1.40E-01	7.30E-01
OREX	Ssc.15983.1.S1_at	5.10E-01	1.60E-01	2.00E-01	7.50E-01	6.30E-01	2.60E-01	8.40E-01
PACA	Ssc.27598.1.S1_at	1.10E-01	3.30E-01	2.30E-02	2.20E-01	7.40E-01	7.40E-01	8.00E-01
PAHO	Ssc.456.1.S1_at	6.10E-01	6.60E-01	9.80E-01	3.50E-01	3.00E-01	1.40E-01	5.80E-01
PCSK1	Ssc.17429.1.S1_at	7.50E-02	8.80E-01	9.90E-01	5.40E-01	9.80E-01	3.00E-01	3.70E-05

Supplementary Table S2. (cont.)

Symbol	Affy Probe ID	Brain and Central Nervous System					Reproduction	
		GSE16855	GSE12604	GSE14739-H	GSE14739-T	GSE14739-A	GSE11590	GSE14739-G
PDGFA	Ssc.6173.3.S1_a_at	1.20E-03	3.00E-01	9.90E-02	5.40E-01	1.90E-01	9.60E-02	1.10E-01
PDYN	Ssc.121.1.S1_at	4.90E-05	8.80E-01	3.50E-01	5.00E-01	5.10E-01	1.20E-01	7.20E-01
PENK	Ssc.11281.1.A1_at	9.30E-01	6.50E-01	4.50E-01	7.50E-01	5.00E-01	4.30E-01	2.00E-05
	Ssc.11281.2.S1_at	2.40E-01	3.00E-01	6.80E-01	3.60E-01	4.60E-01	3.10E-01	7.20E-02
PNOC	Ssc.15910.1.A1_at	3.80E-01	5.20E-01	8.60E-01	7.00E-01	3.80E-01	1.60E-01	4.90E-01
PTHR	Ssc.15910.1.S1_at	7.20E-01	7.90E-01	4.60E-01	5.00E-01	1.40E-01	1.60E-01	6.80E-01
PTHY	Ssc.9991.1.S1_at	2.60E-04	9.20E-02	1.80E-01	4.70E-01	1.80E-01	1.50E-03	7.10E-10
REL1	Ssc.668.1.S1_at	5.60E-01	8.20E-01	5.30E-01	2.20E-01	6.10E-02	1.50E-01	5.70E-01
SCG1	Ssc.162.1.S1_at	3.00E-01	8.90E-01	4.60E-01	2.80E-01	1.60E-01	2.40E-01	4.90E-01
SCG2	Ssc.15718.1.A1_at	8.30E-01	6.70E-01	5.20E-01	1.70E-01	9.80E-01	1.40E-01	7.80E-01
SCG3	Ssc.13645.1.A1_at	6.30E-01	6.70E-01	4.80E-01	1.80E-01	1.30E-01	1.60E-01	6.80E-01
SECR	Ssc.6770.1.A1_at	8.70E-01	4.90E-01	5.80E-03	4.00E-01	3.80E-02	1.00E-03	4.60E-01
SMS	Ssc.710.1.S1_at	6.30E-01	1.90E-01	9.00E-02	2.90E-01	4.10E-01	9.30E-01	5.50E-01
TKN1	Ssc.19520.1.A1_at	5.00E-01	7.20E-01	6.70E-01	3.00E-01	1.60E-01	2.40E-01	6.90E-01
	Ssc.18075.1.A1_at	7.30E-01	1.60E-01	9.30E-01	5.60E-01	1.80E-01	1.30E-01	6.60E-01
TKN4	Ssc.23153.1.S1_at	9.20E-01	2.90E-01	8.00E-01	4.30E-01	2.70E-01	1.30E-01	5.10E-01
TKNK	Ssc.19565.1.S1_at	8.00E-01	5.40E-01	8.10E-01	4.80E-01	2.50E-01	1.50E-01	2.50E-02
	Ssc.19565.2.A1_at	1.90E-01	3.00E-01	7.00E-01	8.00E-01	3.30E-01	3.30E-01	1.40E-02
UTS2	Ssc.437.1.S1_a_at	4.70E-01	4.40E-01	8.90E-01	4.00E-01	1.20E-01	8.90E-02	5.50E-01
VEGFC	Ssc.12790.1.A1_at	1.50E-09	5.30E-01	5.50E-02	3.80E-01	1.50E-01	1.50E-01	4.40E-01
VEGFD	Ssc.29289.1.A1_at	3.70E-01	3.90E-01	3.00E-01	7.80E-01	4.60E-01	5.90E-01	7.80E-02
Prohormone Convertase								
PCSK1	Ssc.141.1.S1_at	6.60E-01	5.00E-01	6.20E-02	9.10E-02	9.90E-03	6.50E-01	2.50E-01
PCSK2	Ssc.109.1.S1_at	4.20E-01	3.60E-01	4.60E-01	6.30E-01	5.10E-01	1.00E-01	5.70E-01
PCSK7	Ssc.5628.1.S1_at	1.00E-02	6.90E-01	1.30E-01	9.90E-01	5.30E-01	2.00E-01	1.10E-01

Supplementary Table S2. (cont.)

Total P-value in Column <0.005	8	0	2	2	0	3	4
Totals by Group					12		7

Symbol	Affy Probe ID	Experiments Muscle						
		GSE18653	GSE19275	GSE8974	GSE14643	GSE15211	GSE21096	GSE16348-D
ADM5	Ssc.26627.1.A1_at	8.10E-01	1.70E-02	1.20E-01	9.80E-01	1.30E-01	1.10E-01	2.50E-01
ADML	Ssc.314.1.S1_at	6.60E-01	2.90E-02	6.30E-04	2.80E-01	7.50E-01	2.20E-01	3.30E-01
ANF	Ssc.16245.1.S1_at	4.60E-01	2.50E-01	9.80E-01	1.20E-01	3.50E-01	2.30E-04	5.20E-02
ANFB	Ssc.629.1.S1_at	6.10E-01	8.20E-02	2.40E-02	1.20E-01	3.80E-01	1.10E-03	9.60E-02
ANFC	Ssc.23867.1.A1_at	1.40E-01	2.40E-02	9.70E-01	9.40E-01	1.20E-01	1.00E-02	7.30E-01
AUGN	Ssc.22487.1.S1_at	6.00E-01	8.80E-01	3.90E-01	6.90E-01	2.90E-01	1.70E-04	7.20E-01
CART	Ssc.15900.1.S1_at	6.50E-01	4.60E-01	6.00E-01	4.30E-01	2.10E-02	1.00E-02	7.60E-02
CCKN	Ssc.717.1.S1_at	1.40E-01	2.80E-02	9.40E-01	5.80E-01	3.80E-01	2.10E-04	5.90E-02
CMGA	Ssc.4653.1.S1_at	1.00E+00	9.80E-02	4.30E-01	8.30E-01	8.40E-01	9.80E-02	1.00E-01
COLI	Ssc.14556.1.S1_at	1.10E-01	3.40E-01	8.20E-01	4.90E-01	1.60E-01	5.70E-02	1.00E-01
CRSP1	Ssc.3741.1.S1_at	7.90E-01	1.60E-01	2.60E-01	7.80E-01	2.80E-01	2.10E-02	4.50E-01
CRSP2	Ssc.18558.1.S1_at	1.60E-01	1.80E-02	7.00E-02	6.00E-01	4.80E-02	2.10E-02	7.80E-02
CRSP3	Ssc.17879.1.S1_at	8.10E-02	6.00E-02	4.80E-01	8.20E-01	1.60E-01	5.30E-02	6.50E-02
EDN1	Ssc.9364.1.S1_at	9.90E-01	1.00E-02	3.40E-01	4.30E-01	2.10E-01	3.20E-01	1.00E-01
GALA	Ssc.713.1.S1_at	1.00E-01	3.90E-02	9.40E-01	5.10E-01	2.70E-01	3.00E-02	1.10E-01
GALP	Ssc.4875.1.S1_at	1.90E-01	8.30E-04	5.50E-01	6.10E-01	4.90E-01	7.20E-03	3.80E-01
GAST	Ssc.644.1.S1_at	9.70E-01	3.70E-01	4.80E-01	9.00E-01	1.10E-01	6.80E-02	1.50E-01
GHRL	Ssc.440.1.S1_at	7.50E-01	1.30E-02	7.40E-01	4.90E-01	1.40E-01	3.40E-02	1.60E-01
GLUC	Ssc.17225.1.S1_at	7.90E-02	4.70E-01	7.50E-01	2.90E-01	3.50E-01	7.50E-03	1.20E-01

Supplementary Table S2. (cont.)

Symbol	Affy Probe ID	Muscle						
		GSE18653	GSE19275	GSE8974	GSE14643	GSE15211	GSE21096	GSE16348-D
GON1	Ssc.16310.1.S1_at	3.20E-01	2.00E-02	8.30E-01	3.30E-01	3.10E-02	4.10E-02	5.80E-02
HEPC	Ssc.376.1.S1_at	3.70E-01	2.90E-02	9.30E-01	2.10E-01	2.70E-01	3.10E-02	1.30E-01
IAPP	Ssc.8324.1.A1_at	1.20E-01	5.10E-02	7.70E-01	2.00E-01	8.00E-01	5.20E-02	5.80E-02
IGF1	Ssc.16231.1.S1_a_at	3.00E-01	5.30E-01	1.30E-01	5.70E-01	9.30E-01	2.20E-02	5.20E-02
	Ssc.16231.2.A1_a_at	4.80E-01	1.70E-01	6.20E-01	2.30E-01	1.40E-01	6.00E-02	3.10E-01
	Ssc.16231.3.S1_a_at	6.00E-01	7.10E-01	8.00E-02	1.00E+00	9.10E-01	2.10E-02	3.20E-01
IGF2	Ssc.9365.1.S1_at	4.10E-01	3.50E-01	6.60E-01	3.90E-01	3.10E-01	1.80E-02	2.70E-01
	Ssc.9365.2.S1_a_at	5.10E-01	6.20E-02	1.10E-01	7.50E-01	6.30E-01	7.90E-02	4.20E-01
	Ssc.9365.3.S1_a_at	7.30E-01	1.00E-01	9.50E-01	7.50E-01	3.60E-01	4.30E-02	1.60E-01
	Ssc.9365.3.S1_x_at	5.60E-01	1.50E-01	2.00E-01	5.30E-01	4.10E-01	3.50E-02	2.50E-01
	Ssc.9365.4.S1_a_at	5.70E-01	2.60E-02	8.50E-01	1.80E-01	9.90E-02	4.10E-02	2.60E-01
	Ssc.9365.5.A1_at	4.10E-01	6.80E-02	8.90E-01	9.70E-01	1.60E-01	7.90E-02	2.50E-01
	Ssc.9365.5.S1_at	5.70E-01	2.60E-01	4.10E-01	4.30E-01	1.20E-01	2.80E-02	3.10E-01
	Ssc.9365.5.S1_a_at	3.80E-01	1.90E-01	8.40E-01	6.40E-01	7.30E-01	1.10E-01	3.80E-01
	Ssc.9365.6.A1_a_at	2.60E-01	7.30E-01	7.60E-01	7.60E-01	7.00E-01	9.10E-02	3.40E-01
	Ssc.9365.6.A1_x_at	2.60E-01	3.60E-01	7.30E-01	9.40E-01	5.30E-01	5.00E-03	1.70E-02
	Ssc.9365.6.S1_x_at	3.30E-01	3.40E-01	7.30E-01	3.20E-01	2.40E-01	8.50E-03	1.10E-01
	Ssc.9365.7.A1_x_at	4.50E-01	4.60E-01	8.70E-01	9.40E-01	8.10E-02	1.20E-01	3.30E-02
INS	Ssc.583.1.S1_at	3.80E-01	2.70E-02	8.60E-01	4.10E-01	1.00E+00	1.20E-02	1.00E-02
INSL3	Ssc.11990.1.S1_at	1.60E-01	2.60E-01	9.20E-01	3.50E-01	2.60E-01	7.90E-02	4.00E-01
MCH	Ssc.3287.1.S1_at	9.80E-01	1.80E-02	2.30E-01	8.80E-01	3.60E-01	7.80E-02	7.00E-02
MOTI	Ssc.714.1.S1_at	2.40E-01	1.00E-02	9.70E-01	2.70E-01	3.40E-01	3.80E-02	9.70E-02
NEU1	Ssc.15668.1.A1_at	1.20E-01	1.30E-02	7.40E-01	3.70E-01	5.10E-02	6.70E-03	4.80E-02

Supplementary Table S2. (cont.)

Symbol	Affy Probe ID	Muscle						
		GSE18653	GSE19275	GSE8974	GSE14643	GSE15211	GSE21096	GSE16348-D
NEU2	Ssc.4210.1.S1_at	8.00E-02	3.50E-02	7.80E-01	4.60E-01	7.10E-02	2.90E-03	1.40E-02
NMB	Ssc.2083.1.A1_at	5.40E-01	2.10E-02	2.60E-01	6.20E-01	6.90E-02	4.40E-02	7.10E-02
NMU	Ssc.12508.1.A1_at	1.20E-01	1.30E-02	7.80E-01	2.00E-01	7.80E-01	5.00E-02	6.40E-02
NPW	Ssc.15796.1.S1_at	1.10E-01	2.20E-01	7.30E-01	8.50E-01	1.50E-01	3.90E-02	2.70E-01
NPY	Ssc.15981.1.A1_at	1.80E-01	1.90E-01	4.50E-01	5.80E-01	3.80E-01	3.00E-03	1.30E-01
	Ssc.15981.1.S1_at	3.40E-01	2.90E-02	5.70E-01	4.40E-01	5.50E-01	3.40E-02	9.50E-02
OREX	Ssc.15983.1.S1_at	6.50E-01	1.70E-01	7.40E-01	2.80E-01	1.10E-01	9.90E-01	9.40E-02
PACA	Ssc.27598.1.S1_at	2.80E-01	2.30E-02	8.10E-01	2.00E-01	2.60E-01	2.30E-01	5.60E-01
PAHO	Ssc.456.1.S1_at	2.70E-01	3.10E-02	8.70E-01	3.10E-01	2.00E-01	3.20E-02	9.30E-02
PCSK1	Ssc.17429.1.S1_at	3.50E-01	9.30E-01	5.10E-02	1.80E-01	9.30E-02	9.40E-02	1.30E-01
PDGFA	Ssc.6173.3.S1_a_at	1.90E-01	7.70E-01	2.90E-01	2.60E-01	9.10E-01	1.40E-01	2.70E-02
PDYN	Ssc.121.1.S1_at	1.60E-01	4.60E-02	1.90E-01	2.50E-01	8.00E-01	2.50E-01	3.10E-02
PENK	Ssc.11281.1.A1_at	4.30E-02	3.50E-01	5.60E-01	2.70E-01	7.40E-01	6.20E-02	9.40E-01
	Ssc.11281.2.S1_at	3.90E-01	8.60E-01	9.40E-01	3.90E-01	1.00E-01	1.90E-01	1.10E-01
PNOC	Ssc.15910.1.A1_at	6.30E-01	1.20E-01	8.70E-01	7.70E-01	9.70E-01	1.90E-01	3.50E-01
	Ssc.15910.1.S1_at	4.10E-01	2.10E-02	9.90E-01	3.60E-01	5.40E-01	1.80E-01	1.90E-01
PTHR	Ssc.9991.1.S1_at	3.80E-02	1.80E-01	2.10E-01	3.80E-01	1.90E-01	3.20E-01	4.80E-01
PTHY	Ssc.668.1.S1_at	2.20E-01	7.90E-03	7.80E-01	6.10E-01	5.90E-01	2.50E-02	4.90E-02
REL1	Ssc.162.1.S1_at	4.70E-01	9.70E-03	9.50E-01	2.70E-01	4.90E-01	1.00E-01	1.40E-01
SCG1	Ssc.15718.1.A1_at	2.10E-01	3.90E-01	2.80E-01	3.10E-01	3.70E-01	2.40E-02	9.50E-02
SCG2	Ssc.13645.1.A1_at	1.50E-01	3.50E-01	2.40E-01	2.70E-01	3.90E-01	8.00E-03	1.60E-01
SCG3	Ssc.6770.1.A1_at	1.70E-01	1.00E-01	8.80E-01	2.90E-01	4.50E-01	6.20E-03	1.10E-01
SECR	Ssc.710.1.S1_at	4.70E-01	3.70E-01	2.60E-01	6.40E-01	1.70E-01	4.20E-02	8.00E-01
SMS	Ssc.19520.1.A1_at	4.70E-01	8.80E-01	2.50E-02	2.10E-01	3.30E-01	4.90E-02	2.80E-03
TKN1	Ssc.18075.1.A1_at	2.90E-01	6.30E-03	9.00E-01	1.60E-01	6.60E-01	7.20E-02	2.70E-02
	Ssc.18075.2.S1_at	2.30E-01	9.10E-03	6.90E-01	5.50E-01	1.00E-01	2.50E-02	4.40E-02

Supplementary Table S2. (cont.)

Symbol	Affy Probe ID	Muscle						
		GSE18653	GSE19275	GSE8974	GSE14643	GSE15211	GSE21096	GSE16348-D
TKN4	Ssc.23153.1.S1_at	3.60E-01	4.80E-02	4.40E-01	9.40E-01	5.50E-01	5.70E-01	1.00E-01
TKNK	Ssc.19565.1.S1_at	1.60E-01	1.20E-02	7.50E-01	5.50E-01	2.10E-01	1.40E-02	9.70E-02
	Ssc.19565.2.A1_at	8.20E-01	3.50E-01	8.10E-01	2.30E-01	2.30E-01	4.50E-02	6.00E-02
UTS2	Ssc.437.1.S1_a_at	3.00E-01	1.10E-02	8.60E-01	3.60E-01	3.70E-01	3.50E-02	5.60E-02
VEGFC	Ssc.12790.1.A1_at	8.90E-01	1.90E-01	1.20E-01	9.50E-01	7.40E-03	4.30E-06	6.10E-02
VEGFD	Ssc.29289.1.A1_at	5.60E-01	5.30E-02	4.40E-01	4.60E-01	4.20E-01	2.90E-02	1.80E-01
Prohormone Convertase								
PCSK1	Ssc.141.1.S1_at	1.50E-01	1.30E-02	8.50E-01	2.20E-01	2.80E-01	7.90E-02	3.40E-02
PCSK2	Ssc.109.1.S1_at	2.10E-01	1.40E-01	8.40E-01	1.90E-01	2.90E-01	1.30E-02	8.80E-02
PCSK7	Ssc.5628.1.S1_at	1.60E-01	1.30E-01	1.90E-01	2.40E-01	1.00E-01	2.40E-02	3.30E-01
Total P-value in Column <0.005		0	1	1	0	0	7	1
Totals by Group								10

Symbol	Affy Probe ID	Experiments							
		Fat							
		GSE17309	GSE14373	GSE14739-B	GSE9333	GSE18359-A	GSE18359-L	GSE13528-F	GSE13528-L
ADM5	Ssc.26627.1.A1_at	4.60E-01	9.90E-01	7.00E-02	4.80E-01	4.00E-01	3.10E-01	2.00E-01	3.30E-01
ADML	Ssc.314.1.S1_at	6.00E-01	3.40E-01	7.60E-01	8.60E-01	2.00E-01	3.40E-01	1.60E-01	2.60E-01
ANF	Ssc.16245.1.S1_at	6.30E-01	6.20E-01	1.50E-01	7.10E-01	2.20E-01	6.60E-01	4.20E-01	4.50E-01
ANFB	Ssc.629.1.S1_at	3.10E-01	3.60E-01	2.30E-01	1.80E-01	3.60E-01	9.40E-01	6.40E-01	9.20E-01
ANFC	Ssc.23867.1.A1_at	5.90E-01	5.80E-01	7.70E-01	3.70E-01	6.50E-01	8.80E-01	8.20E-02	9.70E-01
AUGN	Ssc.22487.1.S1_at	6.40E-01	3.80E-01	3.40E-01	9.10E-01	5.20E-01	1.20E-01	2.70E-04	3.60E-01

Supplementary Table S2. (cont.)

Symbol	Affy Probe ID	Fat							
		GSE17309	GSE14373	GSE14739-B	GSE9333	GSE18359-A	GSE18359-L	GSE13528-F	GSE13528-L
CART	Ssc.15900.1.S1_at	4.60E-01	9.40E-01	1.60E-01	7.90E-01	7.60E-01	7.40E-01	1.70E-01	5.00E-01
CCKN	Ssc.717.1.S1_at	5.20E-01	2.50E-01	8.20E-01	3.90E-01	5.30E-01	6.90E-01	1.80E-01	1.20E-02
CMGA	Ssc.4653.1.S1_at	7.20E-01	7.40E-01	3.30E-01	3.30E-01	4.30E-01	7.80E-01	1.20E-03	3.30E-01
COLI	Ssc.14556.1.S1_at	6.20E-01	9.50E-02	8.70E-01	3.60E-01	4.00E-01	5.70E-01	4.00E-01	9.20E-01
CRSP1	Ssc.3741.1.S1_at	4.40E-01	1.10E-01	1.70E-01	4.40E-01	6.10E-01	5.50E-01	6.00E-01	3.30E-01
CRSP2	Ssc.18558.1.S1_at	6.00E-01	6.10E-01	2.00E-01	8.00E-01	6.30E-01	9.90E-01	3.10E-01	9.10E-01
CRSP3	Ssc.17879.1.S1_at	3.80E-01	9.00E-01	5.00E-02	5.10E-01	9.20E-01	9.90E-01	5.00E-01	8.60E-01
EDN1	Ssc.9364.1.S1_at	5.80E-01	9.60E-01	9.90E-01	7.90E-01	6.80E-01	3.40E-01	3.70E-03	2.90E-01
GALA	Ssc.713.1.S1_at	5.60E-01	6.70E-01	2.30E-01	7.60E-01	7.40E-01	9.80E-01	7.80E-01	9.50E-01
GALP	Ssc.4875.1.S1_at	3.30E-01	7.20E-01	9.00E-02	6.10E-01	4.30E-01	6.60E-01	8.40E-01	7.90E-01
GAST	Ssc.644.1.S1_at	7.40E-01	7.30E-01	6.60E-01	1.60E-01	5.10E-01	8.50E-01	2.30E-01	3.60E-01
GHRL	Ssc.440.1.S1_at	8.20E-01	5.30E-01	5.40E-01	6.20E-01	9.30E-01	9.90E-01	3.30E-01	3.20E-01
GLUC	Ssc.17225.1.S1_at	4.10E-01	4.60E-01	3.10E-01	4.90E-01	6.20E-01	8.60E-01	4.10E-01	9.10E-01
GON1	Ssc.16310.1.S1_at	6.00E-01	5.20E-01	1.30E-01	5.70E-01	9.70E-01	9.10E-01	5.70E-01	5.20E-01
HEPC	Ssc.376.1.S1_at	1.50E-01	7.80E-01	8.30E-02	1.00E+00	4.50E-01	8.90E-02	3.30E-01	3.30E-01
IAPP	Ssc.8324.1.A1_at	6.10E-01	6.80E-01	1.00E-01	5.30E-01	9.10E-01	9.10E-01	5.10E-01	7.60E-01
IGF1	Ssc.16231.1.S1_a_at	7.40E-02	5.30E-01	4.40E-01	6.50E-01	5.60E-03	7.10E-03	6.10E-03	2.70E-01
	Ssc.16231.2.A1_a_at	6.70E-01	9.10E-01	3.80E-01	9.70E-01	5.10E-01	8.30E-01	2.10E-01	9.80E-01
	Ssc.16231.3.S1_a_at	1.50E-01	7.50E-01	2.80E-01	4.90E-01	3.50E-02	7.50E-02	1.30E-02	2.90E-01
IGF2	Ssc.9365.1.S1_at	5.20E-01	8.20E-01	6.50E-01	7.80E-01	5.40E-01	7.80E-01	2.90E-01	8.70E-01
	Ssc.9365.2.S1_a_at	2.70E-01	1.80E-02	4.20E-03	6.20E-01	8.40E-01	8.40E-01	3.40E-02	1.80E-02
	Ssc.9365.3.S1_a_at	5.80E-01	5.00E-03	8.20E-02	5.90E-01	6.40E-01	9.00E-01	1.10E-01	3.80E-02

Supplementary Table S2. (cont.)

Symbol	Affy Probe ID	Fat							
		GSE17309	GSE14373	GSE14739-B	GSE9333	GSE18359-A	GSE18359-L	GSE13528-F	GSE13528-L
	Ssc.9365.3.S1_x_at	4.90E-01	9.30E-03	2.40E-02	5.70E-01	7.50E-01	7.80E-01	3.10E-02	2.60E-02
	Ssc.9365.4.S1_a_at	2.20E-01	8.90E-01	3.30E-01	7.30E-01	8.30E-01	9.10E-01	4.00E-01	2.30E-01
	Ssc.9365.5.A1_at	9.10E-01	9.70E-01	5.10E-02	4.70E-01	5.70E-01	8.10E-01	2.90E-01	9.40E-01
	Ssc.9365.5.S1_at	4.30E-01	3.40E-01	2.30E-01	4.50E-01	8.40E-01	6.20E-01	5.40E-01	5.30E-01
	Ssc.9365.5.S1_a_at	4.40E-01	3.70E-03	1.20E-01	6.50E-01	6.10E-01	9.30E-01	6.40E-02	3.70E-02
	Ssc.9365.6.A1_a_at	9.10E-01	9.20E-01	8.10E-01	6.00E-01	9.80E-01	7.60E-01	2.60E-01	9.30E-01
	Ssc.9365.6.A1_x_at	4.40E-01	4.60E-01	3.30E-01	2.80E-01	7.20E-01	9.90E-01	4.90E-01	6.80E-01
	Ssc.9365.6.S1_x_at	3.70E-01	2.00E-02	9.70E-02	9.20E-01	5.20E-01	7.80E-01	5.30E-02	5.30E-02
	Ssc.9365.7.A1_x_at	9.80E-01	9.30E-01	3.80E-01	4.90E-01	8.20E-01	8.40E-01	5.00E-01	9.90E-01
INS	Ssc.583.1.S1_at	8.80E-01	6.30E-01	3.30E-02	9.10E-01	5.10E-01	9.80E-01	2.40E-01	7.60E-01
INSL3	Ssc.11990.1.S1_at	7.50E-01	5.90E-01	8.00E-01	9.10E-01	5.30E-01	7.10E-01	6.80E-01	7.40E-01
MCH	Ssc.3287.1.S1_at	4.90E-01	7.50E-01	1.10E-01	9.60E-01	8.10E-01	9.60E-01	4.70E-01	7.50E-01
MOTI	Ssc.714.1.S1_at	4.80E-01	9.90E-01	5.60E-01	4.10E-01	6.40E-02	5.40E-01	2.30E-01	3.90E-01
NEU1	Ssc.15668.1.A1_at	9.10E-01	5.50E-01	8.50E-02	4.60E-01	6.80E-01	9.60E-01	3.10E-01	7.60E-01
NEU2	Ssc.4210.1.S1_at	7.30E-01	9.10E-01	1.50E-01	7.70E-01	3.20E-01	6.40E-01	4.60E-01	5.50E-01
NMB	Ssc.2083.1.A1_at	4.70E-01	1.90E-01	1.20E-01	3.00E-01	1.80E-01	9.50E-01	6.50E-02	9.00E-01
NMU	Ssc.12508.1.A1_at	8.40E-01	4.80E-01	7.20E-01	9.60E-01	5.80E-01	9.90E-01	7.60E-01	7.70E-01
NPW	Ssc.15796.1.S1_at	8.40E-01	9.80E-01	6.70E-01	8.00E-01	6.80E-01	9.00E-01	2.80E-01	5.20E-01
NPY	Ssc.15981.1.A1_at	6.30E-01	7.70E-01	2.30E-02	7.00E-01	5.00E-01	9.60E-01	2.10E-01	9.00E-01
	Ssc.15981.1.S1_at	5.00E-01	5.80E-01	8.40E-02	8.80E-01	7.30E-01	9.50E-01	5.10E-01	6.60E-01

Supplementary Table S2. (cont.)

Symbol	Affy Probe ID	Fat							
		GSE17309	GSE14373	GSE14739-B	GSE9333	GSE18359-A	GSE18359-L	GSE13528-F	GSE13528-L
OREX	Ssc.15983.1.S1_at	8.30E-01	6.90E-01	2.00E-02	6.50E-01	4.10E-01	1.90E-01	2.70E-01	8.00E-01
PACA	Ssc.27598.1.S1_at	6.00E-01	8.00E-01	5.60E-03	7.10E-01	4.00E-01	1.80E-01	4.40E-01	4.20E-01
PAHO	Ssc.456.1.S1_at	4.90E-01	8.40E-01	1.60E-01	6.20E-01	8.60E-01	9.60E-01	4.50E-01	6.00E-01
PCSK1	Ssc.17429.1.S1_at	4.60E-01	6.70E-01	6.20E-01	9.90E-01	7.50E-01	9.40E-01	2.50E-01	1.40E-01
PDGFA	Ssc.6173.3.S1_a_at	1.10E-01	3.50E-01	6.10E-02	4.30E-01	4.50E-01	2.40E-01	1.00E+00	4.40E-01
PDYN	Ssc.121.1.S1_at	5.00E-01	9.90E-01	5.10E-02	7.40E-01	4.60E-01	9.90E-01	6.20E-01	7.40E-01
PENK	Ssc.11281.1.A1_at	5.60E-01	7.20E-01	6.90E-01	5.80E-01	7.60E-01	1.00E+00	5.80E-05	9.00E-01
	Ssc.11281.2.S1_at	6.10E-01	5.20E-01	4.50E-01	4.90E-01	3.60E-01	9.40E-01	6.40E-02	9.70E-01
PNOC	Ssc.15910.1.A1_at	4.40E-01	1.80E-01	4.90E-01	4.90E-01	4.10E-01	2.10E-01	9.90E-02	2.40E-01
	Ssc.15910.1.S1_at	6.20E-01	7.40E-01	1.70E-01	5.70E-01	6.30E-01	8.60E-01	1.30E-01	7.80E-01
PTHR	Ssc.9991.1.S1_at	9.90E-01	6.30E-01	4.50E-01	8.70E-01	4.30E-01	9.80E-01	3.20E-01	7.80E-01
PTHY	Ssc.668.1.S1_at	6.20E-01	8.80E-01	1.20E-01	9.00E-01	7.60E-01	1.00E+00	5.90E-01	9.00E-01
REL1	Ssc.162.1.S1_at	6.20E-01	6.10E-02	3.20E-02	5.10E-01	4.90E-01	1.00E+00	9.30E-01	1.00E+00
SCG1	Ssc.15718.1.A1_at	7.10E-01	8.70E-01	6.90E-01	7.40E-01	3.80E-01	9.90E-01	4.90E-01	9.90E-01
SCG2	Ssc.13645.1.A1_at	3.60E-03	8.40E-01	6.20E-01	8.40E-01	9.10E-01	9.50E-01	6.00E-01	8.30E-01
SCG3	Ssc.6770.1.A1_at	6.40E-01	3.50E-01	4.50E-01	2.90E-01	5.70E-01	9.70E-01	6.20E-01	6.20E-01
SECR	Ssc.710.1.S1_at	6.70E-01	2.10E-02	2.80E-01	5.30E-01	2.10E-01	3.90E-02	4.30E-01	6.40E-02
SMS	Ssc.19520.1.A1_at	9.80E-01	1.70E-01	1.00E-01	7.60E-01	7.80E-01	9.80E-01	5.50E-01	4.40E-01
TKN1	Ssc.18075.1.A1_at	7.50E-01	6.50E-01	7.70E-01	6.00E-01	5.40E-02	9.30E-01	5.20E-01	8.20E-01
	Ssc.18075.2.S1_at	6.40E-01	9.20E-01	2.20E-01	7.90E-01	7.50E-01	9.70E-01	1.20E-01	9.40E-01
TKN4	Ssc.23153.1.S1_at	9.40E-01	7.60E-01	5.00E-01	6.10E-01	9.00E-01	8.80E-01	4.10E-01	1.90E-01
TKNK	Ssc.19565.1.S1_at	5.50E-01	6.30E-01	4.20E-01	7.30E-01	5.20E-01	9.90E-01	6.60E-01	9.70E-01

Supplementary Table S2. (cont.)

Symbol	Affy Probe ID	Fat							
		GSE17309	GSE14373	GSE14739-B	GSE9333	GSE18359-A	GSE18359-L	GSE13528-F	GSE13528-L
	Ssc.19565.2.A1_at	6.30E-01	1.00E+00	6.30E-01	5.30E-01	2.30E-01	8.30E-01	3.10E-01	3.30E-01
UTS2	Ssc.437.1.S1_a_at	8.20E-01	9.10E-01	4.60E-02	4.60E-01	8.70E-01	9.40E-01	8.30E-01	9.40E-01
VEGFC	Ssc.12790.1.A1_at	6.70E-01	4.40E-01	5.70E-01	8.20E-01	6.20E-03	1.50E-01	5.30E-02	3.80E-01
VEGFD	Ssc.29289.1.A1_at	6.10E-01	8.70E-01	6.50E-01	7.00E-01	1.10E-02	5.70E-01	2.10E-01	4.40E-01
Prohormone Convertase									
PCSK1	Ssc.141.1.S1_at	3.10E-01	7.80E-01	4.70E-01	2.80E-01	4.10E-01	1.00E+00	1.20E-01	8.40E-01
PCSK2	Ssc.109.1.S1_at	3.10E-01	6.50E-01	2.50E-01	6.30E-01	4.30E-01	3.30E-01	5.10E-01	8.40E-01
PCSK7	Ssc.5628.1.S1_at	4.30E-01	1.00E-01	5.90E-01	4.90E-01	9.20E-01	4.40E-01	5.70E-01	5.40E-01
Total P-value in Column <0.005		1	1	1	0	0	0	4	0
Totals by Group									7

Symbol	Affy Probe ID	Experiments	
		Gut	
		GSE13457	GSE15256
ADM5	Ssc.26627.1.A1_at	3.90E-01	4.90E-02
ADML	Ssc.314.1.S1_at	3.00E-01	4.40E-04
ANF	Ssc.16245.1.S1_at	7.80E-02	5.50E-01
ANFB	Ssc.629.1.S1_at	6.00E-01	5.50E-01
ANFC	Ssc.23867.1.A1_at	5.90E-01	1.20E-01
AUGN	Ssc.22487.1.S1_at	7.90E-01	3.40E-01
CART	Ssc.15900.1.S1_at	5.90E-01	2.60E-01

Supplementary Table S2. (cont.)

Symbol	Affy Probe ID	Gut	
		GSE13457	GSE15256
CCKN	Ssc.717.1.S1_at	2.30E-01	1.10E-01
CMGA	Ssc.4653.1.S1_at	9.80E-01	7.60E-04
COLI	Ssc.14556.1.S1_at	1.00E-01	2.00E-01
CRSP1	Ssc.3741.1.S1_at	6.80E-01	4.40E-01
CRSP2	Ssc.18558.1.S1_at	3.30E-02	7.10E-01
CRSP3	Ssc.17879.1.S1_at	1.80E-01	4.50E-01
EDN1	Ssc.9364.1.S1_at	2.80E-01	2.40E-01
GALA	Ssc.713.1.S1_at	5.70E-01	9.40E-04
GALP	Ssc.4875.1.S1_at	1.70E-01	1.50E-01
GAST	Ssc.644.1.S1_at	5.30E-01	6.80E-01
GHRL	Ssc.440.1.S1_at	3.30E-01	4.70E-02
GLUC	Ssc.17225.1.S1_at	2.80E-01	3.50E-03
GON1	Ssc.16310.1.S1_at	1.20E-01	4.00E-01
HEPC	Ssc.376.1.S1_at	3.40E-01	2.90E-01
IAPP	Ssc.8324.1.A1_at	6.60E-01	2.20E-01
IGF1	Ssc.16231.1.S1_a_at	4.30E-01	3.50E-01
	Ssc.16231.2.A1_a_at	5.80E-01	7.00E-01
	Ssc.16231.3.S1_a_at	1.10E-01	6.30E-02
IGF2	Ssc.9365.1.S1_at	9.00E-01	1.70E-01
	Ssc.9365.2.S1_a_at	1.80E-01	2.50E-02
	Ssc.9365.3.S1_a_at	4.80E-01	2.80E-02
	Ssc.9365.3.S1_x_at	4.70E-02	6.50E-03
	Ssc.9365.4.S1_a_at	8.70E-02	5.00E-01
	Ssc.9365.5.A1_at	4.70E-02	4.40E-01
	Ssc.9365.5.S1_at	8.60E-02	3.60E-01
	Ssc.9365.5.S1_a_at	9.50E-02	2.00E-02

Supplementary Table S2. (cont.)

Symbol	Affy Probe ID	Gut	
		GSE13457	GSE15256
	Ssc.9365.6.A1_a_at	5.50E-02	7.10E-01
	Ssc.9365.6.A1_x_at	5.40E-01	2.80E-01
	Ssc.9365.6.S1_x_at	1.40E-01	5.30E-03
	Ssc.9365.7.A1_x_at	9.20E-02	1.90E-01
INS	Ssc.583.1.S1_at	3.80E-02	1.00E-01
INSL3	Ssc.11990.1.S1_at	8.10E-02	7.20E-01
MCH	Ssc.3287.1.S1_at	6.10E-01	2.00E-01
MOTI	Ssc.714.1.S1_at	6.60E-01	5.20E-01
NEU1	Ssc.15668.1.A1_at	4.30E-02	2.40E-01
NEU2	Ssc.4210.1.S1_at	4.20E-02	7.20E-01
NMB	Ssc.2083.1.A1_at	1.20E-01	3.30E-01
NMU	Ssc.12508.1.A1_at	7.40E-01	8.70E-01
NPW	Ssc.15796.1.S1_at	5.60E-02	4.40E-01
NPY	Ssc.15981.1.A1_at	6.20E-02	3.00E-01
	Ssc.15981.1.S1_at	7.20E-02	2.20E-01
OREX	Ssc.15983.1.S1_at	5.70E-01	3.60E-01
PACA	Ssc.27598.1.S1_at	9.30E-02	1.00E-01
PAHO	Ssc.456.1.S1_at	1.70E-01	3.70E-01
PCSK1	Ssc.17429.1.S1_at	8.10E-01	1.10E-02
PDGFA	Ssc.6173.3.S1_a_at	6.90E-01	7.80E-07
PDYN	Ssc.121.1.S1_at	3.40E-01	1.40E-01
PENK	Ssc.11281.1.A1_at	2.80E-01	1.80E-08
	Ssc.11281.2.S1_at	3.40E-02	3.00E-01
PNOC	Ssc.15910.1.A1_at	9.00E-01	5.00E-01
	Ssc.15910.1.S1_at	8.30E-01	1.60E-01
PTHR	Ssc.9991.1.S1_at	3.60E-01	1.60E-01

Supplementary Table S2. (cont.)

Symbol	Affy Probe ID	Gut	
		GSE13457	GSE15256
PTHY	Ssc.668.1.S1_at	9.40E-01	1.90E-01
REL1	Ssc.162.1.S1_at	2.00E-01	8.50E-01
SCG1	Ssc.15718.1.A1_at	9.00E-01	4.60E-09
SCG2	Ssc.13645.1.A1_at	7.90E-01	6.60E-06
SCG3	Ssc.6770.1.A1_at	4.10E-01	1.20E-01
SECR	Ssc.710.1.S1_at	4.80E-01	1.20E-01
SMS	Ssc.19520.1.A1_at	7.30E-01	2.90E-01
TKN1	Ssc.18075.1.A1_at	5.40E-01	5.20E-01
	Ssc.18075.2.S1_at	5.20E-02	5.10E-01
TKN4	Ssc.23153.1.S1_at	7.70E-01	7.00E-02
TKNK	Ssc.19565.1.S1_at	8.70E-01	6.20E-01
	Ssc.19565.2.A1_at	6.80E-01	1.80E-01
UTS2	Ssc.437.1.S1_a_at	1.40E-01	4.60E-01
VEGFC	Ssc.12790.1.A1_at	5.10E-01	3.70E-03
VEGFD	Ssc.29289.1.A1_at	6.40E-02	6.70E-01
Prohormone Convertase			
PCSK1	Ssc.141.1.S1_at	3.40E-01	6.50E-07
PCSK2	Ssc.109.1.S1_at	7.40E-01	5.40E-01
PCSK7	Ssc.5628.1.S1_at	8.40E-01	1.20E-03
Total P-value in Column <0.005		0	11
Totals by Group			11
Total for all Groups	116		

[1]Experiment Class

[2]Experiment Identification