# SINGLE NUCLEOTIDE POLYMORPHISM ANALYSIS OF GENES EXPRESSED IN PORCINE ADIPOSE TISSUE

By

Y.A.M.S.WICKRAMASINGHE

Bachelor of Veterinary Medicine

University of Peradaniya

Sri Lanka

2003

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degrees of
MASTER OF SCIENCE
July, 2006

SINGLE NUCLEOTIDE POLYMORPHISM

ANALYSIS OF GENES EXPRESSED

IN PORCINE ADIPOSE

TISSUE

Thesis Approved

Dr. Udaya DeSilva
ThesisAdviser

Dr. Richard Essenberg

Dr. David Buchanan

Dr. Gordon Emslie
Dean of the Graduate College

# Acknowledgements

I have learned valuable lessons on genetics and life during the time spent in Oklahoma state university. It would have been an impossible task to achieve with out the guidance, support and inspiration given to me by a wonderful group of people.

First of all I want to thank my advisor, Dr. Udaya DeSilva who gave me the opportunity to work on this project. Dr. Desilva, you are one of the nicest human being that I met in my life. I am truly blessed to have an advisor like you. Thank you very much for all the support, understanding and encouragement given to me. I would also like to thank my master's program committee members Dr. David Buchanan and Dr. Richard Esseneberg for the advices, support and guidance they have given. A very special 'Thank You' goes to Dr. Raluca Mateescu for representing Dr. Buchanan during the thesis defense.

I would like to thank Xiaoping Guo who taught me a, b, c of molecular biology techniques. You were a great teacher and I will never forget the basics I learned from you. My gratitude goes to my two fabulous lab mates, Asitha and Charaka for all the support, cooperation and suggestions given to me through the years.

Also I want to thank our collaborators in this project Dr. Catherine W. Ernst and Nancy Raney from Michigan State University who were involved in collection and transport of tissue samples.

cDNA library was sequenced at University of Oklahoma's advanced center for genome technology and my gratitude goes to Dr Bruce. A. Roe and his staff, and especially to Dr Fraes Najar and Steve Kenton who helped with Polyphred program and setting up of the Linux server. My gratitude goes to Dr. Patricia Ayoubi who spared time amidst her busy schedule to help me getting GO annotations. I want to acknowledge Sashank Bhide for his help given in setting up the Linux program, Polyphred and consed.

I want to share this moment with my best friend in life, my fiancé Eranda Rajapaksha who taught me how to stand on my own feet. Also there is my little sister back in Sri Lanka who encouraged me with her mischievous remarks and bright smile.

And finally I would like to thank the two most important people in my life, my father and mother who held me up taking the burden in their shoulder. You are the inspiration in my life!

Table of Contents

# List of Tables

Table Page

# List of Figures

Figure Page

# CHAPTER 1

Literature Review

## Introduction

The pork industry supplies about thirty eight percent of the daily meat protein intake in the world and plays a major role in the socio-economic structure in many developed and developing countries (www.nppc.org/, 2006).

The pork industry has diversified to utilize multiple breeds of pigs that have specific attributes depending on the customer preference. Animals are being manipulated both nutritionally and genetically to produce the ideal pork according to consumer demand. Conventional selective breeding programs based on phenotypic characteristics have improved the lean growth rate but decreased the quality of meat (Meisinger, 1999). For example, 2003 National Pork Quality Audit revealed that there had been a 5.3% increase in the frequency of Pale, Soft and Exudative (PSE) pork during the period of 1996-2003 (Bates et al., 2003). Enhancement of production efficiency and quality are essential factors in the pork industry and this can be accomplished by incorporating the knowledge on genes responsible for meat quality traits and their association with lean growth in the selective breeding programs. As it is difficult and expensive to collect meat quality data in field conditions, resource populations have to be developed for data collection and for identification of markers and genes associated with meat quality traits (Williams, 1995). This literature review starts with a brief introduction of the United States pork industry and pork quality. Then, the genetic control of pork quality and the importance of identification and mapping of Quantitative Trait Loci (QTL) will be discussed. Finally, the importance of genetic markers in QTL mapping and marker assisted selection will be described with special emphasis on Single Nucleotide Polymorphism markers.

## Pork Industry

The domestic pig belongs to the family Suidae which is one of the three surviving families of sub order Suiformes (Ruvinsky and Rothschild, 1998). Fossils found from swamps and forests in Europe and Asia date back the history of the pig to 40 million years (www.porkboard.org/docs/pkfacts2002.pdf, 2003). It is thought that pigs were first domesticated in China around 4900 B.C and then in Europe by 1500 B.C (www.porkboard.org/docs/pkfacts2002.pdf, 2003). Today, pork is the most popular meat in the world and pigs are found in majority of countries around the world except in places where religious restriction prohibit consumption of pork (www.nppc.org/about/pork_today.html, 2004).

Hernando de Soto first introduced domesticated pigs to United States on his voyage to Tampa Bay, Florida in 1539 (www.porkboard.org/docs/pkfacts2002.pdf, 2003). Pork industry developed rapidly in United States and by 17$^{th}$ century a typical farmer owned around four or five pigs. With the growth of pig herds, more processing facilities and transport methods were needed. Shortly after the civil war, live pigs were transported via rail roads to consumers nationwide (www.porkboard.org/porkfacts/default.html, 2004). In the beginning, mid west states producing surplus of feed grain became major pork producers and due to that, 'Corn Belt' was also known as 'Hog Belt' (www.porkboard.org/porkfacts/default.html, 2004).

Due to advances in genetics, better pig raising methods and proper control of diseases, states like North Carolina that were not in the 'Corn Belt ' were able to increase the pork production dramatically (www.porkboard.org/porkfacts/default.html, 2004). Today, North Carolina is the second largest pork producing state after Iowa

(www.porkboard.org/porkfacts/default.html, 2004). United States meat certification program conducted from 1953-1973 was largely responsible for converting US swine industry from producing 'lard-type' pigs to 'meat-type' (Jones, 1998) and today, United States is one of the leading pork producing countries in the world, with production accounting for 10.2% of the total world supply. Currently, USA is the second largest pork exporter, with Canada in the number one place ( www.nppc.org/, 2006).

United States' pure bred seed stocks consist of nine major swine breeds, which are: Yorkshire, Duroc, Hampshire, Landrace, Berkshire, Spotted, Chester White, Poland China and Pietrain. Market pigs with increased heterosis are produced by cross breeding of these pure bred breeds (www.porkboard.org/porkfacts/default.html, 2004). Two major breeding systems practiced in United States are the rotational breeding system which involves successive use of boars of different breeds with the retention of the gilts with superior characters and the terminal breeding system where boar lines selected specially for carcass and production traits are crossed with gilt lines selected for reproductive traits (www.porkboard.org/porkfacts/default.html, 2004). Consumer preference for lean meat and the economic importance of reproductive efficiency has shifted the pig breeding systems largely toward the terminal breeding system and today, the majority of commercial pigs in USA are produced by this method (www.porkboard.org/porkfacts/default.html, 2004). Most modern pork production systems have a 'closed herd concept', where the breeding females are produced in-house and semen from superior boars are brought from outside (www.porkboard.org/porkfacts/default.html, 2004).

## Pork Quality

Meat quality is a variable trait driven and defined by consumers (Russell, 1997). Due to the competitive nature of the market, quality of pork has become a main issue of concern for the pork industry. According to the variation in quality, pork is divided into three groups as Reddish Firm Non-exudative (RFN), Pale Soft Exudative (PSE) and Dark Firm Dry pork (DFD). High quality pork with more consumer demand belongs to RFN group (Meisinger, 1999). Four primary attributes of pork quality identified by National Pork Producers Council (NPPC) are color, water holding capacity or drip loss, intramuscular fat and palatability (Meisinger, 1999).

Color of pork is determined by the hematin containing globular protein myoglobin. Status of globin and iron will determine the color of meat. The bright red color in fresh pork is produced by oxygenated form of myoglobin and oxidation of myoglobin gives the brown color to pork (van Heugten, 2000).

Intramuscular fat or marbling is responsible for sensory qualities such as juiciness, flavor and tenderness of pork and it is subjectively measured by a standard chart ranging from 1 to 5 (Meisinger, 1999). Consumers in United States prefer a marbling score of 2, while Japanese consumers, who are the leading US pork importers, prefer pork with higher marbling score of 3 to 4 (Meisinger, 1999). Reduced intramuscular fat percentage leads to inferior consistency of meat which is insipid, strawy and dry (Affentranger et al., 1996). Low quality of subcutaneous fat is a cause for fat dependent perishability of pork (Affentranger et al., 1996) with higher water content in subcutaneous fat leading to insufficient processing firmness after cooling. High percentage of unsaturated fatty acids

in the subcutaneous fat has a negative effect on the taste and processing of pork (Affentranger et al., 1996).

A study conducted in France in 1999 showed that even though the consumer demand for pork increased by up to 2.5-3.5% with high intramuscular fat, this positive effect decreased when there was high intermuscular fat (Fernandez et al., 1999). A major objective of modern pig breeding programs is decreasing of intermuscular fat or back fat content without reducing the intramuscular fat depots which are positively correlated with meat quality.

Genetic regulation of intramuscular fat deposition and its association with pork quality was first reported by Duniec (1961) and later by Jonsson, in a Spanish swine population (Jonsson, 1963). Affentranger et al., (1996) tested meat and fat quality properties of Duroc, Pietrain and Large White breeds under different feeding regimes and found out that both genotypes and feeding regimes have an effect on lean meat content and fat quality traits. The authors concluded that genetics is a major player in determining the quality of meat while the feeding regime affects more on intramuscular fat content and the fat quality (Affentranger et al., 1996).

Meat quality traits as a whole are moderately heritable, with heritability ($h^2$) values falling in the range of 0.1- 0.3. However marbling traits seem to have a higher heritability, with $h^2$ value around 0.50. A strong genetic relationship is observed between live measurements of back fat thickness and carcass lean percentage (Sellier, 1998). This has been largely exploited in the pork breeding industry. Compared to pigs from 1950s that had 2.86 inches of back fat, present day pigs have less than 0.75 inches of back fat, a reduction of 75% (www.porkboard.org/porkfacts/default.html, 2004).

Both environment and genetics play a complex role in defining the phenotype of an animal (Archibald and Haley, 1998). In livestock animals, humans have attempted using both of these to manipulate the phenotypes according to consumer demand and profit. Like many other livestock industries, pork industry utilizes genetic information in breeding programs.

Two genes known to have an effect on pork quality are halothane (HAL) or *ryr1* and Napole (RN) genes. Malignant Hypothermia (MH) is a condition seen in swine that is homozygous for defective gene *ryr1* leading to malfunction of ryanodine receptor that results in pale, soft, exudative pork (Fujii et al., 1991). Fujii et al. in 1991 identified the cause for MH as a single point mutation of C to T in the nucleotide 1843 of *ryr1* or halothane gene. Defective halothane gene is in the process of being eliminated from breeding stocks by PCR based DNA screening for homozygous and carrier animals. Napole gene was first reported by Le Roy et al. (1990). Dominant RN gene leads to reduced meat yield and water holding capacity with increased drip loss and cooking loss (Le Roy et al., 1990). In $RN^{(-)}$ individuals there is increased intramuscular glycogen content that has a negative affect on muscle pH (Sellier, 1998). This locus has been mapped to porcine chromosome 15 (Mariani et al., 1996; Sellier, 1998) but the responsible candidate gene is yet to be identified.

Even though the identification of Halothane and Napole genes had a positive effect on pork industry, increased emphasis given to lean and faster growing animals in selection of seed stock industry has led to reduction in meat quality (Le Roy et al., 1990; Cagnazzo et al., 2006).

## Quantitative Trait Loci (QTL)

Unlike in humans, identification of monogenic trait loci in farm animals is of lesser importance as such traits could be easily eliminated by selective breeding (Andersson, 2001).

Most commercially important traits such as milk yield, meat quality, and reproductive performance in farm animals show a continuous distribution of phenotypic values rather than a discrete qualitative value. These traits are controlled by more than one gene and the environment, and such traits are known as quantitative traits (Andersson, 2001). One of the main goals in farm animal genetics is identification and mapping of QTL, or region of the genome that harbors one or more genes affecting a quantitative trait (Geldermann, 1975).

Two main strategies of finding QTL are association tests using candidate genes and genome scan using anonymous DNA markers (Andersson, 2001). In association based studies, the frequency of an allele at polymorphic loci in affected and control populations is studied. If the candidate gene is a true causative gene for the trait of interest, it would be possible to detect even loci with small effects, but as there are many candidate genes for a particular trait of interest this method is more time consuming than a genome scan. For a genome scan to successfully find the map location of a trait locus, an accurate genetic model has to be postulated in a reasonable sample size with markers covering the whole genome (Andersson, 2001). In addition to genetic markers, a resource population that has been measured for the traits of interest and in which the traits of interest is segregating is needed for QTL analysis (Williams, 2005).

## QTL Mapping

Experimental cross breeding of widely divergent lines, breeds or animals has been shown to increase the genetic and phenotypic variation available for heritability studies. It is also important to identify pleiotropic interactions between traits to determine whether the selection would have a negative or positive effect on other important traits. One example of such a resource population is the Holestin and Charolais herds at Roslin Institute where QTL studies are conducted for both beef and dairy related traits (de Koning et al., 1999). Such resource populations were also developed in pigs, by crossing wild boars to domestic breeds and also by crossing diverse breeds such as Chinese Meishan to European breeds such as Large White and Land Race. Several divergent swine cross resource populations are developed in US such as, Pietrain x Duroc at Michigan State University, Yorkshire x Large White at Ohio State University , Meishan x Duroc, Meishan x Hampshire, Meishan x Landrace, Minzhu x Landrace, and Minzhu x Hampshire at Iowa State University (Paszek et al., 2006; Young, 1995). In pig breeding Meishan breed, known to have high fecundity, is used in pig breeding to introduce genes for litter size in Western pig breeds. In order to investigate the potential of introducing the prolific Meishan gene to Europe, a Dutch pig breeding company developed a resource population by cross breeding Meishan and European "White" pig lines (Janss et al., 1997). They analyzed this population for genes responsible for growth, back fat and litter size by segregation analysis and they were able to identify a dominant gene affecting litter size and back fat thickness in $F_1$ and $F_2$ generations.

Unexpected outcomes were observed in some of these crosses with diverse phenotypes. One such example is that in the Meishan x Land Race $F_2$ population where a QTL on

chromosome 7 influencing the back fat thickness originated from the Land Race population, (de Koning et al., 1999). This has been previously reported by Rohrer and Keele (1998) who found that most of the Meishan alleles produced fatter pigs, except the QTL on chromosomes 7 and 10, which showed a positive effect for carcass lean. Currently, there are several programs, such as the USDA-MARC project (Rohrer et al., 1994), European PiGMaP consortium, (Archibald et al., 1995) the Nordic map consortium (Marklund et al., 1996) and NIAI program in Japan (Mikawa et al., 1999) that are involved in development of porcine linkage maps for QTL studies.

A study by Andersson et al. in 1994 crossing European wild boars with domesticated Large White pigs identified loci clustered in chromosome 4 that are responsible for growth and fatness (Andersson, 1994). A paternally expressed QTL that has a major effect on lean meat content in ham was identified in chromosome 2 in a pedigree created by intercrossing European wild boar and the Large White domestic pigs (Jeon et al., 1999). This is one of the first identified QTL for muscle growth, fat deposition and heart growth of the pig and furthermore this locus was associated with IGF2 gene. This QTL was confirmed later byNezer et al. , (1999) by running a whole genome wide scan on a Large White x Pietrain resource population. Subsequently segregation analysis of the IGF2 locus confirmed that the QTL is due to a nucleotide substitution in an evolutionarily conserved CpG island in intron 3 of IGF2 gene (Van Laere et al., 2003). The authors also found that pigs inheriting this mutation paternally have 3 fold higher IGF2 mRNA expression in postnatal muscle. This is one example of phenotypic variation controlled by a regulatory mutation.

An experimental cross between Meishan and Dutch Large White lines revealed a QTL with a major effect on back fat thickness on chromosomes 7 and 2 (de Koning et al., 1999). Additionally, QTL for intra muscular fat were isolated on chromosomes 2, 4, 6 and 7. QTL for fat deposition are valuable not only for improving the pork quality but identification of these QTL in pig could have an implication on understanding the genetics of human obesity that has becomes a major concern in developed counties. After localization of a QTL for a particular trait, to a region ofa chromosome, fine mapping of the QTL is achieved by using additional markers and individuals (Williams, 2005). Afterwards, the causative gene within the QTL region could be identified by a candidate gene or positional cloning approach (Williams, 2005).

## Genetic Markers

Genetic markers are defined as "an observable genetically controlled variation that follows a Mendelian pattern of inheritance" (Williams, 2005). One of the earliest forms of DNA marker is the Restriction Fragment Length Polymorphism (RFLP). Most of the RFLP markers are diallelic and correspond to presence or absence of a restriction enzyme digestion site (Archibald and Haley, 1998). Donis-Keller et al. in 1987 constructed the first large scale RFLP linkage map. One important example of RFLP usage in livestock is the screening of carriers of bovine leukocyte adhesion deficiency (Williams, 2005). Variable Number Tandem Repeats (VNTR) that consist of minisatellite repeated sequences of 20-50bp length, often repeated in many loci and occurring at 10 to 100 different sites in the genome, are used to identify relationships in wildlife populations, verification of pedigrees and in genetic mapping (Williams, 2005). Even though they are

hyper polymorphic, main disadvantages of VNTR are their proximity to the telomeric region of the chromosome and the difficulty in PCR amplification due to large repeat size and high repeat number (Moran, 1998).

Microsatellite markers or Simple Tandem Repeats (STR) contain 5-20 copies of a short sequence of 2-4bp in tandem repeats, with the number of repeat units varying between individuals. Microsatallite markers replaced RFLP because the relatively large number of alleles in single microsatellite locus reduced the number of reference families needed for the mapping processes. Compared to RFLP method, microsatellite loci are easy to genotype by simple PCR amplification followed by separation of the alleles according to the size in a polyacrylamide gel (Vignal, 2002). Most of the human genetic maps are based on microsatellite markers. Microsatellite based maps are developed for livestock species with the major ones being cow, pig, chicken, sheep, goat and horse (Vignal, 2002).

Two other types of genetic markers of interest are RAPDs (Random Amplification of Polymorphic DNA) and AFLPs (Amplified Fragment Length Polymorphism). Both RAPDs and AFLPs are bi- allelic dominant markers and have an advantage in reducing the effort, time and consumables in large scale genotyping because once the technique is optimized the same reagents and conditions can be used for different species. The disadvantage of these markers is their dominancy and random generation (Vignal, 2002). Another problem in RAPDs is the low reproducibility during PCR conditions (Vignal, 2002). AFLP markers can rapidly generate more markers that can be replicated, and are still used for QTL mapping and diversity studies of species such as rabbit that are devoid of dense marker maps (Van Haeringen et al., 2001).

## Single Nucleotide Polymorphisms (SNPs)

Most recently emerged marker to the field of molecular mapping is Single Nucleotide Polymorphisms (SNPs), which are single base changes in a DNA sequence. Botstein et al. in 1980 used SNPs indicating the presence or absence of restriction enzyme cutting sites in construction of genetic linkage maps of human. RFLPs were replaced by Simple Tandem Repeats in 1990s and SNPs emerge back as favorites by the beginning of 21$^{st}$ century.

Difference between a point mutation and a SNP is that, for a base change to be categorized as a polymorphism, the least frequent allele in SNP should have a frequency of 1% or greater in the selected population (Vignal, 2002). Even though four possible nucleotide bases can be present at one position SNPs are usually bi- allelic and the reason for this is the low probability of two independent changes occurring at a single position (Vignal, 2002).

SNPs are categorized into two types; transition, when there is a purine-purine or pyrimidine-pyrimidine exchange, and transversion, when the exchange is purine-pyrimidine or pyrimidine-purine (Freese, 1959). Even though transition to transversion ratio had to be 0.5, in human, mouse and chicken, SNPs show higher number of transitions compared to transversions (Collins and Jukes, 1994; Picoult- Newberg et al.,1999; Smith et al., 2001). Reasons for this high percentage of transition is thought to be the spontaneous deamination of 5-methyl cytosine, the post synthetic modified product of cytosine to thymine and the inefficient repair mechanism for this mutation (Holliday and Grigg, 1993).

SNPs can be classified into two categories according to the location in the genome as coding and non-coding SNPs (Chakravarti, 2001). SNPs within the coding region can be silent polymorphisms that do not have an effect on the resulting amino acid. These types of SNPs are called as synonymous. Non-synonymous SNPs are functional polymorphisms that cause an amino acid change in the protein (Williams, 2005). Change in an amino acid due to the SNP might alter the conformation of the protein and could cause changes such as change in an enzymatic activity. Synonymous SNPs can also create splice sites and cause changes in the expressed gene (Richard and Beckmann, 1995). Non-coding SNPs can also be in the regulating regions, such as in promoters, enhancers or silencers of a gene, and can influence the phenotype (Loots et al., 2000). The popularity of SNPs is largely due to their abundance and relatively even distribution in the genome (Chen and Sullivan, 2003). It is estimated that there are 3.2 billion nucleotides in the human genome and the number of SNPs range between 1.6-3.2 million. It is also estimated that, by comparing DNA sequences from two individuals, a SNP can be found in every 1000-2000bp (Li and Sadler, 1991). Due to this abundance in nature, SNPs can be used as markers to create high density genetic maps for complex traits such as disease resistance in humans and production traits in livestock. Additionally, when compared to STRs, low mutation rates of SNPs are more favorable in population studies. Unlike STRs, many SNPs can have a functional significance if they are located in the coding or the regulatory regions of a gene and could be invaluable in association studies (Gray et al., 2000). SNPs are compatible with high throughput genotyping techniques (Sobrino et al., 2005). Because SNPs are usually bi-allelic, the detection process is easy and less costly when compared to tri and tetra nucleotide markers (Ranade et al., 2001).

The major limitation of SNPs as markers is that, when compared with microsatellites, more SNPs are needed to have a good discrimination of loci. It is estimated that four times more SNPs are required to have a similar discrimination power as STR (Brenner, 1999), but the abundance of SNPs, and efficiency in genotyping techniques mask this disadvantage.

SNPs are emerging in many aspects of human medicine from identification of disease susceptibility, drug tolerance, and mechanism of aging to forensics (Bessenyei et al., 2004). A SNP map of the human genome with 1.42 million SNPs distributed in 23 pairs of chromosomes are created by The International SNP Map working group in 2001 (The International SNP Map Working Group, 2001).

SNP based association studies can be carried out in two ways. One is the direct testing of a SNP for association with a trait of importance. The other method is the usage of SNPs as markers in linkage disequilibrium studies by comparing the frequency of SNPs in unrelated cases and controls (Gray et al., 2000; Suh and Vijg, 2005). When a SNP is located in a gene, it might be directly responsible for a phenotypic change by coding for a genetic variant that changes the function or expression of a protein (Suh and Vijg, 2005). Rather than the whole genome approach, more statistical significance can be obtained when a SNPs located in genes are used as a marker in identification of candidate genes (Suh and Vijg, 2005).

In livestock industry, where AFLP and STR were the predominant method for animal identification and parentage testing the SNPs are emerging as a low cost genotyping technique for paternity testing with the development of high-throughput genotyping

platforms like Illumina Bead array and GenomeLab SNP genotyping methods (Glowatzki-Mullis,1995; Heaton et al., 2002 ).

## Marker Assisted Selection

Information gained from Marker Assisted Selection (MAS) increase the accuracy of selection and genetic gain in selective breeding process.

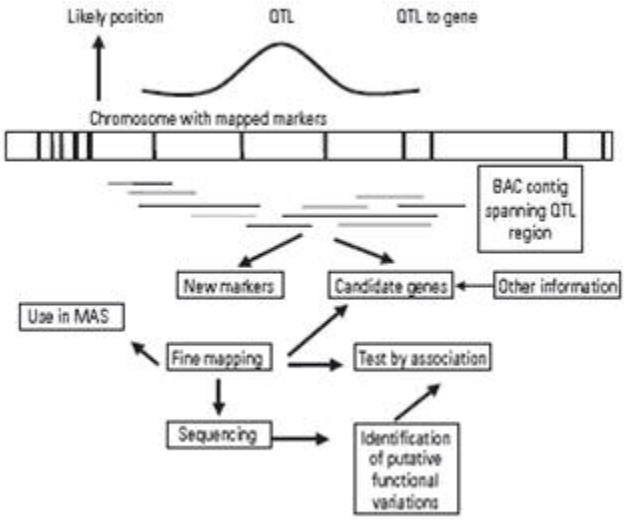
Marker Assisted Selection is a way to transfer genes between breeds and introgression of desirable genes between populations. One such example is identification of disease resistance QTL and incorporation of those loci into commercial populations by MAS cross breeding studies between indigenous and commercial breeds.

There are two major types of markers considered in MAS process. First type is the linked marker, located close to the trait gene and inherited with the trait gene. Second type is the direct marker, a functional polymorphism in the trait gene. Direct markers are more useful in predicting a phenotypic variation in a population because they have 100% association with expressed genes (Williams, 2005).

During the MAS process, if more markers dispersed across the genome are used to characterize the animal for breeding rather than looking only at genes that control a particular trait it would be possible to maintain a wide genetic diversity in resulting progeny (Williams, 2005).

Identical by descent mapping is a method used for high resolution mapping of loci and in this method, DNA samples collected from individuals that inherited a certain allele at a trait locus from a common ancestor are screened with genetic markers (Andersson, 2001).

Even though a marker to a QTL for a particular trait can be used in MAS, it is important to identify the underlying genes that control the variation. Illustration by Williams, (1995) given below, explains the procedure of identifying a trait gene for MAS.




First, the QTL region needs to be localized to a chromosomal region. Then the QTL region is fine mapped using additional individuals, markers or identical by descent mapping. If there is an absence of previously identified candidate genes, the QTL region is cloned and sequenced to gain information on genes and the present variation. Then the sequenced regions are compared between individuals with different phenotypes for identification of putative functional variations. Identified regions with functional variations are subjected to association tests for confirmation of any trait specific gene before applying it in MAS.

## Reference List

Affentranger,P., C.Gerwig, G.J.F.Seewer, D.Schworer, and N.Kunzi. 1996. Growth and carcass characteristics as well as meat and fat quality of three types of pigs under different feeding regimens. *Livestock Production Science* **45**: 187-196.

Andersson,L., C.S.Haley, H.Ellegren, S.A.Knott, M.Johansson, K.Anderson, Andersson-Eklund L, Edfors-Lilja I, Fredholm M and Hansson I. 1994. Genetic mapping of quantitative trait loci for growth and fatness in pigs. *Science* **263**: 1771-1774.

Andersson,L. 2001.Genetic dissection of phenotypic diversity in farm animals. *Nat Rev Genet* **2**: 130-138.

Archibald,A.L. and C.S.Haley. 1998. Genetic Linkage Maps. in *The Genetics of the pig* (ed. M F Rothschild and A Ruvsinksy), pp. 265-294. CAB International, New York.

Archibald,A.L., C.S.Haley, J.F.Brown, S.Couperwhite, H.A.McQueen, D.Nicholson, W.Coppieters, A.Van de Weghe, Stratil A., and A.K.Wintero. 1995. The PiGMaP consortium linkage map of the pig (Sus scrofa). *Mamm Genome* **6**: 157-175.

Bates,R.O., C.W.Ernst, M.E.Doumit, D.B.Edwards, and N.E.Raney. 2003. Development of a Duroc x Pietrain Swine Resource Population. *NSIF conference proceedings*. Des Moines, IA.

Bessenyei,B., M.rka, M.Zeher, and I.Semsei. 2004. Single nucleotide polymorphisms: aging and diseases. *Biogerontology* **5**: 291-303.

Botstein,D., R.L.White, M.Skolnick, and R.W.Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet.* **32**: 314-331.

Brenner,C.H. 1999. The power of SNP's even with out population data.*10th Promega Symposium on Human Identification.* Orlando, Florida.

Cagnazzo,M., M.F.W.te Pas, J.Priem, A.A.C.de Wit, M.H.Pool, R.Davoli, and V.Russo. 2006. Comparison of prenatal muscle tissue expression profiles of two pig breeds differing in muscle characteristics. *J. Anim Sci.* **84**: 1-10.

Chakravarti,A. 2001. Single nucleotide polymorphisms. . .to a future of genetic medicine. *Nature* **409**: 822-823.

Chen,X. and P.F.Sullivan. 2003. Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput. *Pharmacogenomics J.* **3**: 77-96.

Collins,D.W. and T.H.Jukes. 1994. Rates of Transition and Transversion in Coding Sequences since the Human-Rodent Divergence. *Genomics* **20**: 386-396.

de Koning,D.J., L.L.G.Janss, A.P.Rattink, P.A.M.van Oers, B.J.de Vries, M.A.M.Groenen, J.J.van der Poel, P.N.de Groot, E.W.Brascamp, and J.A.M.van Arendonk. 1999. Detection of Quantitative Trait Loci for Backfat Thickness and Intramuscular Fat Content in Pigs (Sus scrofa). *Genetics* **152**: 1679-1690.

Donis-Keller,H., P.Green , C.Helms, S.Cartinhour , B.Weiffenbach, K.Stephens , T.P.Keith , D.W.Bowden, D.R.Smith, and E.S.Lander . 1987. A genetic linkage map of the human genome. *Cell* **51**: 319-337.

Duniec,H., 1961. Heritability of chemical fat content in the loin muscle of baconers. *Anim. Prod* **3**: 195-198.

Fernandez,X., G.Monin, A.Talmant, J.Mourot, and B.Lebret. 1999. Influence of intramuscular fat content on the quality of pig meat - 2. Consumer acceptability of m. longissimus lumborum. *Meat Science* **53**: 67-72.

Freese,E. 1959. The difference between spontaneous and base-analogue induced muations of phage T4. *Proc Natl Acad Sci U S A* **45**: 622-633.

Fujii,J., K.Otsu, F.Zorzato, S.de Leon, V.K.Khanna, J.E.Weiler, P.J.O'Brien, and D.H.MacLennan. 1991. Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science* **253**: 448-551.

Geldermann,H. 1975. Investigations of quantitative characters in animals by gene markers. in the *TAG Theoretical and Applied Genetics* pp. 319-330. Springer Berlin, Heidelberg.

Glowatzki-Mullis,M.L., Galliard.C, Wigger.G and Fries.R. 1995. Microsatellite-based parentage control in cattle. *Anim Genet* **26**: 7-12.

Gray,I.C., D.A.Campbell, and N.K.Spurr. 2000. Single nucleotide polymorphisms as tools in human genetics. *Hum. Mol. Genet.* **9**: 2403-2408.

Heaton,M.P., G.P.Harhay, G.L.Bennett, R.T.Stone, W.M.Grosse, E.Casas, J.W.Keele, T.P.L.Smith, C.G.Chitko-McKown, and W.W.Laegreid. 2002. Selection and use of SNP markers for animal identification and paternity analysis in U.S. beef cattle. *Mammalian Genome* **13**: 272-281.

Holliday,R. and G.W.Grigg. 1993. DNA methylation and mutation. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **285**: 61-67.

Janss,L.L., J.A.Van Arendonk, and E.W.Brascamp. 1997. Segregation analyses for presence of major genes affecting growth, backfat, and litter size in Dutch Meishan crossbreds. *J. Anim Sci.* **75**: 2864-2876.

Jeon,J.T., O.Carlborg, A.Tornsten, E.Giuffra, V.Amarger, P.Chardon, L.ndersson-Eklund, K.Andersson, I.Hansson, K.Lundstrom, and L.Andersson. 1999. A paternally

expressed QTL affecting skeletal and cardiac muscle mass in pigs maps to the IGF2 locus. *Nat Genet* **21**: 157-158.

Jones,G.F. 1998. Genetic Aspects of Domestication, Common Breeds and their Origin. in *The Genetics of the pig* (ed. M F Rothschild and A Ruvsinksy), pp. 17-51. CAB International, New York.

Jonsson,P. 1963. Danish pig progeny testing results. *Zeitschrift für Tierzüchtung und Zuchtüngsbiologie* **78**: 205-252.

Le Roy,P., J.Naveau, J.M.Elsen, and P.Sellier. 1990. Evidence for a new major gene influencing meat quality in pigs. *Genet Res.* **55**: 33-40.

Li,W.H. and L.A.Sadler. 1991. Low Nucleotide Diversity in Man. *Genetics* **129**: 513-523.

Loots,G.G., R.M.Locksley, C.M.Blankespoor, Z.E.Wang, W.Miller, E.M.Rubin, and K.A.Frazer. 2000. Identification of a Coordinate Regulator of Interleukins by Cross-Species Sequence Comparisons. *Science* **288**: 136-140.

Mariani,P., K.Lundstrom, U.Gustafsson, C.Enfalt, R.K.Juneja, and L.Andersson. 1996. A major locus affecting muscle glycogen content is located on pig Chromosome 15. *Mammalian Genome* **7**: 52-54.

Marklund,L., M.Johansson, M.Hoyheim, B.Davies, W.Fredholm, M.Juneja, R.K.Mariani , P.Coppieters, W.H.Ellegren, and L.Andersson. 1996. A comprehensive linkage map of the pig based on a wild pig-Large White intercross. *Anim Genet.* **27**: 255-269.

Meisinger, D. 1999. A system for assuring pork quality. *National Pork Producers Council*, Des Moines, IA.

Mikawa,S., T.Akita, N.Hisamatsu, Y.Inage, Y.Ito, E.Kobayashi, H.Kusumoto, T.Matsumoto, H.Mikami, M.Minezawa, M.Miyake, S.Shimanuki, C.Sugiyama, Y.Uchida, Y.Wada, S.Yanai, and H.Yasue. 1999. A linkage map of 243 DNA markers in an intercross of Gottingen miniature and Meishan pigs. *Animal Genetics* **30**: 407-417.

Moran,C. 1998. Molecular Genetics. in *The Genetics of the pig* (ed. M F Rothschild and A Ruvsinksy), pp. 135-162. CAB International, New York.

Nezer,C., L.Moreau, B.Brouwers, W.Coppieters, J.Detilleux, R.Hanset, L.Karim, A.Kvasz, P.Leroy, and M.Georges. 1999. An imprinted QTL with major effect on muscle mass and fat deposition maps to the IGF2 locus in pigs. *Nat Genet* **21**: 155-156.

Paszek,A.A., P.J.Wilkie, G.H.Flickinger , L.M.Miller , C.F.Louis , G.A.Rohrer , L.J.Alexander , C.W.Beattie, and L.B.Schook. 2006. Interval mapping of carcass and meat quality traits in a divergent swine cross. *Animal Biotechnology* **12**: 155-165.

Picoult-Newberg,L., T.E.Ideker, M.G.Pohl, S.L.Taylor, M.A.Donaldson, D.A.Nickerson, and M.Boyce-Jacino. 1999. Mining SNPs From EST Databases. *Genome Res.* **9**: 167-174.

Ranade,K., M.S.Chang, C.T.Ting, D.Pei, C.F.Hsiao, M.Olivier, R.Pesich, J.Hebert, Y.D.Chen, V.J.Dzau, D.Curb, R.Olshen, N.Risch, D.R.Cox, and D.Botstein. 2001. High-Throughput Genotyping with Single Nucleotide Polymorphisms. *Genome Res.* **11**: 1262-1268.

Richard,I. and J.S.Beckmann. 1995. How neutral are synonymous codon mutations? *Nat Genet* **10**: 259.

Rohrer,G.A., L.J.Alexander, J.W.Keele, T.P.Smith, and C.W.Beattie. 1994. A Microsatellite Linkage Map of the Porcine Genome. *Genetics* **136**: 231-245.

Rohrer,G.A. and J.W.Keele. 1998. Identification of quantitative trait loci affecting carcass composition in swine: I. Fat deposition traits. *J. Anim Sci.* **76**: 2247-2254.

Russell,R. 1997. Meat (Pork) Quality / A Total Overview in practical meat quality technology, What is meat quality?. *NSIF Proceedings*. Washington, DC.

Ruvinsky,A. and M.F.Rothschild. 1998. Systematics and Evolution of the Pig. in *The Genetics of the pig* (ed. M F Rothschild and A Ruvsinksy), pp. 1-16. CAB International, New York.

Sellier,P. 1998. Genetics of Meat and Carcass Traits. in *The Genetics of the pig* (ed. M F Rothschild and A Ruvsinksy), pp. 463-510. CAB International, New York.

Smith,E.J., L.Shi, P.Drummond, L.Rodriguez, R.Hamilton, S.Ramlal, G.Smith, K.Pierce, and J.Foster. 2001. Expressed Sequence Tags for the Chicken Genome from a Normalized 10-Day-Old White Leghorn Whole Embryo cDNA Library: 1. DNA Sequence Characterization and Linkage Analysis. *J Hered* **92**: 1-8.

Sobrino,B., M.Brion, and A.Carracedo. 2005. SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Science International* **154**: 181-194.

Suh,Y. and J.Vijg. 2005. SNP discovery in associating genetic variation with human disease phenotypes. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **573**: 41-53.

The International SNP Map Working Group. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**: 928-933.

van Heugten,E., 2000.Understanding pork quality. *Swine News* .NCSU Extension Swine Husbandry, North Carolina.

Van Haeringen,W.A., M.Den Bieman, G.F.Gillissen, A.Lankhorst, M.T.R.Kuiper, L.F.M.Van Zutphen, and H.A.Van Lith. 2001. Mapping of a QTL for Serum HDL Cholesterol in the Rabbit Using AFLP Technology. *J Hered* **92**: 322-326.

Van Laere,A.S., M.Nguyen, M.Braunschweig, C.Nezer, C.Collette, L.Moreau, A.L.Archibald, C.S.Haley, N.Buys, M.Tally, G.Andersson, M.Georges, and L.Andersson. 2003. A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature* **425**: 832-836.

Vignal,A., Milan.D, SanCristbal.S and Eggen.A. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. *Genet. Sel. Evol* **34**: 275-305.

Williams,J.L. 2005. The use of marker-assisted selection in animal breeding and biotechnology. *Rev Sci Tech*. **24**: 379-391.

http://www.nppc.org/. 2006. NPPC

www.nppc.org/about/pork_today.html. 2004. Today's U.S. Pork Industry.

www.porkboard.org/docs/pkfacts2002.pdf. 2003. Pork facts 2002/2003.

www.porkboard.org/porkfacts/default.html. 2004. Pork Quick facts.

Young,L.D. 1995. Review of Swine Genetics in the U.S. *NSIF proceedings*. Clive, IA.

# CHAPTER 2

Development of Porcine Subcutaneous Adipose Tissue cDNA Library

## Introduction

Sequencing the complete genome to find genes of an organism is a tedious task. This is especially true in eukaryotic genomes, where only a small portion of the genome is directly involved in gene expression. Analysis of gene expression is facilitated by construction of cDNA libraries. A cDNA library has the same sequence distribution of original mRNA in the cell and represents the genes expressed at the time of tissue collection.

Sequencing of Expressed Sequence Tags (ESTs) from clones derived from a cDNA library of a specific tissue is a popular method for discovering of new genes, gene mapping and identification of coding regions of a genome (Adams and Kelley, 1991). Sequences of ESTs are a valuable resource for the development of high density hybridization arrays for transcriptional profile studies and comparative map development (Fahrenkrug et al., 2002). This method of gene discovery has been applied in livestock, including pig. Fahrenkrug et al. (2002) normalized and sequenced porcine cDNA libraries derived from embryonic and reproductive tissues and deposited more than 50,000 EST sequences in GenBank. Successful mapping of human ESTs to pig genome has significantly increased the number of genes mapped to the porcine genome and this approach can be applied to increase the density of mammalian gene maps ( Lahbib-Mansais et al., 1999).

Even though the number of porcine ESTs in GenBank porcine dbEST has reached 500,000 (www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html, 2006) in January 2006, the contribution of adipose tissue ESTs is very low, with less than 1% of transcripts being adipose derived. Large number of ESTs generated from skeletal muscle (Yao, 2002 ),

embryonic cells, reproductive tissue (Fahrenkrug et al., 2002), spleen, thymus, lymph node, bone marrow (Rink et al., 2002;Uenishi et al., 2004), and brain (Nobis et al., 2003) of pig are deposited in public databases.

The adipocyte is the major cell type in the adipose tissue with blood cells, endothelial cells, pericytes and adipocyte precursor cells making up rest of the tissue (Ailhaud et al., 1992). Even thought the main function of mature adipocytes is energy storage, they also act as an endocrine organ with adipocytic secretions involved in immunological functions, vascular diseases and appetite regulation (Gregoire et al., 1998). Similar to many other domestic animals, pigs reach an age when muscle growth decreases concomitantly with an increase in body fat mass (Anderson and Kauffman, 1973). As adipose tissue is the primary site of fatty acid synthesis in pig (Anderson and Kauffman, 1973), information on genes expressed in adipose tissue are very important in identification of candidate genes for marbling and back fat thickness in swine. Lack of knowledge on genes expressed in adipose tissue creates a drawback in selective breeding for meat quality traits such as marbling and tenderness. First part of this research project is focused on addressing this shortfall by constructing a subcutaneous adipose tissue specific cDNA library from $F_2$ progeny of a Duroc x Pietrain resource population. Duroc breed, originated in eastern United States and the corn belt, is used as a terminal sire line in many breeding programs (www.ansi.okstate.edu/breeds/swine, 1997; Canadian Centre for Swine Improvement, 2003). They are slow growers with excellent muscle quality traits, such as high intra muscular fat which is softer, with more unsaturated fatty acids and redder muscle fibers (Cagnazzo et al., 2006; Sellier, 1998). Pietrain, a breed recognized as one of the leanest and heaviest muscled breeds of pigs,

originated in the village of Pietrain in Belgium (www.ansi.okstate.edu/breeds/swine, 1997). Jones (1998) described Pietrain as the most interesting and controversial swine breed of the 20th century. This breed is included in various breeding programs in Europe to improve the market quality of swine, but is not used much in United States. Pietrain animals are prolific and fast growing, but meat quality is poor, with pale muscle fibers and low intra muscular fat depots ( www.ansi.okstate.edu/breeds/swine, 1997;Cagnazzo et al., 2006). In this breed, subcutaneous dorsal adipose tissue contains more water and fewer lipids, with higher ratio of polyunsaturated fatty acids over saturated fatty acids (Sellier, 1998). When compared with Duroc sired pigs, loin chops of Pietrain sired pigs are less marbled, with lower ultimate pH and decreased water holding capacity (Edwards et al., 2003).

Cagnazzo et al. (2006) state that Duroc and Pietrain breeds represent extremes of modern western pig breeds. Edwards et al. (2003) evaluated the meat and carcass quality measures of Duroc vs. Pietrain sired pigs and their results indicated that traits from both Duroc and Pietrain breeds can be utilized for improvement of US pork industry (Edwards et al., 2003).

According to Cepica et al (1995) heterozygosis increased with the genetic distance between the two parental populations and a linkage divergence is created. This linkage divergence is responsible for creation of unique genomic markers for genotyping performance traits in $F_2$ and later generations (Bates et al., 2003). Heterozygosis will be high in a resource population with Duroc and Pietrain parents that are at extremes for meat quality traits. As information gained from a genetic marker in a given population increases with the degrees of heterozygosity (Archibald and Haley, 1998) a progeny

developed from Duroc x Pietrain founder animals is ideal for identification of molecular markers for QTL mapping of genes responsible for pork quality.

We constructed a subcutaneous adipose tissue cDNA library from the $F_2$ progeny of Duroc x Pietrain resource population. The cDNA library was sequenced and 13,526 high quality sequences were deposited in dbEST division of GenBank. Obtained ESTs were clustered and assembled in to 2080 unique contigs. Three hundred and thirty contigs that have significant similarities to known genes were clustered according to Gene Ontology Consortium classifications.

## Materials and Methods

### Development of the Resource Population

The $F_1$ resource population was developed at Michigan State University by crossbreeding of 15 RYR1 normal Pietrain females to 4 Duroc males from a closed unselected population. From the $F_1$ progeny, 51 female and 6 male animals were retained and crossbred to generate the $F_2$ population. Live measures of weight at birth, weaning at 6 weeks, scanning of tenth rib back fat, last rib back fat and loin eye back fat were obtained from 959 $F_2$ animals born in 11 different groups. Pigs from the $F_2$ population were slaughtered between 23-26 weeks of age over 3 harvest dates per group, and carcass and meat quality measurements were obtained for these animals. Data gathering process from the resource population was completed in February 2004.

Subcutaneous adipose tissue samples from 9 $F_2$ animals were selected for the cDNA library construction. These 9 animals consisted of 3 animals with relatively high back fat thickness, 3 animals with relatively low back fat thickness and 3 animals with intermediate back fat thickness.

### RNA Extraction

One gram of adipose tissue was homogenized in 4ml of TRIzol (Invitrogen, Carlsbad, CA, USA). Afterwards, samples were centrifuged at 9000rpm for 15 minutes at $4^0$C. The lower aqueous layer was transferred to a fresh tube and incubated at room temperature for 5 minutes. After incubation, 1ml of chloroform was added to each 4ml of TRIzol. The tube was then vortexed for 15 seconds, incubated in room temperature for 3 minutes and centrifuged at 7000rpm for 30 minutes at $4^0$C. The upper aqueous

layer was transferred to a fresh tube and the RNA was precipitated by adding 0.8ml of isopropyl alcohol per 1ml of aqueous phase. Afterwards, the mixture was incubated for 30 minutes at -80$^0$C and then centrifuged at 9000rpm for 10 minutes at 4$^0$C. The Supernant was decanted, and the pellet was washed with 3ml of 75% ethanol and centrifuged at 9000rpm for 5 minutes at 4$^0$C. The ethanol was removed and the pellet was air dried for 5 -10 minutes. Finally it was resuspended in 250µl of DEPC treated water. Optical Density (OD) of RNA samples was measured by reading the absorbance in UV spectrophotometer and the quality of RNA was verified by gel electrophoresis of 3µg of RNA on a 1.25% denaturing agarose gel. The total RNA extracted from 9 animals were pooled and mRNA was isolated from pooled RNA.

**Isolation of mRNA from RNA**

Messenger RNA was isolated from RNA using Ambion Poly (A) purist mRNA purification kit (Ambion, Austin, TX, USA). Seven hundred and eighty microliters of 2X binding buffer was added to the total RNA solution, mixed well and the solution was transferred to Oligo(dT) cellulose tube. The tube was heated for 5 minutes at 75$^0$C to denature secondary structures and then the tube was shaken gently at room temperature for 1.5 hours. Afterwards, it was centrifuged at 6000rpm for 3 minutes at room temperature. The supernant was decanted and Oligo(dT) cellulose was washed two times with 0.5ml wash solution 1 and three times with 0.5ml of wash solution 2. The Oligo(dT) cellulose column was placed in a new tube and 200µl of pre warmed (80$^0$C) RNA storage solution was added and vortexed. After that, the tube was centrifuged at 7900rpm for 2 minutes at room temperature to elute mRNA. The sample was then eluted

for a second time with another 200µl of The RNA storage solution. The elute was precipitated by adding 40µl of 5M ammonium acetate, 1µl of glycogen and 1.1ml of 100% ethanol and incubating at -80$^0$C for 30 minutes. After incubation, the sample was centrifuged at 14000rpm for 30 minutes at 4$^0$C. The supernant was decanted and the pellet was washed with 400µl of 75% ethanol and centrifuged for 10 minutes at 4$^0$C at 14000rpm. Finally, the pellet was dried at room temp for 5minutes and resuspended in 12µl of nuclease free water. OD of the mRNA was measured by UV the absorbance in UV spectrophotometer and the quality of mRNA was verified by gel electrophoresis of 2µg of mRNA on a 1.25% denaturing agarose gel.

**cDNA Library Construction**

Subcutaneous adipose tissue cDNA library was constructed according to the following procedure using Invitrogen superscript plasmid system (Carlsbad, CA, USA).

**First Strand cDNA Synthesis**

One micro liter of DEPC treated water was added to 4µl of m-RNA solution to bring the volume to 5µl. Then, 2µl of Not 1 primer adapter was added to the solution and the mixture was heated for 10 minutes to denture secondary m-RNA structures. After 10 minutes the tube was quickly chilled on ice and 4µl of 5x first strand buffer, 2µl of 0.1 M DTT, 1µl of 10mM dNTP mix and 1µl of [α-$^{32}$P] dCTP (1µCi/µl) were added, briefly vortexed and placed at 37$^0$C for 2 minutes for equilibrium of temperature. After 2 minutes, 5µl of Superscript II RT was added to the tube, gently mixed and incubated for one hour in room temperature. The reaction was terminated after one hour by placing the

tube on ice. Two microliters of reaction was removed and added to a tube containing 43µl of 20mM EDTA at pH 7.5 and 5µl of yeast tRNA and this was used for analysis of cDNA products. The rest of the reaction was used for second strand cDNA synthesis.

**Second Strand cDNA Synthesis**

Ninety three microliters of distilled water, 30µl of 5x $2^{nd}$ strand buffer, 3µl of 10mM dNTP mix, 1µl *of E.coli* DNA ligase (10u/µl), 4µl of *E.coli* DNA polymerase (10u/µl) and 1µl *of E.coli* RNAaseH were added to 18µl of first strand cDNA. The tube was gently vortexed and incubated at $16^0$C for 2 hours. After 2 hours, 2µl of 10U $T_4$ DNA polymerase was added and incubated for another 5 minutes at $16^0$C. Reaction was placed on ice and 10µl of 0.5M EDTA was added to stop the reaction. Then, 150µl of phenol: chloroform: isoamyl alcohol (25: 24: 1) was added, vortexed and centrifuged at 14000 x g in room temperature for 5 minutes to separate the phases. 140µl of upper aqueous layer was carefully removed and transferred in to a fresh 1.5µl micro centrifuge tube and cDNA was precipitated by adding 70µl of 7.5M $NH_4OAC$ and 0.5ml of 100% ethanol at $-20^0$C and centrifuging at 14000 x g at room temperature for 20 minutes.
The pellet was washed again with 70% ethanol at $-20^0$C and centrifuged for 2 minutes at 14000 x gat room temperature. The supernatant was removed carefully and the cDNA pellet was dried at $37^0$C for 5 minutes to evaporate the residual ethanol.

**Sal 1 Adapter Ligation**

The cDNA pellet was dissolved in 25µl of DEPC treated water and 10µl of 5x $T_4$ DNA ligase buffer, 10µl of Sal 1 adapter and 5µl of $T_4$ DNA ligase were added to this ; gently

mixed and incubated at $16^0$C overnight. 50µl of phenol: chloroform: isoamyl alcohol (25: 24: 1) was added to the overnight incubated reaction and centrifuged at 14000 x g for 5 minutes at room temperature for separation of phases. Upper aqueous layer was carefully removed and cDNA was precipitated adding 25µl of 7.5M $NH_4OAc$ and 150µl of $-20^0$C absolute alcohol and centrifuging at room temperature for 20 minutes at 14000 x g. Finally, the pellet was washed with 0.5ml of $-20^0$C 70% ethanol by centrifuging 2 minutes at 14000 x g in room temperature and dried at $37^0$C for 10 minutes.

**Analysis of cDNA Products**

The supernant removed from the first strand synthesis was spotted in duplicate of 10µl aliquots on to glass fiber filters. One filter was dried at room temperature and used to determine the specific activity of the dCTP reaction. The other film was washed three times in sequence, for 5 minutes each time in a beaker containing 50ml of fresh, ice cold 10% (w/v) TCA containing 1% (w/v) sodium pyrophosphate. The filter was then washed once again with 50ml of 95% ethanol at room temperature. This filter was used to determine the yield of first strand cDNA.

Both filters were counted in Beckman LS 6000 scintillation counter (Beckman Coulter Inc, CA, USA) to determine the amount of $^{32}$P in the reaction and the amount of $^{32}$P that was incorporated to the reaction.

**Not 1 Digestion**

Forty one microliters of DEPC treated water, 5µl of REACT 3 buffer and 4µl of Not 1 were added to the cDNA pellet and incubated at $37^0$C for 2 hours. The cDNA was

separated into upper phase by centrifuging for 5 minutes at room temperature at 14000 x g with 50µl of phenol:chloroform: isoamyl alcohol (25: 24: 1). Forty five microliters of upper aqueous layer was precipitated by adding 25µl of 7.5M $NH_4OAc$ and 150µl of (-$20^0$C)absolute ethanol and centrifuging at 14000 x g for 20minutes. The pellet was washed with 0.5ml of 70% ethanol and dried at $37^0$C for 10 minutes. Twenty three microliters of first strand cDNA and 10µl of cDNA from the second strand were concentrated to 5µl in DNA Speed Vac and 3µl of sample were loaded on a 1% agarose gel with 2µl of loading dye to check the quality of cDNA.

**Column Chromatography**

Column Chromatography was performed to optimize the size fractionation of the cDNA and to remove residual adaptors and fragments released from Not 1 digestion. The cDNA was dissolved in 100µl of TEN buffer and eluted through a column that has a flow rate of 0.8ml/13min with drop size of 25µl. The column was first washed with 20% ethanol and then rinsed 3 times by allowing 0.8ml TENbuffer to drain through the column. Then the entire cDNA (100µl) was added to the center of the top grid and single drop fractions were collected to individual tubes labeled 1-20. The volume of each drop in separate tubes was measured and Cerenkov counts for each fraction were obtained by placing the tubes in a scintillation counter. OD of the samples was measured by UV spectrophotometer and cDNA concentration was calculated. Three microliters of fractions 7-17 with 2µl of DNA loading dye were loaded on a 1% agarose gel to check the quality and size distribution of the fractions. The gel was then transferred to a film

cassette with 8x10 Kodak Bio mass film; exposed overnight and developed next day to visualize the quality and the size of the cDNA fragments.

**Ligation of cDNA to the Vector**

Fragment 9 which had sizes of cDNA between 0.7kb-6kb was selected and ligated to the Not1/Sal1 cut vector pSPORT1 (Invitrogen, Carlsbad, CA, USA) according to the following procedure. Four microliters of 5x $T_4$ DNA ligase buffer, 1µl of pSPORT1 vector, 7.2µl of cDNA (10.08ng) and 6.8µl of DEPC treated water was mixed well in a sterile 1.5ml microcentrifuge tube and then 1µl of $T_4$ DNA ligase was added and mixed by pipetting. The mixture was incubated 1 hour at room temperature and $4^0$C overnight for the ligation reaction to occur.

**Transformation of Ligated Vector to DH5α™ Competent Cells**

Five microliters of ligation reaction was mixed with 100µl of MAX Efficiency® DH5α™ competent cells (Invitrogen, Carlsbad, CA, USA) and incubated on ice for 30 minutes. Afterward, the cells were heat shocked for 45 seconds in $42^0$C water bath. The cells were then transferred back on ice for 2 minutes. After incubation of on ice, 0.9ml room temperature S.O.C medium was added to the cells and incubated at $37^0$C for one hour in a shaker at a speed of 225rpm.

**Analysis of Sizes of Inserts in the cDNA Library**

Five microliters of transformation mixture was mixed with 100µl of S.O.C medium and plated on a YT plate containing 100µg/ml Ampicillin and 40µl of X-gal (40mg/ml).

Plates were incubated over night at $37^0$C. The number of colonies on the plate was counted and 91 colonies were transferred for a separate overnight growth at $37^0$C in 150µl of YT media with Ampicillin (200µg/ml) and 8% glycerol. Fifty microliters of 91 overnight grown colonies were transferred separately to 4ml of YT broth with Ampicillin (200µg/ml) and incubated at $37^0$C over night in the shaker at 225rpm.

DNA was extracted from 91 overnight grown colonies by QIAGEN plasmid DNA extraction kit (Qiagen, Valencia, CA, USA) according to the manufacturers' instructions. Inserts were PCR amplified by using $SP_6$ and $T_7$ primers. Twenty five micro liter PCR reactions were performed by mixing 18.7µl of distilled water, 2.5µl of 10x PCR buffer (Applied Biosystems, Foster city, CA), 1.5µl of 25mmol $MgCl_2$, 0.5µl of 10µmol dNTP, 0.5µl of 25µmol $T_7/SP_6$ primer mixture, 0.25µl of 5U/µl Taq DNA Polymerase and 0.5µl of plasmid DNA. PCR was performed in MJ Dyad thermal cycler (MJ Research Inc., Waltham, MA, USA) with first cycle of $95^0$C for 4 minutes followed by $94^0$C for 30 seconds, annealing temperature $53^0$C for 30 seconds, $72^0$C for 2 minutesfor 30 additional cycles with a final incubation at $72^0$C for 8 minutes. The size of the inserts was estimated by gel electrophoresis of 10µl of the product in 0.8% agarose gel stained with 6µl of ethidium bromide (10mg/ml).

**Sequencing of the cDNA Library**

Remaining volume of transformed cells were sent to Advanced Center for Genome Technology at University of Oklahoma, where the cells were plated on YT plates containing Ampicillin (100µg/ml) and 40µl of X-gal (40mg/ml) and incubated over night

at $37^0$C. Fifteen thousand clones were robotically picked up for sequencing from the 3' terminal in an ABI 3700 capillary sequencer (Applied Biosystems, Foster city, CA). Trace files of 13,958 sequences were assembled and clustered into contigs.

**Analysis of Expressed Sequence Tags**

Blast searches were performed between 2020 translated conitg sequences and the UniProt data base (www.pir.uniprot.org/, 2006). From the results of 1334 protein sequences, most significant blast results were selected and queried in gene-association.go-uniprot tool (version 060506) from Gene Ontology (GO) database to obtain the GO annotations (www.geneontology.org/,2006). Annotations were found for 330 proteins and GO annotation numbers of these proteins were selected and arranged in a Microsoft Excel spread sheet. This data was processed by Genesis software for clustering of gene products according to Gene Ontology consortium classifications (Sturn et al., 2002).

## Results

We were able to extract high quality RNA without genomic DNA or protein contamination (see Figure 1 and Figure 2). Sizes of isolated mRNA ranged from approximately 300bp to 4kbp and the concentration calculated from OD value was 1.53µg/µl. Quality of mRNA is an essential factor because mRNA used as the template influence the yield and size distribution of first strand cDNA. Amount of cDNA in the first strand reaction was calculated as 1.948µg representing 38% yield relative to 5µg of starting material of mRNA.

Size fractionation of the cDNA was achieved by column chromatography and the size dispersion of cDNA in fractions was visualized in the gel picture (see Figure 3). Fractions 8, 9 and 10 contained cDNA of size range 0.6-6kbp, fraction 11 contained cDNA of size range 0.6-6kbp and fraction 16 and 17 had cDNA below 1kbp size.

PCR amplification of the 91 inserts from the resulting clones showed that 46 clones (50.5%) had inserts >1kb, 18 clones (19.7%) had inserts of 1kb, 18 clones (19.7%) had inserts < 1kb and 3 (3.2%) clones were without an insert (see Figure 4).

Fifteen thousand clones were picked for 3'end-sequencing. Sequence analysis showed that there were 10 clones with no adaptors, 479 clones with short inserts and 603 clones without an insert. There were 327 clones with ribosomal RNA inserts, 27 clones with *E-coli* DNA insert and a one clone with mitochondrial DNA insert. These clones accounted for 9.4 % of total number of clones sequenced.

Thirteen thousand nine hundred and fifty eight sequences were assembled and clustered in to 2080 unique clusters. Of these, 726 (35%) clusters consisted of only one sequence read (singlets) and 1354 (65%) were with at least 2 reads. From all the sequences, 13447

were deposited in GenBank EST database as Porcine Adipose Tissue cDNA library (PADT) Sus scrofa cDNA 3' mRNA sequences (accession numbers EB424599-EB411152).

Three hundred and thirty contigs that had significant similarities to known genes were clustered according to Gene Ontology Consortium classifications. In the categorization of biological process, 235 genes with 3663 GO relations were annotated. The majority of these genes encode for proteins belonging to physiological and cellular processes. Genes belonging to biological processes of reproduction, growth and pigmentation were not found in the porcine adipose tissue cDNA library (see Figure 5). In the categorization of cellular component 161 genes with 776 GO relations were annotated and majority of gene products were classified as located in the cell (see Figure 6).

In the categorization of molecular function, 309 genes with 854 GO annotations were annotated (see Figure 7). The majority of these gene products involve selective stoiechiometric interactions with other molecules and catalysis. Approximately 60% of annotated gene products categorized as binding interact with nucleotides; nucleic acids or ion binding proteins (see Figure 8a). From the annotated gene products of catalysis reactions, 38% were oxidoreductases, 25% were hydrolases and 22% were transferases (see Figure 8b). Electron carrier proteins were high among oxidoreductases (see Figure 9a). Acid anhydrases constitute 34%, whilst esterases and peptidases constitute 45% of hydrases (see Figure 9b). Trasferases acting on phosphorus transfer group constituted 48% of transferase enzymes (see Figure 9c). Variability was highest in oxidoreductases which had 11 different categories. Majority of annotated signal transducers were

combining with an extracellular or intracellular messenger to initiate a change in cell activity (see Figure 10).

**Figure 1**

Picture of ethidium bromide stained 1.25% denaturing RNA gel of subcutaneous adipose tissue RNA extracted from 9 $F_2$ animals. Three micro grams of RNA was loaded in lanes 1-9.

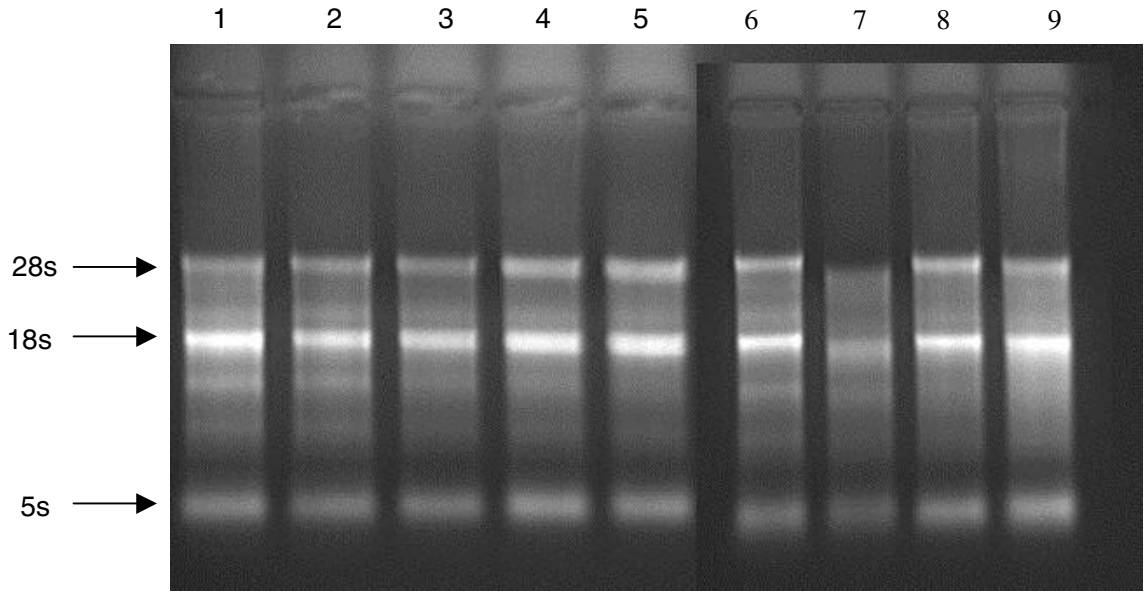

1
2
3
4
5
6
7
8
9
28s
18s
5s

**Figure 2**

Picture of Ethidium bromide stained 1.25% denaturing gel picture of subcutaneous adipose tissue mRNA extracted from pooled RNA samples of 9 $F_2$ animals. Lanes 1, 2 and 3 represent RNA marker, 2µg of mRNA and 2µg positive control total RNA in third lane.

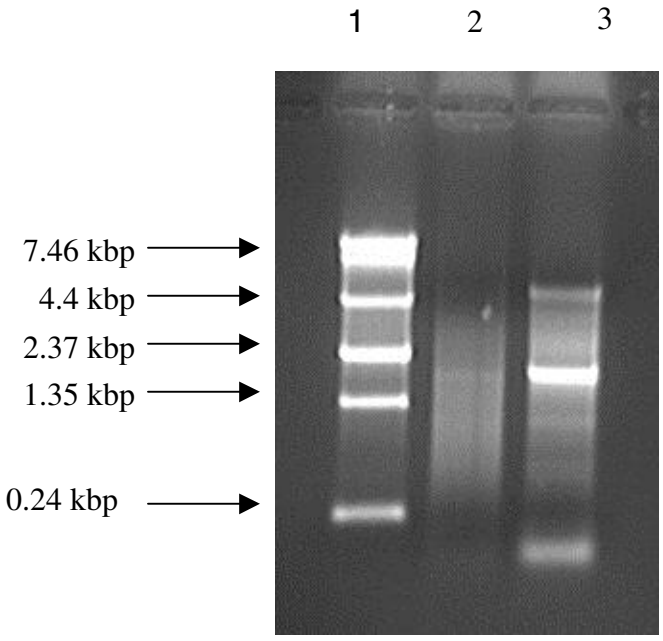

1
2
3
7.46 kbp
4.4 kbp
2.37 kbp
1.35 kbp
0.24 kbp

**Figure 3**

Electrophoretic analysis results of size fractioned cDNA. Three microliters of fractions 7-17 were loaded on Ethidium bromide stained 1% agarose gel and electrophoresed at 10V. DNA marker (1kb) was loaded to estimate the size dispersion of cDNA in the fractions.

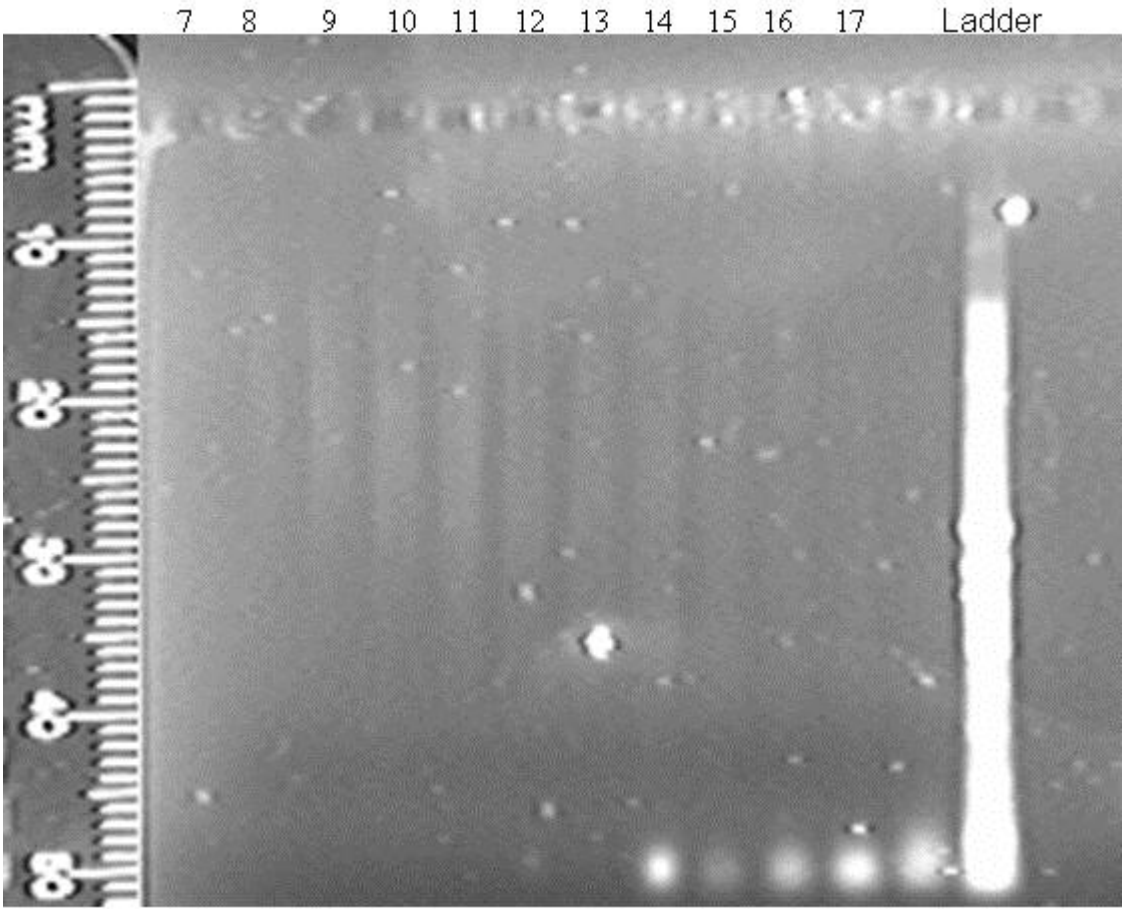

7
8
9
10
11
12
13
14
15
16
17
Ladder

**Figure 4**

Picture of ethidium bromide stained 0.8% agarose gel of PCR amplified inserts from 91 cDNA clones. Sizes of the inserts were estimated by comparing to 1kb DNA ladder.

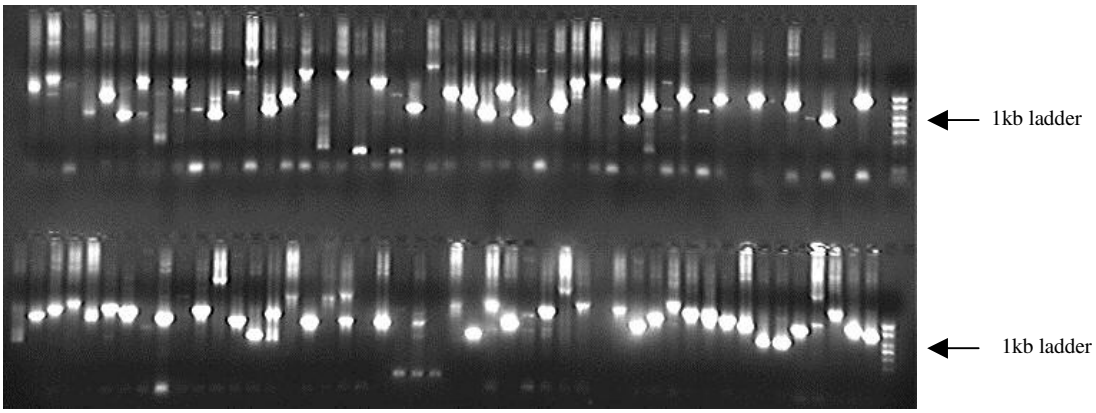
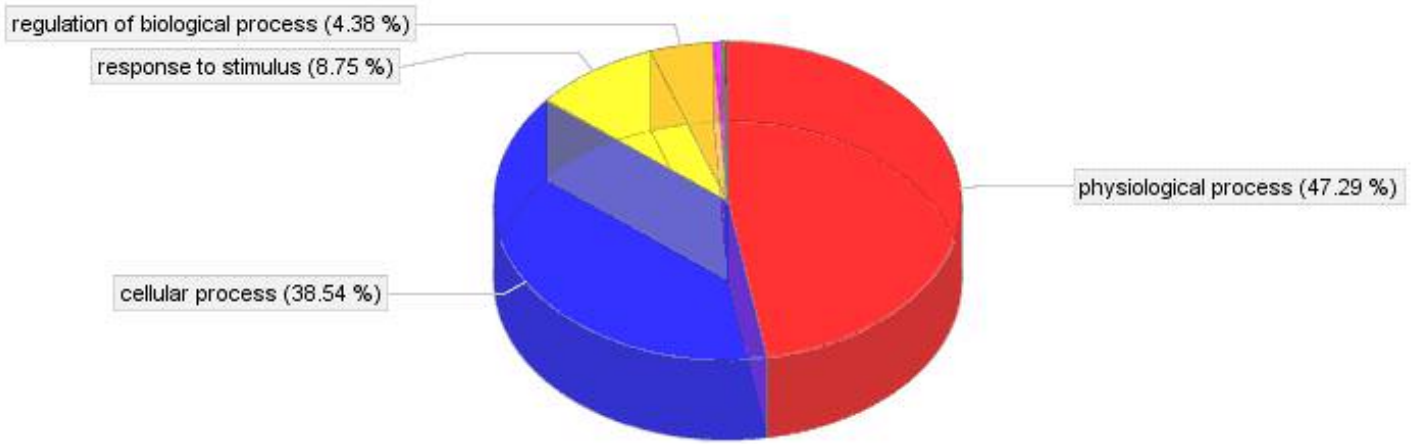

1kb ladder
1kb ladder

**Figure 5**

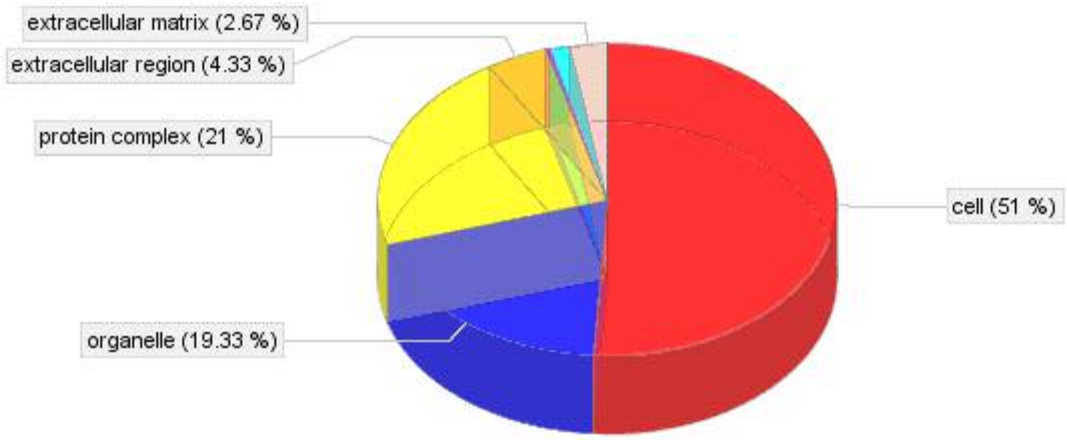
Gene products annotated in biological process group according to Gene Ontology Consortium classifications (Sturn et al., 2002). There were 235 genes with 3863 GO relations clustered in biological process, with the majority of gene products involved in physiological and cellular process




| | | |
|---|---|---|
| physiological process | 227 | 47.292 % |
| cellular process | 185 | 38.542 % |
| biological process unknown | 0 | 00.000 % |
| response to stimulus | 42 | 08.750 % |
| regulation of biological process | 21 | 04.375 % |
| development | 3 | 00.625 % |
| reproduction | 0 | 00.000 % |
| growth | 0 | 00.000 % |
| interaction between organisms | 1 | 00.208 % |
| viral life cycle | 1 | 00.208 % |
| pigmentation | 0 | 00.000 % |

**Figure 6**

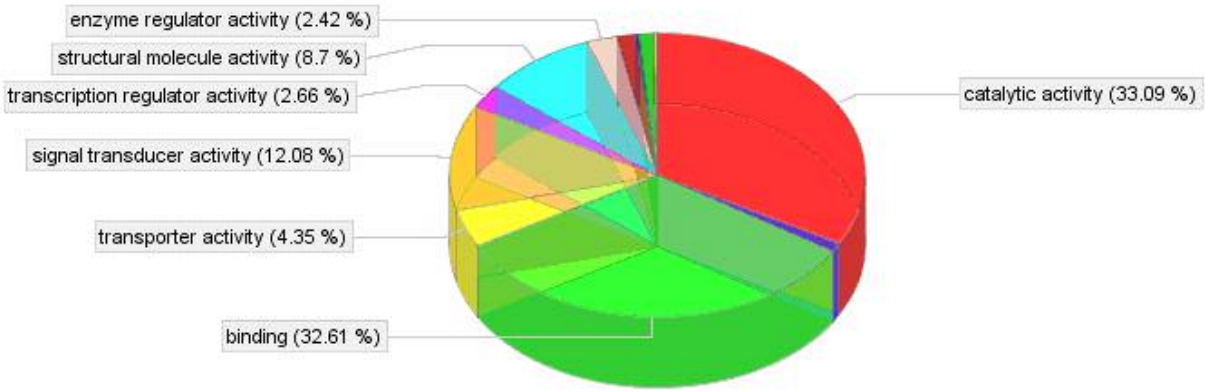
Gene products annotated in cellular component group according to Gene Ontology Consortium classifications (Sturn et al., 2002). There were 161 gene products with 776 GO relations clustered in cellular component with majority of gene products localized in the cell.




| | | | |
|---|---|---|---|
| cell | | 153 | 51.000 % |
| organelle | | 58 | 19.333 % |
| cellular component unknown | | 0 | 00.000 % |
| protein complex | | 63 | 21.000 % |
| extracellular region | | 13 | 04.333 % |
| membrane-enclosed lumen | | 1 | 00.333 % |
| envelope | | 4 | 01.333 % |
| extracellular matrix | | 8 | 02.667 % |
| synapse | | 0 | 00.000 % |
| virion | | 0 | 00.000 % |

**Figure 7**

Gene products annotated in molecular function group according to Gene Ontology Consortium classifications (Sturn et al., 2002). There were 309 genes with 854 GO relations clustered under molecular function with majority of gene products involved in molecular functions of catalysis, binding and signal transduction.

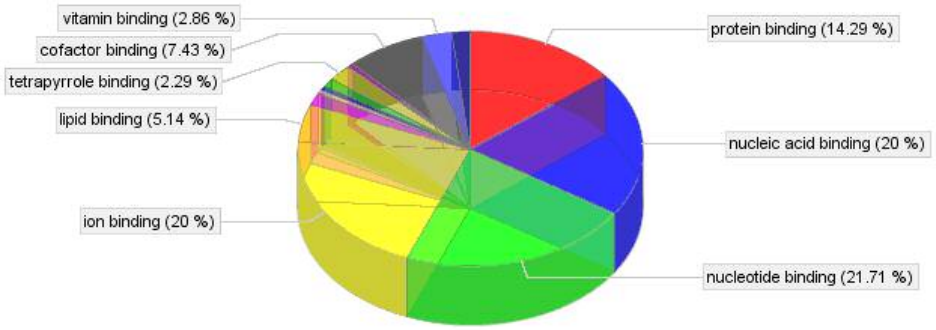



| | | | |
|---|---|---|---|
| catalytic activity | | 137 | 33.092 % |
| molecular function unknown | | 4 | 00.966 % |
| binding | | 135 | 32.609 % |
| transporter activity | | 18 | 04.348 % |
| signal transducer activity | | 50 | 12.077 % |
| transcription regulator activity | | 11 | 02.657 % |
| structural molecule activity | | 36 | 08.696 % |
| enzyme regulator activity | | 10 | 02.415 % |
| triplet codon-amino acid adaptor activity | | 0 | 00.000 % |
| translation regulator activity | | 6 | 01.449 % |
| motor activity | | 1 | 00.242 % |
| antioxidant activity | | 5 | 01.208 % |
| nutrient reservoir activity | | 1 | 00.242 % |
| chaperone regulator activity | | 0 | 00.000 % |
| protein tag | | 0 | 00.000 % |
| chemoattractant activity | | 0 | 00.000 % |
| chemorepellant activity | | 0 | 00.000 % |
| energy transducer activity | | 0 | 00.000 % |

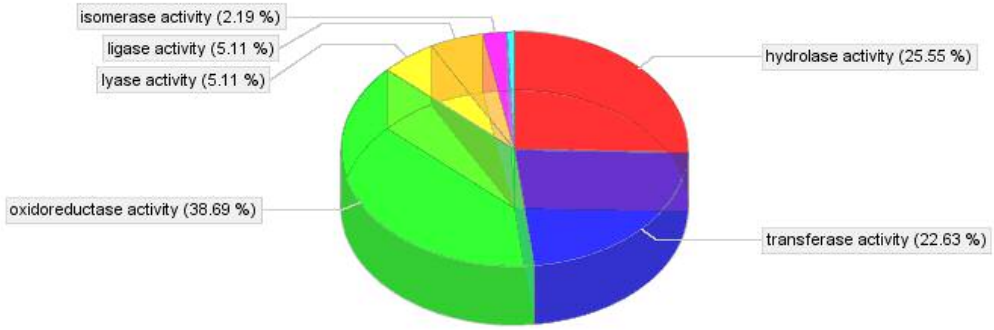
**Figure 8**

(a) Gene products annotated in binding category of molecular function according to Gene Ontology Consortium classifications (Sturn et al., 2002). There were 131 genes with 299 GO relations clustered in molecular function of binding with majority of gene products involved in interacting with nucleotides and nucleic acids.

(b) Gene products annotated in catalytic activity category of molecular function according to Gene Ontology Consortium classifications (Sturn et al., 2002). There were 131 genes with 304 GO relations clustered in molecular function of catalysis with majority of gene products in oxidoreductase, hydrolase and transferase activities.


vitamin binding (2.86 %)
cofactor binding (7.43 %)
tetrapyrrole binding (2.29 %)
lipid binding (5.14 %)
ion binding (20 %)
protein binding (14.29 %)
nucleic acid binding (20 %)
nucleotide binding (21.71 %)

**(a)**


isomerase activity (2.19 %)
ligase activity (5.11 %)
lyase activity (5.11 %)
oxidoreductase activity (38.69 %)
hydrolase activity (25.55 %)
transferase activity (22.63 %)

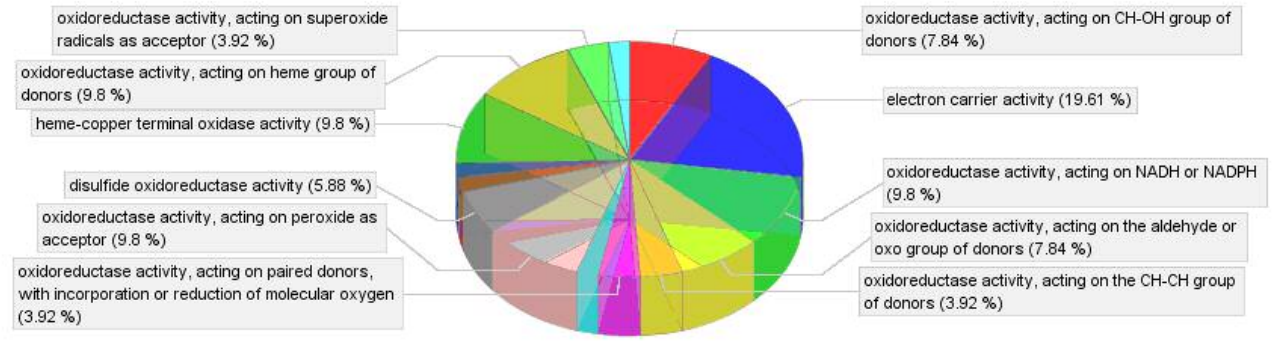
**(b)**

**Figure 9**

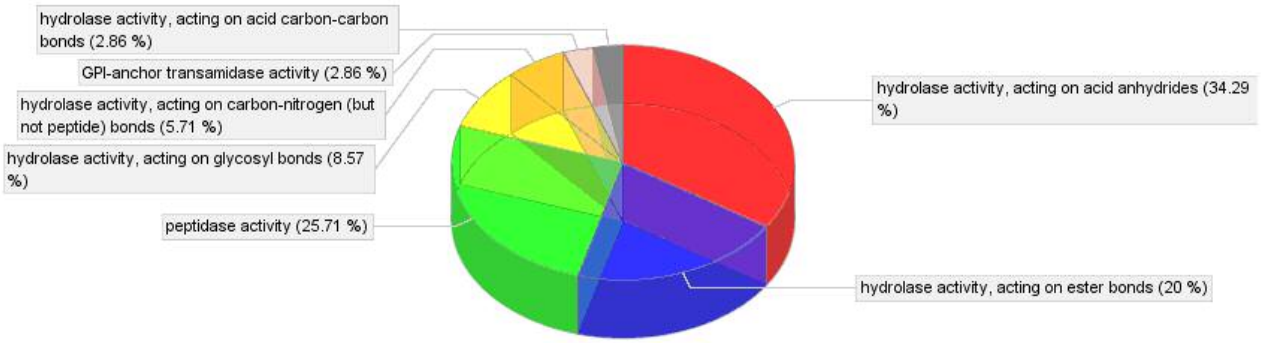
(a) Gene products annotated in oxidoreductase activity group of catalysis according to Gene Ontology Consortium classifications (Sturn et al., 2002). There were 37 genes with 75 GO relations clustered under oxidoreduction activity.

(b) Gene products annotated in hydrolase activity group of catalysis according to Gene Ontology Consortium classifications (Sturn et al., 2002). There were 35 genes with 61 GO relations clustered under hydrolase activity.

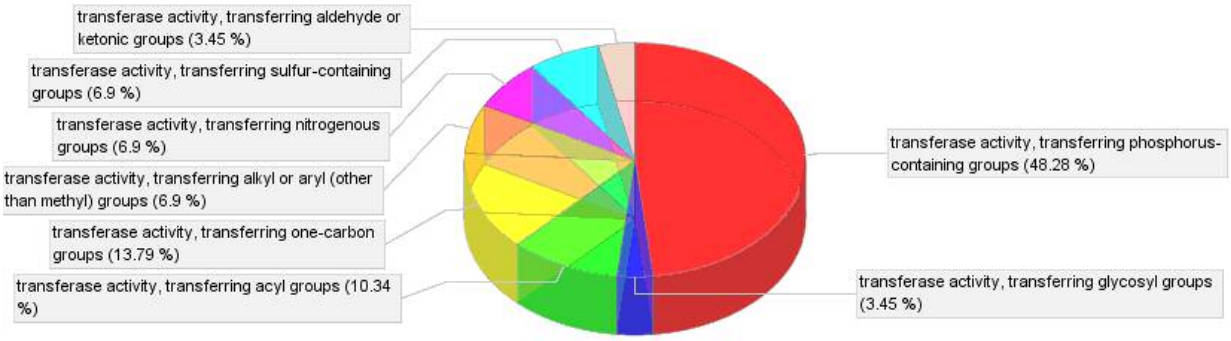
(c) Gene products annotated in transferase activity group of catalysis according to Gene Ontology Consortium classifications (Sturn et al., 2002). There were 28 genes with 73 GO relations clustered under transferase activity.


oxidoreductase activity, acting on superoxide radicals as acceptor (3.92 %)
oxidoreductase activity, acting on heme group of donors (9.8 %)
heme-copper terminal oxidase activity (9.8 %)
disulfide oxidoreductase activity (5.88 %)
oxidoreductase activity, acting on peroxide as acceptor (9.8 %)
oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen (3.92 %)
oxidoreductase activity, acting on CH-OH group of donors (7.84 %)
electron carrier activity (19.61 %)
oxidoreductase activity, acting on NADH or NADPH (9.8 %)
oxidoreductase activity, acting on the aldehyde or oxo group of donors (7.84 %)
oxidoreductase activity, acting on the CH-CH group of donors (3.92 %)

**(a)**


hydrolase activity, acting on acid carbon-carbon bonds (2.86 %)
GPI-anchor transamidase activity (2.86 %)
hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds (5.71 %)
hydrolase activity, acting on glycosyl bonds (8.57 %)
peptidase activity (25.71 %)
hydrolase activity, acting on acid anhydrides (34.29 %)
hydrolase activity, acting on ester bonds (20 %)

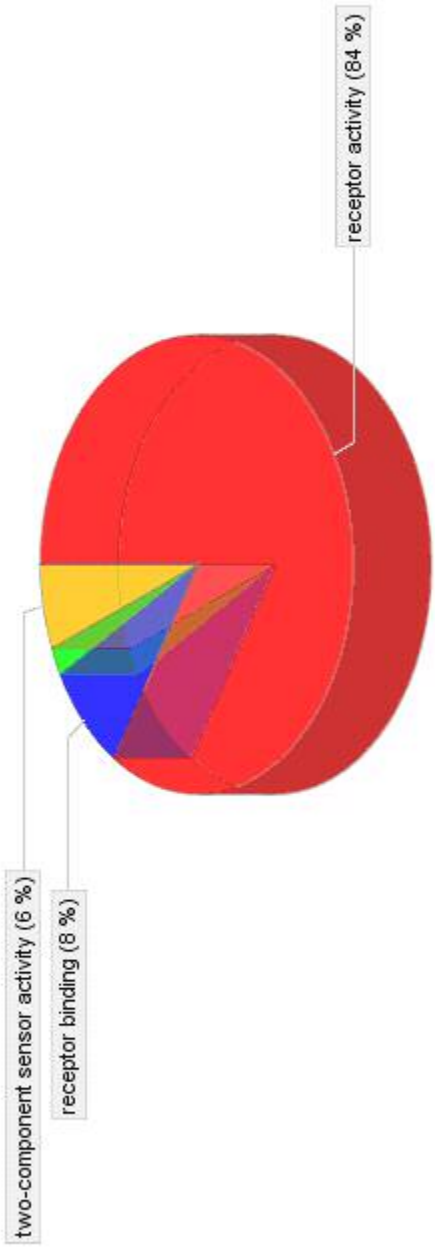
**(b)**


transferase activity, transferring aldehyde or ketonic groups (3.45 %)
transferase activity, transferring sulfur-containing groups (6.9 %)
transferase activity, transferring nitrogenous groups (6.9 %)
transferase activity, transferring alkyl or aryl (other than methyl) groups (6.9 %)
transferase activity, transferring one-carbon groups (13.79 %)
transferase activity, transferring acyl groups (10.34 %)
transferase activity, transferring phosphorus-containing groups (48.28 %)
transferase activity, transferring glycosyl groups (3.45 %)

**(c)**

**Figure 10**

Gene products annotated in signal transducer activity group of molecular function according to Gene Ontology Consortium classifications (Sturn et al., 2002). There were 49 genes with 78 GO relations clustered under molecular function of signal transduction.


receptor activity (84 %)
two-component sensor activity (6 %)
receptor binding (8 %)

## Discussion

The majority of clones in adipose tissue cDNA library had inserts between 0.5 -1.7kb sizes. Approximately 3% of the clones had short inserts. In an ideal cDNA library inserts have to be near full length copies of mRNA's from which they were derived and a minimum number of clones should have small inserts or inserts below 500bp (Invitrogen, 2003) and the porcine subcutaneous adipose tissue library fulfilled this criterion.

The cDNA library was non-normalized and as a consequence, redundancy of the library was high, with 65% of ESTs found more than once. About 4.3% of ESTs were observed more than 20 times. Carninci et al., (2000) classified mRNAs into 3 categories depending on their function as superprevalent, intermediate and rare. In a normal cell, 5-10 species of superprevalent mRNA comprise 20% of the mass of mRNA, 500-2000 species of mRNA comprise 40-60% of mass of mRNA and 10,000-20,000 of rare mRNA comprise 20-40% of mRNA mass (Carninci et al., 2000). The main problem associated with a non-normalized cDNA library is the low representation of these rarely expressed transcripts compared to intermediate and highly expressed transcripts (Carninci et al., 2000). The ability of identifying rare transcripts was hindered in the non-normalized adipose tissue cDNA library due redundancy of intermediate and highly expressed transcripts.

Kim et al., (2006) report 72% redundancy in a non-normalized porcine backfat cDNA library. They state that redundancy rate in the backfat tissue library is high when compared with other cDNA libraries constructed according to the same protocol. This low complexity of porcine adipose tissue gene expression profile was an unexpected finding because adipose tissue is recognized as a highly active metabolic and endocrine

organ (Ailhaud et al., 1992). Over the past decade, many studies are being done *in vivo* and *in vitro* to understand the complex nature of adipose tissue (Ailhaud et al., 1992). A bovine adipose tissue cDNA library developed in our laboratory according to same procedure showed only a 30% redundancy rate (unpublished data). Better understanding of complexity of gene expression profile in porcine subcutaneous adipose tissue could be obtained by comparing the diversity of the cDNA library with other porcine tissue cDNA libraries as well as bovine and human adipose tissue cDNA libraries. A skeletal muscle cDNA library was developed at Michigan State University from same animal population and the complexity and gene diversity of the two tissues could be analyzed by comparison of adipose tissue and skeletal muscle cDNA libraries.

Transcript for light chain (L) subunit of ferritin comprised the largest cluster with 101 clones and it was the most abundantly found insert in the library. Ferritin L chain was the third most abundant transcript found in bovine adipose tissue cDNA library. Festa et al. (2000) demonstrated that during differentiation of 3T3-L1 cells to adipocytes, there is an increase in mRNA of both H and L ferritin subunits. It is thought that ferritin involved in iron metabolism protects the adipose tissue from highly concentrated oxidable substrates and due to this reason it is reasonable that ferritin subunits are highly expressed in adipose tissue.

Transcripts for type I collagen pre-pro-alpha1 (I) chain, translation elongation factor 1α, ribosomal protein S2, vimentin, apolipoprotein E, glutathione peroxidase, hormone sensitive lipase, cytochorme c oxidase sub unit III, peptidylprolyl isomerase, ATPase subunit 6, galectin-I, fatty acid synthase, annexin A2, beta 2-microglobulin and thioredoxin interacting proteins were abundant in the library, with each of these

transcripts found in more that 30 clones. These genes play role in energy production, cell structure and in basic cellular functions. Translation elongation factor 1α, vimentin and annexin A2 transcripts were abundant in the backfat cDNA library constructed by Kim et al., (2006) and the bovine adipose tissue cDNA library (unpublished data), but there was a marked difference in other redundant transcripts between these libraries.

Only 1334 translated queries of contigs showed a significant homology to UniProt protein sequences. Sequences without a significant homology were either unknown proteins or those ESTs were not genuine protein coding sequences.

From the total of 1334 protein sequences only 330 were annotated according to GO consortium classifications. These 330 sequences contained high percentage of binding and catalytic proteins. High percentages of oxiodoreducatse, hydrolase and transferase activities indicate porcine subcutaneous adipose tissue to be an active energy utilizing tissue. In contrast to the results obtained by Kim et al. (2006), there was only contig with GO term of virion and viral life cycle associated genes in thecDNA library . Their library had a significant number of gene products classified in virion and viral life cycle group. Physiological status of the animals at the time of sample collection might be a reason for this difference between the two libraries. High percentage of gene products with receptor activity in signal transduction category indicates porcine subcutaneous adipose tissue to be a dynamic tissue that interacts with extracellular or intracellular messengers.

An adipose tissue specific porcine cDNA microarray will be printed in the future by selecting unique amplicons from the porcine subcutaneous tissue cDNA library. This would be very valuable for transcriptional profiling studies on $F_2$ animals that are extremes in fat deposition traits.

## Reference List


Adams,M.D. and J.M.Kelley. 1991. Complementary DNA Sequencing: Expressed Sequence Tags and Human Genome Project. *Science* **252**: 1651.

Ailhaud,G., P.Grimaldi, and R.Negrel. 1992. Cellular and Molecular Aspects of Adipose-Tissue Development. *Annual Review of Nutrition* **12**: 207-233.

Anderson,D.B. and R.G.Kauffman. 1973. Cellular and enzymatic changes in porcine adipose tissue during growth. *J. Lipid Res.* **14**: 160-168.

Archibald,A.L. and C.S.Haley. 1998. Genetic Linkage Maps. in *The Genetics of the pig* (ed. M F Rothschild and A Ruvsinksy), pp. 265-294. CAB International, New York.

Bates,R.O., C.W.Ernst, M.E.Doumit, D.B.Edwards, and N.E.Raney. 2003. Development of a Duroc x Pietrain Swine Resource Population.*NSIF conference proceedings.* Des Moines, IA.

Cagnazzo,M., M.F.W.te Pas, J.Priem, A.A.C.de Wit, M.H.Pool, R.Davoli, and V.Russo. 2006. Comparison of prenatal muscle tissue expression profiles of two pig breeds differing in muscle characteristics. *J. Anim Sci.* **84**: 1-10.

Canadian Centre for Swine Improvement. 2003. Duroc as a terminal sire line. *Review by the Canadian Centre for Swine Improvement*. Ontario, Canada.

Carninci,P., Y.Shibata, N.Hayatsu, Y.Sugahara, K.Shibata, M.Itoh, H.Konno, Y.Okazaki, M.Muramatsu, and Y.Hayashizaki. 2000. Normalization and Subtraction of Cap-Trapper-Selected cDNAs to Prepare Full-Length cDNA Libraries for Rapid Discovery of New Genes. *Genome Res.* **10**: 1617-1630.

Cepica S, Wolf J, Hojny J, Vacková I, and Schröffel J. 1995. Relations between genetic distance of parental pig breeds and heterozygosity of their F1 crosses measured by genetic markers. *Anim Genet* **26**: 135-140.

Edwards,D.B., R.O.Bates, and W.N.Osburn. 2003. Evaluation of Duroc- vs. Pietrain-sired pigs for carcass and meat quality measures. *J. Anim Sci.* **81**: 1895-1899.

Fahrenkrug,S.C., T.P.L.Smith, B.A.Freking, J.Cho, J.White, J.Vallet, T.Wise, G.Rohrer, G.Pertea, R.Sultana, J.Quackenbush, and J.W.Keele. 2002. Porcine gene discovery by normalized cDNA-library sequencing and EST cluster assembly. *Mammalian Genome* **13**: 475-478.

Festa,M., G.Ricciardelli, G.Mele, C.Pietropaolo, A.Ruffo, and A.Colonna. 2000. Overexpression of H Ferritin and Up-regulation of Iron Regulatory Protein Genes during Differentiation of 3T3-L1 Pre-adipocytes. *J. Biol. Chem.* **275**: 36708-36712.

Gregoire,F.M., C.M.Smas, and H.S.Sul. 1998. Understanding Adipocyte Differentiation. *Physiol. Rev.* **78**: 783-809.

Invitrogen. 2003. SuperScript™ Plasmid System with Gateway® Technology for cDNA Synthesis and Cloning, pp.3. Invitrogen Corporation, Carlsbad, CA.

Jones,G.F. 1998. Genetic Aspects of Domestication, Common Breeds and their Origin. in *The Genetics of the pig* (ed. M F Rothschild and A Ruvsinksy), pp. 17-51. CAB International, New York.

Karsenty,E., E.Barillot, G.Tosser-Klopp, Y.Lahbib-Mansais, D.Milan, F.Hatey, S.Cirera, M.Sawera, C.B.Jorgensen, B.Chowdhary, M.Fredholm, K.Wimmers, S.Ponsuksili, R.Davoli, L.Fontanesi, S.Braglia, P.Zambonelli, D.Bigi, S.Neuenschwander, and J.Gellin. 2003. The GENETPIG database: a tool for comparative mapping in pig (Sus scrofa). *Nucl. Acids Res.* **31**: 138-141.

Kim,T.H., N.S.Kim, D.Lim, K.T.Lee, J.H.Oh, H.S.Park, G.W.Jang, H.Y.Kim, M.Jeon, B.H.Choi, H.Y.Lee, H.Y.Chung, and H.Kim. 2006. Generation and analysis of large-scale expressed sequence tags (ESTs) from a full-length enriched cDNA library of porcine backfat tissue. *BMC Genomics* **7**: 36

Lahbib-Mansais, Y, G.Dalias, D.Milan , M.Yerle, A.Robic, G.Gyapay and J.Gellin. 1999. A successful strategy for comparative mapping with human ESTs: 65 new regional assignments in the pig. *Mammalian Genome* **10**: 145-153.

Nobis,W., X.Ren, S.P.Suchyta, T.R.Suchyta, A.J.Zanella, and P.M.Coussens. 2003. Development of a porcine brain cDNA library, EST database, and microarray resource. *Physiol. Genomics* **16**: 153-159.

Rink,A., E.M.Santschi, and C.W.Beattie. 2002. Normalized cDNA libraries from a porcine model of orthopedic implant-associated infection. *Mammalian Genome* **13**: 198-205.

Sellier,P. 1998. Genetics of Meat and Carcass Traits. in *The Genetics of the pig* (ed. M F Rothschild and A Ruvsinksy), pp. 463-510. CAB International , New York.

Sturn,A., J.Quackenbush, and Z.Trajanoski. 2002. Genesis: cluster analysis of microarray data. *Bioinformatics* **18**: 207-208.

Uenishi,H., T.Eguchi, K.Suzuki, T.Sawazaki, D.Toki, H.Shinkai, N.Okumura, N.Hamasima, and T.Awata. 2004. PEDE (Pig EST Data Explorer): construction of a database for ESTs derived from porcine full-length cDNA libraries. *Nucl. Acids Res.* **32**: D484-D488.

www.ansi.okstate.edu/breeds/swine. 1997. Breeds of livestock.

www.geneontology.org/. 2006. Gene Ontology

www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html. 2006. NCBI.

www.pir.uniprot.org/.2006. UniProt

Yao,J., P.Coussens, P.Saama, S.Suchyta, and C.Ernst. 2002. Generation of expressed sequence tags from a normalized porcine skeletal muscle cDNA library. *Animal Biotechnology* **13**: 211-222.

# CHAPTER 3

Identification and Segregation Analysis of SNPs

## Introduction

The cDNA library was constructed from RNA samples pooled from 9 animals that might harbor polymorphisms in the coding regions of the genome. Comparison of sequences within overlapping regions of clones in clusters of the cDNA library is a good method for identification of these coding SNPs. Increased redundancy of the library is a positive factor in SNP detection process because the chance of detecting a polymorphism is high with more transcripts per cluster.

Four common methods used for SNP detection are direct DNA sequencing, heteroduplex analysis, single strand conformation polymorphisms and variant detector arrays (Gray et al., 2000). Direct DNA sequencing is the most widely used method from above four because of its ease of application and the ability to provide information on location of the polymorphism as well as the DNA sequence flanking it(Nickerson et al., 1997).

One of the negative factors in direct sequencing is that, some times, it is difficult to distinguish heterozygote sites in sequence trace files due to variability in fluorescence signal and inconsistency of base calling (Nickerson et al., 1997). PolyPhred is a software that helped to overcome this pitfall of differentiating heterozygotes from homozygotes in sequence trace files and it can be applied for detection and scoring of SNPs (Nickerson et al., 1997). PolyPhred is a combination of two sequence analysis applications. The base calling program Phred was developed by Phil Green and Ewing in 1998 and the assembling program Phrap was developed by Phil Green in 1994 (Ewing and Green, 1998; Nickerson et al., 1997).

In PolyPhred program, Phred does the base-calling, quality checking and peak size information of the sequence trace files, and Phrap assembles the input to relevant contigs

while deriving a consensus sequence for each contig. A site of SNP appears in the sequence trace as two overlapping peaks with reduced peak heights, with area under the peak nearly being the same. The height of the peak in heterozygous is half of the height of peak in homozygous peak at the same position. During the analysis process the PolyPhred program identifies the SNP sites by comparing peaks of sequence traces with the consensus sequence and it assigns a score ranging from 0-99 (Nickerson et al., 1997). The quality score is divided into six point ranking systems and PolyPhred converts the scores into ranks and assigns color coded tags to the ranks which can be visually analyzed by the program Consed (Gordon et al., 1998).

When a polymorphism is located in a coding region of a gene it might be responsible for a phenotypic change, by creation of a genetic variant that changes the function or expression of a protein (Suh and Vijg, 2005). As the cDNA library was constructed from $F_2$ outliers for back fat thickness, if there is segregation of a SNP exclusively in one of the outlier groups, that SNP would be a valuable marker for identification of candidate gene for back fat thickness.

We identified 362 cSNPs in the cDNA library with Phred quality scores above 20. They were categorized as synonymous and non synonymous SNPs and the segregation pattern of some of the non synonymous SNPs in $F_0$ and $F_2$ population was studied. We have also initiated the mapping of genes with confirmed SNPs to porcine genome by RH mapping.

## Materials and Method

### Identification of SNPs from cDNA Library Sequence Trace Files

Chromatogram files of 13,958 contigs were transferred to a UNIX workstation and base-called with bioinformatics software Phred (version 0.020425). The trace files were, then assembled with Phrap (version 0.990319) and scanned by PolyPhred (version 5.0) program (Ewing et al., 1998; Ewing and Green, 1998; Nickerson et al., 1997). Results were visualized by the bioinformatics software Consed (version 12.0) (Gordon et al., 1998). Areas of sequence trace files that were flagged red in the Consed output were visually analyzed, and polymorphisms that have a Phred quality score above 20 were selected for further analysis.

BLAST searches were conducted against nr and dbEST databases at NCBIto determine the gene identities of consensus sequences of contigs with SNPs (Altschul et al., 1990). NCBI blastx search was used to identify synonymous and non synonymous cSNPs (Altschul et al., 1990).

### Segregation Analysis of the Identified Coding SNPs in $F_0$ Population

### PCR Amplification of Regions with Polymorphisms

Twenty three non synonymous cSNPs identified from the adipose tissue EST trace files were traced back to $F_0$ population. For this process, regions with polymorphisms were PCR amplified from genomic DNA of $F_0$ animals and the PCR products were sequenced. Primers were designed for PCR amplification of regions with SNPs by using the Primer3 software and IDT oligoanalyser (Steve and Skalets, 2006; www.idtdna.com/analyzer/Applications/OligoAnalyzer/, 2006).

As the pig genome is not fully sequenced, EST sequences were aligned with homologous regions of *Homo sapiens* and *Bos Taurus* genomes during the primer designing process (www.ncbi.nlm.nih.gov/projects/genome/guide/cow/, 2006; (www.ncbi.nlm.nih.gov/projects/genome/guide/human/, 2006).

Regions with SNPs were separately PCR amplified from genomic DNA of 4 Duroc and 15 Pietrain animals. Reaction cocktail of 20µl contained 9.55µl of distilled water, 2µl of 10x PCR buffer (Applied Biosystems, Foster city, CA), 1.2µl of 25mmol $MgCl_2$, 0.4µl of 10µmol dNTP, 0.8µl of 10µmol Forward Primer, 0.8µl of 10µmol Reverse Primer, 0.25µl of 5U/µl Taq DNA polymerase and 5µl of 2ng/µl genomic DNA (total of 10ng). PCR was performed in a MJ Dyad thermal cycler and cycling conditions were, 1 cycle of $95^0$C for 4 minutes followed by 32 cycles of $94^0$C for 30 seconds, annealing temperature for 30 seconds, $72^0$C for 30 seconds-1 minute for 32 cycles. Annealing temperature varied according to the Tm of the primers and the extension time was changed according to the product size (See table.1). Quality of the PCR products were visualized by electrophoresis of 10µl of PCR product on 1.5% agarose gel stained with 6µl of ethidium bromide (10mg/ml).

Sequencing was performed from pooled DNA samples because the major goal of the project is identification of informative markers and not rare polymorphisms. To increase the accuracy and intensity of the sequence reaction, a second PCR was carried out from the products of the first PCR.

2 µl of $1^{st}$ PCR product from the animals were pooled during the second PCR. There were three pools of DNA, with pool one consisted of female Pietrain animals 1 to 8, pool two with female Pietrain animals 9 to16 and third one with 4 duroc males. Second PCR

was performed by mixing 13.55µl of distilled water, 2µl of 10x PCR buffer (Applied Biosystems, Foster city, CA), 1.2µl of 25mmol $MgCl_2$, 0.4µl of 10µmol dNTP, 0.8µl of 10µmol Forward Primer, 0.8µl of 10µmol Reverse Primer, 0.25µl of 5U/µl Taq DNA Polymerase and 1µl of 1:100 dilution of 1st PCR product. PCR was performed in MJ Dyad thermal cycler with 1 cycle of $95^0$C for 4 min followed by 32 cycles of $94^0$C for 30 seconds, annealing temperature for 30 seconds, $72^0$C for 30 seconds-1 minute for 32 cycles. Annealing temperature varied with the pair of primers and the extension time was changed according to the product size (See table.3). Quality of the PCR products were visualized by electrophoresing 10µl of the product in 1.5% agarose gel stained with 6µl of ethidium bromide (10mg/ml).

**Sequencing of Regions with the SNP**

Before performing the sequence PCR reaction, excess dNTPs and unincorporated primers were removed from the PCR products. This was achieved by adding 0.5µl of 1U/µl shrimp Alkaline Phosphatase and 10U/µl Exonulcease -1 (USB, Cleveland, OH, USA) to 4µl of PCR product and incubating it at $37^0$C for 30minutes and $85^0$C for 15 minutes in MJ Dyad PCR thermal cylcer.

Afterward, sequence PCR (Sanger et al., 1977) of 10µl reaction was carried out by adding 3µl of distilled water, 2µl of BigDye terminator mix (Applied Biosystems, Foster City, CA, USA), 1µl of 10µmol sequencing primer (see table.3), 2µl of 5x sequencing buffer (400mM Tris pH 9, 10mM $MgCl_2$), 2µl of PCR product and thermal cycling in MJ Dyad PCR thermal cylcer for 1 cycle of $95^0$C for 30 seconds followed by 35 cycles of $96^0$C for 10 seconds, $50^0$C for 5 seconds and $60^0$C for 4 minutes.

The resulting PCR product was cleaned up of unincorporated dye particles by adding 10µl of distilled water, 2µl of 3M Na acetate and 57µl of 100% ethanol into 10µl of sequence PCR product and precipitating the DNA by centrifuging at 4000rpm for 30 minutes at room temperature. The precipitate was then washed by adding 70µl of 70% ethanol and centrifuged at 4000rpm for 10 minutes at room temperature. Afterwards ethanol was discarded, the tubes were air dried and the pellet was resuspended in 4µl of sequence gel loading dye. Just before loading to thepolyacrylamide sequence gel , the mixture was denatured to disrupt the secondary structures of DNA by heating at $95^0$C for 3 minutes and then immediately transferred to ice.

Sequencing was carried out in ABI PRISM® 377 DNA Sequencer (Applied Biosystems, Foster City, CA, USA) with reaction based on dye labeled terminator chemistry (Sanger et al., 1977; Prober et al., 1987 ). After the laser excitement and detection of fluorescent dyes by the scanner the, gel images were analyzed and trace files were created in a MacG3 computer. Obtained trace files were uploaded and run in the Polyphred program (version 5.0) and the output was graphically visualized by Consed (version 12.0).

**Segregation Analysis of Identified Polymorphisms in 9 $F_2$ Animals**

Genomic DNA was obtained from 9 $F_2$ animals that contributed for subcutaneous adipose tissue cDNA library. Regions with the confirmed non synonymous SNPs were separately PCR amplified using the same primers and conditions described in the previous section (see table 1 and 2). PCR products were sequenced in ABI PRISM® 377 DNA Sequencer according to the procedure described in the previous section. Trace files obtained were uploaded and analyzed in Polyphred program (version 5.0) and the output was graphically visualized by using Consed (version 12.0).

After the confirmation of polymorphisms in the $F_0$ and $F_2$ populations, information on the regions of amino acid substitution were extrapolated by checking for homologous regions in NCBI Conserved Domain Database (CDD) (Marchler-Bauer et al., 2005). Putative locations of genes with confirmed SNPs were determined according to INRA human-pig comparative map (Milan et al., 2000).

**Radiation Hybrid Mapping of Genes with Confirmed Non Synonymous SNPs**

Radiation Hybrid (RH) mapping is a rapid and powerful approach to quickly localize a sequence to a genome (www.mun.ca/biology/scarr/4241_Mar05,07/GG10_RHM.jpg ,2005). During the generation of clones for RH panel, thymidine kinase positive ($TK^+$) swine fibroblast cells are irradiated and broken in to fragments. These chromosomal fragments are fused with chromosomes in $TK^-$ Chinese hamster cell lines. Cells with porcine chromosome fragments are separated by selecting cells that express thymidine kinase and Then ninetyor h undred and eighteen clone RH panel is constructed by selection of cells that express thymidine kinase due to $TK^+$ activity of incorporated porcine chromosomal fragments. Cell lines included in the RH panel should have single insert of sufficient size with identifiable banding pattern (www.mun.ca/biology/scarr/4241_Mar05,07/GG10_RHM.jpg, 2005).

Primers were developed and optimized for radiation hybrid mapping of genes with non synonymous SNPs to the porcine genome. Primers designed for RH mapping process should amplify the product only from porcine genomic DNA and the primers must not amplify a product from hamster genomic DNA. Because, primers developed for three sets of genes amplified a product in hamster genomic DNA and we were unable to map

these three genes to the porcine genome. Other six genes with confirmed non synonymous SNPs were mapped to the porcine genome by using INRA-Minnesota 7000rads radiation hybrid panel (IMpRH). Radiation hybrid mapping was performed in 90 clone format hamster-porcine hybrid cell lines (Rachel et al., 1999). Ninety clones were selected according to IMpRH instructions (Milan et al., 2000) and each clone was used as a template in a PCR reaction performed in the MJ Dyad thermal cycler. Twenty micro liter PCR reaction contained 13.55µl of distilled water, 2µl of 10x PCR buffer (Applied Biosystems, Foster city, CA) 1.2µl of 25mmol $MgCl_2$, 0.4µl of 10µmol dNTP, 0.8µl of 10µmol Forward Primer, 0.8µl of 10µmol Reverse Primer , 0.25µl of 5U/µl Taq DNA Polymerase and 1µl of template DNA. PCR reactions were also performed with porcine genomic DNA as a positive control and hamster DNA as a negative control. Cycling conditions were one cycle of 1 minute at 95ºC, 30 seconds at annealing temperature, and 30 seconds at 72 ºC followed by 34 additional cycles of 30 seconds at 94ºC, 30 seconds at annealing temperature and 30 seconds at 72ºC (see Table 4). PCR products were visualized by electrophoresis in 2% agarose gels with 500ng/ml of ethidium bromide. Experiments were repeated once again and each clone was scored twice. Scoring, data analysis and chromosomal assignment were performed using the software available at IMpRH server (Milan et al., 2000).

**Table-1.**Primers and conditions used for 1st PCR amplification of regions with SNPs

| Contig | Forward Primer | Reverse Primer | Annealing temp ($C^0$) | Extension time | Product size |
|---|---|---|---|---|---|
| 2009 | GAACGCCTGTCCTCCTACTTCA | CAGCGGTGTGGTTGTAGAAA | 62 | 40 sec | 312bp |
| 1994 | GCTCTCAGACCCCTTGTGAA | GCTCTCTCGACGATGGCTTT | 51.2 | 30 sec | 338bp |
| 1991 | TCATCGGCTCAACAAACAT | GGAATTAATAGGGCGGGT | 51.2 | 30 sec | 184bp |
| 1990 | CTCCTCCGACTCCTCCTTCT | ATCGGCTCAACAAGGTCATC | 58 | 30 sec | 203bp |
| 1989 | TCGACTCCTGGTTCAATACT | TTTGGAGAAGGTGTGTGCTG | 62 | 30 sec | 290bp |
| 1952 | CGTTTGTCATTACCAAAAAGAG | CATCCGGTTTAGTCATCCAC | 52.3 | 30 sec | 198bp |
| 1935 | GCACCAAGACCGTGAAGAAG | GGGCACGTAGTTGTCTCTCC | 52.3 | 30 sec | 244bp |
| 1927 | GATGGAACCCTTCGCCTCT | TAATGGGATTGCTGCTGTTG | 60 | 30 sec | 480bp |
| 1898 | AGGTGCATTTATTTCCTCCTG | CATCTCATACCCGCTCCATT | 56 | 30 sec | 247bp |
| 1859 | GTTTATGCAACCAGGGCAGA | GACGACGTGAAGGAGCAGA | 62 | 30 sec | 368bp |
| 1854 | TTCTGGTGCTGGAGGAAGAT` | GGCTTATGGGAGAGGGAGAG | 62 | 30 sec | 236bp |
| 1685 | GAACGACACAGTCACTATCC | CCTGTGTTTGGGCTCATT | 65 | 30 sec | 268bp |
| 1639 | CCAAAAGCCTGCCCAATG | CTCGCTCAGACTTGCTGCT | 62 | 30 sec | 137bp |
| 1602 | AGAGGTTGACCGTGGGAG | ACAAACAACCGCCCGACT | 62 | 30 sec | 206bp |
| 1549 | CAAGAACTGCCTGCTGACC | CCATCGTTTTTGCCATCC | 62 | 30 sec | 243bp |
| 1424 | CCCCACTCCCTGAGCTATTT | GCCTCGCTCTGGTTGTAGTAG | 52 | 30 sec | 263bp |
| 1310 | TGACGATGTGGGTGTTTGAG | TCACAGATCTTGCCGATGAA | 60 | 1 min | 1080bp |
| 1190 | TCTTTTCGTCGCTTGCTG | GATGTGGGAGCTATTGTGACC | 62 | 30 sec | 119bp |
| 1184a | GGATGAAGACGGGAAAGGAC | TACAGGACCCAGGAGAGGC | 62 | 30 sec | 371bp |
| 1184b | GCCTCTCCTGGGTCCTGTA | TAGGCTCACGGTCCCTCA | 62 | 30 sec | 96bp |
| 994 | GCTCATCTGCAAACGTCCTG | CACTGTTGCTGCTGTCCAAG | 59.2 | 30 sec | 219bp |
| 789 | TCCAAATTCTCTACAACGAACA | GGAGATACAGGCACCGAGTT | 52.3 | 30 sec | 184bp |

**Table-2.**Primers and conditions used for 2nd PCR amplification of regions with SNPs

| Contig | Forward Primer | Reverse Primer | Annealing temp (C') | Extension time | Product Size |
|---|---|---|---|---|---|
| 2009 | GAACGCCTGTCCTCCTACTTCA | CTCCATTACTGCCGCTTG | 60 | 30 sec | 228bp |
| 1994 | AAGTTCATCCGGCACCAGT | GCTCTCTCGACGATGGCTTT | 51.2 | 30 sec | 291bp |
| 1991 | TCATCGGCTCAACAAACAT | GGAATTAATAGGGCGGGT | 51.2 | 30 sec | 184bp |
| 1990 | CTCCTCCGACTCCTCCTTCT | ATCGGCTCAACAAGGTCATC | 58 | 30 sec | 203bp |
| 1989 | TCGACTCCTGGTTCAATACT | TTTGGAGAAGGTGTGTGCTG | 62 | 30 sec | 290bp |
| 1952 | CGTTTGTCATTACCAAAAAGAG | CATCCGGTTTAGTCATCCAC | 52.3 | 30 sec | 198bp |
| 1935 | GCACCAAGACCGTGAAGAAG | GGGCACGTAGTTGTCTCTCC | 52.3 | 30 sec | 244bp |
| 1927 | GATGGAACCCTTCGCCTCT | TAATGGGATTGCTGCTGTTG | 60 | 30 sec | 480bp |
| 1898 | AGGTGCATTTATTTCCTCCTG | CATCTCATACCCGCTCCATT | 50 | 30 sec | 247bp |
| 1859 | GTTTATGCAACCAGGGCAGA | GACGACGTGAAGGAGCAGA | 62 | 30 sec | 368bp |
| 1854 | TTCTGGTGCTGGAGGAAGAT` | GGCTTATGGGAGAGGGAGAG | 62 | 30 sec | 236bp |
| 1685 | GAACGACACAGTCACTATCC | GTCTTGCCACCACCAAAATG | 62.9 | 30 sec | 204bp |
| 1639 | CCAAAAGCCTGCCCAATG | CTCGCTCAGACTTGCTGCT | 62 | 30 sec | 137bp |
| 1549 | CAAGAACTGCCTGCTGACC | CCATCGTTTTTGCCATCC | 62 | 30 sec | 243bp |
| 1190 | TCTTTTCGTCGCTTGCTG | GATGTGGGAGCTATTGTGACC | 62 | 30 sec | 119bp |
| 1184a | GGATGAAGACGGGAAAGGAC | TACAGGACCCAGGAGAGGC | 62 | 30 sec | 371bp |
| 1184b | GCCTCTCCTGGGTCCTGTA | TAGGCTCACGGTCCCTCA | 62 | 30 sec | 96bp |
| 994 | GCTCATCTGCAAACGTCCTG | AGCGACCGAGCCCTTCTT | 58 | 30 sec | 168bp |
| 789 | TCCAAATTCTCTACAACGAACA | GGAGATACAGGCACCGAGTT | 52.3 | 30 sec | 184bp |

**Table-3.**Primers used for sequence analysis of SNPs

| Contig | Sequencing Primer |
|---|---|
| 2009 | GAACGCCTGTCCTCCTACTTCA |
| 1994 | AAGTTCATCCGGCACCAGT |
| 1991 | TCATCGGCTCAACAAACAT |
| 1990 | ATCGGCTCAACAAGGTCATC |
| 1989 | TTTGGAGAAGGTGTGTGCTG |
| 1952 | CATCCGGTTTAGTCATCCAC |
| 1935 | GCACCAAGACCGTGAAGAAG |
| 1927 | GATGGAACCCTTCGCCTCT |
| 1898 | CATCTCATACCCGCTCCATT |
| 1859 | GACGACGTGAAGGAGCAGA |
| 1854 | GGCTTATGGGAGAGGGAGAG |
| 1685 | GTCTTGCCACCACCAAAATG |
| 1639 | CTCGCTCAGACTTGCTGCT |
| 1602 | GTCCCCAGTCGCTTCTCTG |
| 1549 | CCATCGTTTTTGCCATCC |
| 1424 | CTCCGTAAGTCTGTGCGTTG |
| 1310 | TGACGATGTGGGTGTTTGAG |
| 1190 | TCTTTTCGTCGCTTGCTG |
| 1184a | TACAGGACCCAGGAGAGGC |
| 1184b | TAGGCTCACGGTCCCTCA |
| 994 | AGCGACCGAGCCCTTCTT |
| 789 | GGAGATACAGGCACCGAGTT |

**Table-4 PCR** Primers and PCR conditions used for Radiation Hybrid panel mapping

| GenBank Accession number | Forward Primer | Reverse Primer | Annealing temp ($C^0$) | Extension time | Product size |
|---|---|---|---|---|---|
| EB422656 | TTCTGGTGCTGGAGG AAGAT` | GGCTTATGGGAGAGG GAGAG | 62 | 30 sec | 200bp |
| EB412932 | GAACGACACAGTCAC TATCC | CCTGTGTTTGGGCTC ATT | 65 | 30 sec | 268bp |
| EB418674 | TTAGTCAGGGTTCCC AAGACTTCC | ATCTGACAGTCCTTT CCCGTCTTC | 62 | 30 sec | 371bp |
| EB412767 | GCTCATCTGCAAACG TCCTG | CACTGTTGCTGCTGT CCAAG | 59.2 | 30 sec | 219bp |
| EB419129 | AACCCAGAGGCATTG ACAAC | ATGGCTTTGCGGT TCTTG | 55 | 30 sec | 223bp |
| EB423906 | GTGACTTCAACCT GGGCTGT | TTTGGAGAAGGTGT GTGCTG | 62 | 30sec | 202bp |

## Results

There were a total of 362 SNPs above Phred quality score of 20 in the porcine subcutaneous adipose tissue library. These polymorphisms were dispersed between 264 of G/A (C/T), 39 of C/G, 21 of G/T, 21 of A/C and 17 of T/A polymorphisms (see Figure 11a). According to the type of nucleotide substitution, these SNPs could be categorized into 264 transitions and 98 transversions (see Figure 11b). Putative identifications of genes with the SNPs were performed by searching in NCBI non redundant and EST databases and results with E values less than $1 \times 10^{-22}$ were considered as a significant match. Identified SNPs were dispersed among 243 unique sequences (see Appendix I). EST for pig poly-ubiquitin mRNA had 12 SNPs and that was the highest number of SNPs per consensus sequence. MHC class I antigen EST which had 9 SNPs was the second highest. NCBI blastx search identified 48 SNPs from the total of 362 cSNPs as non synonymous SNPs (see Appendix II).

From the 23 selected non synonymous SNPs that were further analyzed 11 were confirmed by sequencing of genomic DNA from 9 $F_2$ animals and 19 $F_0$ animals (see Table 5).

Confirmed SNPs were analyzed in detail for location of the substituted amino acid in the encoded protein and for the location of the gene with the SNP in the porcine genome. Detailed analyses of each of the confirmed SNPs are given below.

Consensus sequence of contig 994 (GenBank accession number - EB412767) was identified as RTP801, transcript variant and an A/T polymorphism was identified at 273$^{rd}$ nucleotide. Genomic DNA sequence trace files of pooled Pietrain animals had both A and T at 273$^{rd}$ nuceltotide whilst in Duroc animals' allele A was detected. One animal

with relatively low back fat percentage and one animal with relatively intermediate back fat percentage in $F_2$ population were heterozygous for alleles A and T. Others were homozygous for allele T at 273rd nucleotide.

Replacement of T with A at 273rd nucleotide created an amino acid substitution from arginine to threonine. Active motif of the protein encoded by this transcript could not be found in the NCBI conserved domain data base. In humans the RTP801 gene is mapped to human chromosome 10q24.33 ( Corradetti et al., 2005), and based on human –pig comparative map it was postulated that the pig RTP801 gene should localize to porcine chromosome 14 (Goureau et al. 1996).

Consensus sequence of contig 1854 (GenBank accession number - EB422656) was identified as similar to Cell Death Activator CIDE-3 (Cell death-inducing DFFA-like effecter protein C or Fat-specific protein FSP27 homolog), transcript, variant 6 mRNA. Single nucleotide polymorphism of allele C to G was detected at the 555th nucleotide. Presence of this polymorphism was confirmed by genomic sequences of $F_0$ and $F_2$ animals. Pietrain animals were heterozygous for alleles C and G and Duroc animals were homozygous for the allele C at 555th nucleotide. This polymorphism segregated in all three $F_2$ animal groups.

This non-synonymous SNPof C to G created an amino acid substitution from threonine to serine. It was found by searching in the conserved domain data base that this change occurs at the 132nd amino acid encoded by the transcript and this site is located 6 amino acids away from the conserved domain region of CIDE-3 protein. CIDE-3 protein is located in human chromosome 3p25 (Liang L, 2003) and based on human –pig

comparative map it was postulated that CIDE-3 gene is in porcine chromosome 13 (Goureau et al. 1996).

Consensus sequence of contig 1310 (GenBank accession number - EB421233) was identified as transcript for glycerol-3-phosphate dehydrogenase 1 (soluble GPD1). A SNP of allele G to A was identified at 258$^{th}$ nucleotide. The polymorphism was confirmed by sequencing of genomic DNA from $F_0$ and $F_2$ populations and both alleles G and A were found to be present in Duroc and Pietrain population. This polymorphism segregated in all three $F_2$ animal groups.

Replacement of G with A at nucleotide 258 of contig 1310 created a substitution in 65$^{th}$ amino acid in glycerol-3-phosphate dehydrogenase 1 from iso-leucine to valine. The gene for glycerol-3-phosphate dehydrogenase 1 protein is located in human chromosome 12q12-q13 ( Prasad et al., 1997) and based on human –pig comparative map It was postulated that this gene is located in porcine chromosome 5 (Goureau et al. 1996).

Consensus sequence of contig 1184 (GenBank accession number - EB422656) was identified as similar to cytochrome P450 4F6 isoform 2. A SNP of allele G to T was detected at 288$^{th}$ nucleotide. Pooled samples of both Duroc and Pietrain animals showed alleles G and T at nucleotide 288. All nine animals checked in $F_2$ population were heterozygous for this polymorphism.

Replacement of G with T at 288$^{th}$ nucleotide contig 1184 created an amino acid substitution in 92$^{nd}$ amino acid in the encoded protein from glutamic acid to aspartic acid. Gene for Cytochrome P450 4F6 protein is located in human chromosome 19q13.2 (Fernandez-Salguero et al., 1995) and based on human –pig comparative map it was

postulated that P450 4F6 isoform 2 gene is in porcine chromosome 2 (Goureau et al. 1996).

Consensus sequence of contig 1994 (GenBank accession number - EB419511) was identified as similar to 60S ribosomal protein L32. A SNP of allele G to A at 177$^{th}$ nucleotide and a SNP of allele C to G at 286$^{th}$ nucleotide was detected in contig 1994. Presences of these polymorphisms were confirmed by genomic sequences of $F_0$ and $F_2$ animals. Pooled samples of both Duroc and Pietrain animals showed these two SNPs and the two polymorphisms segregated in all 3 $F_2$ animal groups. Replacement of G with A at 177$^{th}$ nucleotide of contig 1994 created an amino acid substitution of asparagine to aspartic acid in 52$^{nd}$ amino acid and replacement of C with G at 286$^{th}$ nucleotide creates an amino acid substitution of valine to leucine in 88$^{th}$ amino acid of 60S ribosomal protein L32. Amino acid changes created by these polymorphisms are located in the conserved domain region of the protein. Gene for 60S ribosomal protein L32 is located in human chromosome 3 (Kenmochi et al., 1998) and based on human –pig comparative map it was postulated that 60S ribosomal protein L32 gene is located in porcine chromosome 13 (Goureau et al. 1996).

Consensus sequence of contig 1989 (GenBank accession number- EB423506) was identified as similar to 40S ribosomal protein S20. A SNP of allele G to A was detected at 251$^{st}$ nucleotide. Presence of this polymorphism was confirmed by genomic sequences of $F_0$ and $F_2$ animals. Pooled samples of both Duroc and Pietrain animals showed alleles G and A at 251$^{st}$ nucleotide. All 9 animals checked in $F_2$ population were heterozygous for the alleles G and A. Replacement of G with A at 251$^{st}$nucleotide created an amino acid substitution from leucine to valine at 60$^{th}$ amino acid in ribosomal protein S20. The

amino acid change created by the polymorphism is located in the conserved domain region of the protein. The gene for 40S ribosomal protein S20 is in human chromosome 8 (Kenmochi et al., 1998) and based on human –pig comparative map it was postulated that 40S ribosomal protein S20 gene is located in porcine chromosome 4 (Goureau et al. 1996).

Consensus sequence of contig 1685 (GenBank accession number- EB412932) was identified as similar solute carrier family 12, member 3 mRNA. A SNP of allele C to T was detected at 142$^{nd}$ nucleotide. Presence of this polymorphism was confirmed by genomic sequences of $F_0$ and $F_2$ animals. Pooled samples of Pietrain animals showed both alleles C and T at 142$^{nd}$ nucleotide. Duroc animals were homozygous for allele T. One animal with intermediate thick back fat percentage in $F_2$ population was heterozygous for alleles C and T while others were homozygous for allele T at the142$^{nd}$ nucleotide.

This polymorphism of T to C created an amino acid substitution and signal for praline instead of leucine as the 28$^{th}$ amino acid in the encoded protein. The amino acid change created by the polymorphism is located in the conserved domain region of the protein. Gene for solute carrier family 12, member 3 is located in human chromosome 16 (Mastroianni et al., 1996)and based on human –pig comparative map it was postulated that solute carrier family 12, member 3 gene is located in porcine chromosome 6 (Goureau et al. 1996).

Consensus sequence of contig 1927 (GenBank accession number EB423721) was identified as transcript similar to mRNA encoding G-beta like protein. A SNP of allele A to C was detected at the 522$^{nd}$ nucleotide. Presence of this polymorphism was confirmed

by genomic sequences of $F_0$ and $F_2$ outliers. One pooled sample of eight Pietrain animals showed alleles A and C at the 522$^{nd}$ nucleotide. Other pooled samples of Pietrain and Duroc animals had only the alleles A at this position. Only one animal from the $F_2$ population was heterozygous for the alleles A and C. All others were homozygous for allele A.

Replacement of A with C at522 $^{nd}$ nucleotide of this gene created an amino acid substitution of tyrosine to serine at the 170$^{th}$ amino acid in WD40 domain of the encoded protein. The amino acid change created by the polymorphism is located in the conserved domain region of the protein. The gene for G-beta like protein is located in human chromosome 5 (Treacher Collins Syndrome Collaborative Group, 1996) and based on human –pig comparative map it was postulated that this gene is located in porcine chromosome 2 (Goureau et al. 1996).

Consensus sequence of contig 1549 (GenBank accession number - EB424152) was identified as transcript similar to thyroid hormone-responsive protein (THRSP) mRNA. A SNP of allele T to A was detected at 149$^{th}$ nucleotide. Presence of this polymorphism was confirmed by genomic sequences of $F_0$ and $F_2$ outliers. Pooled samples of both Duroc and Pietrain animals showed alleles T and A at 149$^{th}$ nucleotide.

This polymorphism segregated in all three $F_2$ animal groups. This SNP created an amino acid substitution from glutamic acid to aspartic acid at the 39$^{th}$ amino acid of thyroid hormone-responsive protein and region with this substitution is located in a conserved domain of the protein. Gene for thyroid hormone-responsive protein is located in human chromosome 11q13.5 (Liu and Towle,1994) and based on human – pig comparative map it was postulated that this gene is located in porcine chromosome 2 (Goureau et al. 1996).

Consensus sequence of contig 1991 (GenBank accession number - EB421930) was identified as *Sus scrofa* mitochondrial ATPase 6 mRNA, L transcript. A polymorphism of C to T was identified at the 438$^{th}$ nucleotide. This is homologous to the 9474$^{th}$ nucleotide of *Sus scrofa* mitochondrion, complete genome. In Pietrain animals pooled genomic DNA sequence trace files polymorphism C was dominant with a higher peak compared to polymorphism T. Only the nucleotide T could be detected in genomic DNA sequence trace files of Duroc animals. Eight $F_2$ animals had nucleotide C whilst one relatively fat $F_2$ animal had nucleotide T at the 9474$^{th}$ position in mitochondrial genome. This SNP created an amino acid substitution from histidine to tyrosine at 119$^{th}$ amino acid of protein ATPase subunit 6. The amino acid change is located in the conserved domain region of the protein ATPase 6.

**RH mapping results**

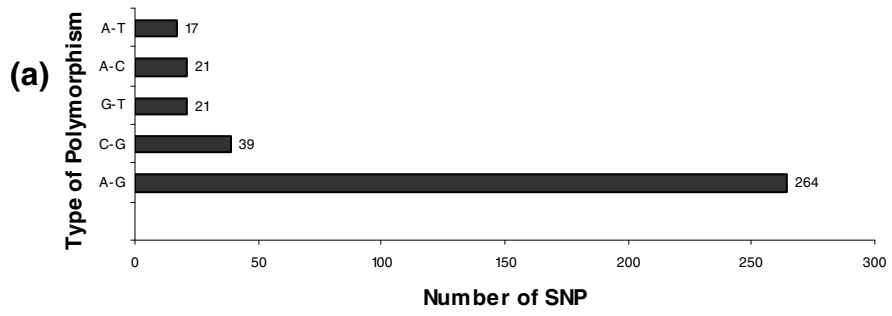
PCR reactions designed for Glycerol-3-phosphate dehydrogenase 1, G-beta like protein and thyroid hormone responsive protein amplified a product in hamster genomic DNA. Even though the PCR reaction performed for 60S ribosomal protein L32 and 40S ribosomal protein S20 did not amplify hamster genomic DNA, a significant Logarithm of Odds (LOD) value could not be obtained after RH map scoring process.

For each gene, PCR results visualized in the agarose gel were scored twice, according to the IMpRH mapping tool instructions as positive, negative and doubtful. Representative PCR panels for four successfully mapped genes are shown in Figures 12, 14, 16 and 18. During the PCR scoring process, positive results in both PCRs were marked (1), doubtful results in both PCRs were marked (?) and negative results in both PCRs were marked as

(0). Obtained scores (see Tables 6, 8, 10, 12 ) were uploaded to IMpRH server (imprh.toulouse.inra.fr/, 2006) and two point analysis was performed to identify an already mapped marker that is closest to the gene of interest. CIDE-3 gene showed linkage with LOD of 5.57 to marker S0223 in porcine chromosome 13 (see Table-7). Location of marker S0322 in 90 clone map by Hawken et al. (1999) is given in Figure 13. Solute carrier family 12, member 3 transcript showed linkage with LOD of 5.45 to marker SWR726 in porcine chromosome 6 (see Table-9). Location of marker SWR726 in 90 clone map by Hawken et al. (1999) is given in Figure 15. Cytochrome P450 4F6 isoform 2 transcript showed linkage with LOD of 8.31 to marker SW834 in porcine chromosome 2 (see Table-11). Location of marker SW834 in 90 clone map by Hawken et al. (1999) is given in Figure 17. RTP801 gene showed linkage with LOD of 6.44 to marker SW1536 in porcine chromosome 14 (see Table-13). Location of marker SW1536 in 90 clone map by Hawken et al. (1999) is given in Figure 19.

**Figure 11a** – Graphical output of type of coding SNPs identified in porcine adipose tissue cDNA library

**Figure 11b** – Graphical output of coding SNPs identified in porcine adipose tissue library categorized according to the substitution type


(a)
Type of Polymorphism
A-T
A-C
G-T
C-G
A-G
17
21
21
39
264
0
50
100
150
200
250
300
Number of SNP

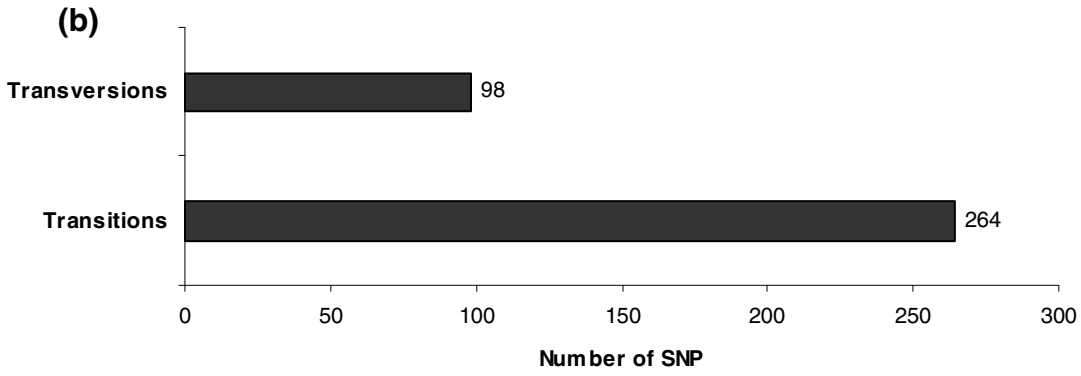
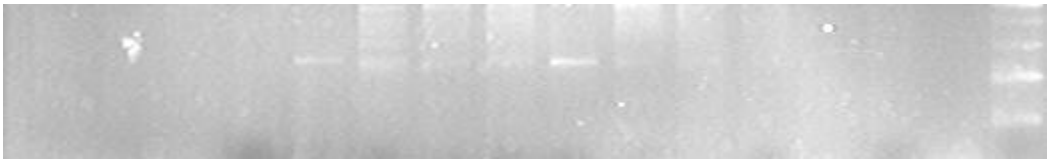

(b)
Transversions
Transitions
98
264
0
50
100
150
200
250
300
Number of SNP

**Table-5** Summary of confirmed SNPs and their segregation in $F_0$ and $F_2$ population

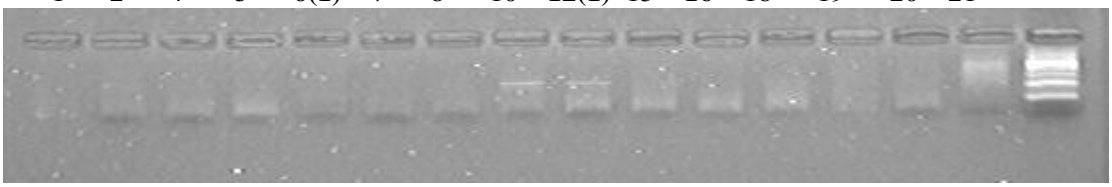
| GenBank Accession Number | Contig | $F_0$ population | $F_2$ population |
|---|---|---|---|
| EB419511 | 1994 | SNP in Duroc and Pietrain | SNP in all three groups |
| EB419511 | 1994 | SNP in Duroc and Pietrain | SNP in one animal |
| EB421930 | 1991 | SNP in Pietrain. | C in 8 animals, T in fat animal |
| EB423506 | 1989 | SNP in Duroc and Pietrain | SNP in all 9 animals |
| EB423721 | 1927 | SNP in Pietrain | SNP in one animal |
| EB422656 | 1854 | SNP in Pietrain | SNP in all three groups |
| EB412932 | 1685 | SNP in Pietrain | SNP in one animal |
| EB424152 | 1549 | SNP in Pietrain and Duroc | SNP in all three groups |
| EB421233 | 1310 | SNP Duroc and Pietrain | SNP in all three groups |
| EB418674 | 1184 | SNP Duroc and Pietrain | SNP in all 9 animals |
| EB412767 | 994 | SNP in Pietrain | SNP in all three groups |

**Figure 12**

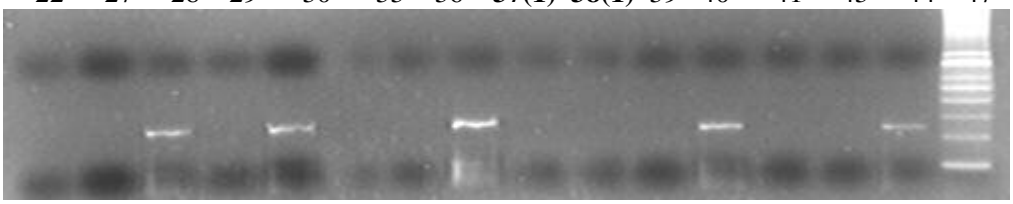
A representative PCR panel showing results obtained by scoring Pig-hamster 90 clone radiation hybrid panel for the gene CIDE-3 (imprh.toulouse.inra.fr/, 2006)



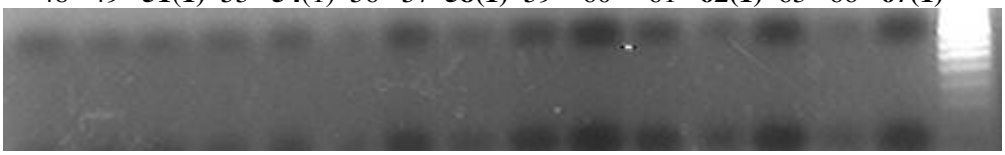
1 2 4 5 **6(1)** 7 8 10 **12(1)** 13 **16** 18 19 20 21



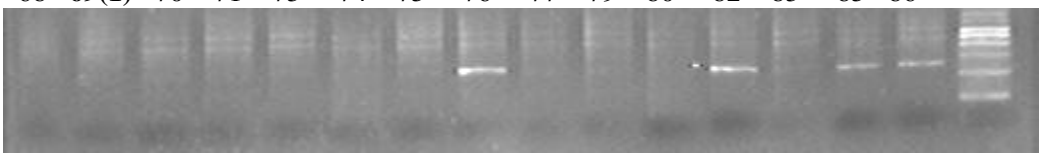
22 27 28 29 30 33 36 **37(1)** **38(1)** 39 40 41 43 44 47



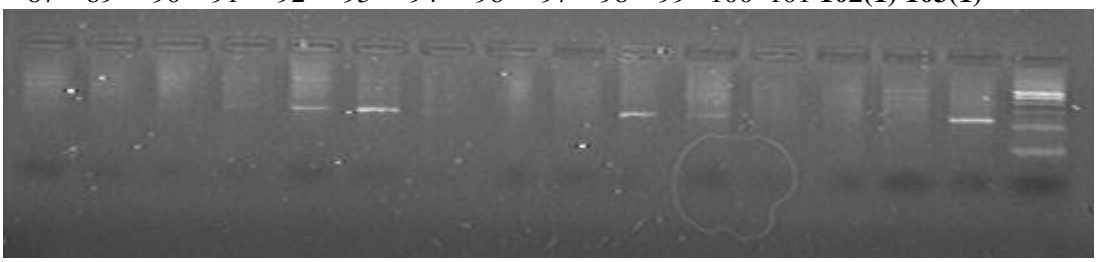
48 49 **51(1)** 53 **54**(1) 56 57 **58(1)** 59 60 61 **62(1)** 63 66 **67(1)**



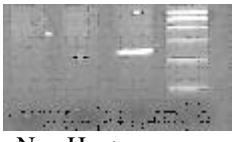
68 **69(1)** 70 71 73 74 75 76 77 79 80 82 83 85 86



87 89 90 91 92 93 94 96 97 98 99 100 101 **102(1)** **103(1)**



104 105 106 107 **108(1)** **109(1)** 110 111 112 **113(1)** 114 115 116 117 **118(1)**



Neg H +ve

**Table-6** RH map scoring panel for gene CIDE-3.

| Clones in 90 panel | Score | Clones in 90 panel | Score |
|---|---|---|---|
| 1 | 0 | 68 | 0 |
| 2 | 0 | 69 | 1 |
| 4 | 0 | 70 | 0 |
| 5 | 0 | 71 | 0 |
| 6 | 1 | 73 | 0 |
| 7 | 0 | 74 | 0 |
| 8 | 0 | 75 | 0 |
| 10 | 0 | 76 | 0 |
| 12 | 1 | 77 | 0 |
| 13 | 0 | 79 | 0 |
| 16 | 0 | 80 | 0 |
| 18 | 0 | 82 | 0 |
| 19 | 0 | 83 | 0 |
| 20 | 0 | 85 | 0 |
| 21 | 0 | 86 | 0 |
| 22 | 0 | 87 | 0 |
| 27 | 0 | 89 | 0 |
| 28 | 0 | 90 | 0 |
| 29 | 0 | 91 | 0 |
| 30 | 0 | 92 | 0 |
| 33 | 0 | 93 | 0 |
| 36 | 0 | 94 | 0 |
| 37 | 1 | 96 | 1 |
| 38 | 1 | 97 | 0 |
| 39 | 0 | 98 | 0 |
| 40 | 0 | 99 | 0 |
| 41 | 0 | 100 | 1 |
| 43 | 0 | 101 | 0 |
| 44 | 0 | 102 | 1 |
| 47 | 0 | 103 | 1 |
| 48 | 0 | 104 | 0 |
| 49 | 0 | 105 | 0 |
| 51 | 1 | 106 | 0 |
| 53 | 0 | 107 | 0 |
| 54 | 1 | 108 | 1 |
| 56 | 0 | 109 | 1 |
| 57 | 0 | 110 | 0 |
| 58 | 1 | 111 | 0 |
| 59 | 0 | 112 | 0 |
| 60 | 0 | 113 | 1 |
| 61 | 0 | 114 | 0 |
| 62 | 1 | 115 | 0 |
| 63 | 0 | 116 | 0 |
| 66 | 0 | 117 | 0 |
| 67 | 1 | 118 | 1 |

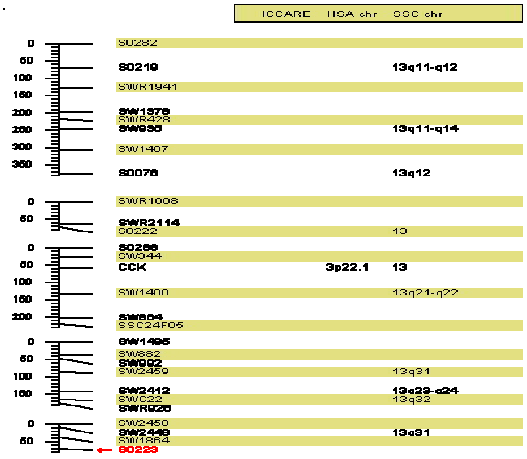
**Table-7** RH map result output obtained from RH mapping software available at IMpRH (http://imprh.toulouse.inra.fr) database. Results show a linkage of CIDE-3 gene to marker S0223 in chromosome 13 with a LOD value of 5.57.

| Chromosome | Marker | Retention fraction | Distance (ray) | LOD score |
|---|---|---|---|---|
| 13 | S0223 | 16 | 0,5 | 5,57 |

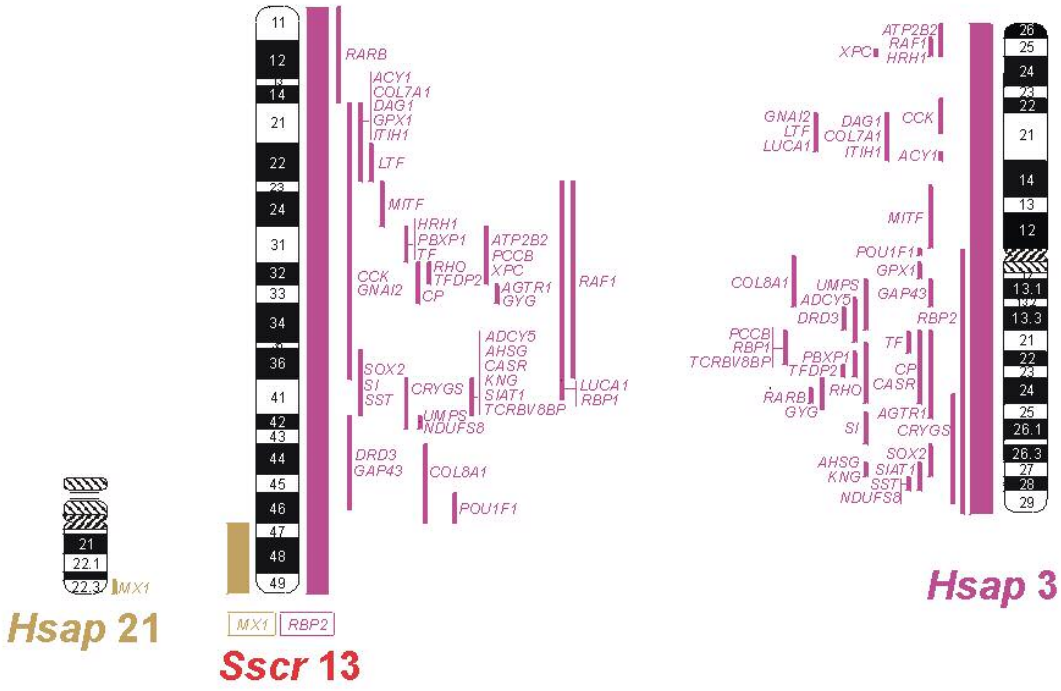
**Figure 13**

(a) Chromosomal location of marker S0223 in porcine chromosome 13 according to the 90 clone map byHawken et al. (1999)

(b) Chromosomal location of marker S0223 in porcine chromosome 13 (imprh.toulouse.inra.fr/, 2006)


ICCARE HGA chr CCC chr
SU252
SO219
13q11-q12
SWR1941
SW1579
SWR427R
SW950
13q11-q14
SW1407
SO078
13q12
SWR1008
SWR2114
SO222
13
SO269
SW344
CCK
3p22.1
13
SW1400
13q21-q22
SW804
SSC24F05
SW1495
SW862
SW902
SW2450
13q21
SW2412
13q23-q24
SW822
13q32
SWR520
SW2450
SW2449
13q31
SWR1884
SO229

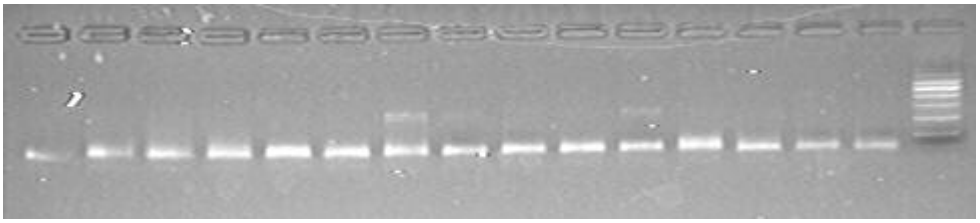
**(a)**


Hsap 21
MX1
Sscr 13
MX1
RBP2
Hsap 3
RARB
ACY1
COL7A1
DAG1
GPX1
ITIH1
LTF
MITF
HRH1
PBXP1
TF
RHO
TFDP2
CP
CCK
GNAI2
ATP2B2
PCCB
XPC
AGTR1
GYG
RAF1
ADCY5
AHSG
CASR
KNG
SIAT1
TCRBV8BP
LUCA1
RBP1
SOX2
SI
SST
CRYGS
UMPS
NDUFS8
DRD3
GAP43
COL8A1
POU1F1
ATP2B2
RAF1
XPC
HRH1
GNAI2
LTF
LUCA1
DAG1
COL7A1
ITIH1
CCK
ACY1
MITF
POU1F1
GPX1
COL8A1
UMPS
ADCY5
GAP43
DRD3
RBP2
PCCB
RBP1
TCRBV8BP
TF
PBXP1
TFDP2
CP
CASR
RARB
RHO
GYG
AGTR1
SI
CRYGS
AHSG
SOX2
KNG
SIAT1
SST
NDUFS8

**(b)**

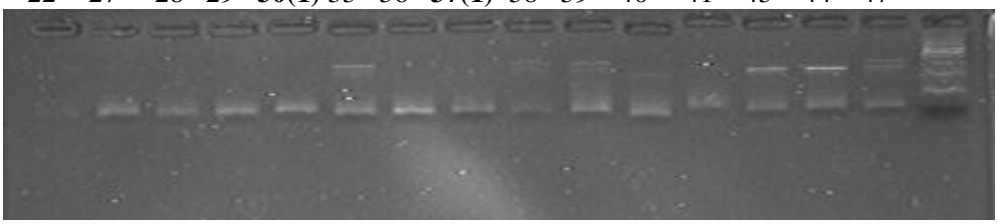
**Figure 14**

A representative PCR panel showing results obtained by scoring Pig-hamster 90 clone radiation hybrid panel for the gene, solute carrier family 12, member 3 (imprh.toulouse.inra.fr/, 2006).

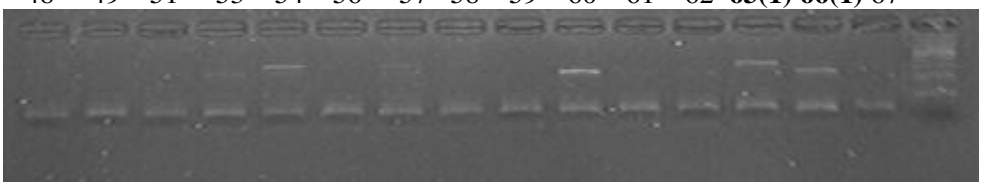


1 2 4 5 6 7 **8(1)** **10(1)** 12 13 **16(1)** 18 19 20 21

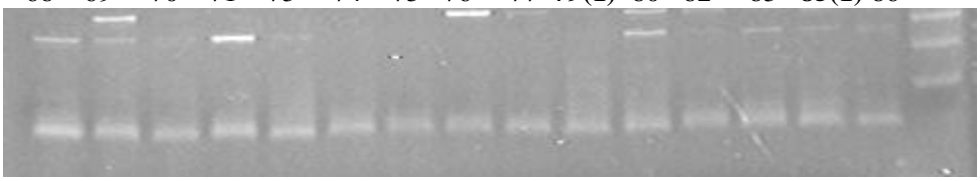
22 27 28 29 **30(1)** 33 36 **37(1)** 38 39 40 41 43 44 47



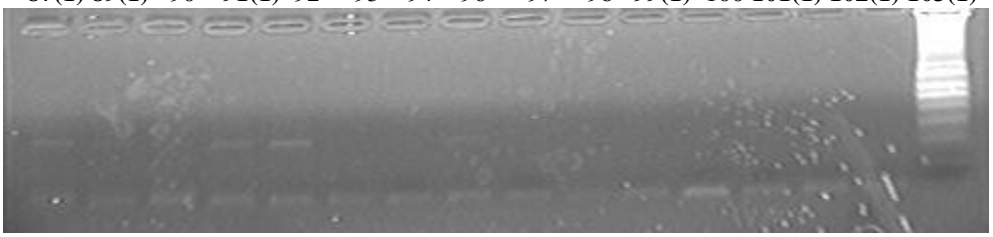
48 49 51 53 54 56 57 58 59 60 61 62 **63(1)** **66(1)** 67



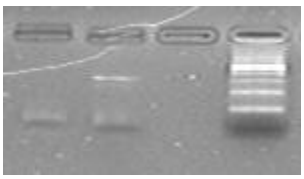
68 69 70 71 73 74 75 76 77 **79(1)** 80 82 83 **85(1)** 86



**87(1)** **89(1)** 90 **91(1)** 92 93 94 96 97 98 **99(1)** 100 **101(1)** **102(1)** **103(1)**



**104(1)** 105 106 **107(1)** **108(1)** 109 110 111 112 113 114 115 116 117 118



H + Neg

**Table-8** RH map scoring panel, solute carrier family 12, member 3

| Clones in 90 panel | Score | Clones in 90 panel | Score |
|---|---|---|---|
| 1 | 0 | 68 | 0 |
| 2 | 0 | 69 | 0 |
| 4 | 0 | 70 | 0 |
| 5 | 0 | 71 | 0 |
| 6 | 0 | 73 | 0 |
| 7 | 0 | 74 | 0 |
| 8 | 1 | 75 | 0 |
| 10 | 1 | 76 | 0 |
| 12 | 0 | 77 | 0 |
| 13 | 0 | 79 | 1 |
| 16 | 1 | 80 | 0 |
| 18 | 0 | 82 | 0 |
| 19 | 0 | 83 | 0 |
| 20 | 0 | 85 | 1 |
| 21 | 0 | 86 | 0 |
| 22 | 0 | 87 | 1 |
| 27 | 0 | 89 | 1 |
| 28 | 0 | 90 | 0 |
| 29 | 0 | 91 | 1 |
| 30 | 1 | 92 | 0 |
| 33 | 0 | 93 | 0 |
| 36 | 0 | 94 | 0 |
| 37 | 1 | 96 | 0 |
| 38 | 0 | 97 | 0 |
| 39 | 0 | 98 | 0 |
| 40 | 0 | 99 | 1 |
| 41 | 0 | 100 | 0 |
| 43 | 0 | 101 | 1 |
| 44 | 0 | 102 | 1 |
| 47 | 0 | 103 | 1 |
| 48 | 0 | 104 | 1 |
| 49 | 0 | 105 | 0 |
| 51 | 0 | 106 | 0 |
| 53 | 0 | 107 | 1 |
| 54 | 0 | 108 | 1 |
| 56 | 0 | 109 | 0 |
| 57 | 0 | 110 | 0 |
| 58 | 0 | 111 | 0 |
| 59 | 0 | 112 | 0 |
| 60 | 0 | 113 | 0 |
| 61 | 0 | 114 | 0 |
| 62 | 0 | 115 | 0 |
| 63 | 1 | 116 | 0 |
| 66 | 1 | 117 | 0 |
| 67 | 0 | 118 | 0 |

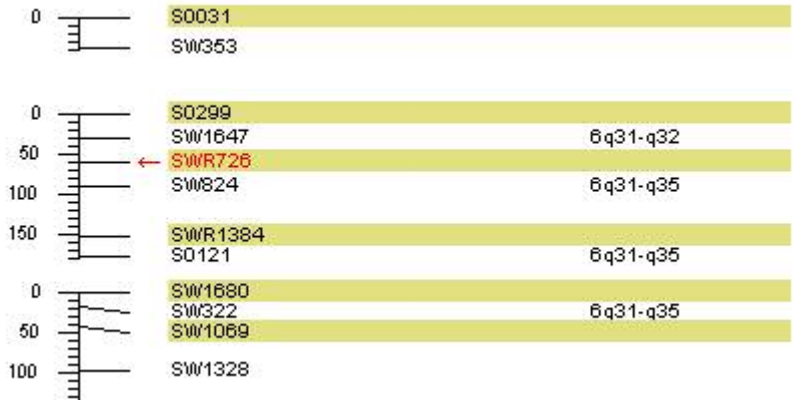
**Table-9** RH map result output obtained from RH mapping software available at IMpRH (http://imprh.toulouse.inra.fr) data base. Results show a linkage of solute carrier family 12, member 3 gene to marker SWR726 in chromosome 6 with a LOD score of 5.45.

| Chromosome | Marker | Retention fraction | Distance (ray) | LOD score |
|---|---|---|---|---|
| 6 | SWR726 | 26 | 0,55 | 5,45 |

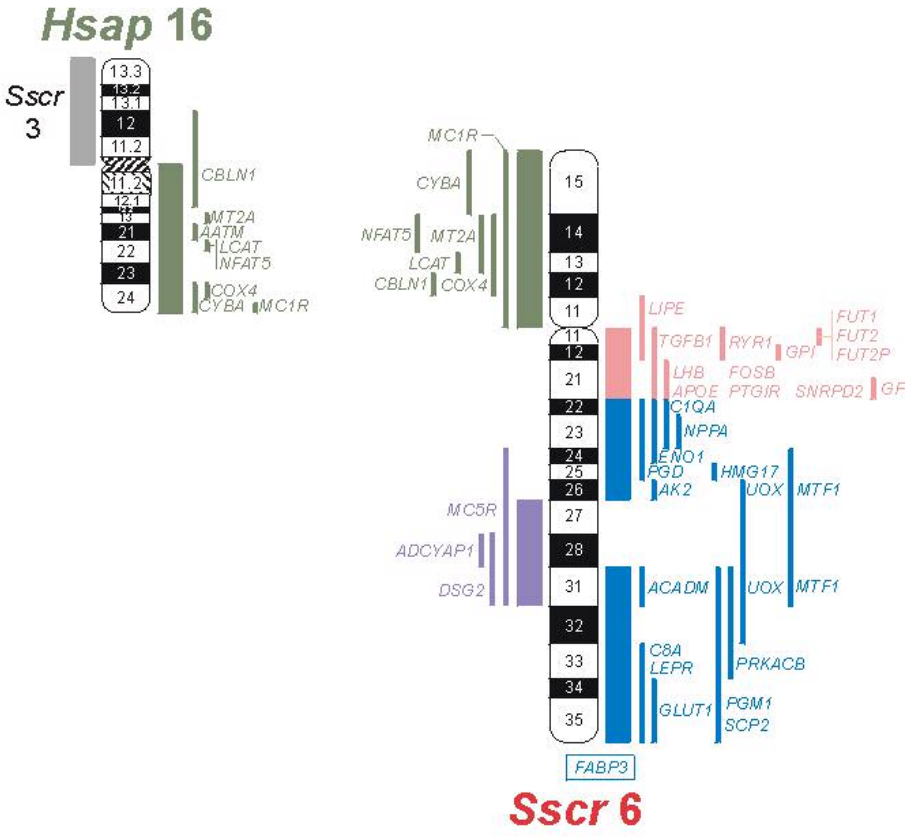
**Figure 15**

**(a)** Chromosomal location of marker SWR726 in porcine chromosome 6 according to the 90 clone map byHawken et al. (1999)

**(b)** Chromosomal location of marker SWR726 in porcine chromosome 6 in IMpRH server (imprh.toulouse.inra.fr/, 2006)


0
S0031
SW353
0
S0299
SW1647
6q31-q32
50
SWR726
100
SW824
6q31-q35
150
SWR1384
S0121
6q31-q35
0
SW1680
SW322
6q31-q35
50
SW1069
100
SW1328

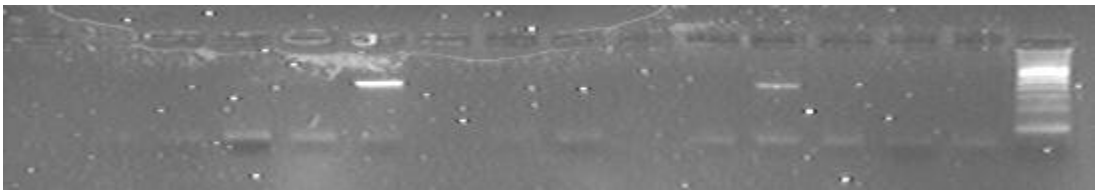
**(a)**


Hsap 16
Sscr 3
13.3
13.2
13.1
12
11.2
11.1
12.1
13
21
22
23
24
CBLN1
MT2A
AATM
LCAT
NFAT5
COX4
CYBA
MC1R
MC1R
CYBA
NFAT5
MT2A
LCAT
CBLN1
COX4
15
14
13
12
11
11
12
21
22
23
24
25
26
27
28
31
32
33
34
35
LIPE
TGFB1
RYR1
GPI
FUT1
FUT2
FUT2P
UHB
FOSB
APOE
PTGIR
SNRPD2
GF
C1QA
NPPA
ENO1
PGD
AK2
HMG17
UOX
MTF1
MC5R
ADCYAP1
DSG2
ACADM
UOX
MTF1
C8A
LEPR
PRKACB
GLUT1
PGM1
SCP2
FABP3
Sscr 6

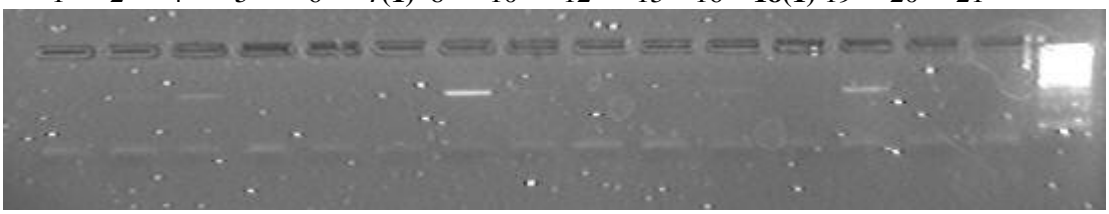
**(b)**

**Figure 16**

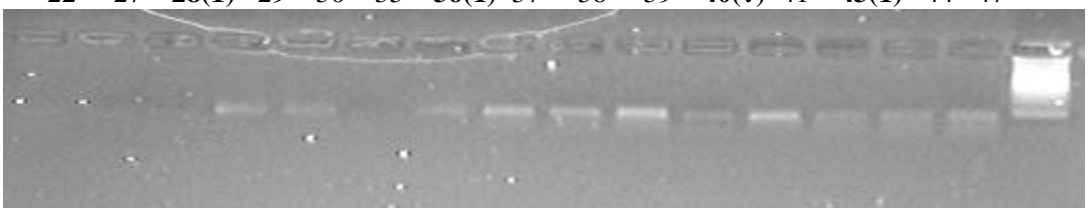
A representative PCR panel showing results obtained by scoring Pig-hamster 90 clone radiation hybrid panel for the gene cytochrome P450 4F6 isoform 2 (imprh.toulouse.inra.fr/, 2006).



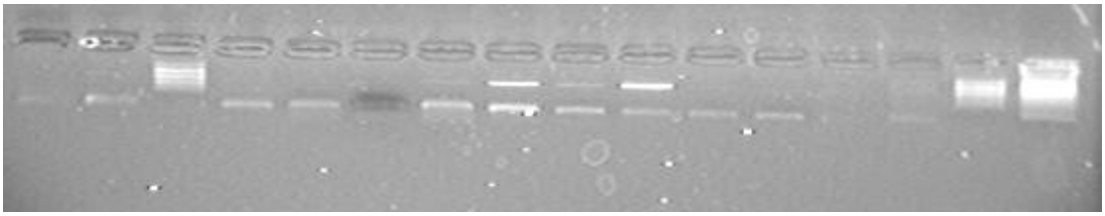
1 2 4 5 6 **7(1)** 8 10 12 13 16 **18(1)** 19 20 21



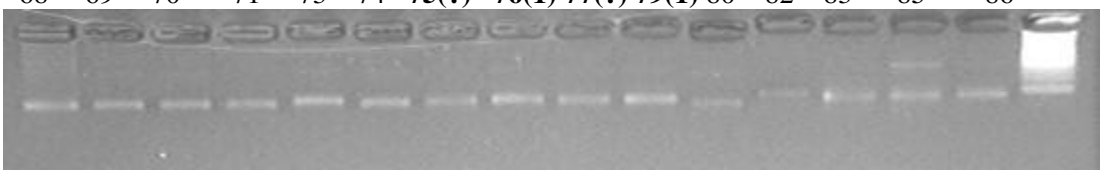
22 27 **28(1)** 29 30 33 **36(1)** 37 38 39 **40(?)** 41 **43(1)** 44 47



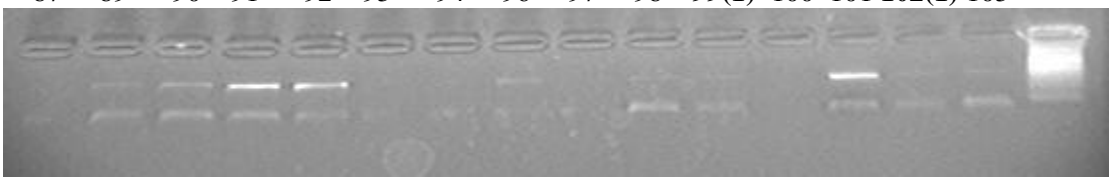
48 49 51 53 54 56 57 58 59 60 61 62 63 66 67



68 69 70 71 73 74 **75(?)** **76(1)** **77(?)** **79(1)** 80 82 83 85 86



87 89 90 91 92 93 94 96 97 98 **99(1)** 100 101 **102(1)** 103



104 **105(1)** **106(1)** **107(1)** **108(1)** 109 110 **111(1)** 112 **113(?)** **114(?)** 115 **116(1)** **117(1)** **118(?)**

Neg + Hamster

**Table -10** RH map scoring panel of cytochrome P450 4F6 isoform 2

| Clones in 90 panel | Score | Clones in 90 panel | Score |
|---|---|---|---|
| 1 | 0 | 68 | 0 |
| 2 | 0 | 69 | 0 |
| 4 | 0 | 70 | 0 |
| 5 | 0 | 71 | 0 |
| 6 | 0 | 73 | 0 |
| 7 | 1 | 74 | 0 |
| 8 | 0 | 75 | ? |
| 10 | 0 | 76 | 1 |
| 12 | 0 | 77 | ? |
| 13 | 0 | 79 | 1 |
| 16 | 0 | 80 | 0 |
| 18 | 1 | 82 | 0 |
| 19 | 0 | 83 | 0 |
| 20 | 0 | 85 | 0 |
| 21 | 0 | 86 | 0 |
| 22 | 0 | 87 | 0 |
| 27 | 0 | 89 | 0 |
| 28 | 1 | 90 | 0 |
| 29 | 0 | 91 | 0 |
| 30 | 0 | 92 | 0 |
| 33 | 0 | 93 | 0 |
| 36 | 1 | 94 | 0 |
| 37 | 0 | 96 | 0 |
| 38 | 0 | 97 | 0 |
| 39 | 0 | 98 | 0 |
| 40 | ? | 99 | 1 |
| 41 | 0 | 100 | 0 |
| 43 | 1 | 101 | 0 |
| 44 | 0 | 102 | 1 |
| 47 | 0 | 103 | 0 |
| 48 | 0 | 104 | 0 |
| 49 | 0 | 105 | 1 |
| 51 | 0 | 106 | 1 |
| 53 | 0 | 107 | 1 |
| 54 | 0 | 108 | 1 |
| 56 | 0 | 109 | 0 |
| 57 | 0 | 110 | 0 |
| 58 | 0 | 111 | 1 |
| 59 | 0 | 112 | 0 |
| 60 | 0 | 113 | ? |
| 61 | 0 | 114 | ? |
| 62 | 0 | 115 | 0 |
| 63 | 0 | 116 | 1 |
| 66 | 0 | 117 | 1 |
| 67 | 0 | 118 | ? |

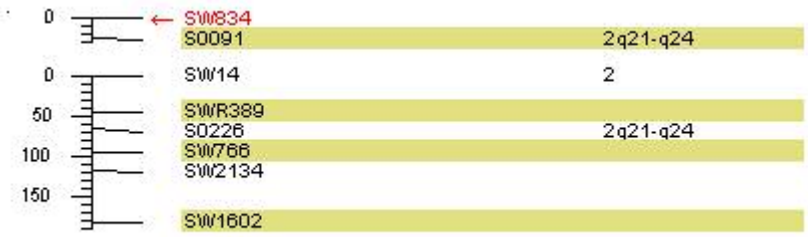
**Table-11** RH map result output obtained from RH mapping software available at IMpRH (http://imprh.toulouse.inra.fr) data base. Results show a linkage of cytochrome P450 4F6 isoform 2, gene to marker SW834 in chromosome 2 with a LOD score of 7.35.

| Chromosome | Marker | Retention fraction | Distance (ray) | LOD score |
|---|---|---|---|---|
| 2 | SW834 | 32 | 0,35 | 7,35 |

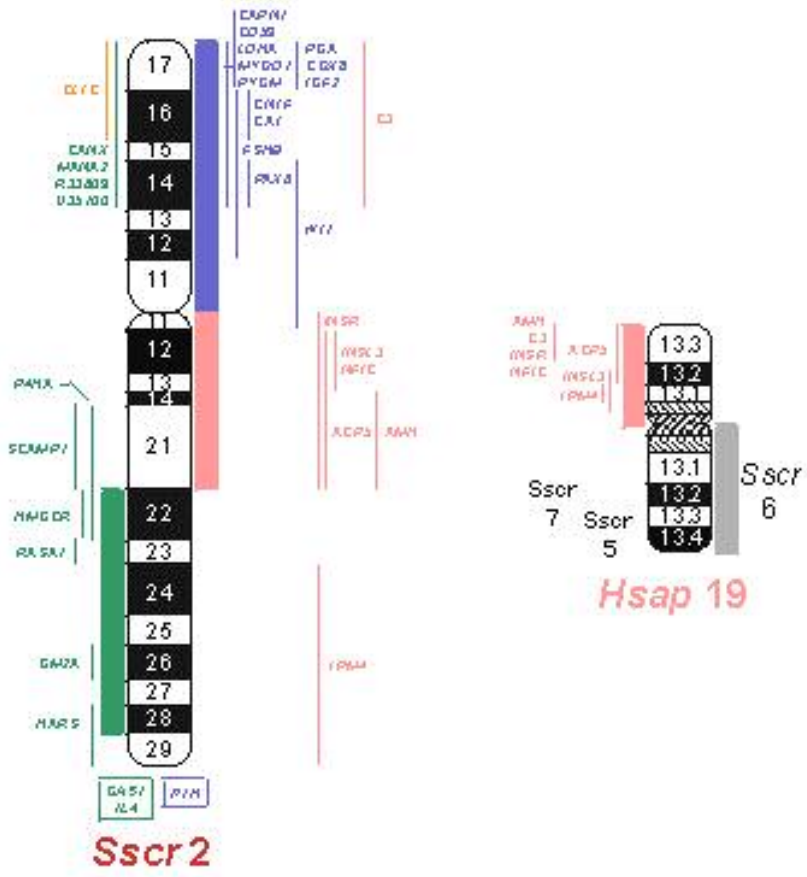
**Figure 17**

(a) Chromosomal location of marker SW834 in porcine chromosome 2 according to the 90 clone map byHawken et al. (1999)

(b) Chromosomal location of marker SW834 in porcine chromosome 2 in (imprh.toulouse.inra.fr/, 2006)


0
SW834
S0091
2q21-q24
0
SW14
2
50
SW R389
S0226
2q21-q24
100
SW766
SW2134
150
SW1602

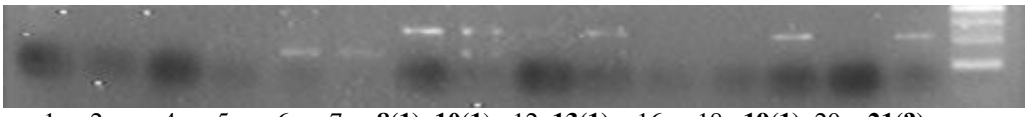
**(a)**


17
16
15
14
13
12
11
12
13
14
21
22
23
24
25
26
27
28
29
Sscr 2
13.3
13.2
13.1
13.2
13.3
13.4
Sscr 7
Sscr 5
Sscr 6
Hsap 19

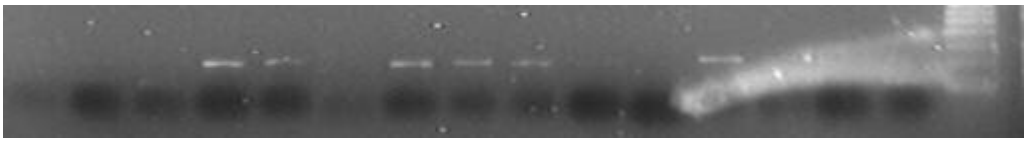
**(b)**

**Figure 18**

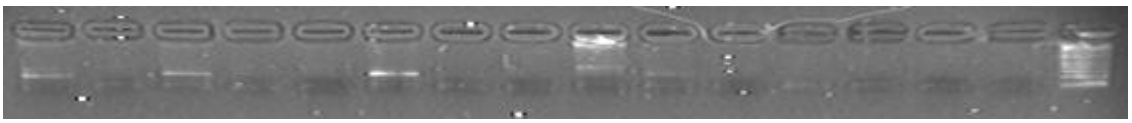
A representative PCR panel showing results obtained by scoring Pig-hamster 90 clone radiation hybrid panel for the gene RTP801 (imprh.toulouse.inra.fr/, 2006).



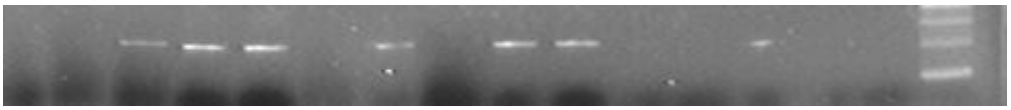
1 2 4 5 6 7 **8(1)** **10(1)** 12 **13(1)** 16 18 **19(1)** 20 **21(?)**



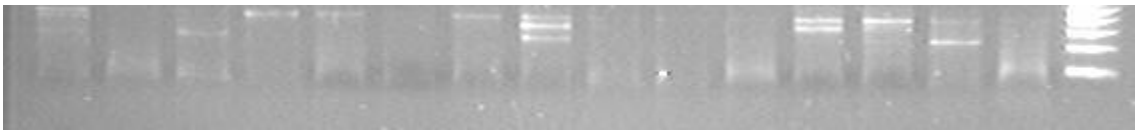
22 27 28 **29(1)** **30(1)** **33** **36(1)** **37(1)** **38(1)** 39 40 **41(?)** 43 44 47



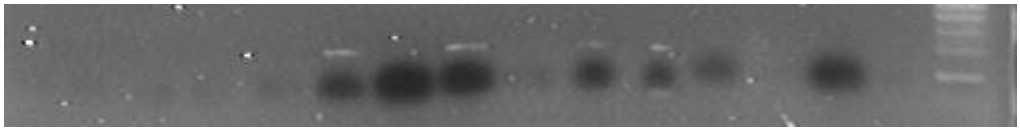
**48(1)** 49 **51(1)** 53 54 **56(1)** 57 58 59 **60(1)** 61 62 63 66 67



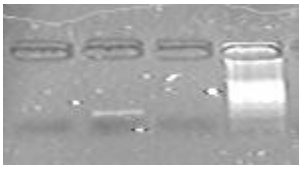
68 69 **70(1)** **71(1)** **73(1)** 74 **75(?)** 76 **77(1)** **79(1)** 80 82 **83(1)** 85 86



**87(1)** 89 **90(1)** 91 92 93 94 **96(1)** 97 98 **99(?)** 100 **101(?)** **102(1)** **103(?)**



104 105 106 107 109 **108 (1)** 110 **111 (1)** 112 **113 (1)** **114 (?)** 115 116 117 118



Neg + Hamster

**Table-12** RH map scoring panel of RTP801

| Clones in 90 panel | Score | Clones in 90 panel | Score |
|---|---|---|---|
| 1 | 0 | 68 | 0 |
| 2 | 0 | 69 | 0 |
| 4 | 0 | 70 | 1 |
| 5 | 0 | 71 | 1 |
| 6 | 0 | 73 | 1 |
| 7 | 0 | 74 | 0 |
| 8 | 1 | 75 | 1 |
| 10 | 1 | 76 | 0 |
| 12 | 0 | 77 | 1 |
| 13 | 1 | 79 | 1 |
| 16 | 0 | 80 | 0 |
| 18 | 0 | 82 | 0 |
| 19 | 1 | 83 | 1 |
| 20 | 0 | 85 | 0 |
| 21 | ? | 86 | 0 |
| 22 | 0 | 87 | 1 |
| 27 | 0 | 89 | 0 |
| 28 | 0 | 90 | 1 |
| 29 | 1 | 91 | 0 |
| 30 | 1 | 92 | 0 |
| 33 | 0 | 93 | 0 |
| 36 | 1 | 94 | 0 |
| 37 | 1 | 96 | 1 |
| 38 | 1 | 97 | 0 |
| 39 | 0 | 98 | 0 |
| 40 | 0 | 99 | 1 |
| 41 | ? | 100 | 0 |
| 43 | 0 | 101 | ? |
| 44 | 0 | 102 | 1 |
| 47 | 0 | 103 | ? |
| 48 | 1 | 104 | 0 |
| 49 | 0 | 105 | 0 |
| 51 | 1 | 106 | 0 |
| 53 | 0 | 107 | 0 |
| 54 | 0 | 108 | 0 |
| 56 | 1 | 109 | 1 |
| 57 | 0 | 110 | 0 |
| 58 | 0 | 111 | 1 |
| 59 | 0 | 112 | 0 |
| 60 | 1 | 113 | 1 |
| 61 | 0 | 114 | ? |
| 62 | 0 | 115 | 0 |
| 63 | 0 | 116 | 0 |
| 66 | 0 | 117 | 0 |
| 67 | 0 | 118 | 0 |

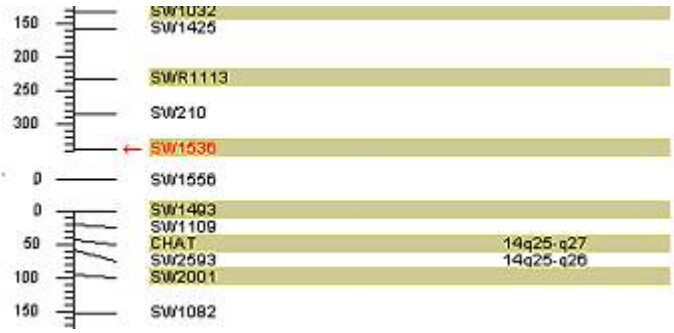
**Table-13** RH map result output obtained from RH mapping software available at IMpRH (http://imprh.toulouse.inra.fr) data base. Results show a linkage of RTP801 gene to marker SW1536 in chromosome 14 with a LOD score of 6.44.

| Chromosome | Marker | Retention fraction | Distance (ray) | LOD score |
|---|---|---|---|---|
| 14 | SW1536 | 27 | 0,47 | 6,44 |

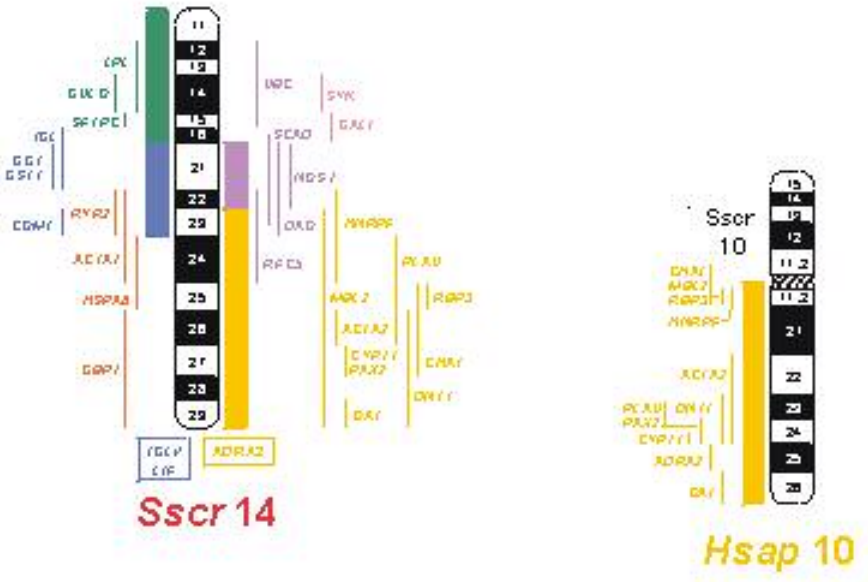
**Figure 19**

(a) Chromosomal location of marker SW1536 in porcine chromosome 14 according to the 90 clone map byHawken et al. (1999)

(b) Chromosomal location of marker SW1536 in porcine chromosome 14 in IMpRH server (imprh.toulouse.inra.fr/, 2006)


150
200
250
300
0
0
50
100
150
SW1032
SW1425
SWR1113
SW210
SW1536
SW1556
SW1493
SW1109
CHAT
SW2593
SW2001
SW1082
14q25-q27
14q25-q26

**(a)**


Sscr 14
Sscr 10
Hsap 10

**(b)**

## Discussion

There were 362 coding SNPs with Phred quality score above 20 in the porcine subcutaneous adipose tissue cDNA library. Even though most of the identified cSNPs might not determine a change in a phenotype, there will be100% association of these SNPs with trait genes and they would be useful for candidate gene isolation in marker assisted selection process.

Transitions were 2.6 times higher than transversions. This bias towards transitions was observed in previous studies. Results obtained from a study of human SNPs in dbEST database by Picoult-Newberg et al. in 1999 showed transitions to transversion ratio of 1.7 and a study on chicken SNPs showed this ratio to be 2.3 (Smith et al., 2001). Reasons for the high percentage of transitions is thought to be the spontaneous deamination of 5-methyl cytosine, the post synthetic modified product of cytosine to thymine and the inefficient repair mechanism to correct this mutation (Holliday and Grigg, 1993). Cargill et al., (1999) reported that non synonymous SNPs in humans occur at a lower rate and allele frequency. In accordance with this observation, the majority of cSNPs (88%) were synonymous and did not change the amino acid sequence of the encoded protein. Only 48 cSNPs (12%) among the total of 362 were non-synonymous SNPs.

Poly-ubiquitin transcripts that encoded for the protein ubiquitin, which is involved in ATP dependent selective degradation of cellular proteins had the highest number of SNPs and MHC class I transcript had the second highest number of polymorphisms in the library. dbEST showed 26 polymorphisms for poly-ubiquitin gene from human genome sequence results with 10 SNPs in the coding region with one non-synonymous SNP (www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html, 2006). Dworkin-Rastl et al,

(1984) reported increased percentage of population polymorphism in ubiquitin sequences both at DNA and RNA levels in *Xenopus* species. In PEDE pig cSNP data base, 22 different ubiquitin related sequences showed cSNPs (pede.dna.affrc.go.jp/csnp/csnp_main.php, 2006). These findings in literature indicate that ubiquitin gene is prone to have high amount of SNPs. None of the observed polymorphisms in ubiquitin resulted in an amino acid substitution and this is similar to results obtained on human ubiquitin by Wilborg et al. in 1985. They have stated that ubiquitin is a much conserved protein through evolution. In literature both MHC classes I and II are reported to be highly polymorphic and it is stated that the MHC genes have the highest number of polymorphism among mammalian genes (Parham and Ohta, 1986; Hughes and Yeager, 1998). Reasons for high number of polymorphisms in MHC class I might be due to maternal-fetal incompatibility, over dominant selection, frequency dependent selection and interlocus conversion (Parham and Ohta, 1986; Hughes and Yeager, 1998).

Only 11 out of 23 selected non synonymous SNPs were confirmed to be positives by sequencing of genomic DNA of $F_0$ and $F_2$ population. This is a confirmation of 48% of non-synonymous SNPs. This difference in SNPs between the adipose tissue cDNA library ESTs and genomic sequences could be due to sequence artifacts, contig assembly errors and differences in duplicated sequences due to amplification of nearly identical DNA sequences found in various parts of the genome.

Five SNPs out of 12 unconfirmed polymorphisms occurred only in one clone in the contig. Even though they had high Phred quality scores, these 5 SNPs might be sequence artifacts. In very rare instances mismatches could arise due to base calling errors or

errors generated during cDNA synthesis or propagation in Escherichia coli (Picoult-Newberg et al., 1999b).

RNA editing is an interesting emerging concept that explains some of these mismatches in polymorphisms between genomic DNA and EST sequences. In RNA editing the RNA sequence is altered from the one that is encoded by the genome, during or after the transcription, by nucleotide insertion, deletion or modification mechanisms (Wedekind et al., 2003). This process might enable a different protein to be expressed. Even though there are many examples of mRNA editing in plants and protozoan organelles, currently there are relatively few instances of RNA editing reported in mammals. There are few examples of RNA editing in mammals by formation of inosine through adenosine deamination. In cytosine to inosine RNA editing genomic DNA shows A whilst EST sequence at the edited site shows G (Levanon et al., 2004). There is also the example of cytosine to uracil mRNA editing in apolipo- protein B(Wedekind et al., 2003). C ytosine to uracil RNA editing result in a dG to dA transition and it might be evaluated as a SNP. Levanon et al. in 2004 characterized human RNA editing sites and they found that 92% of sites occur within Alu repeats. Even though it is stated that editing is seen only one in 83 EST sequences (Levanon et al., 2004), this possibility can not be ruled out as a reason for the mismatches in the cDNA library.

Even if 40% of the polymorphisms found in the cDNA library are confirmed to be positives there are156 cSNP that could be used as direct markers that are directly associated with segregation of genes. None of the confirmed SNPs segregated exclusively in an outlier group for back fat thickness. Even though these 11 SNPs cannot

be applied as markers in identification of candidate genes for back fat thickness, they might segregate with some other specific trait genes.

Genome wide scan on growth, carcass merit and performance quality traits of $F_2$ resource population will be conducted at Michigan State University and the confirmed cSNPs can be included as markers in this genotyping process. Subcutaneous adipose tissue and skeletal muscle cDNA microarrays will be developed from the unique sequences obtained from adipose tissue cDNA library and skeletal muscle cDNA library constructed at Michigan State University. Then, gene expression profiles of genes with confirmed SNPs can be studied in detail in outliers for many other traits.

Brief discussion on the 11 genes with confirmed non synonymous SNPs and their importance are given below.

Shoshani et al in 2002 cloned the RTP801 gene from a cDNA library of rat and identified the human ortholog by searching EST database. It was revealed from their experiments that there is a strong up regulation of RTP801 due to hypoxia during ischemic stroke. Even though several secondary structures are predicted for the encoded protein functional motifs are yet unknown (Ellisen et al., 2002). Shoshani et al., (2002) showed that RTP801 participates in both protective and detrimental effects of hypoxia. RTP801 act as a negative regulator in mammalian target of Rapamycin pathway which plays an essential role in cell growth and control (Corradetti et al., 2005). RTP801 knockout mice showed significant reduction in ischemic proliferative retinopathy, which is a complication of diabetes mellitus (Brafman et al., 2004). Their studies suggest that RTP801 could be regarded as a novel therapeutic target for treatment of proliferative retinopathies.

The SNP identified in RTP801 created an amino acid substitution from threonine to arginine. Both arginine and threonine are hydrophilic amino acids. Arginine has an R group that is positively charged and they bind to molecules that have negative charges in the surface ( Lehninger et al., 1993). Threonine is an amino acid with a side chain ending with a hydroxyl group which acts as a hydrogen donor in many enzymatic reactions. This amino acid substitution might influence the structure and function of the RTP801 protein and it is needed to identify the functional motifs of the protein to predict on the substitution effect. The radiation hybrid mapping information indicates RTP801 gene to be located in porcine chromosome 14 and this is consistent with the results obtained by bi-directional chromosome painting (Goureau et al. 1996).

A non synonymous SNP was identified in gene CIDE-3 that belongs to cell death inducing DNA fragmentation factor in apoptosis. Liang et al. (2003) cloned and isolated CIDE-3 gene from human adipose and breast tissues. FSP27, the mouse homolog of CIDE-3, is reported to be an adipocyte specific marker gene in mouse (Danesch et al., 1992) but in the study conducted by Linag et al. (2003) they have demonstrated the expression of this gene in many other cell lines such as in a hepatocellular carcinoma cell line and a malignant melanoma cell line. Pig homolog for CIDE-3 gene is not available in NCBI database and searching of the NCBI EST database showed similar sequences reported mainly from adipose tissue cDNA libraries. CIDE-1 is expressed in high levels in brown adipose tissue and CIDE-1 null mice were lean and resistant to diet induced obesity and diabetes (Zhou et al., 2003). No detailed research has been done on CIDE-3 and its functions on adipocytes but this gene might have a regulatory role in energy balance and adiposity.

The polymorphism found in the gene similar to CIDE-3 created an amino acid substitution from threonine to serine. Both these amino acids are hydrophilic and have similar biochemical properties. Substitution between these two might not have major effect on the structure of a protein that might affect the functional properties. The radiation hybrid mapping information indicates CIDE-3 gene to be located in porcine chromosome 13 and this is consistent with the results obtained by bi-directional chromosome painting (Goureau et al. 1996).

Cytochrome P450 F6 isoform catalyze the hydroxylation of eicosanoids and a polymorphism was detected in this gene that substitutes aspartic acid for glutamic acid. These two amino acids have similar properties with both being hydrophilic and acidic and a substitution between two of them might not have a major effect on the functional properties of the protein. Radiation hybrid mapping indicates that cytochrome P450 F6 isoforms transcript to be located in porcine chromosome 2 and this is consistent with the results obtained by bi-directional chromosome painting (Goureau et al. 1996).

Solute carrier family 12, member 3 is a sodium/chloride transporter and this gene is very highly expressed in the kidney. Polymorphism detected in this gene substitute proline to leucine. Proline can disrupt protein folding pattern especially in α helices and force a kink in the protein structure. Amino acid substitution might influence the structure and function of this transporter protein and it is needed to identify the functional motifs of the protein to predict more on the effect of amino acid substitution. The radiation hybrid mapping information indicates solute carrier family 12, member 3 transcript to be located in porcine chromosome 6 and this is consistent with the results obtained by bi-directional chromosome painting (Goureau et al. 1996).

A non synonymous SNP was confirmed in thyroid hormone responsive protein. This gene is expressed in liver and adipocytes and plays a role in lipogenesis. Gene expression studies by Clarke et al. (1990) and Liu and Towlele (1994) in rat and mouse showed nutritional and hormonal control in expression pattern of this protein. Eighty percent of the amino acid sequence of this protein is found to be conserved among human, rat and mouse (Clarke et al., 1990; Liu and Towle, 1994). Detected polymorphism in thyroid hormone responsive protein gene created substitution of aspartic acid for glutamic acid that has similar biochemical properties. A substitution between these two amino acids might not have a major effect on the structure of the protein.

Polymorphism in glycerol-3-phosphate dehydrogenase 1 created a substitution of amino acid valine for iso-leucine. Both these amino acids have large aliphatic hydrophobic side chains. A substitution between these two amino acids might not have a major effect on hydrophobic interactions and folding pattern of the protein

Polymorphism found in mRNA encoding G-beta like protein created an amino acid substitution from tyrosine to serine. Serine is a hydrophilic amino acid and participates in hydrogen donor reactions. Tyrosine, a hydrophobic amino acid has different biochemical properties. Created polymorphism might be having an effect on the structure of G-protein and signal transduction mechanisms and it is needed to identify the functional motifs of the protein to predict more on the effect of amino acid substitution.

Eukaryotic ribosome consist of at least 80 proteins ( Kenmochi et al., 1998) that function in regulation of protein synthesis process. Polymorphisms found in two ribosomal protein encoding genes caused substitution of bio chemically similar amino acids and

these polymorphisms might not have a significant effect on the functions of encoded proteins.

Non synonymous SNP in mitochondrial genome changed the 119$^{th}$ amino acid in ATPsynthase 6 subunit from histidine to tyrosine. These two amino acids have different biochemical properties with hydrophilic histidine having a positively charged basic R group of a pKa near neutrality and hydrophobic and non polar tyrosine having an aromatic side chain. Some proteins control the properties and conformation of polypeptides in acidic medium by protonation of nitrogen atom in histidine. Tyrosine on the other hand can participate in hydrophobic interactions and hydroxyl group of tyrosine can form hydrogen bonds and can act as important functional group in some enzymatic activities (Lehninger et al., 1993). When the biochemical properties of these two amino acids are considered there is a high possibility of a conformational change occurring in the structure of ATPsynthase6 subunit by the SNP. T to G transversion at nucleotide 8993 in mitochondrial ATPase gene in human was found in some isolated cases of mental retardation and ataxia without retinitis pigmentosa (de Coo et al., 1996). Substitutions of amino acids in mtATPase 6 show defects like Leber optic atrophy, cardiac conduction defects and Leigh syndrome, a severe neurological disorder. Polymorphism of C to T found in the cDNA library sequence aligned with human nucleotide 8879 and allele T was present at that site in human sequence. No reports were found on any polymorphisms or related diseases at nucleotide 8879 of human mitochondrial genome. But in the future it might be useful to check on any polymorphisms at this site in human genome.

In humans, mitochondrial DNA show a remarkable regional variation that is attributed to genetic drift and natural selection (Mishmar et al., 2003). Hundred and four complete human mtDNA sequences from humans in all regions and lineages were analysed by Mishmar et al., (2003) and it was found out that African mtDNA variation did not deviate much from the neutral sequence while Asian, European, Siberian and Native American lineages showed a considerable deviation. They did an analysis on 13 genes encoded by mitochondrial genome and it was revealed that ATPase 6 gene which encodes a more conserved mitochondrial protein has the highest number of amino acid sequence variation and these amino acid variation found to differ with the climatic zone (Mishmar et al., 2003).

It is hypothesized that mitochondrial DNA mutations allow changes due to dietary and environmental conditions (Marchler-Bauer et al., 2005). Because mitochondrial genome encodes 13 polypeptides of oxidative phosphorylation, genes of the mitochondria play a vital role in cellular energy production for generation of ATP and heat. Energy generated from the oxidative phosphoryaltion is utilized for ATP synthesis and heat production. The balance between these two functions is determined by efficiency in the generation of electrochemical gradient between mitochondrial matrix and the inner membranes. Mitochondrial ATPase6 variants might act as partial uncouplers and reduce the electrochemical gradient thus there will be less production of ATP and higher basal metabolic rate.

When nucleotide at 9474 of the *Sus scrofa* mitochondrial genome was analyzed it was revealed that Duroc, Hampshire and Swedish wild boar to have allele T at 9474$^{th}$ position. Chinese breeds (Taoyuan, Lanyu, Large white, Wannanhua, Jiangquhai,

Yushanhei, Dahuabai, Diannan short ear, Yimenghei, Wuzhishan, Qingping, Min, Jinhua, Tongcheng, Erhualian, Rongchang, Xiang, Ningxiang, Zang, Meishan, Nuogu), Large white, Korean wild boar and Berkshire breed had allele C at this position. Landrace animals showed both C and T and no previous reports were found on Pietrain breed. These results show a geographical demarcation in the polymorphism between Asian and Western breeds. We can hypothesize that same as in human environmental selection might have led to this regional specific polymorphism. Chinese native pig breeds have slow growth rate, low dressing percentage and low lean meat percentage while Duroc, and Hampshire breeds show more positive attributes in these qualities. Mitochondrial DNA variations specially in protein like ATPase6 and ATPase8 subunits might be contributing to these changes and it would be very important to study the polymorphisms in mitochondrial genome in breeding stocks and identify candidate genes in mitochondrial genome that are responsible for the pork quality.

Chromosomal location of 4 genes with confirmed non synonymous cSNPs were confirmed by RH mapping technique. Experiments conducted by Rettenberger et al., (1995) and Goureau et al., (1996) demonstrated that there is conservation of synteny between human and pig genome (Rettenberger et al., 1995; Goureau et al., 1996). Comparative genetic mapping has shown that porcine and human genomes are more similarly organized than human and mouse genome (Ellegren et al., 1994; Rettenberger et al., 1995). Therefore these cSNPs that were mapped to the porcine genome and confirmed to be in conserved syntenic regions of the porcine and human chromosome might be used for identification of a potential candidate gene for QTL studies in human.

Glycerol-3-phosphate dehydrogenase 1, G-beta like protein, thyroid hormone responsive protein, 60S ribosomal protein L32 and 40S ribosomal protein S20 genes could not be mapped to the porcine genome. PCR reactions designed for Glycerol-3-phosphate dehydrogenase 1, G-beta like protein and thyroid hormone responsive protein amplified hamster genomic DNA. Alignment of porcine mRNA sequences of these three genes with mouse mRNA sequences revealed more than 90% homology between the sequences. It is essential to design and optimize primers from non conserved regions of porcine and mouse sequences for prevention of hamster genomic DNA amplification during RH mapping process.

Even though the PCR reaction performed for 60S ribosomal protein L32 and 40S ribosomal protein S20 did not amplify hamster genomic DNA, high percentage of positive scores (above 90%) were obtained from RH map scoring process. Literature search on ribosomal proteins revealed most ribosomal proteins to have high number of silent pseudogenes dispersed through out the human genome (Dudov and Perry, 1984). Amplification of pseudogenes of 60S ribosomal protein L32 and 40S ribosomal protein S20 would be the reason for high percentage of positive scores obtained during RH panel mapping process. For identification of non conserved regions between the functional gene and pseudogenes, PCR amplified products of the gene of interest needed to cloned and sequenced. Then, the cloned sequences need to be aligned with EST sequence for identification of pseudogenes and primers for RH mapping must be developed from non homologous regions of EST found after aligning with cloned sequences.

It is essential to calculate the frequency of these confirmed 11 SNPs in larger populations as polymorphisms that only exhibit variation in very small percentage of a population is

not useful for linkage studies. The power of an association study based on SNPs depends on the number of SNPs used and the quality of the polymorphism, such as the likelihood of the SNP locus actually being polymorphic in the population of interest (Eisenberg et al., 2005). Because of that it would be interesting to study these polymorphisms in other swine breeds.

## Reference List

Altschul,S.F., W.Gish, W.Miller, E.W.Meyers, and D.J.Lipman. 1990. Basic Local Alignment Search Tool. *Journal of Molecular Biology* **215**: 403-410.

Buetow,K.H., M.N.Edmonson, and A.B.Cassidy. 1999. Reliable identification of large numbers of candidate SNPs from public EST data. *Nat Genet* **21**: 323-325.

ca.expasy.org/cgi-bin/niceprot.pl?P00846. 2006. ExPasy.

Cargill M., D.Altshuler and J.Ireland .1999.Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nature Genetics **22**, 231–8.

Clarke,S.D., M.K.Armstrong, and D.B.Jump. 1990. Nutritional control of rat liver fatty acid synthase and S14 mRNA abundance. *J Nutr* **120**: 218-224.

Collins,D.W. and T.H.Jukes. 1994. Rates of Transition and Transversion in Coding Sequences since the Human-Rodent Divergence. *Genomics* **20**: 386-396.

Corradetti,M.N., K.Inoki, and K.L.Guan. 2005. The Stress-inducted Proteins RTP801 and RTP801L Are Negative Regulators of the Mammalian Target of Rapamycin Pathway. *J. Biol. Chem.* **280**: 9769-9772.

Danesch,U., W.Hoeck, and G.M.Ringold. 1992. Cloning and transcriptional regulation of a novel adipocyte-specific gene, FSP27. CAAT-enhancer-binding protein (C/EBP) and

C/EBP-like proteins interact with sequences required for differentiation-dependent expression. *J. Biol. Chem.* **267**: 7185-7193.

de Coo,I.F., H.J.Smeets, F.J.Gabreels, N.Arts, and B.A.van Oost. 1996. Isolated case of mental retardation and ataxia due to a de novo mitochondrial T8993G mutation. *Am J Hum Genet.* **58**: 636-638.

Dudov, K.P. and R.P. Perry. 1984. The gene family encoding the mouse ribosomal protein L32 contains a uniquely expressed intron-containing gene and an unmutated processed gene. *Cell* **37:** 457–468.

Dworkin-Rastl,E., A.Shrutkowski, and M.B.Dworkin. 1984. Multiple ubiquitin mRNAs during *xenopus laevis*development contain tandem repeats of the 76 amino acid coding sequence. *Cell* **39**: 321-325.

Eisenberg,E., K.Adamsky, L.Cohen, N.Amariglio, A.Hirshberg, G.Rechavi, and E.Y.Levanon. 2005. Identification of RNA editing sites in the SNP database. *Nucl. Acids Res.* **33**: 4612-4617.

Ellegren,H., B.P.Chowdhary, M.Johansson, L.Marklund, M.Fredholm, I.Gustavsson, and L.Andersson. 1994. A Primary Linkage Map of the Porcine Genome Reveals a Low Rate of Genetic Recombination. *Genetics* **137**: 1089-1100.

Ellisen,L.W., K.D.Ramsayer, C.M.Johannessen, A.Yang, H.Beppu, K.Minda, J.D.Oliner, F.McKeon, and D.A.Haber. 2002. REDD1, a Developmentally Regulated Transcriptional Target of p63 and p53, Links p63 to Regulation of Reactive Oxygen Species. *Molecular Cell* **10**: 995-1005.

Ewing,B. and P.Green. 1998. Base-Calling of Automated Sequencer Traces Using Phred.II.Error Probabilities. *Genome Res.* **8**: 186-194.

Ewing,B., L.Hillier, M.C.Wendl, and P.Green. 1998. Base-Calling of Automated Sequencer Traces Using Phred.I.Accuracy assessment. *Genome Res.* **8**: 175-185.

Fernandez-Salguero, P, S.M.G. Hoffman, S. Cholerton, H. Mohrenweiser, H. Raunio, A. Rautio, O.Pelkonen, J.Huang, W.E. Evans, J.R. Idle and F.J. Gonzalez.1995.
A genetic polymorphism in coumarin 7-hydroxylation: sequence of the human CYP2A genes and identification of variant CYP2A6 alleles. *Am. J. Hum. Genet.* **57**: 651-660.

Fronicke,L., B.P.Chowdhary, H.Scherthan, and I.Gustavsson. 1996. A comparative map of the porcine and human genomes demonstrates ZOO-FISH and gene mapping-based chromosomal homologies. *Mammalian Genome* **7**: 285-290.

Gordon,D., C.Abajian, and P.Green. 1998. Consed: A Graphical Tool for Sequence Finishing. *Genome Res.* **8**: 195-202.

Goureau,A., M.Yerle, A.Schmitz, J.Riquet, D.Milan, P.Pinton, G.Frelat, and J.Gellin. 1996. Human and Porcine Correspondence of Chromosome Segments Using Bidirectional Chromosome Painting. *Genomics* **36**: 252-262.

Gray,I.C., D.A.Campbell, and N.K.Spurr. 2000. Single nucleotide polymorphisms as tools in human genetics. *Hum. Mol. Genet.* **9**: 2403-2408.

Guryev,V., E.Berezikov, R.Malik, R.H.A.Plasterk, and E.Cuppen. 2004. Single Nucleotide Polymorphisms Associated With Rat Expressed Sequences. *Genome Res.* **14**: 1438-1443.

Hawken, R.J., J.Murtaugh, G.H.Flickunger, M.Yerle, A.Robic, D.Miland, J.Gellin, C.W.Beattie, L.Schook and L.J.Alexander. 1999. A first-generation porcine whole-genome radiation hybrid map. *Mammalian Genome* 10, 824-830.

Hughes,A.L. and M.Yeager. 1998. Natural selection at major histocompatibility complex loci of vertebrates. *Annual Review of Genetics* **32**: 415-435.

imprh.toulouse.inra.fr/. 2006. INRA.

James M.Prober, George L.Trainor, Rudy J.Dam, Frank W.Hobbs, Charles W.Robertson, Robert J.Zagursky, Anthony J.Cocuzza, Mark A.Jensen, and Kirk Baumeister. 1987. A System for Rapid DNA Sequencing with Fluorescent Chain-Terminating Dideoxynucleotides. *Science, New Series* **238**.

Kenmochi N., T.Kawaguchi, S.Rozen, E.Davis, N.Goodman, T.J Hudson, T. Tanaka and D.C.Page.1998. A map of 75 human ribosomal protein genes. *Genome Res.* **8**:509-523.

Lehninger,N.C, D.L.Nelson, M.M.Cox. 1993. Amino acid and peptides in *Principles of Biochemistry* pp. 109-133. W H Freeman & Co, New York.

Levanon,E.Y., E.Eisenberg, R.Yelin, S.Nemzer, M.Hallegger, R.Shemesh, Z.Y.Fligelman, A.Shoshan, S.R.Pollock, D.Sztybel, M.Olshansky, G.Rechavi, and M.F.Jantsch. 2004. Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nat Biotech* **22**: 1001-1005.

Liang L, M Zhao, Z Xu, KK Yokoyama and T Li. 2003. Molecular cloning and characterization of CIDE-3, a novel member of the cell-death-inducing DNA-fragmentation-factor (DFF45)-like effector family. *Biochem J* **370**: 195-203.

Liu,H.C. and H.C.Towle. 1994. Functional synergism between multiple thyroid hormone response elements regulates hepatic expression of the rat S14 gene. *Mol Endocrinol.* **8**: 1021-1037.

Marchler-Bauer,A., J.B.Anderson, P.F.Cherukuri, C.Weese-Scott, L.Y.Geer, M.Gwadz, S.He, D.I.Hurwitz, J.D.Jackson, Z.Ke, C.J.Lanczycki, C.A.Liebert, C.Liu, F.Lu, G.H.Marchler, M.Mullokandov, B.A.Shoemaker, V.Simonyan, J.S.Song, P.A.Thiessen, R.A.Yamashita, J.J.Yin, D.Zhang, and S.H.Bryant. 2005. CDD: a Conserved Domain Database for protein classification. *Nucl. Acids Res.* **33**: D192-D196.

Mastroianni, N., M. De Fusco, M. Zollo, G.Arrigo, O.Zuffardi, A.Bettinelli, A. Ballabio and G.Casari. 1996. Molecular cloning, expression pattern, and chromosomal localization of the human Na-Cl thiazide-sensitive cotransporter (SLC12A3). *Genomics* **35**: 486-493.

Milan,D., R.Hawken, C.Cabau, S.Leroux, C.Genet, Y.Lahbib, G.Tosser, A.Robic, F.Hatey, L.Alexander, C.Beattie, L.Schook, M.Yerle, and J.Gellin. 2000. IMpRH Server: an RH mapping server available on the Web. *Bioinformatics* **16**: 558-559.

Mishmar,D., E.Ruiz-Pesini, P.Golik, V.Macaulay, A.G.Clark, S.Hosseini, M.Brandon, K.Easley, E.Chen, M.D.Brown, R.I.Sukernik, A.Olckers, and D.C.Wallace. 2003. Natural selection shaped regional mtDNA variation in humans. *PNAS* **100**: 171-176.

Nickerson,D.A., V.O.Tobe, and S.L.Taylor. 1997. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucl. Acids Res.* **25**: 2745-2751.

Parham,P. and T.Ohta. 1986. Population Biology of Antigen Presentation by MHC Class I Molecules. *Science* **272**: 67-74.

pede.dna.affrc.go.jp/csnp/csnp_main.php. 2006. cSNP.

Picoult-Newberg,L., T.E.Ideker, M.G.Pohl, S.L.Taylor, M.A.Donaldson, D.A.Nickerson, and M.Boyce-Jacino. 1999b. Mining SNPs From EST Databases. *Genome Res.* **9**: 167-174.

Prasad, R., A. B. Zhadanov, Y. Sedkov, F. Bullrich, T. Druck, R. Rallapalli, T.Yano, H. Alder, C. M. Croce, K. Huebner, A. Mazo and E. Canaani. 1997. Structure and expression pattern of human ALR, a novel gene with strong homology to ALL-1 involved in acute leukemia and to Drosophila trithorax. *Oncogene* **15**: 549-560.

Rachel,J.H., M.Jeremy, H.F.Gail, Y.Martine, R.Anne, M.Denis, G.Joel, W.B.Craig, B.S.Lawrence, and J.A.Leeson. 1999. A first-generation porcine whole-genome radiation hybrid map. *Mammalian Genome* **10**: 824-830.

Rettenberger,G., C.Klett, U.Zechner, J.Kunz, W.Vogel, and H.Hameister. 1995. Visualization of the conservation of synteny between humans and pigs by heterologous chromosomal painting. *Genomics* **26**: 372-378.

Sanger,F., S.Nicklen, and A.R.Coulson. 1977. DNA Sequencing with Chain-Terminating Inhibitors. *PNAS* **74**: 5463-5467.

Shoshani,T., A.Faerman, I.Mett, E.Zelin, T.Tenne, S.Gorodin, Y.Moshel, S.Elbaz, A.Budanov, A.Chajut, H.Kalinski, I.Kamer, A.Rozen, O.Mor, E.Keshet, D.Leshkowitz, P.Einat, R.Skaliter, and E.Feinstein. 2002. Identification of a Novel Hypoxia-Inducible

Factor 1-Responsive Gene, RTP801, Involved in Apoptosis. *Mol. Cell. Biol.* **22**: 2283-2293.

Steve,R. and H.J.Skalets. 2006. Primer3 on the WWW for general users and for biologist programmers. *Bioinformatics Methods and Protocols: Methods in Molecular Biology* 365-386.Humana Press, Totowa,NJ.

Suh,Y. and J.Vijg. 2005. SNP discovery in associating genetic variation with human disease phenotypes. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **573**: 41-53.

Treacher Collins Syndrome Collaborative Group.1996. Positional cloning of a gene involved in the pathogenesis of Treacher Collins syndrome. *Nature Genet.* 1**2:** 130-136.

Wedekind,J.E., G.S.C.Dance, M.Sowden, and H.C.Smith. 2003. Messenger RNA editing in mammals: new members of the APOBEC family seeking roles in the family business. *Trends in Genetics* **19**: 207-216.

Wiborg, O, M.S.Pedersen , A.Wind , L.E.Berglund , K.A.Marcker and J.Vuust .1985.The human ubiquitin multigene family: some genes contain multiple directly repeated ubiquitin coding sequences. EMBO J: **4**, 755–759

www.idtdna.com/analyzer/Applications/OligoAnalyzer/. 2006. IDT.

www.mun.ca/biology/scarr/4241_Mar05,07/GG10_RHM.jpg. 2005.RH

www.ncbi.nih.gov/. 2006. NCBI.

www.ncbi.nih.gov/SNP/. 2006. NCBI.

www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html. 2006. NCBI.

www.ncbi.nlm.nih.gov/Omim/

www.ncbi.nlm.nih.gov/projects/genome/guide/cow/. 2006. NCBI.

www.ncbi.nlm.nih.gov/projects/genome/guide/human/. 2006. NCBI.

www.porkboard.org/docs/pkfacts2002.pdf. 2003. Pork facts 2002/2003.

Zhou,Z., S.Yon Toh, Z.Chen, K.Guo, C.Peng Ng, S.Ponniah, S.C.Lin, W.Hong, and P.Li. 2003. Cidea-deficient mice have lean phenotype and are resistant to obesity. *Nat Genet* **35**: 49-56.

# Appendix I

## Identified cSNPs in Porcine Adipose Tissue cDNA Library

| Contig | GenBank acession No. | Phred score | SNP site in the contig |
|---|---|---|---|
| 181 | EB420279 | 40 | 283 T C |
| 225 | EB417654 | 28 | 225 T C |
| 229 | EB414959 | 40 | 227 G A |
| | | 27 | 263 A G |
| 250 | EB415618 | 29 | 164 G A |
| 256 | EB415709 | 46 | 223 C T |
| 271 | EB411973 | 35 | 127 A G |
| 310 | EB412079 | 45 | 384 A G |
| | | 27 | 510 G A |
| | | 21 | 423 C T |
| | | 48 | 432,433 TG CC |
| 331 | EB421901 | 42 | 139 C T |
| 335 | EB415716 | 42 | 124 A G |
| | | 35 | 158 C T |
| 341 | EB412180 | 35 | 112 C T |
| | | 25 | 146 C T |
| 353 | EB417235 | 27 | 322 C A |
| 356 | EB418198 | 45 | 74 T C |
| 365 | EB415217 | 56 | 219 G A |
| 371 | EB421979 | 45 | 185 T C |
| 437 | EB422832 | 39 | 96 C T |
| 444 | EB416948 | 39 | 151 A G |
| 447 | EB420267 | 29 | 222 T G |
| 459 | EB416891 | 40 | 244 G A |
| | | 39 | 321 C T |
| 471 | EB421648 | 25 | 218 G A |
| 492 | EB411404 | 24 | 33 G A |
| 496 | EB419250 | 25 | 293 C T |
| | | 40 | 314 A G |
| 624 | EB419525 | >29 | 37,38 GG CC |
| | | | 45,47 GG CC |
| 632 | EB420544 | 29 | 61 A G |
| 651 | EB416452 | 25 | 75 A G |
| 658 | EB417075 | 25 | 45, 46 GG CC |
| 671 | EB419138 | 56 | 168 G A |
| | | 39 | 196 A G |
| 736 | EB424171 | 25 | 72 C T |
| | | 27 | 46 G A |
| 754 | EB423927 | 33 | 43 G T |

| | | | |
|---|---|---|---|
| | | 33 | 46 A G |
| 785 | EB416068 | 31 | 54 G A |
| 789 | EB418518 | 29 | 68 A G |
| 793 | EB413838 | 42 | 80 C T |
| | | 25 | 247 A T |
| 817 | EB416702 | 29 | 278 C G |
| 819 | EB421163 | 22 | 57 CG GA |
| 900 | EB424046 | 31 | 14 G C |
| | | >20 | 22 GG CC |
| 932 | EB420393 | 28 | 201 A G |
| 933 | EB419471 | 40 | 196 C A |
| 969 | EB416015 | >30 | 25,26 CC GG |
| | | 39 | 33 C G |
| 994 | EB412767 | 40 | 273 A T |
| 1013 | EB417318 | 51 | 175 G A |
| 1019 | EB411280 | 31 | 123 G A |
| 1091 | EB420993 | 37 | 64 C T |
| | | 56 | 73 T A |
| 1158 | EB412333 | 28 | 190 T C |
| 1168 | EB412292 | 56 | 222 C T |
| 1169 | EB414270 | 29 | 53 G C |
| 1184 | EB418674 | 36 | 289 G T |
| | | 32 | 298 C A |
| 1190 | EB423798 | 40 | 199 A G |
| 1193 | EB411546 | 42 | 147 A G |
| 1198 | EB417324 | 48 | 246 C T |
| 1222 | EB412994 | 32 | 274 G A |
| 1226 | EB411180 | 28 | 319 T G |
| 1246 | EB412482 | 33 | 171 G A |
| | | 46 | 90 T C |
| 1249 | EB421718 | 45 | 138 C A |
| | | 42 | 186 A G |
| 1265 | EB417892 | 28 | 74 G T |
| 1269 | EB420672 | 25 | 206 C G |
| 1271 | EB424463 | 45 | 185 CT |
| 1281 | EB413055 | 28 | 233 T C |
| 1286 | EB415145 | 29 | 257 G A |
| 1295 | EB422552 | 29 | 461 A G |
| 1297 | EB424353 | 45 | 62 G A |
| 1310 | EB421233 | 37 | 258 A G |
| 1323 | EB414030 | 27 | 132 G A |
| 1327 | EB421029 | 40 | 284 C T |
| 1332 | EB416263 | 56 | 150 G A |
| 1345 | EB412737 | 42 | 94 G A |
| 1354 | EB422516 | 28 | 188 A C |
| 1356 | EB421839 | 45 | 206 T C |
| 1362 | EB416712 | 25 | 12 C A |
| 1364 | EB421981 | 35 | 64 T C |

| | | | |
|---|---|---|---|
| 1375 | EB419118 | 37 | 124 G T |
| 1377 | EB414310 | 46 | 134 T C |
| 1382 | EB424362 | 37 | 273 T A |
| 1387 | EB421865 | 21 | 72 G A |
| 1397 | EB413221 | 33 | 84 A G |
| 1400 | EB416654 | 27 | 188 C G |
| 1413 | EB416865 | >30 | 20 GG CC |
| 1424 | EB419951 | 45 | 115 A C |
| | | 36 | 149 T C |
| 1427 | EB417156 | 27 | 232 A T |
| | | 27 | 259 T C |
| 1429 | EB414958 | 40 | 97 T C |
| 1441 | EB420362 | 45 | 101 C T |
| 1446 | EB412345 | 27 | 241 C G |
| 1447 | EB411278 | 42 | 952 A C |
| 1455 | EB415789 | 34 | 22 GG CC |
| 1456 | EB422783 | 40 | 503 G C |
| | | 40 | 503 G C |
| 1458 | EB412084 | 34 | 32 C A |
| | | 33 | 35, 36 CA AG |
| | | 42 | 48 A G |
| | | 53 | 53 T C |
| | | 44 | 59 T C |
| | | 42 | 62 G A |
| | | 56 | 66 T C |
| 1476 | EB414494 | 51 | 125 T C |
| 1485 | EB413832 | 29 | 291 C T |
| | | >25 | 70 CA TG |
| | | 46 | 181 A G |
| | | 46 | 186 T A |
| 1492 | EB412433 | 35 | 107 T C |
| 1507 | EB422865 | 40 | 175 176 CA TC |
| | | 40 | 175 176 CA TC |
| | | 42 | 262 T G |
| 1508 | EB422256 | 42 | 65 C T |
| | | 29 | 89 A C |
| | | 27 | 286 287 TG CA |
| 1520 | EB416738 | 40 | 280 CA TG |
| | | 35 | 159 C T |
| 1522 | EB419221 | 34 | 22 T C |
| 1531 | EB419164 | 37 | 238 C T |
| 1532 | EB424314 | 45 | 127 A G |
| 1549 | EB424152 | 30 | 149 T A |
| | | 25 | 329 T C |
| | | 30 | 155 G A |
| | | 30 | 329 T C |
| 1551 | EB419774 | 42 | 115 C G |
| 1552 | EB424419 | 26 | 97 G A |

| | | | |
|---|---|---|---|
| 1553 | EB421204 | 33 | 388 G A |
| | | 40 | 479 G A |
| 1555 | EB424413 | 46 | 116 G A |
| 1562 | EB411668 | 29 | 291 C T |
| 1567 | EB424555 | 45 | 97 C T |
| 1572 | EB424375 | 40 | 22 23 CC GG |
| 1580 | EB414413 | 42 | 89 90 TG CA |
| | | 35 | 94 G A |
| | | 30 | 104 T C |
| | | 42 | 107 A C |
| 1591 | EB421192 | 29 | 149 G A |
| 1602 | EB421157 | 31 | 589 T C |
| | | 40 | 606 G T |
| | | 40 | 608 A G |
| | | 29 | 613 GA CC |
| | | 29 | 653 A G |
| | | 29 | 667 G C |
| | | 32 | 673 G A |
| 1616 | EB415426 | 35 | 89 T C |
| | | 40 | 138 T C |
| 1639 | EB420466 | 37 | 253 G A |
| 1642 | EB423378 | 29 | 33 A G |
| | | 24 | 306 C T |
| | | 23 | 226 C T |
| 1649 | | 47 | 230 A G |
| | | 40 | 235 G A |
| | | >35 | 170 171 GT CC |
| 1652 | EB423134 | 36 | 1018 T C |
| 1657 | EB422280 | 40 | 208 T C |
| | | 40 48 | 239 240 CA TG |
| 1660 | EB422673 | 51 | 897 A G |
| 1668 | EB424235 | 46 | 242 G A |
| | | 26 | 292 G C |
| 1671 | EB420082 | 46 | 135 T C |
| 1674 | EB423385 | 55 | 55 A G |
| | | 42 | 412 AG |
| 1678 | EB424446 | 35 | 35 G A |
| | | 29 | 91 C T |
| | | 42 | 35 G A |
| | | 45 | 291 T A |
| | | 45 | 357 A G |
| 1685 | EB412932 | 42 | 142 T C |
| 1687 | EB419485 | 35 | 111 C T |
| | | 35 | 171 T C |
| 1688 | EB417170 | 45 | 469 A G |
| 1690 | EB421019 | 42 | 128 A G |
| 1701 | EB423068 | 27 | 83 84 AA TT |
| | | 44 | 71 CT |

| | | | |
|---|---|---|---|
| | | 37 | 128 AC |
| 1709 | EB419160 | 35 | A G 271 |
| 1722 | EB422631 | 29 | 327 A T |
| 1723 | EB424530 | 31 | T C 59 |
| 1726 | EB424436 | 39 | 78 G T |
| 1729 | EB420314 | 29 | 92 G A |
| | | 40 | 176 C T |
| 1732 | EB417515 | 35 | 461 T C |
| 1737 | EB423246 | 35 | 57 G A |
| | | 40 | 62 G C |
| | | 37 | 65 T G |
| | | 40 | 102 103 CG TA |
| | | 40 | 102 103 CG TA |
| 1739 | EB423122 | **21** | 67 T G |
| 1748 | EB419065 | 37 | 658 G C |
| 1754 | EB417726 | 30 | 149 T G |
| | | 40 | 176 T C |
| 1757 | EB417359 | 29 | 271 A C |
| 1772 | EB418141 | 37 | 220 G A |
| | | 33 | 749 T C |
| | | 44 | 220 G A |
| | | 48 | 253 A G |
| 1776 | EB423558 | 48 | 494 C G |
| | | 56 | 274 G C |
| 1779 | EB421151 | 37 | 48 T C |
| | | 35 | 49 T C |
| | | 45 | 71 T C |
| 1780 | EB423377 | 45 | 230 T C |
| | | 40 | 268 G C |
| | | 40 | 334 T C |
| 1784 | EB422716 | 34 | 99 A G |
| 1785 | EB418557 | 51 | 73 T C |
| 1790 | EB423275 | 56 | 455 C G |
| 1793 | EB422724 | 35 | 124 G A |
| 1797 | EB418023 | 42 | 168 A G |
| 1798 | EB424592 | 39 | 102 T C |
| 1801 | | 40 | 443 G A |
| | | 38 | 474 G A |
| | | 35 | 556 T A |
| 1803 | EB423670 | 45 | 753 G A |
| 1804 | EB422801 | 31 | 122 A G |
| 1808 | EB424279 | 46 48 | 568 TG CA |
| 1809 | EB423207 | 20 | 268 GA |
| | | 25 | 280 G A |
| 1810 | EB415586 | 36 | 60 G A |
| 1811 | EB416810 | 29 | 581 T C |
| 1815 | EB423064 | 37 | 136 G A |
| | | 50 | G A 198 |

| | | | |
|---|---|---|---|
| 1816 | EB424115 | 32 | 511 G A |
| | | 45 | 700 C T |
| | | 35 | 753 T C |
| | | 90 | 842 T C |
| 1818 | EB421701 | 42 | 568 C T |
| 1824 | EB415443 | 29 | G A 740 |
| 1830 | EB415792 | 42 | 111 A G |
| 1833 | EB422781 | 39 | 118 A G |
| 1834 | EB424337 | 35 | 72 A G |
| 1836 | EB423166 | 37 | 341 C G |
| 1837 | EB419622 | 43 | 802 CT |
| | | 51 | 829 C T |
| 1843 | EB419622 | 39 | 141 G C |
| | | 27 | 1410 C T |
| | | 42 | 1426 C G |
| | | 42 | 1426 C G |
| 1844 | EB420483 | 40 | 938 G A |
| 1848 | EB418225 | 51 | 233 A C |
| 1854 | EB422656 | 35 | 396 A G |
| | | 27 | 555 C G |
| 1856 | EB424099 | >30 | 283 284 CA TG |
| | | 35 | 292 T C |
| 1858 | EB422632 | 24 | 146 A G |
| | | 40 | 158 C G |
| | | >30 | 165-169 CCAC AGTC |
| | | 28 | 173 A G |
| 1859 | EB423205 | 45 | 101 G A |
| | | 42 | 106 G C |
| 1860 | EB419245 | 32 | 27 G C |
| 1862 | EB422742 | 40 | 107 G T |
| 1864 | EB415761 | 45 | C T 61 |
| | | 23 | 301 G C |
| | | 44 | 343 T C |
| 1865 | EB415239 | 42 | 1205 G A |
| 1866 | EB419990 | 27 | 290 C T |
| 1873 | EB424137 | 27 | 925 G A |
| | | 35 | 940 A G |
| 1874 | EB415351 | 31 | 445 A G |
| | | 46 | 451 T A |
| | | 44 | 487 C G |
| | | 26 | 493 C T |
| | | >25 | 473 CT TG |
| 1876 | EB422887 | 51 | 190 G A |
| | | 42 | 271 A G |
| | | 46 | 624 A C |
| | | 35 | 675 T C |
| 1877 | EB423208 | 37 | 924 A C |
| | | 48 | 930 A T |

| | | | |
|---|---|---|---|
| | | 51 | 958 T A |
| 1880 | EB418468 | 40 | G A 146 |
| 1883 | EB424147 | 34 | 73 C T |
| 1884 | EB422669 | 35 | 20 21 CC GG |
| | | 44 | 280 C T |
| 1886 | EB422296 | 27 | 565 G C |
| 1891 | EB424558 | 46 | C A 120 |
| 1892 | EB419146 | 33 | 850 C T |
| 1893 | EB422389 | 42 | 133 G A |
| 1898 | EB424285 | 33 | 116 G A |
| | | 29 | 923 G T |
| 1899 | EB424547 | 29 | 71 T G |
| 1900 | EB423963 | 45 | 394 G A |
| | | 30 | 1378 G A |
| 1902 | EB423436 | 29 | 563 G A |
| 1905 | EB415941 | 27 | 966 T C |
| 1907 | EB422497 | >35 | 12 CC GG |
| 1908 | EB414246 | 42 | 412 G A |
| 1909 | EB424542 | 33 | 187 C G |
| | | 37 | 430 A T |
| 1912 | EB424596 | 30 | 437 G A |
| | | 21 | 866 C T |
| | | 30 | 872 T G |
| | | 29 | 875 T C |
| | | 32 | 878 A G |
| | | 40 | 885 G C |
| | | 40 | 893 G A |
| | | 32 | 899 G A |
| | | 32 | 909 G A |
| | | 29 | 908 G A |
| | | 29 | 953 C T |
| | | 37 | 956 G A |
| 1916 | EB423576 | 32 | 713 G C |
| 1917 | EB422530 | 40 | 905 T C |
| 1922 | EB424430 | 37 | 149 T C |
| | | 27 | 177 T G |
| | | 39 | 358 G A |
| 1923 | EB420751 | 40 | 1041 C T |
| 1926 | EB419012 | 29 | 276 G A |
| 1927 | EB421895 | 35 | 41 T C |
| | | 33 | 522 A C |
| 1933 | EB423752 | 45 | 288 C T |
| | | 34 | 830 A T |
| 1934 | EB423484 | 56 | 225 T C |
| 1935 | EB424199 | 45 | 187 C T |
| 1939 | EB423731 | 29 | 692 A C |
| 1943 | EB423501 | 32 | 397 C T |
| 1944 | EB423600 | 51 | 116 T G |

| | | | |
|---|---|---|---|
| | | 39 | 231 A C |
| 1945 | EB423004 | 40 | 185 G A |
| 1947 | EB423442 | 34 | 316 A G |
| 1949 | EB423137 | 48 | 1347 A T |
| 1951 | EB423290 | 32 | 367 C T |
| 1952 | EB424568 | 51 | 307 A T |
| | | 56 | 310 A T |
| 1953 | EB424528 | 29 | 372 C T |
| 1954 | EB424437 | 40 | 158 C G |
| 1956 | EB422828 | 26 | 790 C T |
| 1957 | EB422123 | **45** | 100 G A |
| 1958 | EB423762 | 46 | 219 G A |
| | | 42 | 644 G A |
| 1959 | EB413777 | 51 | 290 A G |
| | | 26 | 504 C T |
| 1961 | EB424178 | 51 | 213 C G |
| | | 32 | 1132 T C |
| 1965 | EB422787 | 30 | 763 T C |
| 1966 | EB424145 | 29 | 172 C T |
| 1967 | EB422754 | 22 | 196 C G |
| 1972 | EB422077 | 34 | 61 G C |
| | | 51 | 670 T C |
| 1973 | EB424208 | 44 | G A 132 |
| 1976 | EB424442 | 42 | 224 A G |
| 1980 | EB422872 | 37 | 66 C T |
| | | 35 | 666 C A |
| 1982 | EB423646 | 23 | G C 193 |
| 1989 | EB423906 | 20 | 251 G A |
| 1990 | EB424375 | 31 | 207 G A |
| | | 22 | 325 T G |
| 1991 | EB424369 | 40 | 438 C T |
| 1994 | EB419129 | 37 | 286 C G |
| | | 33 | 177 G A |
| 1996 | EB424530 | 20 | 481 A G |
| | | 29 | 523 524 CC GG |
| 1997 | EB421955 | 40 | 440 A G |
| 1998 | EB424345 | 40 | 631 T C |
| | | 35 | 808 G A |
| 2000 | EB419283 | 46 | 476 C T |
| 2001 | EB423879 | 37 | 36 C A |
| 2002 | EB424587 | 28 | 206 A G |
| 2007 | EB421715 | 48 | 348 T C |
| | | 51 | 1913 C T |
| | | 56 | 1916 C T |
| 2008 | EB420457 | 36 | 766 C T |
| 2009 | EB421259 | 34 | 108 T G |
| | | 29 | 594 A G |
| | | 51 | 784 T C |

| | | | |
|---|---|---|---|
| | | 40 | 812 C T |
| | | 44 | 830 G A |
| 2010 | EB424450 | 33 | 215 G A |
| | | 35 | 528 C T |
| 2011 | EB411744 | 37 | 538 T A |
| | | 33 | 539 G C |
| | | 29 | 612 G T |
| | | 28 | 615 C T |
| | | 27 | 618 C T |
| | | 35 | 661 C G |
| | | 35 | 661 C T |
| | | 35 | 664 C T |
| | | 42 | 723 T C |
| | | 39 | 730 C T |
| 2012 | EB424333 | 29 | 237 C A |
| 2013 | EB424216 | 41 | 250 G A |
| 2014 | EB424473 | 43 | 312 A G |
| 2016 | EB424555 | 45 | 149 C T |
| | | 42 | 843 G A |
| 2017 | EB424293 | 29 | 433 T G |
| | | 40 | 1662 C T |
| 2019 | EB423033 | 56 | 225 G C |
| 2020 | EB424501 | 30 | 161 G A |
| | | 32 | 305 G C |

# Appendix II

## List of Genes with Non Synonymous SNPs and the Created Amino Acid Substitution

| Contig | Gene Name | Amino Acid Substitution |
|---|---|---|
| 969 | PREDICTED: Canis familiaris similar to ribosomal protein L36 (LOC476738), mRNA | D/E |
| 994 | PREDICTED: Canis familiaris similar to DNA-binding protein inhibitor | R/T |
| 1169 | PREDICTED: Bos taurus similar to cell death-inducing DFFA-like effector a isoform 2 | E/Q |
| 1184 | PREDICTED: Bos taurus similar to Cytochrome P450 4F6 | D/E |
| 1190 | Sus scrofa clone Clu_54069.scr.msk.p1.Contig1, mRNA sequence | L/F |
| 1198 | Sus scrofa clone Clu_5579.scr.msk.p1.Contig1, mRNA sequence | L/F |
| 1246 | Bos taurus prosaposin (PSAP), mRNA | P/S |
| 1265 | Human DNA sequence from clone RP11-284F21 on chromosome 1 | R/H |
| 1281 | Bos taurus ATP synthase, H+ transporting, mitochondrial F1 complex, | L/F |
| 1310 | Bos taurus glycerol-3-phosphate dehydrogenase 1 (soluble) (GPD1), mRNA, complete cds | I/V |
| 1413 | Sus scrofa integral membrane protein 2B mRNA, complete cds | G/E |
| 1424 | Sus scrofa MHC class I antigen (SLA-1) mRNA, SLA-1*mel7 allele, partial cds | S/Y |
| 1475 | Homo sapiens fumarylacetoacetate hydrolase (fumarylacetoacetase) (FAH), mRNA | D/E |
| 1549 | Bos taurus thyroid hormone-responsive protein (THRSP) mRNA, complete | D/E |
| 1552 | Sus scrofa clone Clu_27494.scr.msk.p1.Contig2 | Q/E |
| 1602 | Sus domesticus Ig rearranged lambda chain mRNA V-J-C-regions, 3' end | I/T, F/W, IL/TI |
| 1639 | Sus scrofa complement C1r mRNA, partial cds | D/E |
| 1649 | Sus scrofa clone rcki02_o9.y1.abd, mRNA sequence | E/G |
| 1678 | PREDICTED: Bos taurus similar to flightless I homolog, transcript variant 1 (LOC514446), mRNA | I/L |
| 1685 | Sus scrofa clone rill301b_d24.y1.abd, mRNA sequence | L/P |
| 1726 | Bos taurus cDNA clone IMAGE:30957751 | D/Y |
| 1776 | Gekko japonicus BS011P (BS011) mRNA, complete cds | A/P |
| 1808 | Sus scrofa breed Meishan MHC class I antigen (SLA-1) mRNA, SLA-1*wxd allele, complete cds | D/E |
| 1810 | Bos taurus ribosomal protein S14 (RPS14), mRNA | R/H |
| 1834 | PREDICTED: Canis familiaris aldolase C, transcript variant 1 (LOC480622), mRNA | Y/C |
| 1854 | PREDICTED: Bos taurus similar to Cell death activator CIDE-3 transcript variant | R/Q, S/T |
| 1859 | Sus scrofa clone rnbm14b_e9.y1.abd, mRNA sequence | P/A |
| 1860 | Bos taurus RPL13 protein-like, mRNA (cDNA clone MGC:127884 IMAGE:7961743),complete cds | L/V |
| 1862 | Canis familiaris similar to nascent polypeptide-associated complex alpha polypeptide mRNA | Q/H |
| 1898 | Sus scrofa COX7A1 gene, CAPNS1 gene, CKAP1 gene, POLR2I gene and CLIPR-59 gene | T/N |
| 1927 | S.scrofa mRNA encoding G-beta like protein | Y/S |
| 1934 | Sus scrofa partial mRNA for heart fatty acid-binding protein (FABP3 gene) | T/I |
| 1935 | Sus scrofa 40S ribosomal protein S17 mRNA, complete cds | A/V |
| 1952 | PREDICTED: Canis familiaris similar to nuclease sensitive element binding protein 1 mRNA | FF/YY |
| 1972 | PREDICTED: Bos taurus similar to cell death-inducing DFFA-like effector a isoform 2 mRNA | E/D |
| 1989 | Sus scrofa 40S ribosomal protein S20 mRNA, partial cds | L/V |
| 1990 | Bos taurus ribosomal protein, large P2 (RPLP2), mRNA | D/E |
| 1991 | Sus scrofa mitochondrion, complete genome | Y/H |
| 1994 | Sus scrofa 60S ribosomal protein L32 mRNA, complete cds | N/D, V/L |
| 1997 | Sus scrofa mitochondrion, complete genome | Q/R |
| 2009 | Sus scrofa glutathione peroxidase 3 (GPX3) mRNA, partial cds | R/L, H/R |
| 2016 | Sus scrofa ferritin heavy-chain (FTH1), mRNA | GPP |
| 2017 | Sus scrofa partial mRNA for elongation factor 1-alpha 1 (EF1A1 gene) | L/R |

VITA

**Y.A.M.S Wickramasinghe**

**Candidate for the Degree of**

**Master of Science**

Thesis: SINGLE NUCLEOTIDE POLYMORPHISM ANALYSIS OF GENES EXPRESSED IN PORCINE ADIPOSE TISSUE

Major Field: Animal science

**Education**

**2004-2006** **:** Completed the requirements for Master of Science degree with a major in Animal Science at Oklahoma State University in July 2006

**1998-2003** : Bachelor of Veterinary medicine degree from University of Peradeniya Sri Lanka *(Second classs –Honors)*

**Experience :**

**Feb 2003-July 2004** Lecturer in department of Basic Veterinary Sciences, Faculty of Veterinary Medicine and Animal Science, University of Peradeniya, Sri Lanka.

**January 10$^{th}$ 2005 – May 5$^{th}$ 2005**: Graduate Teaching Assistant in Animal Genetics

**August 22$^{nd}$ 2005 –Dec 09 2005:** Graduate Teaching Assistant in Introduction to the Animal Sciences

**August 23$^{rd}$ 2004 –to date**: Graduate Research Assistant in research Laboratory of Dr Udaya DeSilva at Oklahoma State University.

**Awards and Honors**

: Life time member of Phi Kappa Phi National Honor society (2005)

: Dr. Dewansa Senevirarne Memorial Gold Medal awarded for best performance in Veterinary Public Health (2000)

: D.Seneviratne Memorial Award for Excellence in Veterinary Public Health (2000)

Name: Y.A.M.S.Wickramasinghe Date of Degree: July, 2006

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: SINGLE NUCLEOTIDE POLYMORPHISM ANALYSIS OF GENES EXPRESSED IN PORCINE ADIPOSE TISSUE

Pages in Study: 150 Candidate for the Degree of Master of Science

Major Field: Animal Science

Scope and Method of Study: Main objective of this study was identification of SNPs that can be used as direct markers in QTL studies and high throughput genotyping programs. A subcutaneous adipose tissue cDNA library was developed from $F_2$ progeny of Duroc x Pietrain resource population. Fifteen thousand clones were sequenced from this library and 13958 sequence trace files were clustered in to 2080 unique clusters. Gene Ontology classifications were obtained for 330 of these sequences. Sequence trace files were analyzed for SNPs by bioinformatics programs, Polyphred, Phrap and Consed. Twenty three non synonymous SNPs were further analyzed for their segregation patterns in $F_0$ and $F_2$ populations. Four genes with confirmed SNPs were mapped to the swine genome using INRA Radiation hybrid (RH) mapping tool.

Findings and Conclusions: Non normalized porcine adipose tissue cDNA library showed 65% redundancy. Gene Ontology analysis indicated the library to be significantly enriched with gene products for catalysis, binding and signal transduction. There were total of 362 coding SNPs above Phred quality score of 20 in the adipose tissue cDNA library. Transition polymorphisms were 2.6 times higher than the trnasversions and 88% of SNPs were synonymous. Eleven non synonymous coding SNPs were confirmed by resequencing and none of them segregated exclusively in an outlier group for back fat thickness. RTP801, CIDE-3, cytochrome P456 F6 subunit and solute carrier family 12 member 3 genes were successfully mapped to the porcine genome by RH mapping technique. After validation and calculation of frequencies in large swine populations these SNPs can be used as direct markers in high throughput genotyping programs.

ADVISER'S APPROVAL: Dr. Udaya DeSilva