Institut für Tierwissenschaften, Abt. Tierzucht und Tierhaltung

der Rheinischen Friedrich – Wilhelms – Universität Bonn

**Identification and validation of functional candidate genes related to the
inverted teat defect in pigs**

I n a u g u r a l – D i s s e r t a t i o n

zur Erlangung des Grades

Doktor der Agrarwissenschaft

(Dr. agr.)

der

Hohen Landwirtschaftlichen Fakultät

der

Rheinischen Friedrich – Wilhelms – Universität

zu Bonn

vorgelegt im Oktober 2008

von

Saowaluck Yammuen-Art

aus

Tak, Thailand

Dedicated to my beloved parents and all members of my family

Identifizierung und Validierung funktioneller Kandidatengene für den Stülpzitzendefekt beim Schwein

Die Stülpzitze ist der häufigste Erbfehler des Gesäuges von Schweinen. Die Anzahl der Gene, die an der Entwicklung dieses Defektes beteiligt sind, ist unbekannt. Das Ziel dieser Untersuchung war die Erstellung eines genomweiten Genexpressionsprofils in Geweben des Gesäuges von Sauen während der Laktation, die Auswertung der Unterschiede zwischen normalen und Stülpzitzen und die Charakterisierung von funktionellen Kandidatengenen für den Stülpzitzendefekt beim Schwein. Proben von normalen und Stülpzitzen wurden von laktierenden Sauen gesammelt. Die RNA wurde extrahiert und für die Microarray Analysen mit genomweiten porcinen Affymetrix-Arrays verwendet. Proben wurden weiterhin mit einer real-time PCR untersucht, um die Ergebnisse der Expressionanalyse von ausgewählten Genen zu verifizieren. Zusätzlich wurden drei Gene aufgrund ihrer beschriebenen Funktion in Publikationen ausgewählt. Ihre Expressionsprofile wurden erstellt und die Assoziation mit der Gesäugequalität in experimentellen und kommerziellen Schweinepopulationen analysiert.

Die genomweite Expressionanalyse ergab beim Vergleich von normalen und Stülpzitzen 1253 unterschiedlich exprimierte Transkripte, von diesen waren 695 Transkripte stärker und 558 geringer exprimiert in normalen im Vergleich zu Stülpzitzen. Die Validierung von fünf ausgewählten Genen mit eine real-time PCR zeigte die signifikant stärkere Expression des *CTGF* (Connective tissue growth factor) und des *IGF-II* (Insulin-like growth factor 2) Gens in Stülpzitzen, während *GDF8* (Growth differentiation factor 8) in normalen Zitzen stärker exprimiert war. Für die beiden Gene *EGF* (Epidermal growth factor) und *EGFR* (Epidermal growth factor receptor) konnte die unterschiedliche Expression nicht bestätigt werden. Zusammenfassend zeigt diese Untersuchung die mögliche Bedeutung der Kandidatengene *CTGF*, *IGF-II* und *GDF8* für den Stülpzitzendefekt beim Schwein. Es konnte ebenfalls mit einer real-time PCR nachgewiesen werden, dass zwei der aus der Literatur ausgewählte Gene *RLN3* (Relaxin 3) und *GPCR135* (G-protein coupled receptor 135) in Zitzen von betroffenen Sauen stärker exprimiert waren im Vergleich zu Geweben von Sauen ohne Gesäugedefekt. Es wurden jeweils zwei Polymorphismen in den Sequenzen von *RLN3* und *GPCR135*, detektiert, eine signifikante Assoziation wurde zwischen den Haplotypen des *RLN3* Gens zu der Stülpzitze nachgewiesen.

Identification and validation of functional candidate genes related to the inverted teat
defect in pigs

The inverted teat defect is the most common disorder of the teat in pigs. The number of
genes involved in the development of this disorder is unknown. The aim of this study
was to investigate the genome-wide gene expression profile in porcine teats at the
lactating stage, to evaluate the differences between normal and defect teats in lactating
sows, and to investigate functional candidate genes for the inverted teat defect in pigs.
Samples of normal and defect teats were collected from lactating sows (n=2). RNA was
extracted and further used for microarray analysis performed using genome-wide
porcine Affymetrix arrays. Real-time PCR was performed to validate the expression of
selected genes. Three additional genes were selected from their described function in
the literature to analyse their expression profiles and to perform an association study
using samples of animals from an experimental and commercial pig population.

The genome-wide expression analysis revealed 1253 differentially expressed transcripts
between normal and inverted teats, of which 695 transcripts were found being higher
and 558 transcripts lower expressed in normal compared to inverted teats. Validation of
five selected genes using real-time PCR revealed a significantly higher expression of
connective tissue growth factor (*CTGF)* and insulin-like growth factor 2 *(IGF-II)* in
inverted teats, whereas growth differentiation factor 8 (*GDF8)* was highly expressed in
normal teats. For both epidermal growth factor *(EGF)* and epidermal growth factor
receptor *(EGFR)* no differentially expression could be verified. In conclusion, this study
promotes the functional candidate genes *CTGF*, *IGF-II*, and *GDF8* as candidates for the
inverted teat defect in pigs. Also for two of the genes selected from the literature,
relaxin 3 (*RLN3)* and G-protein coupled receptor 135 (*GPCR135*) were higher
expressed in teats from affected sows compared to normal sow using real-time PCR.
Two polymorphisms were detected in *RLN3* and *GPCR135*, a significant association
between the haplotype of *RLN3* and the inverted teat defect was found.

Contents   page

List of abberations

2-DE	: Two-dimensional gel electrophoresis
A260	: Absorbance at 260 nm wavelength (UV light)
ADAM17	: A Disintegrin and Metalloproteinase Protein
AREG	: Amphiregulin
BMP	: Bone morphogenetic
bp	: Base pairs
BSA	: Bovine serum albumin
BTC	: Betacellulin
C/EBP	: CCAAT/enhancer binding protein
cAMP	: Cyclic adenosine monophosphate
CCL28	: Chemokine ccl28
CCND1	: Cyclin D1
CLU	: Complement Cytolysis Inhibitor
CSF-1	: Colony-stimulating factor 1
CSN1S1	: Alpha-S1 casein
CSN1S2	: Alpha-S2 casein
Csn2	: Beta casein
CTGF	: Connective tissue growth factor
CXCL12	: CXCL12 chemokine
CYR61	: Cysteine-rich, angiogenic inducer 61
ddH$_2$O	: Distilled & deionized water
Dkk1	: Dickkopf 1
DMSO	: Dimethyl sulfoxide
DNA	: Deoxynucleic acid
dNTP	: Deoxyribonucleoside triphosphate
dNTP	: Dideoxyribonucleoside triphosphate
DEPC	: DiethylenePyrocarbonate
DTT	: Dithiothreitol
E	: Estrogen
E.coli	: Escherichia coli
ECM	: Extracellular matrix

EDTA	:	Ethylenediaminetetraacetic acid
EEF1A1	:	Eukaryotic translation elongation factor 1 alpha 1
EGF	:	Epidermal growth factor
EGFR	:	Epidermal growth factor receptor
ELISA	:	Enzyme linked immunosorbent assay
EPGN	:	Epigen
EPIR	:	Epiregulin
ERK	:	Extracellular signal-regulated kinase
ESR	:	Estrogen receptor
FBAT	:	Family based association tests
FBN1	:	Fibrillin 1
FGF	:	Fibroblast growth factor
FGF2	:	Fibroblast growth factor 2
FGFR2B	:	Fibroblast growth factor receptor 2B
FSHB	:	Follicle-stimulating hormone beta
FSHR	:	Follicle-stimulating hormone receptor gene
g	:	Gravity
GC	:	Vitamin D binding protein (Group-specific component)
GDF8	:	Growth differentiation factor 8
GDP	:	Guanosine diphosphate
GH	:	Growth hormone
GHR	:	Growth hormone receptor
GHRHR	:	Growth hormone releasing hormone receptor
GPCR135	:	G protein coupled receptor 135
GPCR142	:	G protein coupled receptor 142
GPRC5B	:	G protein-coupled receptor
GTP	:	Guanosine-5'-triphosphate
GTPγS	:	Guanosine 5'-O-[gamma-thio]triphosphate
HB-EGF	:	Heparin-binding epidermal growth factor
HMWP	:	High molecular weight proteins
HRF	:	Hista-mine-releasing factor
HUVE	:	Human umbilical vein endothelial
IGFBP3	:	Insulin like growth factor BP3

IGF-II	: Insulin-like growth factor 2
IL10	: Interleukin 10
IL-18	: Interleukin -18
INS	: Insulin
INSL	: Insulin-like peptides
IPTG	: Isopropylthio-ß-D-galactoside
JNK	: C-jun N-terminal kinase
LEF1	: Lymphoid enhancer factor 1
LEPR	: Leptin receptor
LGR7	: Leucin-rich repeat-containing G protein-coupled receptor 7
LH	: Luteinizing hormone
LYZ	: Lysozyme
MAPK	: Mitogen-activated protein kinase
MC2R	: Adrenocorticotropic hormone receptor
MFGE8	: Milk fat globule-EGF factor 8 (pp47 protein)
MLCK	:.Myosin light chain kinase
MP	: Metalloproteinase
mRNA	: Messenger RNA
MS	: Mass spectrometry
MSTN	: Myostatin
MyoD	: Myogein differentiationantigen
NaCl	: Sodium chloride
NOV	: Nephroblastoma overexpressed
P	: Progesterone
PCR	: Polymerase chain reaction
PDGF	: Platelet-derived growth factor
PGRMC1	: Steroid membrane binding protein
PIGR	: Poly-Ig receptor
PI-PLCb	: Phosphoinositide-specific phospholipase C
PKC	: Protein kinase C
PRL	: Prolactin
PRLR	: Prolactin receptor
pro*TGFA*	: Precursor of transforming growth factor-α

PTGDS	: Prostaglandin D2 synthase
PTHLH	: Parathyroid-hormone-like hormone gene
PTHR1	: Parathyroid hormone receptor
PTMs	: Post-translational modifications
PVN	: Paraventricular nucleus
QTL	: Quantitative trait loci
RANKL	: Receptor activator for nuclear factor κ B ligand
RFLP	: Restriction fragment length polymorphism
RLN1	: Relaxin 1
RLN3	: Relaxin 3
RNA	: Ribonucleic acid
RTK	: Receptor tyrosine kinase
S27a	: Ribosomal protein
SDS	: Sodium dodecyl sulfate
SIFT	: Sorting intolerant from tolerant
SLA	: Swine leucocyte antigen
SLS	: Sample loading solution
SNP	: Single nucleotide polymorphism
SOD1	: Superoxide dismutase 1
SOD2	: Superoxide dismutase (Mn type)
SON	: Supraoptic nucleus
STAT3	: Signal transducer and activator of signalling-3
STAT5	: Signal transducer and activator of signalling-5
STE	: Estrogen sulfotransferase
TAE	: Tris-acetate buffer
TBE	: Tris- borate buffer
TBX3	: T-box 3
TE	: Tris- EDTA buffer
TEBs	: Terminal end buds
TGFA	: Transforming growth factor alpha
TGFB1	: Transforming growth factor beta1
TGFBR3	: Transforming growth factor-beta type III receptor
TM	: Transmembrane domain

TSHB	: Thyroid stimulating hormone beta
U	: Units
UV	: Ultra-violet light
VEGF-A	: Vascular Endothelial Growth Factor
VIPR1	: Vasoactive intestinal peptide receptor
W	: Watts
WAP	: Whey acidic protein
Wnt	: Wingless-type MMTV integration site family
X-gal	: 5-Bromo-4-chloro-3-indolyl-beta-D-galactoside

List of tables

List of figures

1   Introduction

The inverted teat defect has a considerably negative impact on the economical efficiency of pig production in Germany (Oltmanns 2003). This heritable defect is characterized by the failure of nipples to protrude from the udder surface (Große et al. 1996). The inverted teat defect obviously decreases the number of functional teats and results in an increase of the mortality rate of piglets, especially in large litters. It has also a negative impact on the possibility of selection regarding other economically important traits. The number of genes involved in the defect is unknown, it is important to get more knowledge of the genetic cause of the inverted teat defect.

Candidate genes can be identified in QTL regions as positional candidates, or selected as functional candidates based on their biochemical and physiological properties and function in metabolism (Stratil and Geldermann 2004, Tabor et al. 2002, Zhu and Zhao 2007). A number of candidate genes are already identified using QTL analyses and the direct candidate gene approach (Jonas et al. 2008, Oltmanns 2003, Ün 2002). The association of some genes with the inverted teat defect could be described in pigs of an experimental and commercial populations such as relaxin 1 (*RLN1*) (Chomdej 2005). However these association studies could not reveal a strong evidence for any of the analyzed genes being a major gene affecting this heritable defect in pigs (Chomdej 2005, Trakooljul 2004, Yammuen-art et al. 2007b). The application of functional genomics using expression profiling is perhaps the most promising approach to identify functional candidate genes. Understanding patterns of gene expression is expected to provide an insight into the complex regulatory networks and will most probably lead to the identification of genes relevant to new biological processes, or genes implicated in disease (Nygard et al. 2007).

The following two different strategies functional candidate genes for the inverted teat defect were selected from expression profiling of samples of the porcine mammary gland and due to their biological metabolism as described in the literature. Using microarray analysis to detect functional candidate genes, differentially expressed genes were identified using samples of normal and inverted teats from sows with or without defect.

The present study was undertaken with the following objectives:

(1) to identify differentially expressed genes between normal and inverted teats in pigs

(2) to quantify the transcript abundance of differentially expressed genes between normal and inverted teat

(3) to detect single nucleotide polymorphisms and analyze the association of selected candidate genes with the inverted teat defect in pigs

2   Literature review

2.1   The mammary gland

2.1.1   Development of the mammary gland

In the early embryo, three layers can be found including the ectoderm, the mesoderm, and the endoderm. The ectoderm (outer layer) gives rise to the skin (epidermis) and the nervous system; the mesoderm (middle layer) develops to muscles, blood vascular system, and sexual organs; and the endoderm (inner layer) gives rise to the alimentary canal and the digestive glands. The mammary gland is originated from the two layers ectoderm and mesoderm (Schmidt 1971).

The development of the mammary gland begins during the early fetal stage and proceeds beyond the initiation of the lactation. The first sign of the development of the mammary gland is the thickening of the ectodermal cells (embryonic skin) on the ventral surface of the embryo giving rise to the mammary band. Serial changes in the thickened area of the ectoderm are identified as the mammary band, mammary streak, mammary line, mammary crest, mammary hillock, and mammary bud (Figure 1).
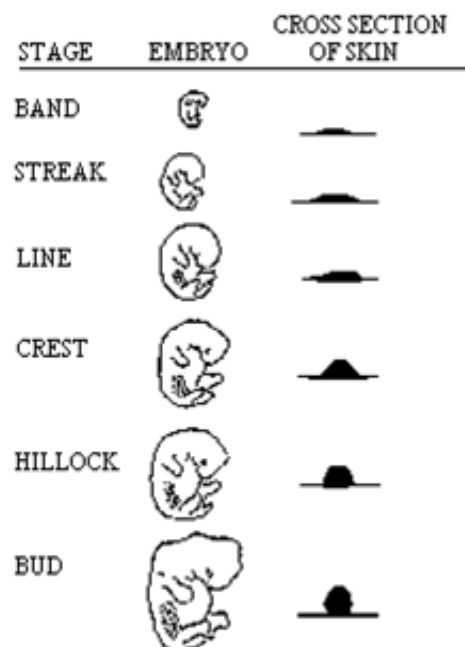
Figure 1:   Developmental stages from the band through the bud of the mammary apparatus of the embryo. The relation of the embryo size to the surface of the skin is presented (adapted from the review of Schmidt (1971))

The first visible indication of the mammary gland development can be seen during the proliferation of areas of epithelial cells in the epidermis of the ventral region. The small dissections which can be found at the upper body side are called mammary band, they are visible at approximatly day 23 in pigs. These areas of proliferation extend in a line between the fetal axilla and the inguinal region and form two indistinct bands called the mammary lines. Starting at day 28, the mammary lines regress back to distinct areas called mammary hillocks, which are visible around day 40 (Willham and Whatley 1963). The mammary hillocks represent solid epithelial masses (the mammary buds) that continue to evaginate into the underlying stroma and become surrounded by a more cellular zone of fibroblast like cells within a dense collagenous mesenchyme. Upon cell multiplication, aggregation and differentiation, four configurations called mammary buds are formed. The mammary buds are representing the end of the teat in the area where in later stages the streak canal can be found. The mammary glands in all species arise from the mammary bud. Each mammary bud will give rise to the primary sprout of the gland through cell differentiation. The primary sprout is further developing to a wide net of sprouts with the opening in the teat channel, which are leading the milk from the cistern to the outside of the body in later stages of the development. These milk leading channels are also known as galactophores. The number of galactophores per teat or nipple varies considerably among mammals. The galactophores are predetermined by the primary sprout growing from the mammary bud. The cells are multiplied and branched from the primary sprout, giving rise to the secondary sprouts. These are the structural features destined to become the mammary ducts. The sprouts become canalized shortly before birth (Schmidt 1971). A schematic overview of the channel development is shown in figure 2. Another necessary part of the successful progression of mammogenesis is an adequately developed mammary fat pad (Schmidt 1971). The fatty pad arises from the mesoderm during the fetal development. The nonglandular portion of the udder, mainly connective tissue, is well developed at birth (Schmidt 1971).

The growth of the mammary gland from the birth to the puberty is in most mammals isometric compared to the growth rate of the body (Fenton 2006).
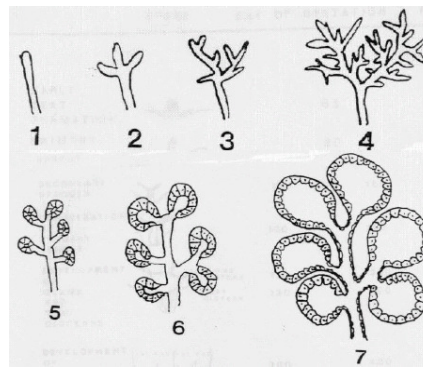
Figure 2:   Schematic overview of the proliferation of the gland from the primary sprout
            (1, 2) and the secondary sprouts (3, 4) before birth to the fully formed gland
            cells (5 to 7) during lactations (Schmidt 1971)

During puberty the ductal epithelium undergoes an active proliferation, ramifying into
the surrounding tissue of the mammary fat pad (MFP). It establishs a ductal network
that ultimately supports pregnancy associated alveolar development. Terminal end buds
mediate the elongation and primary branching of ductal epithelium through the fat pad
at this stage. In the mammary gland of non pregnant females, a large fatty pad still
exists. The major portion of the mammary gland growth occurs during the pregnancy,
especially the last part and it is controlled by different hormones. The growth of the
mammary gland is slow at the beginning of pregnancy, but the rate of growth
accelerates as the pregnancy advances. As pregnancy progresses, the adipose cells of
the pad are gradually replaced by ducts, alveoli, blood vessels, and connective tissue.
Most of the mammary growth during the first half of gestation is mainly ductal growth
and lobular formation; the alveoli are not formed before the pregnancy is established. In
the second half of the gestation, ductal growth continues, but the main part of the
growth is lobuloalveolar. Eventhough with the start of the lactation the mammary gland
is fully formed, the development of the mammary gland will continue during further
reproductive cycles as at the end of the lactation involution leads to an apotosis of the
mammary cells. The mammary gland continues formation, apoptosis and
morphogenesis during every new cycle of pregnancy and lactation (Ford et al. 2003).

2.1.2   The mammary gland in pigs

The mammary glands in pigs are parallel on each side of the ventral midline. The secretory tissue of each teat is separate and independent to the neighbouring teats. There are two lactiferous ducts in each teat complex that branch into smaller ducts and ductules, which in turn terminate in the alveoli. Each visible gland is composed of two simple glands, each with a streak canal in the teat. Therefore the teat of the pig has two orifices which are aligned anterior-posterior in the teat end (Schmidt 1971). Each simple gland is separated from the other, thus implements that the secretory tissue is independent from the adjacent gland (Schmidt 1971).

The sow has in average 12 to 14 mammary gland complexes (6 to 7 teats on each side). The numbers of nipples can range in various breeds between 6 and 32 teats (Schmidt 1971). In europe mainly commercially used dam lines Large White and different Landrace lines were found with an average of 13 to 14 teats per animal (Clayton et al. 1981). Asian breeds have often a higher number of teats, in Meishan and Jiaxing pigs an average of 17 and 19.8 teats were found, respectively (Fernandez et al. 2004, Haley et al. 1987).

2.1.3   Genetic background of teat characteristic

A range of the heritability estimated for the number of teats in pigs is wide. Heritabilities between 0.07 and 0.79 are published, most were detected in an interval from 0.30 to 0.50 (Hirooka et al. 2001, Lee and Wang 2001, Seo et al. 1996, Wang et al. 2000, Zhang et al. 2000). In another study the heritability for the number of teats on the left side was estimated around 0.20 while the heritability of the number of teats on the right side was around 0.18 (Borchers et al. 2002). The estimated values may be considerable different as the heritabilities are population and model dependent. The phenotypic correlation between the number of teats on the left and the right side range between 0.50 and 0.60 (Borchers et al. 2002, Seo et al. 1996).

2.1   The inverted teat defect in pigs

2.2.1   Description

The inverted teat defect has a considerable negative impact on the economical efficiency of pig production in Germany (Oltmanns 2003). This heritable defect decreases the number of functional nipples and results in a high mortality rate of piglets, especially in large litters. Due to the nessecity of selection against this defect in dam lines, it has also a considerable negative impact on the possibility of selection to other economically important traits.

The development of the mammary gland in pigs depends on the proliferation of the mesenchyme surrounding the epithelial buds, which raise the epithelial bud up from the surface. The insufficient proliferation of mesenchyme at the teat ground during the fetal development may lead to the development of inverted teat (Günther 1984). The inverted teat is characterized by a failure of nipples to protrude from the udder surface (Große et al. 1996). It has deformity of the short protruding part of the teat. Because the teat canal is held inward and forming a small crater, the milk can not flow out (Figure 3).
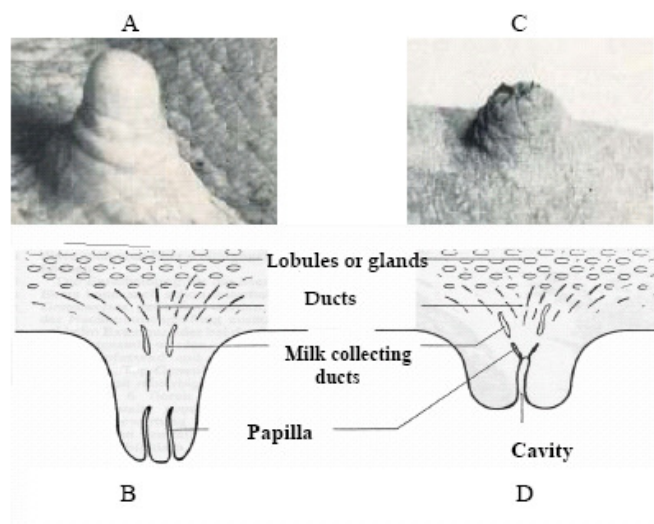
Figure 3:   The differences between a normal teat (A, B) and inverted teat (C, D)
           (Beilage et al. 1996, Steffens 1993)

2.2.2   Genetic background of inverted teat defect

The mode of inheritance of the inverted teat defect is not fully understand. This disorder has a genetic cause, but the number of genes involved is unknown. It is most likely that a major gene and a few modifying minor genes are involved. The estimated heritabilities of the inverted teat range from 0.2 to 0.5 (Clayton et al. 1981, Meyers et al. 2005, Molenat and Thibault 1977). The heritability of the number of functional teats is lower than the heritability of the total number of teats, but the regression of the performance traits on the number of functional teats is larger in absolute magnitude than the corresponding regression on the total number of teats (Enfield and Rempel 1961).

2.3   Candidate gene approach

The candidate gene approach is a useful tool to determine the association of a genetic variant with a disorder and to identify the genes of modest effect (Kwon and Goate 2000). The candidate genes can be searched in identified linkage regions of the chromosome as 'positional candidate genes'. Further genes can be selected based on their biochemical and physiological properties and their function in the metabolism as 'functional candidate genes' (Stratil and Geldermann 2004, Tabor et al. 2002, Zhu and Zhao 2007). The selection of candidate genes is followed by the identification of polymorphisms within the gene and the analysis of association of the genotype at the candidate gene loci with the phenotype (Stratil and Geldermann 2004).

The positional candidate gene approach combines the linkage information of a particular trait and the mapping information of a gene. The genetic markers are used to identify regions that affect traits which are considered to be quantitative traits. The method to detect these loci is called the quantitative trait loci (QTL) analysis. Following the analysis, genes can be detected which are mapped in the region of the QTL. Additional information may also be obtained using comparative mapping. This enables the identification of further candidate genes which are mapped in other species in a region of conserved synteny which corresponds to the QTL region.

The functional candidate gene approach is based on the knowledge of physiology, biochemistry or pathology which clearly indicates the mechanism of the trait. The identified genes may be structural genes or genes in regulatory or biochemical pathways affecting the trait expression. Functional genes can be selected from their described function in literature, whereas the information of knock-out or silencing experiments is

very useful to identify possible candidate genes. Moreover, the identification of functional candidate genes by expression profiling is perhaps the most promising application of functional genomics. Gene expression analysis is increasingly important in many fields of biological research. Understanding patterns of expressed genes is expected to provide an insight into complex regulatory networks and will most probably lead to the identification of genes relevant to biological processes, or implicated in disease. Two methods to measure transcript abundance have gained much popularity and are frequently applied. Microarray analysis allows the parallel analysis of thousands of genes in two differentially labeled RNA populations, it is a tool with a major impact to understand the transcriptional basis of complex biological systems. The expression of thousands of genes in a given tissue or cell type can be measured simultaneously in two or more biological conditions, such as 'infected' and 'non infected', and compared to identify differentially expressed transcripts (Hiendleder et al. 2005). The real-time PCR provides the simultaneous measurement of gene expression in many different samples for a limited number of genes, and is especially suitable when only a small number of cells are available (Nygard et al. 2007).

Using a combination of the positional and functional candidate gene approachs, a number of genes were already identified as possible candidate genes for the development of the inverted teat defect in pigs. For some functional candidate genes such as the androgene receptor (*AR*), no significant association could be detected (Trakooljul et al. 2004). For other genes such as the transforming growth factor beta1 (*TGFB1*), *RLN1* and parathoroid hormone like hormone (*PTHLH*), effects on the heritable defect were detected (Chomdej 2005). Further positional and functional genes were found having a significant effect such as the growth hormone (*GH*), leucin-rich repeat-containing G protein-coupled receptor 7 (*LGR7*), as well as *RLN3* and *GPCR135* for which the results are described in this work (Jonas et al. 2007, Yammuen-art et al. 2007a).

2.3.1   Gene expression in the mammary gland

The prenatal stage of the mammary development is regulated by sequential and reciprocal signaling between the epithelium and the mesenchyme (Robinson 2004). A number of recently published studies in human and mouse have identified some of the signals that control specific steps during the development, using targeted gene deletion and transgenic expression (Robinson 2004). There is a number of efficient methods for the study of gene expression profiles known such as the oligonucleotide DNA chip, the cDNA microarray, serial analysis of gene expression (SAGE) and large-scale expressed sequence tags (ESTs) sequencing (Su et al. 2006).

The mammary gland development is initiated during the embryonic development. Several signaling pathways are active since the early stages of bud development (Robinson 2004). Robinson (2004) found in her study that the first indication for an involvement of wingless-type MMTV integration site family (*Wnt*) signals arises from the deletion of the lymphoid enhancer factor 1 (*LEF1*). Among the members of the *Wnt* family, the three genes *Wnt6*, *Wnt10a* and *Wnt10b* start to be expressed in the surface ectoderm around day 11.25 of the fetal development in mouse. The importance of *Wnt* signalling is further illustrated by the finding that the expression of dickkopf 1 (*Dkk1*), an inhibitor of *Wnt* signals, in the mammary anlagen results in an absence of mammary buds. The T-box 3 gene (*TBX3*) encodes a T-box transcription factor that is expressed in the mammary bud and other sites of tissue interactions, such as the limb bud. Mice with a targeted deletion of *TBX3* fail to develop mammary glands and also display limb and yolk sac deficiencies. This signal appears to act at a very early stage in mammary organogenesis as mammary buds and localized expression of *Wnt10b* and *LEF1* are absent in *TBX3* knockout mice (Robinson 2004).

The fibroblast growth factor (*FGF*) signals play an important role in many sites of the epithelial mesenchymal interaction. They are also involved in the mammary bud formation. Different members of the *FGF* family which are further described, are known to play a role during the mammary gland development. The fibroblast growth factor receptor 2B (*FGFR2B*) is expressed in the mammary placodes from embryonic day 11 onward. The earliest expression of *FGF10* is detected at day 15.5 of the embryonic development in mouse. Expression of the *FGF7* occurs at an earlier stage and is thought to compensate for the absence of *FGF10* (Robinson 2004).

The *PTHLH* gene is not only involved in the regulation of the calcium homeostasis, it also mediates the formation of the primary mammary mesenchyme and nipple sheath.

PTHLH is expressed in the mammary bud, and the parathyroid hormone receptor *(PTHR1)* is expressed in the ventraldermal mesenchyme (Robinson 2004).

The mammary glands develop postnatally by a branching morphogenesis creating an arborated ductal system on which secretory lobuloalveoli develop at pregnancy (Gary 2001). It was found that there is a significant increase in the expression of cell cycle regulatory genes and a concomitant increase in nucleic acid and macromolecular synthesis during gestation (Clarkson et al. 2004). Progesterone, placental lactogens, prolactin, and the osteoclast differentiation factor introduce a signal to the alveolar proliferation and differentiation during pregnancy and possibly lactation. Estrogen and progesterone are primarily known for their role in the development and function of the female reproductive system. The need for a functional mammary gland development is dependent on a successful pregnancy (Connor et al. 2007, Hennighausen and Robinson 2001). In addition, *IGF-1*, *EGF*, or related peptides, and elements of the activin/inhibin family were shown to be necessary for ductal growth. The inhibition of ductal growth, and in particular, lateral branching, is necessary to preserve stromal space for later lobuloalveolar development. Excellent evidence that *TGFB1* naturally inhibits this infilling, possibly by blocking hepatocyte growth factor synthesis, is reviewed along with an evidence indicating that the action of *TGFB1* is modulated by its association with the extracellular matrix (Gary 2001).

The proportion of cell cycle genes is diminished during the lactation. These genes are replaced by an increase of genes involved in development and differentiation, including fatty acid biosynthesis and other metabolic processes. These genes are important for the terminal differentiation and transition to a secretory phenotype. A strong statistical relationship exists between involution and immune related genes (Clarkson et al. 2004). The gene expression during involution was found being significant associated with inflammation, the acute phase response, or humoral immunity and innate cellular defence as shown in table 1 (Clarkson et al. 2004).

The majority of significantly induced genes during the involution are also differentially regulated at earlier stages during the pregnancy. This includes a marked increase in inflammatory mediators during involution and at parturition, which correlates with the signalling of the leukaemia inhibitory factor *STAT3* (signal transducer and activator of signalling-3). Before the involution starts, an increase of cell proliferation, biosynthesis and metabolism-related genes was observed by the study of Clarkson et al. (2004). The first 24 hours after weaning are characterized by a transient increase in the expression of

components of apoptosis, inflammatory cytokines and acute phase response genes. Later regulators of intrinsic apoptosis are induced in conjunction with markers of phagocyte activity, matrix proteases, suppressors of neutrophils and soluble components of specific and innate immunity (Clarkson et al. 2004)

Table 1: Relationship between the gene expression pattern and gene function in the 12-point developmental time course (Clarkson et al. 2004). Shown in the cluster, the number of transcripts (Tr.), the biological process and molecular function

Cluster	Tr.	Biological process	Molecular function
V_{SL} (virgin and suppressed lactation)	357	metabolism cell surface signal development extracellular matrix (ECM) and morphogenesis	immunity signal transducer
G (gestation)	349	cell cycle cell growth	cell cycle cytokine
G_{SL} (gestation, suppressed lactation	605	biogenesis cell cycle	cancer enzyme nucleic acid chemokine
L_T (lactation, transient)	35	cell adhesion, metabolism cransport, cell surface and intracellular signal	signal transducer transport cytokine
L (lactation)	144	biogenesis, metabolism, transport, development, ECM and morphogenesis,	cancer transport
L_G (lactation and gestation)	83	cell cycle metabolism development physiological process ECM and morphogenesis	enzyme

Continued Table 1

Cluster	Tr.	Biological process	Molecular function
I_L (involution and lactation)	624	cell death metabolism transport development ECM and morphogenesis	apoptosis immunity signal transducer transport
I_T (involution, transient)	245	biogenesis cell cycle cell growth cell surface signal intracellular signal	signal transducer
I_P (involution and parturition)	156	biogenesis homeostasis transport intracellular signal	immunity signal transducer transport
I_G (involution and gestation)	582	cell cycle metabolism proliferation transport	cell cycle transport
PC (postcoitum)	93	transport ECM and morphogenesis	transport

The gene expression profile of the porcine mammary gland based on the analysis of 28941 ESTs as well as preliminary results of the comparison of the expression profiles using samples of different breeds and within a number of developmental stages present the majority of known 4785 genes expressed in the porcine mammary gland during lactation (Su et al. 2006).
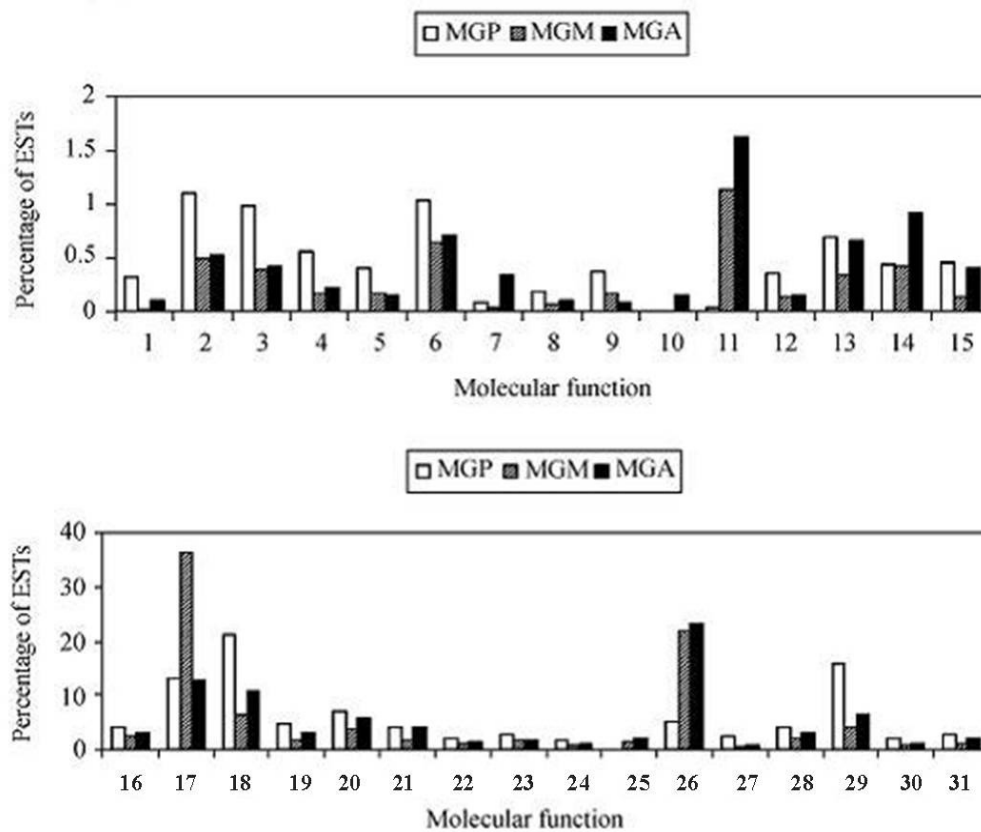
Figure 4:   EST abundance in each major Gene Ontology (GO) molecular function category. X-axis represents the molecular function terms based on GO comprising 1, antioxidant; 2, receptor binding; 3, glycosaminoglycan binding; 4, helicase; 5, lyase; 6, kinase; 7, cell adhesion molecule; 8, DNA repair protein; 9, electron transfer flavoprotein; 10, galactose binding lectin; 11, heterotrimeric G-protein GTPase; 12, high-density lipoprotein; 13, ECM structural constituent; 14, structural constituent of cytoskeleton; 15, structural constituent of muscle; 16, Apoptosis inhibitor; 17, metal ion binding; 18, nucleic acid binding; 19, nucleotide binding; 20, protein binding; 21, hydrolase; 22, oxidoreductase; 23, transferase; 24, chaperone; 25, defense/immunity protein; 26, enzyme inhibitor; 27, protein degradation tagging; 28, signal transducer; 29, constituent of ribosome; 30, transcription regulator; 31, translation regulator. Y-axis is the percentage of ESTs of each library. The gene expression profiles of three different cDNA libraries comprising MGP (Mammary gland - 7 days pre birth), MGM (Mammary gland - 14 days after birth) and MGA (Mammary gland - 7 days after weaning) were constructed and analyzed. (Su et al. 2006)

A remarkable diversity of porcine mammary gene expression was found at different developmental stages. Seven days before parturition, genes of ribosome structural constituent were abundantly expressed in the porcine mammary gland. At the same time, also genes involved in the regulation of transcription and translation, nucleic acid and receptor binding and antioxidants, were abundantly expressed. The genes of the whey acidic protein (*WAP*) and α-lactalbumin were found to be expressed close to the end of the gestation. This might be used to determine the maturity of the mammary epithelial cells. The number of genes being significantly higher expressed during late gestation, was higher than during the other two developmental stages, the end of gestation and lactation observed in the same study (Su et al. 2006). Genes of ribosome structural constituent were abundantly expressed during late gestation, which might be explained by a preparation for the intrinsic protein synthesis during lactation. At the same time point, some proteins involved in the regulation of transcription and translation were also expressed, such as myostatin, thyroid receptor interactor and translation initiation factor. In the middle of the lactation, genes involved in the pathways of the biosynthesis of the milk lipid and proteins were also highly expressed such as the stearoyl CoA desaturase and fatty acid Coenzyme A ligase (take charge of the fatty acid biosynthesis), the phenylalanine hydroxylase (one of the enzymes in the tyrosine metabolism), and the α-lactalbumin (participates in the biosynthesis of lactose). The expression level of other proteins was decreased significantly after weaning, but the β-casein gene was highly expressed at this stage, even higher than in the middle of the lactation. This indicated that the gene regulation system of β-casein might be different from other primary milk proteins. Most of these large clusters arise from milk protein genes, indicating that these genes are highly expressed in the mammary gland during the middle of lactation The most abundant ten genes detected in this study were beta casein (*Csn2*), alpha-S1 casein (*CSN1S1*), beta-lactoglobulin, alpha-S2 casein (*CSN1S2*), hista-mine-releasing factor (*HRF*), kappa casein, ribosomal protein S27a, serum amyloid A-2, eukaryotic translation elongation factor 1 alpha 1 (*EEF1A1*) and alpha-lactalbumin. The mammary gland involution goes through two distinct stages. In the first stage, alveolar cells undergo programmed cell death (PCD), but there is no remodeling of the lobular-alveolar structure. During the second stage, the lobular-alveolar structure of the gland is obliterated as proteinases degrade the basement membrane and ECM. At that time, lysozyme, some kinds of hydrolases and cathepsin are abundantly expressed in the mammary gland. This indicates that during the second

stage of the involution, the lysosomes might be involved in degrading the apoptotic cell fragments as well as the milk proteins. On the contrary as the lobularalveolar structure was gradually destroyed, some membrane protein genes, ECM structural constituents and cell adhesion molecule genes were abundantly expressed. The gene of serum amyloid A-2 protein and of haptoglobin are extremely high expressed seven days after weaning. At the same time, some immunoreaction related genes such as swine leucocyte antigen *(SLA)* class II histocompatibility antigen and beta-2-microglobulin are also higher expressed in the mammary gland (Su et al. 2006).

2.3.2   QTL analysis

Using the linkage analysis to detect QTL for the inverted teat defect, a number of loci could be detected on chromosomes 1, 2, 3, 4, 5, 6, 7, 8, 14, 16 and 18 (Jonas 2006, Oltmanns 2003, Ün 2002). Several genes are known to be mapped in the regions of the QTL effects as show in table 2.

Table 2:   Suggested candidate genes in the QTL regions for the inverted teat defect
            from previous studies

SSC	positional candidate genes	Reference
1	Relaxin1 (*RLN1*)	(Ün 2002)
	Estrogen receptor (*ESR*)	(Jonas 2006)
2	Insulin-like growth factor 2 (*IGF-II*)	(Oltmanns 2003)
	Follicle-stimulating hormone beta (*FSHB*)	
3	Transforming growth factor alpha (*TGFA*)	(Oltmanns 2003)
	Follicle-stimulating hormone receptor gene (*FSHR*)	(Jonas 2006)
4	Thyroid stimulating hormone beta (*TSHB*)	(Jonas 2006)
5	Parathyroid-hormone-like hormone gene (*PTHLH*)	(Jonas 2006)
6	Transforming growth factor B1 (*TGFB1*)	(Ün 2002)
	Leptin receptor (*LEPR*)	(Jonas 2006)
7	Prolactin (*PRL*)	(Jonas 2006)
8	Fibroblast growth factor 2 (*FGF2*)	(Ün 2002)
	Epidermal growth factor (*EGF*)	
	Leucin-rich G-protein-coupled receptor 7 (*LGR7*)	(Jonas 2006)
16	Growth hormone receptor (*GHR*)	(Ün 2002)
18	Insulin like growth factor BP3 (*IGFBP3*)	(Ün 2002)
	Growth hormone releasing hormone receptor (*GHRHR*)	

2.4   Candidate genes for the inverted teat defect

2.4.1   Relaxin 3 (*RLN3*)

The relaxin-like peptide family belongs to the relaxin/insulin superfamily. This
superfamily consists of relaxin-like peptides including *RLN1*, *RLN2*, and *RLN3* (also
called *INSL7*), and insulin-like peptides comprising *INSL3*, *INSL4*, *INSL5*, and *INSL6*.
RLN3 is the most recently discovered member of the relaxin family of peptide
hormones. In contrast to *RLN1* and *RLN2*, whose main functions are associated with
pregnancy, *RLN3* was shown to be involved in neuropeptide signaling in the brain
(Rosengren et al. 2006).

2.4.1.1 Gene structure

The *RLN3* gene contains two exons (Bathgate et al. 2002). It encodes a putative prohormone sequence incorporating the classic two-chain comprising an A- and B-peptide chain linked by disulfide bonds with an intra-chain disulfide bond in the A-chain, analogous to that of insulin (Figure 5A). The RXXXRXX(I/V) motif in the B-chain essential for relaxin receptor binding (Figure 5B) (Sherwood 2004).
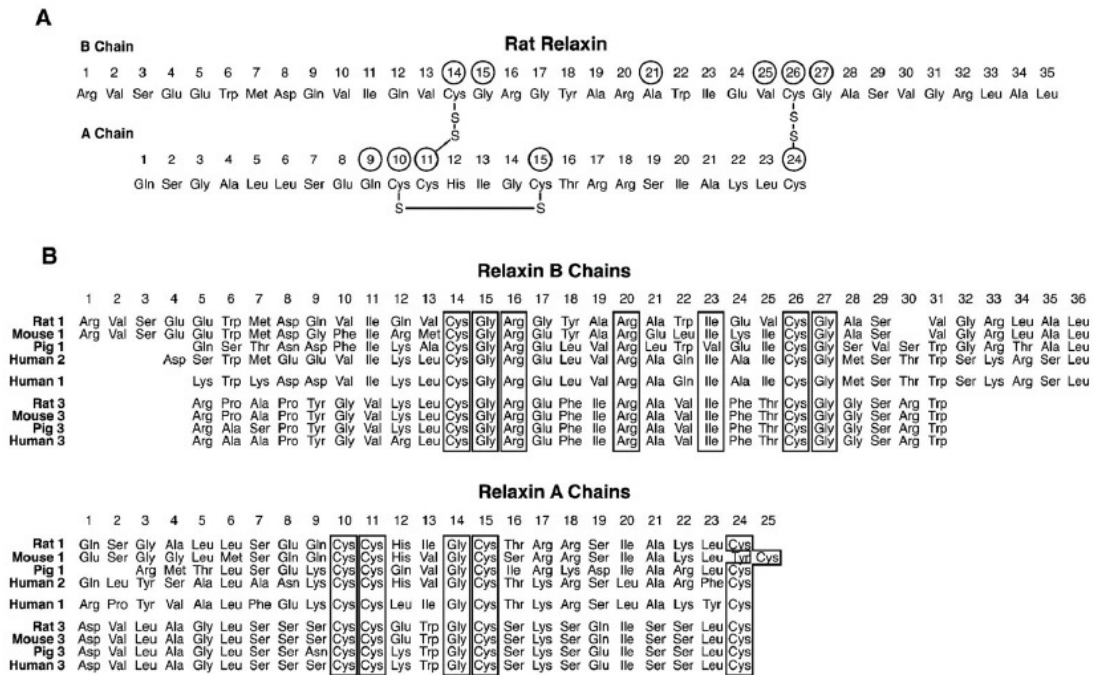
Figure 5:   The structure of the rat relaxin (A) and the structure of the B and A chains of different relaxin genes in comparison (B) (Sherwood 2004)

The human and mouse genes encode deduced proteins of 142 and 141 amino acids, respectively. Both have a predicted 27 amino acid B-chain, an 66 amino acid C-peptide, and an 24 amino acid A-chain (Bathgate et al. 2002).

2.4.1.2 Gene function

The mouse *RLN1* mRNA is expressed in the mammalian liver, lung, thymus, spleen, kidney, skin, testis, epididymis, and myometrium, whereas mouse *RLN3* mRNA is present the brain, thymus, spleen, lung, testis, ovary, and mammary gland. This gene is weakly expressed in the heart, liver, epididymis, prostate, and uterus (Figure 6). Using *in situ* hybridization experiments, it has been clearly demonstrated that the mouse *RLN3* mRNA is expressed in distinct brain regions compared to the *RLN1* mRNA. Thus

suggests the possible different functions of these two *RLN* genes in the brain of the mouse (Bathgate et al. 2002).
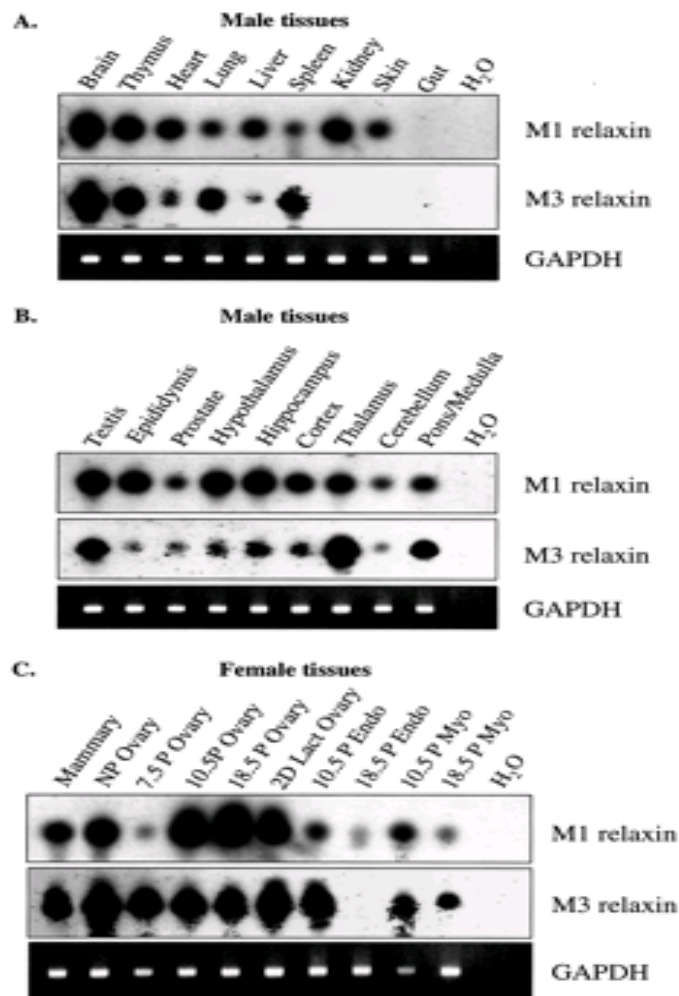
Figure 6: Mouse *RLN1* mRNA (M1) and mouse *RLN3* mRNA (M3) expression were determined in a number of non-reproductive male tissues (A), in reproductive tissues and specific brain regions of the male (B), and in female reproductive tissues at different stages of pregnancy (C) (Bathgate et al. 2002)

Northern blot analysis of human tissues detected weak signals in spleen, thymus, peripheral blood leukocytes, lymph node, and testis. Similar analysis using mouse tissues revealed an 1.2 kb transcript only in the brain. RT-PCR revealed a high expression of *RLN3* in brain, ovary, and testis, moderate expression in thymus, lung, and spleen, very low expression in heart and liver, and no expression in kidney, skin, and gut. *In situ* hybridization using samples of mouse brain showed that the expression is localized in the pons/medulla, with the highest levels in the pars ventromedialis of the dorsal tegmental nucleus. *RLN3* was also expressed at far lower levels in the

hippocampus and olfactory regions. Further use of RT-PCR could detect the expression of *RLN3* only in brain and testis. *In situ* hybridization revealed limited *RLN3* expression in rat brain (Liu et al. 2003b). These results suggesting a still more or less unspecific picture of the expression of the *RLN3* gene, underlined by the still unknown main function of this gene.

The human *RLN3* gene maps to chromosome 19p13.3, close to *INSL3* (19p13.2) whereas the mouse gene could be located to chromosome 8 (Bathgate et al. 2002). Comparative mapping shows the possible positon of the porcine *RLN3* in the q-chain on SSC2 (Meyers et al. 2005).

2.4.2   G-protein coupled receptor 135 (*GPCR135*)

The putative receptors or binding proteins for members of the relaxin/insulin family of peptides comprise four subtypes of receptors/proteins (Liu et al. 2003b). The first are insulin and *IGF* receptor as the single transmembrane cytokine/growth factor-type receptors. The second are the *IGF*-binding proteins, which are secreted soluble binding proteins for *IGF1* and *IGF2*. The third are the leucin-rich repeat containing G-protein coupled receptors 7 (*LGR7*) and *LGR8* as the hormone-type receptors for *RLN1*, *RLN3*, and *INSL3*. The fourth is the typical type I G protein-coupled receptor (*GPCR*) such as *GPCR135,* also called *SALPR*. The *GPCRs* represent the largest known family of receptors interacting at the plasma membrane with extracellular ligands. They are characterised by seven transmembrane domains with an extracellular N-terminus, a cytoplasmic C-terminus, and several conserved structural motifs (Bockaert and Pin 1999, Boels and Chica Schaller 2003). Recently, two closely related *LGRs*, *LGR7* and *LGR8*, have been shown to be receptors for relaxin (Hsu et al. 2002). *INSL3* has been shown to be a selective ligand for *LGR*8 and *RLN3* has been shown to be an additional ligand for *LGR7*. Because crossover activity has been demonstrated for relaxin receptors, it has been tested whether other members of the relaxin/insulin family are also ligands for *GPCR135*. It was found that only *RLN3* can activate *GPCR135*. The porcine relaxin stimulates cyclic adenosine monophosphate (cAMP) accumulation in 293 *LGR7* and *LGR8* transfected cells, whereas *INSL3* only stimulates the cAMP production in *LGR8* transfected cells (Kumagai et al. 2002, Sudo et al. 2003).

LGR7 and *LGR8* belong to the *GPCR* hormone receptor family with significant homology to thyroid stimulatory hormone receptor and luteinizing hormone receptor.

These hormone receptor *GPCRs* have long N-terminal extracellular domains (>300 amino acids) and are known to be involved in stimulation of cAMP. *GPCR135* is not essentially homologous to *LGR7* and *LGR8*. *GPCR135* has a short N-terminal extracellular domain (<100 amino acids) and is coupled to cAMP inhibition. It is a typical neuropeptide-like receptor with a significant homology to somatostatin receptors.

LGR7 and *LGR8* have been identified as the receptors for *RLN1*, both genes belong to the hormone receptor family and share significant homologies with the luteinizing and the thyroid stimulating hormone receptor. The *RLN3* was demonstrated to be an additional ligand for *LGR7*. Recently, the *RLN3* has been reported as a ligand for two related orphan *GPCR*s, *GPCR135* and *GPCR142* (Bathgate et al. 2002, Chen et al. 2005, Liu et al. 2003b, Liu et al. 2003a).

Pharmacological studies indicated that *RLN3* is the only member of this family being able to activate the *GPCR135* gene (Bathgate et al. 2002). Radiolabeled *RLN3* saturably bound *GPCR135* in a monophasic manner with high affinity. The neuroanatomical colocalization of *GPCR135* and *RLN3*, coupled with a clear high affinity interaction, suggest that *GPCR135* is a receptor for *RLN3* (Liu et al. 2003b).

Various ligands use *GPCRs* to stimulate membrane (Figure 7), cytoplasmic and nuclear targets. *GPCRs* interact with heterotrimeric G proteins composed of α-, β- and γ-subunits that are guanosine diphosphate (GDP) bound in the resting state. Agonist binding triggers a conformational change in the receptor, which catalyses the dissociation of GDP from the α subunit followed by guanosine-5'-triphosphate (GTP)-binding to Gα and the dissociation of Gα from Gβγ subunits. The α-subunits of G proteins are divided into four subfamilies: $G\alpha_s$, $G\alpha_i$, $G\alpha_q$ and $G\alpha_{12}$, and a single *GPCR* can couple to either one or more families of Gα proteins. Each G protein activates several downstream effectors. Typically $G\alpha_s$ stimulates adenylyl cyclase and increases levels of cyclic AMP (cAMP), whereas $G\alpha_i$ inhibits adenylyl cyclase and lowers cAMP levels, and members of the $G\alpha_q$ family bind to and activate phospholipase C (PLC), which cleaves phosphatidylinositol bisphosphate (PIP_2) into diacylglycerol and inositol triphosphate (IP_3). The G_β subunits and $G\gamma$ subunits function as a dimer to activate many signalling molecules, including phospholipases, ion channels and lipid kinases. Besides the regulation of these classical second-messenger generating systems, Gβγ subunits and Gα subunits such as $G\alpha_{12}$ and $G\alpha_q$ can also control the activity of key intracellular signal-transducing molecules, including small GTP-binding proteins of the

Ras and Rho families and members of the mitogen-activated protein kinase (MAPK) family of serine-threonine kinases, including extracellular signal-regulated kinase (ERK), c-jun N-terminal kinase (JNK), p38 and ERK5 (Dorsam and Gutkind 2007).
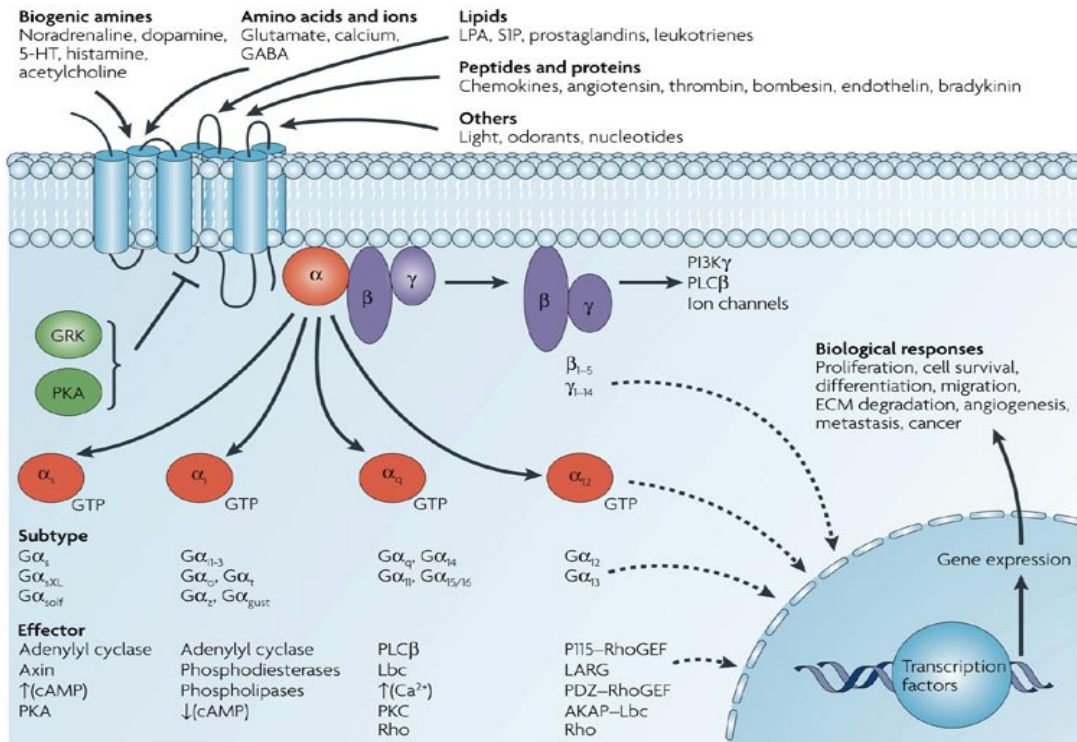
Figure 7: Diversity of G-protein-coupled receptor signaling (Dorsam and Gutkind 2007)

GPCR activation leads to the stimulation of different RTKs and the subsequent activation of the ERK/MAPK cascade. The process is known as *GPCR*–receptor tyrosine kinase (RTK) transactivation and involves different mediators depending on the cell type (Figure 8).
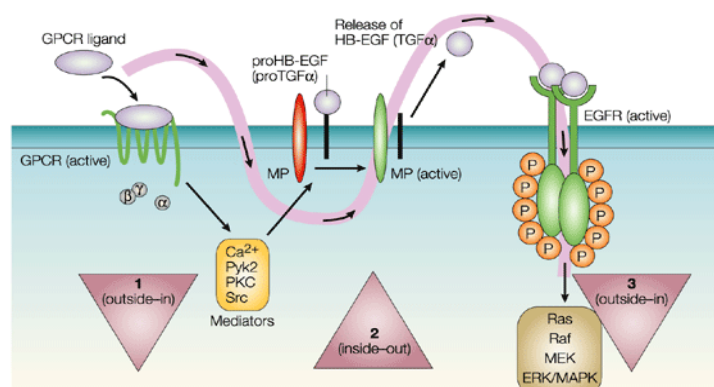
Figure 8: The triple-membrane-passing-signalling model of *GPCR* activation leads to the stimulation of different RTKs and the subsequent activation of the ERK/MAPK cascade (Wetzker and Bohmer 2003)

GPCR transactivation of the epidermal growth factor receptor (*EGFR*) occurs in many cell types through generation of a cognate ligand, the heparin-binding *EGF* (*HB-EGF*), which activates the *EGFR* and subsequently the ERK/MAPK cascade. *HB-EGF* is generated through extracellular proteolytic cleavage of pro*HB-EGF* (a membrane-spanning, latent form of this growth factor) that is mediated by the action of a metalloproteinase (MP). A similar activation might also occur for other latent growth factors, such as the precursor of transforming growth factor-α (pro*TGFA*) (Wetzker and Bohmer 2003).

2.4.2.1 Gene structure

The deduced 469 amino acid protein of *GPCR135* has the characteristic seven transmembrane domain (TM) structure of a *GPCR*, as well as two putative N-glycosylation sites in its N-terminal domain. The *GPCR135* shares the highest sequence similarity with somatostatin receptors and angiotensin receptors. The *GPCR135* gene contains a single exon (Matsumoto et al. 2000). The putative mouse and rat *GPCR135* receptors are 85% and 86% sequence identical to the human *GPCR135*, respectively, whereas they share a higher homology (94%) with each other. The TM, TM2, TM3, and TM6 are conserved among human, mouse, and rat *GPCR135* (Figure 9). Between mouse and rat *GPCR135*, almost all TM domains, excluded TM4, are completely conserved. The rat *GPCR135* cDNA has two putative translational starting codons (ATG). The first ATG in the rat *GPCR135* gene is unique and leads to a seven amino acid (MPKAHLS) addition at the N terminus. The second ATG in the rat *GPCR135* gene is conserved among human, mouse, and rat and corresponds to the apparent translation start site for human and mouse genes. The putative rat *GPCR135* receptor protein derived from using the second ATG as the translational starting codon. Both human *GPCR135* and rat *GPCR135* have 469 amino acids, whereas the mouse *GPCR135* has 472 amino acids with three extra amino acids between TM5 and TM6 (Chen et al. 2005).

```
    Human: MQMADAATIATMNKAAGGDKLAELFSLVPDLLEAANTSGNASLQLPDLWWELGLELPDGAPPGHPPGSGGAESADTEARV 80
    Mouse: MQVASATPAATVRKAAAGDELSEFFALTPDLLEVANASGNASLQLQDLWWELGLELPDGAAPGHPPGGGGGAESTDTEARV 80
      Rat: MQVASATTAAPMSKAAAGDELSGFFGLIPDLLEVANRSSNASLQLQDLWWELGLELPDGAAPGHPPGSGGAESADTEARV 80
Consensus: MQ:A.A:. AT:.KAA:GD.L:E:F:L.PDLLE.AN:SGNASLQL.DLWWELGLELPDGA:PGHPPG:GGAES:DTEARV

    Human: RILISVVYWVVCALGLAGNLLVLYLMKSMQGWRKSSINLFVTNLALTDFQFVLTLPFWAVENALDFKWPFGKAMCKIVSM 160
    Mouse: RILISAVYWVVCALGLAGNLLVLYLMKSKQGWRKSSINLFVTNLALTDFQFVLTLPFWAVENALDFKWPFGKAMCKIVSM 160
      Rat: RILISAVYWVVCALGLAGNLLVLYLMKSKQGWRKSSINLFVTNLALTDFQFVLTLPFWAVENALDFKWPFGKAMCKIVSM 160
Consensus: RILIS.VYWVVCALGLAGNLLVLYLMKS.QGWRKSSINLFVTNLALTDFQFVLTLPFWAVENALDFKWPFGKAMCKIVSM
                    TM1                                          TM2

    Human: VTSMNMYASVFFLTAMSVTRYHSVASALKSHRTRGHGRGDCCGRSLGDSCCFSAKALCVWIWALAALASLPSAIFSTTVK 240
    Mouse: VTSMNMYASVFFLTAMSVARYHSVASALKSHRTRGRGRGDCCGQSLRESCCFSAKVLCGLIWASAALASLPNAIFSTTIR 240
      Rat: VTSMNMYASVFFLTAMSVARYHSVASALKSHRTRGHGRGDCCGQSLGESCCFSAKVLCGLIWASAAIASLPNVIFSTTIN 240
Consensus: VTSMNMYASVFFLTAMSV:RYHSVASALKSHRTRG:GRGDCCG:SL :SCCFSAK.LC  IWA AA:ASLP.:IFSTT::
                    TM3                                          TM4

    Human: VMGEELCLVRFPDKLLGRDRQFWLGLYHSQKVLLGFVLPLGIIILCYLLLVRFIADRRAAGTK---GGAAVAGGRPTGAS 317
    Mouse: VLGEELCLMHFPDKLLGWDRQFWLGLYHLQKVLLGFLLPLSIISLCYLLLVRFISDRRVVGTTDAVGAAAAPGGGLSTAS 320
      Rat: VLGEELCLMHFPDKLLGWDRQFWLGLYHLQKVLLGFLLPLSIISLCYLLLVRFISDRRVVGTT---DGATAPGGSLSTAG 317
Consensus: V:GEELCL::FPDKLLG:DRQFWLGLYH QKVLLGF:LPL:II LCYLLLVRFI:DRR..GT.---.:A..:GG  :.AS
                                                    TM5

    Human: ARRLSKVTKSVTIVVLSFFLCWLPNQALTTWSILIKFNAVPFSQEYFLCQVYAFPVSVCLAHSNSCLNPVLYCLVRREFR 397
    Mouse: ARRRSKVTKSVTIVVLSFFLCWLPNQALTTWSILIKFNAVPFSQEYFQCQVYAFPVSVCLAHSNSCLNPILYCLVRREFR 400
      Rat: ARRRSKVTKSVTIVVLSFFLCWLPNQALTTWSILIKFNVVPFSQEYFQCQVYAFPVSVCLAHSNSCLNPILYCLVRREFR 397
Consensus: ARR SKVTKSVTIVVLSFFLCWLPNQALTTWSILIKFNAVPFSQEYF CQVYAFPVSVCLAHSNSCLNP:LYCLVRREFR
                    TM6                                          TM7

    Human: KALKSLLWRIASPSITSMRPFTATTKPEHEDQGLQAPAPPHAAAEPDLLYYPPGVVVYSGGRYDLLPSSSAY        469
    Mouse: KALKNLLWRIASPSLTNMRPFTATTKPEPEDHGLQALAPLNAAAEPDLIYYPPGVVVYSGGRYDLLPSSSAY        472
      Rat: KALKNLLWRIASPSLTSMRPFTATTKPEPEDHGLQALAPLNATAEPDLIYYPPGVVVYSGGRYDLLPSSSAY        469
Consensus: KALK:LLWRIASPS:T:MRP FTATKPE.ED:GLQA AP :AAAEPDL:YYPPGVVVYSGGRYDLLPSSSAY
```

Figure 9:   Amino acid sequence comparison of human, mouse, and rat *GPCR135* (Chen et al. 2005)

2.4.2.2 Gene function

The *GPCR135* gene is expressed abundantly in the hypothalamus with a discrete expression in the paraventricular nucleus of the hypothalamus and supraoptic nucleus as well as in the cortex, septal nucleus and preoptical area. The *GPCR135* mRNA was detected in the brain, testis, thymus, and adrenal gland (Figure 10). The mRNA expression of *GPCR135* and *RLN3* mRNA using *in situ* hybridization distribution in the rat brain showed that *RLN3* and its receptor are discretely expressed in different areas of the central nervous system (Figure 10). Since the available information regarding the *RLN3* and *GPCR135* genes is limited, their functional roles in mammalian physiology remain still unclear (Liu et al. 2003b).
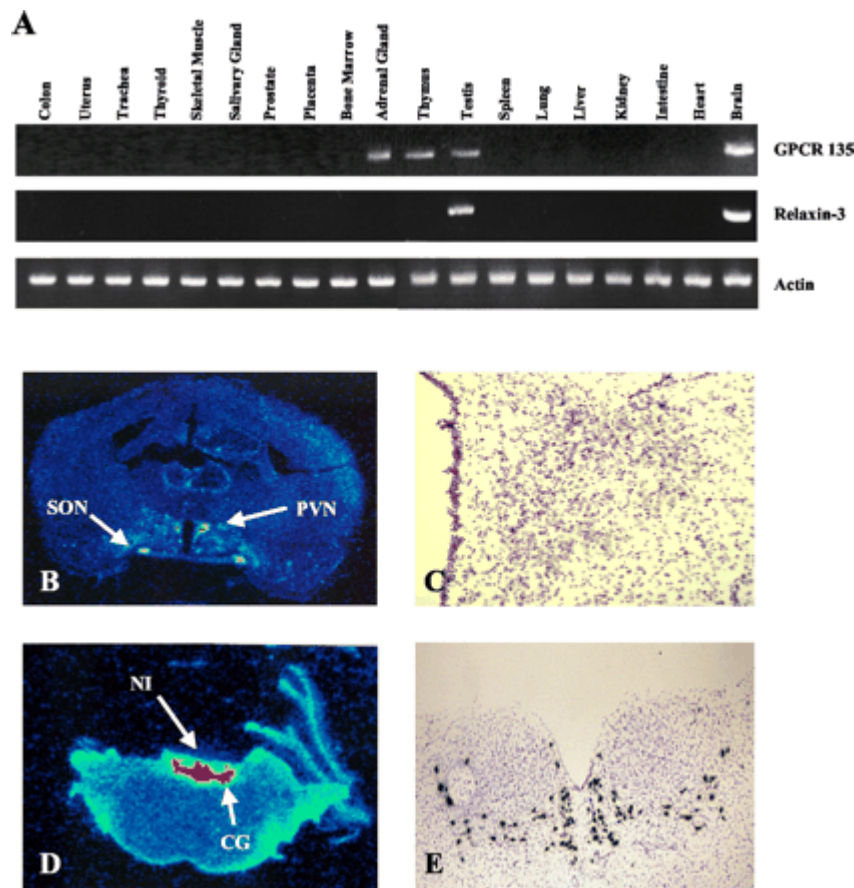
Figure 10:   RT-PCR detection of *GPCR135* and *RLN3* mRNA expression profiles in different human tissues (A); *GPCR135* mRNA distribution is distinct in the paraventricular nucleus (PVN) and supraoptic nucleus (SON) (B); brightfield photomicrograph of the paraventricular nucleus showing expression of *GPCR135* mRNA (C); *RLN3* mRNA distribution in the central gray and nucleus incertus (D); brightfield photomicrograph of central gray (*CG*) and nucleus incertus (*NI*) showing expression of *RLN3* mRNA (E) (Liu et al. 2003b)

Using RT-PCR, the *GPCR135* gene expression could be detected predominantly in brain, with the highest expression in substantia nigra and pituitary, followed by hippocampus, spinal cord, amygdala, caudate nucleus, and corpus callosum; the expression was very low in cerebellum. In peripheral tissues, a relatively high expression was detected in the adrenal gland, and a low expression was detected in pancreas, salivary gland, placenta, mammary gland, and testis (Figure 11) (Liu et al. 2003b, Matsumoto et al. 2000)
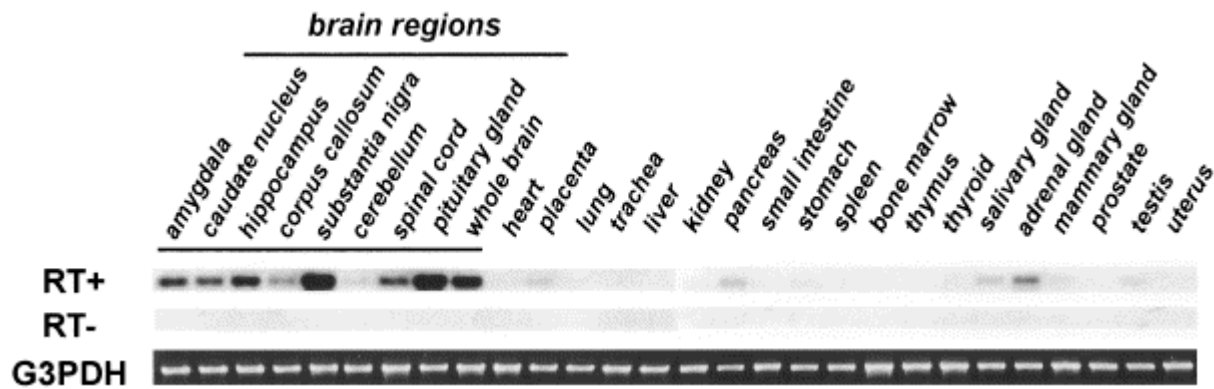
Figure 11:   The RT-PCR analysis of *GPCR135* mRNA in human tissues and brain
             regions and the peripheral tissues. The ethidium-bromide-stained PCR
             products of *G3PDH* are shown as a quantitative control for each cDNA
             (Matsumoto et al. 2000)

The *GPCR135* gene was found being located on chromosome 5p15.1-p14 in human
using radiation hybrid analysis (Matsumoto et al. 2000).

2.4.3   G-protein coupled receptor 142 (*GPCR142*, *GPR100*)

The *GPCR142* gene is a member of the rhodopsin family of the GPCRs (Fredriksson et
al. 2003). Also the human *GPCR142* was reported to be a bradykinin receptor for *RLN3*
(Boels and Chica Schaller 2003).

2.4.3.1 Gene structure

The mouse and human *GPCR142* genes contain a single coding exon (Fredriksson et al.
2003). The *GPCR142* gene shows conserved amino acids or amino acid motifs, typical
for members of the rhodopsin family of *GPCRs* (Figure 12). This includes an
asparagine in the first transmembrane domain, an LXXXD motif in the second
transmembrane domain, tryptophan and proline in the fourth transmembrane domain, an
FXXXWXP motif in the sixth transmembrane domain, an NPXXY motif in the seventh
transmembrane domain, and conserved cysteine residues in the first two extracellular
loops. These cystein residues form an S-S bridge and contribute to protein stability, and
a partially conserved *GPCR* signature triplet sequence (typically DRY, here ARY),
found downstream of the third transmembrane domain.

```
                                                      *10       *20       *30       *40       *50
GPCR142:                                    MPTLNTSASPPTFFWANASGGSVLSADDAPMPVKFLALRLMVALAYGLVGAIGLLGNL
Consensus:                                  :  ..    .   :. A G...    A      . :R:::.:.Y :V A:GL.GNL
GPCR135: MQMADAATIATMNKAAGGDKLABLFSLVPDLLBAANTSGNASLQLPDLWWELGLELPDGAPPGHPPGSGGABSADTEARVRILISVVYWVCALGLAGNL
            *10       *20       *30       *40       *50       *60       *70       *80       *90      *10
                                                                                                    TM1

        *60       *70       *80       *90       *100      *110      *120      *130      *140
GPCR142: AVLWVLSNCARRAPGPPSDTFVFNLALADLGLALTLPPFWAABSALDFHWPPGGALCKMVLTATVLNVYASIPLITALSVARYWVVAMAA---------G
Consensus: .VL:::  .  .     . FV NLAL.D:..:.LTLPPFWA:E.ALDF:WPPGLA:CK:V  .T :N:YAS:F::TA:SV RY. VA A.      G
GPCR135: LVLYLMKSMQG-WRKSSINLFVTNLALTDFQFVLTLPPFWAVENALDFKWPPGKAMCKIVSMVTSMNMYASVFFLTAMSVTRYHSVASALKSHRTRGHGRG
            *110      *120      *130      *140      *150      *160      *170      *180      *190
                      TM2                   TM3

        *150      *160      *170      *180      *190              *200      *210      *220      *230
GPCR142: PGTHLSL-----FWARIATLAVWAAAALVTVPTAVFGVBGEVCGVRLCLLRFP-------SRYWLGAYQLQRVVLAFMVPLGVITTSYLLLLAFL-QRR-
Consensus:        SL    F A:.. :  :WA.AAL.::P:A:F:.. .V G .LCL.RFP       ..::WLG Y: Q:V:L:F::PLG:I. .YLLL: F: :RR
GPCR135: DCCGRSLGDSCCFSAKALCVWIWALAALASLPSAIFSTTVKVMGEELCLVRFPDKLLGRDRQFWLGLYHSQKVLLGFVLPLGIIILCYLLLVRFIADRRA
            *200      *210      *220      *230      *240      *250      *260      *270      *280      *290
                      TM4                                                 TM5

                        *240      *250      *260      *270      *280      *290      *300      *310
GPCR142: -------------QRRQDSRV--VARSVRILVASFFLCWFPNHVVTLWGVLVKFDLVPWNSTFYTIQTYVFPVTTCLAHSNSCLNPVLYCLRREPRQA
Concensus:             .    ..R: V::SV I:V SFFLCW:PN:.:T W::L:KF: VP.: ..: Q.Y.FPV:.CLAHSNSCLNPVLYCL:RRE R:A
GPCR135: AGTKGGAAVAGGRPTGASARRLSKVTKSVTIVVLSFFLCWLPNQALTTWSILIKFNAVPPSQEYFLCQVYAPPVSVCLAHSNSCLNPVLYCLVRREFRKA
            *300      *310      *320      *330      *340      *350      *360      *370      *380      *390
                                                          TM6                             TM7

        *320      *330      *340      *350      *360      *370
GPCR142: LAGTFRDLRSRLWPQGGGWVQQVALKQVGRRWVASNPRESRPSTLLTNLDRGTPG
Consensus: L : ::.: S ...
GPCR135: LKSLLWRIASPSITSMRPFTATTKPEHEDQGLQAPAPPHAAABPDLLYYPPGVVVYSGGRYDLLPSSSAY
            *400      *410      *420      *430      *440      *450      *460
```

Figure 12: Amino acid sequence comparison of human *GPCR135* and rat *GPCR142* (Liu et al. 2003a)

The *GPCR142* gene has N-linked glycosylation sites in the N-terminal extracellular domain and phosphorylation sites, as well as basic amino-acid clusters in the third intracellular loop and in the C-terminal cytoplasmic domain (Boels and Chica Schaller 2003) as shown in figure 13.
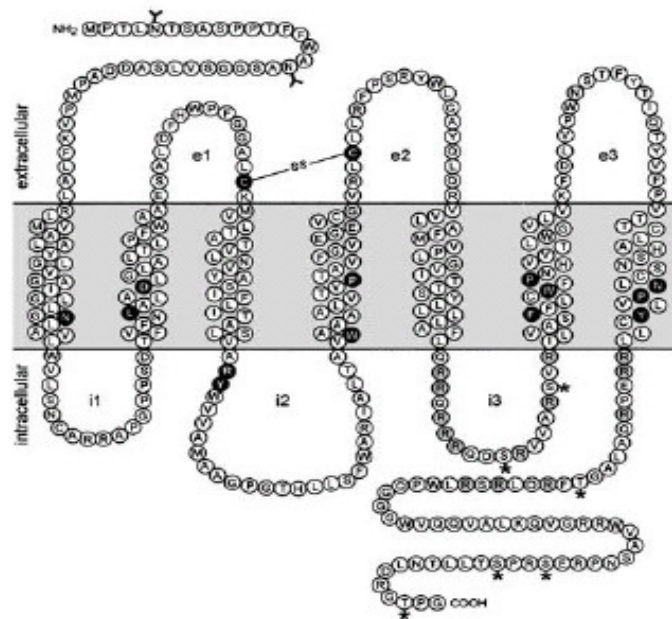
Figure 13: Schematic representation of the *GPCR142* protein showing the seven putative transmembrane domains (Boels and Chica Schaller 2003)

The mouse *GPCR142* shares 74% sequence identity to that of human. Human *GPCR142* has 374 amino acids, whereas the mouse equivalent is longer with 414 amino acids. As one important difference between the human and mouse *GPCR142,* the mouse *GPCR142* has a different and much longer C-terminal tail (Chen et al. 2005).

2.4.3.2 Gene function

The *GPCR142* belongs to the type I *GPCR* family similar to somatostatin and angiotensin receptors. It is also possible that *GPCR142* is activated by a different ligand, for example a peptide that is not member of the insulin/relaxin family (Liu et al. 2003a). Other marked differences between *GPCR142* and *GPCR135* are their mRNA expression patterns. The *GPCR135* mRNA is expressed in restricted tissues with the predominant expression in the brain, whereas the *GPCR142* mRNA is expressed in a broader range of peripheral tissues, suggesting that *GPCR142* may have a different physiological role from that of *GPCR135* (Figure 14). Because *RLN3* is primarily expressed in the brain, the relative abundant expression of *GPCR142* in the peripheral tissues also demonstrates the possible existence of an additional ligand expressed in peripheral tissues.
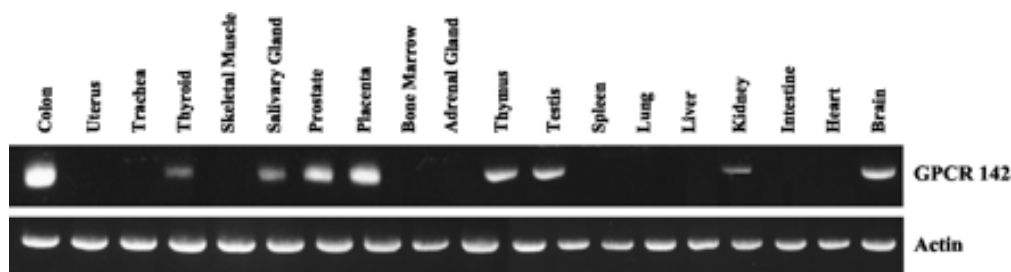
Figure 14: Reverse transcriptase-PCR showed *GPCR142* mRNA expression profiles in different human tissues (Liu et al. 2003a)

The *GPCR142* gene is located on chromosome 1q22 in human and chromosome 3 in mouse (Fredriksson et al. 2003).

2.4.4   Connective tissue growth factor (*CTGF*, *CCN2*)

The insulin-like growth factor binding proteins (*IGFBPs*) are a family of homologous proteins that regulate the biological activities of the *IGFs* and may also be capable of *IGF*-independent actions. A number of characteristics of this protein family are of potential relevance to the *IGFBPs*. It has been suggested for some of the *IGFBPs*, that these proteins may be capable of *IGF*-independent regulation of cell growth. Several of the proteins have been shown to interact with both cell surfaces and extracellular matrix and being capable of binding to heparin, properties also shared by some members of the *IGFBP* family and several genes from this family are induced by *TGFB*, as is the case for some *IGFBPs*.

The *CTGF* gene, also called insulin-like growth factor-binding protein 8 (*IGFBP8*) belongs to the CCN (Cysteine-rich, angiogenic inducer 61) gene family. Members of this family encode cysteine-rich secreted proteins with roles in cell growth and differentiation (Brigstock 2002, Grotendorst et al. 2000, Hurvitz et al. 1999, Lau and Lam 1999). The *CTGF* gene is a cysteine-rich mitogen secreted by human umbilical vein endothelial (HUVE) cells (Bradham et al. 1991). *CTGF* and *CYR61* stimulate cell proliferation, chemotaxis, adhesion, production of ECM components, and play roles in processes such as implantation, placentation, embryogenesis, differentiation and development (Brigstock 1999, Lau and Lam 1999).

2.4.4.1 Gene structure

The *CTGF* gene comprises four exons and shares 28 to 38% amino acid identity with *IGFBP1* to *IGFBP6* (Kim et al. 1997). The *CTGF* gene is one of six different proteins, varying between 348 and 379 amino acids, that are the products of a group of immediate-early genes expressed after induction by growth factors or certain oncogenes (Kim et al. 1997). Although the similarity of the COOH-terminal sequences is low (20%), the NH$_2$-terminal regions are well conserved among these new members and the *IGFBPs*. The *CTGF* gene also contains the conserved *IGFBP* motif (GCGCCXXC) in the NH$_2$ terminus. It can be suggested that the *CTGF* gene shares significant structural homology with these genes as most of the cysteines are conserved in *IGFBP1* to *IGFBP6*. The *CTGF* gene specifically binds *IGF-I* and *IGF-II*, although with relatively low affinity as compared with the *IGFBPs*. The strong changes in the levels of *CTGF*

mRNA occur during follicular and luteal growth, and ovulation which involve processes such as angiogenesis and tissue repair (Wandji et al. 2000).

2.4.4.2 Gene function

The 2.4 kb *CTGF* mRNA could be detected in a broad spectrum of normal human tissues. In particular, the *CTGF* mRNA is expressed at high levels in spleen, ovary, gastrointestinal tract, prostate, heart, and testis (Kim et al. 1997). Proliferative endometrium, epithelial and vascular endothelial cells showed strong *CTGF* immunoreactivity, whereas stromal cells are negative or only weakly positive for the *CTGF* protein. During pregnancy, the decidual, epithelial and endothelial cells of the endometrium are all immunoreactive to the *CTGF* gene expression and localization of its encoded protein in human uterine tissues. The localization of *CTGF* supports a role for this molecule in regulating aspects of uterine cell growth, migration, and/or matrix production during the menstrual cycle and pregnancy. In the rat ovary, the *CTGF* gene is switched on at the very earliest stages of follicular development when gonadotropin-stimulated rat granulosa cells have just begun to proliferate. A similar pattern of expression is observed in pig ovary, where *CTGF* has been hypothesized to promote ovarian cell growth and blood vessel formation during follicular and luteal development. The *CTGF* gene is structurally and functionally related to *PDGF* and the vascular endothelial growth factor, growth factors that are mitogenic for mesenchymally derived cells in blood, muscle, bone/cartilage, and connective tissue (Slee et al. 2001). The presence of *CTGF* expression in human mammary tumors suggests that *CTGF* is involved in the connective tissue stromal proliferation of these tumors (Frazier and Grotendorst 1997).

Martinerie et al. (1992) assigned the *CTGF* gene to human chromosome 6q23.1 by a combination of study of mouse/human somatic cell hybrids and fluorescence *in situ* hybridization.

2.4.4.3 Polymorphism in the *CTGF* gene

A SNP within the *CTGF* promoter region (G945C) located at the position -743 bp from the transcription start site, has recently been published in the dbSNP database (Fonseca et al. 2007). The G945C substitution represses the *CTGF* transcription, and the allele

with the base G is significantly associated with susceptibility to systemic sclerosis. The G allele is strongly linked to transcriptional activation of *CTGF* and to the risk of systemic sclerosis, particularly in patients with antitopoisomerase I autoantibodies and pulmonary fibrosis (Fonseca et al. 2007).

2.4.5   Epidermal growth factor receptor (*EGFR*)

The *EGFR* (also called *ERBB1)* is a member of the type I *RTK* family (Hackel et al. 1999). The *EGFR* gene is a transmembrane tyrosine kinase that is activated upon binding *EGF*, *TGFA*, amphiregulin (*AREG*), *HB-EGF*, betacellulin (*BTC*), epiregulin (*EPIR*) or epigen (*EPGN*), each of which is expressed as a transmembrane precursor that is proteolytically shed from the cell surface (Harris et al. 2003).

2.4.5.1 Gene structure

The *EGFR* molecule has three regions: one projects outside the cell and contains the site for binding *EGF*; the second is embedded in the membrane; the third projects into the cytoplasm of the cell's interior. Each member of the *RTK* family comprises a conserved protein tyrosine kinase domain that resides within the cytoplasm, a transmembrane domain that makes a single pass through the plasma membrane, and a glycosylated, extracellular ligand-binding domain (Bazley and Gullick 2005). In the *EGF* receptor family, this last domain exhibits four subdomains denominated L1, S1 (CR1), L2 and S2 (CR2) (or, more simply, I, II, III and IV respectively) (Lax et al. 1989). Of these domains, S1 and S2 are homologous cysteine-rich regions (CR1 and CR2), while L1 and L2 form the ligand-binding site (Garrett et al. 2002, Ogiso et al. 2002). It is likely that they are derived from an ancient gene duplication event, and it is notable that the cysteine residues do not form disulphide bonds between the two S1/S2 domains (Figure 15).
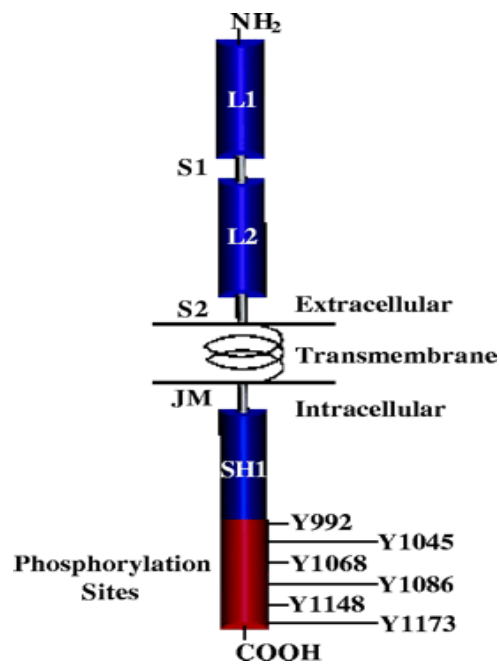
Figure 15: Schematic structure of *EGFR*. The *EGFR* monomer possesses an extracellular domain consisting of two ligand-binding subdomains (L1 and L2) and two cysteine-rich domains (S1 and S2) (Bazley and Gullick 2005)

2.4.5.2 Gene function

The *EGFR* mRNA and the protein are abundant in the mammary stroma (Luetteke et al. 1999, Schroeder and Lee 1998). *EGF* induces *EGFR* phosphorylation in gland-free fat pads (Sebastian et al. 1998). Stromal rather than epithelial *EGFR* is essential for the ductal development (Wiesen et al. 1999). Ductal outgrowth is severely impaired in triple-null mice lacking amphiregulin (*AREG)*, *EGF* and *TGFA* which are lactation incompetent, and variably impaired in mice lacking only *AREG* (Sternlicht et al. 2005). Neither ductal outgrowth nor lactation is affected by the elimination of *EGF*, *TGFA*, *HB-EGF* or *BTC* alone or in various combinations (Jackson et al. 2003, Luetteke et al. 1999, Sternlicht et al. 2005). During mammary development the crucial *EGFR* ligand *AREG* comes from the epithelium. *EGFR* is enriched in the peri-epithelial mammary stroma (Coleman et al. 1988, Schroeder and Lee 1998). It could be shown that the stromal rather than epithelial *EGFR* is required for the mammary epithelial development *in vivo* (Wiesen et al. 1999). *EGFR* signaling promoted steroidogenesis in mouse oocyte-granulosa cell complexes and luteinizing hormone (*LH*), induced steroidogenesis in a mouse leydig cell line (Jamnongjit et al. 2005).

The *EGFR* gene is located on human chromosome 7 (Carlin and Knowles 1982).

2.4.6   Insulin-like-growth factor 2 (*IGF-II*)

The *IGF-I* and *IGF-II* genes are also known as somatomedin C and somatomedin A. They are single chain polypeptides that share amino acid sequences with a homology of 47% with insulin (*INS*) and about 31% with *RLN*. The mature 67 amino acid peptide shares sequence homology with both *INS* and *IGF-I* (O'Dell and Day 1998). *IGF-II* plays a key role during the mammalian growth, influencing foetal cell division and differentiation and possibly metabolic regulation (O'Dell and Day 1998). The functions of *IGF-I* and *IGF-II* include the mediation of growth hormone action, the stimulation of growth of cultured cells, the stimulation of the action of *INS*; these genes are involved in development and growth. They appear to be autocrine regulators of the cell proliferation. A QTL in the region of *IGF-II* had a major effect on muscle mass and fat deposition in pigs (Jeon et al. 1999, Nezer et al. 1999).

2.4.6.1 Gene structure

The human *IGF-II* gene comprises nine exons and four promoters (Brissenden et al. 1984). The exons 7 to 9 encode the prepro *IGF-II* protein; the exons 1 to 6 are non-coding and form alternative 5'-untranslated regions of different RNA molecules which are expressed from the four promoters in a tissue and development specific way. The structure of *IGF-II* preprohormone consists of a signal peptide of 24 residues, 67 amino acids of the mature peptide and 89 amino acids in a carboxy-terminal extension termed the E-domain. Similar to proinsulin, *IGF-II* is divided into A-, B- and C-domains. A- and B-domains are similarly bridged by two inter-domain disulphide bonds, with one internal disulphide bond in the A-domain. The A- and B-domains are connected by a C-domain, which unlike the *INS* C-domain is not proteolytically cleaved during structural maturation. The D-domain is not present in *INS*. The carboxy-terminal sequence (E-domain) of the preprohormone is removed during processing. The region involved in the recognition of the type 1 receptor is at the same site that binds the insulin receptor. The type 2 receptor recognises several residues in the A-domain. Binding of the *IGFBPs* depends on the residues in the first part of the B-domain (Figure 16).
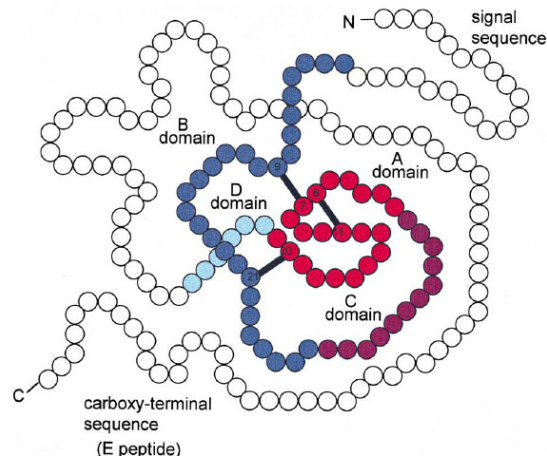
Figure 16:   Structure of *IGF-II* preprohormone divided into A-, B- and C-domains (O'Dell and Day 1998)

2.4.6.2 Gene function

The *IGF* family plays an important role in embryonic, fetal and placental development. The *IGF-II* mRNAs are present in the mouse fetus at higher levels compared to those of *IGF-I*. Although overexpression of the *IGF-II* gene alone did not have a major impact on the fetal growth, in the absence of *IGF-IIR*, elevated *IGF-II* led to severe placental and fetal overgrowth (Moore et al. 2007). *PRL* was shown to induce *IGF-II* mRNA and *IGF-II* to induce the cyclin D1 (*CCND1*) protein expression in mouse mammary epithelial cultures (Brisken et al. 2002). Alveologenesis was retarded in both *IGF-II*- and *CCND1*-deficient cells. *IGF-II* and prolactin receptor (*PRLR*) mRNAs colocalized in the mammary epithelium. *IGF-II* is a mediator of the *PRL*-induced alveologenesis; *PRL*, *IGF-II*, and *CCND1* are components of a developmental pathway in the mammary gland (Brisken et al. 2002). Several genes, including *IGF-II*, are involved in growth and tissue remodeling and are expressed at relatively higher levels in the villus sections of the placenta compared with other tissues (Sood et al. 2006). The loss of *IGF-II* is associated with fetal growth restriction in mice. The precise nature of the loss of uterine lumenal *IGFs* following conceptus elongation suggests that the release of *IGFs* during day 12 and 13 of pregnancy is very critical for subsequent development and survival of pig embryos (Ashworth et al. 2005).

The porcine *IGF-II* maps to chromosome 2 (2p1.7) (Jeon et al. 1999). By *in situ* hybridization (Morton et al. 1986) assigned human *IGF-II* to 11p15. Using cDNA

probes in the analysis of somatic cell hybrids, the *IGF-II* was demonstrated to be located on chromosome 11 (11p15-p11) (Brissenden et al. 1984, Tricoli et al. 1984)

2.4.6.3 Polymorphism in the *IGF-II* gene

A paternally expressed QTL affecting muscle growth, fat deposition, and size of the heart in pigs maps to the *IGF-II* region. This QTL is caused by a nucleotide substitution in intron 3 of *IGF-II* (Van Laere et al. 2003). Gaunt et al. (2001) identified three single-nucleotide polymorphisms (SNPs) in *IGF-II* which were associated with body mass index (BMI) in a cohort of over 2500 middle-aged Caucasoid males. The *IGF-II* genetic variation may be a significant determinant of body weight in middle-aged male.

2.4.7   Epidermal growth factor (*EGF*)

EGF is a part of a complex network of growth factors and receptors which help to modulate the growth of cells. *EGF* is released by cells and picked up either by the cell itself to stimulate its own growth, or by neighboring cells to stimulate their ability to divide. *EGF* has a profound effect on the differentiation of specific cells *in vivo* and is a potent mitogenic factor for a variety of cultured cells of both ectodermal and mesodermal origin (Carpenter and Cohen 1979).

2.4.7.1 Gene structure

The *EGF* gene consists of a single polypeptide chain of amino acid residues. The predicted protein sequence of the porcine *EGF* precursor contains 1214 amino acids, similar to the human *EGF* precursor (1207 amino acids, 81% identity). The location of the three intramolecular disulfide bonds of the primary amino acid sequence of the mouse *EGF* are shown in figure 17.
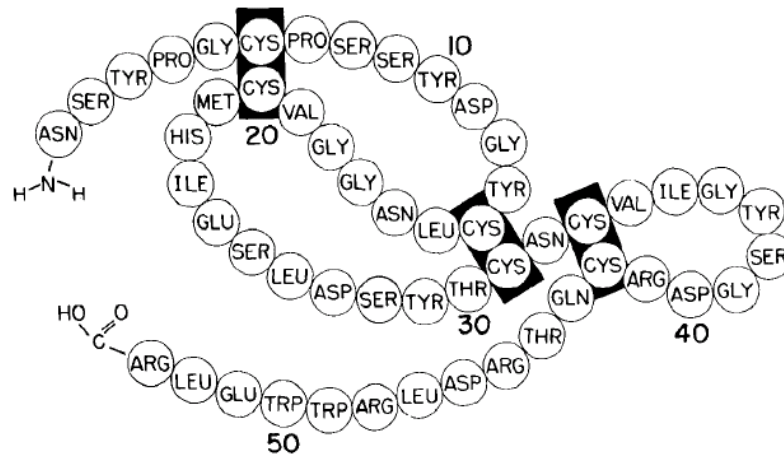
Figure 17: Amino acid sequence of the mouse *EGF* with placement of disulfide bonds
(Carpenter and Cohen 1979)

The disulfide bonds of the mouse *EGF* are required for their biological activity. Studies of the sequences revealed that some residues are completely conserved throughout the six cysteines and the glycines at the positions 18 and 39 (Figure 18). The positions 13, 15, 41 and 47 are at the growth factor receptor interface. There are considerable variations among the sequences, especially at the N- and C-termini.

Figure 18: Structural assignment of the known *EGF* sequences which bind to the *EGFR*
(Campbell et al. 1990)

2.4.7.2 Gene function

The *EGF* gene and the receptor are expressed in many tissues, including reproductive organs. Both genes are involved in angiogenesis, embryo implantation and development as well as in proliferation and differentiation of various cells (Andronowska et al. 2006). The *EGF* expression is regulated during the reproductive cycle and early pregnancy. This pattern of gene expression may be important during the development of the early conceptus.

The mRNA for the *TGFA*, *BTC*, and *HB-EGF* genes are present in the virgin mammary gland, but the levels decrease during pregnancy until the gene expression is virtually

disappearing during the lactation. In contrast, the expression of *EGF*, which is low in the virgin and the pregnant mammary gland, increases dramatically toward the end of the pregnancy and peaks during lactation, with high levels of *EGF* observed in milk. Expression of *EGF* decreases in the involuting mammary gland, while expression of *TGFA*, *BTC*, and *HB-EGF* increases. *EGF* is able to inhibit the apoptosis of apoptotic mouse mammary epithelial cells (Rosfjord and Dickson 1999). Immunohistochemical studies with antibodies to the extreme C-terminal region of the *EGFR* revealed localization in many of the same cell populations of the developing mouse mammary gland, including the cap cell layer (DiAugustine et al. 1997).

EGF may stimulate epithelial synthesis of type IV collagen, a component of the basal lamina which is required for the epithelial attachment and proliferation and is also required for the growth of cultured mammary gland. *EGF* causes the reappearance and growth of involuted mammary end buds in ovariectomized mice (Engelman et al. 1995). The *EGF* mRNA in the mouse mammary gland is increased by lactogenic hormones (Fenton and Sheffield 1993). *PRL* inhibits *EGF*-induced mitogenesis in the mammary gland. *EGF* is able to stimulate DNA synthesis in a normal mammary epithelial cell line, but the lactogenic hormone *PRL* inhibits the mitogenic effect. The inhibition was specific for *EGF*, because *PRL* did not alter the DNA synthesis induced by cholera toxin or *IGF-I* (Fenton and Sheffield 1993). An analysis of *in vivo* effects of *EGF* in 5-week-old ovariectomized mice has shown that *EGF* promotes normal ductal morphogenesis by stimulating proliferation in the end buds (Coleman et al. 1988).

The *EGF* gene is located on chromosome chromosome 8q23-q27 in pigs (Mendez et al. 1999). Brissenden et al. (1984) mapped the *EGF* locus to 4q21-4qter, possibly near the T-cell growth factor *(TCGF)*.

2.4.7.3 Polymorphism in the *EGF* gene

Shahbazi et al. (2002) identified a single-nucleotide substitution from G to A, at position 61 of the *EGF* gene in human. Frequencies of the A and G alleles of *EGF* were 56% and 44%, respectively. Cells from individuals which were homozygous for the A allele produced significantly less *EGF* than cells from the other homozygotes or the A/G heterozygotes. Compared to the A/A genotype, G/G was significantly associated with a risk of malignant melanoma. The A-G polymorphism of *EGF* is also involved in

the occurrence but also in the malignant progression of gastric cancer (Hamai et al. 2005).

2.4.8   Growth differentiation factor 8 (*GDF8*)

The *TGFB* superfamily encompasses a large number of growth and differentiation factors that play important roles in regulating the embryonic development and in maintaining the tissue homeostasis in adult animals. The *GDF8* gene, also called myostatin (*MSTN*) is a member of this superfamily with a role in the control and maintenance of skeletal muscle mass (McPherron et al. 1997).

2.4.8.1 Gene structure

The porcine *GDF8* gene is composed of three exons including 373, 374 and 381 bp. The active form of the protein comprises an 376 amino acids polypeptide that contains all the sequence hallmarks of the *TGFB* superfamily (Stratil and Kopecny 1999). The promoter region of the human *GDF8* gene contains an E-box sequence that binds muscle regulatory factors of the myogein differentiation antigen (*MyoD*) family. The *GDF8* is synthesized as a preprotein activated by two proteolytic cleavages. The removal of the signal sequence is followed by a cleavage at a tetrabasic processing site, resulting in a 26-kD amino-terminal propeptide and a 12.5-kD carboxy-terminal peptide, a dimer of which is the biologically active portion of the protein (Zimmers et al. 2002).

2.4.8.2 Gene function

The *GDF8* gene is expressed specifically in developing and adult skeletal muscle. During the early stages of embryogenesis, the *GDF8* expression is restricted to the myotome compartment of developing somites. At later stages and in adult animals, *GDF8* is expressed in many different muscles throughout the body (McPherron et al. 1997).

Postnatally, the *GDF8* mRNA was detected in the skeletal muscle and mammary gland in pigs. The expression in the tubuloalveolar secretory lobules of the lactating mammary gland is intriguing and indicates the possibility that *GDF8* performs a regulatory role pertaining to gestational or lactational mammary gland growth and development and/or metabolism (Ji et al. 1998).

The porcine *GDF8* gene is mapped to chromosome 15q2.3 by fluorescence *in situ* hybridization (Sonstegard et al. 1998). The *GDF8* gene maps to the chromosomal region 2q33.2 in human (Gonzalez-Cadavid et al. 1998).

2.4.8.3 Polymorphism in the *GDF8* gene

Three SNPs located in the 3' encoding region, the promoter region and the frist intron are already detected in the porcine *GDF8* gene. The mutation frequency for the SNP in the 3' encoding region (C to T) was relatively low. For the SNP in the 5' promoter region (T to A), the allele T dominates in the imported lean-type pig breeds such as Yorkshire, Landrace, Duroc, Hampshire, Pietrain and hybrid pigs. A similar pattern for the SNP in the region of frist intron (G to A) was found with G the dominant allele in Yorkshire, Landrace and their hybrids, while White pigs the frequency of allele A was much higher in Erhualian and Hubei (Jiang et al. 2002).

There were two SNPs in exon 2 at position 480 (G to T) and in third exon at position 1008 (A to G). Two mutations did not change the amino acid but there was a significant relationship between the polymorphism in third exon and back fat thickness. No significant relationship could be detected between the polymorphism in third exon and the lean meat percentage (Li et al. 2002). Subsequent DNA sequencing led to the detection of a nucleotide substitution (C2150T) in third exon of the *GDF8* gene (Cho et al. 2005, Stratil and Kopecny 1999). Although the C to T mutation did not change the type of the encoded amino acid (Stratil and Kopecny 1999), it could influence the mRNA stability (Zhi-Hua and John 1999). The T allele increased the number of muscle fibres increased the eye muscle area without decreasing the backfat thickness (Cho et al. 2005).

3   Material and methods

The aim of this study was to identify functional candidate genes for the inverted teat defect in pigs. In total eight functional candidate genes were investigated in this study. Genes were selected using two different strategies; in a first approach, genes differentially expressed between normal and inverted teats of sows with or without defect, were identified by microarray experiment (Figure 19). Thus first approach leads to the selection of the five functional candidate genes *CTGF*, *EGFR*, *IGF-II*, *EGF* and *GDF8*. The different expression of these genes was further validated using real-time PCR, an efficient method for quantification of mRNA transcription levels due to its high sensitivity, reproducibility and large dynamic range. In addition, real-time PCR is fast, easy to use and provides simultaneous measurement of gene expression in many different samples for a limited number of genes (Nygard et al. 2007, Radonic et al. 2004, Vandesompele et al. 2002). Using the method of real-time PCR the transcript abundance of the differentially expressed genes was quantified between normal and inverted teats.
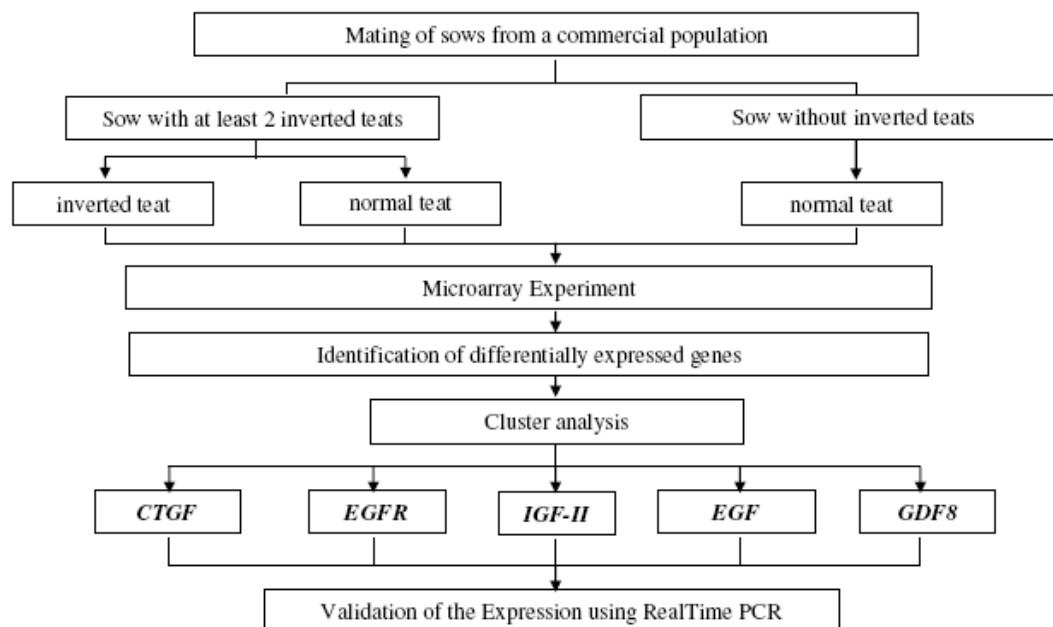
Figure 19: Overview of the first part of the experiment: analysis of the microarray

Secondly, genes were selected due to their described function in the literature (Figure 20). The *RLN3* gene and two of the known receptors, *GPCR135* and *GPCR142* were selected based on their described function in the literature. Previously an other member of the *RLN* family, the *RLN1* gene was already identified as a positional and functional

candidate genes, its role for the development in the mammary gland in pig could be confirmed (Chomdej 2005). To complete the analysis of these genes regarding their function, similar to the expression validation of the genes derived from the first approach, a real-time PCR was also performed for these genes.
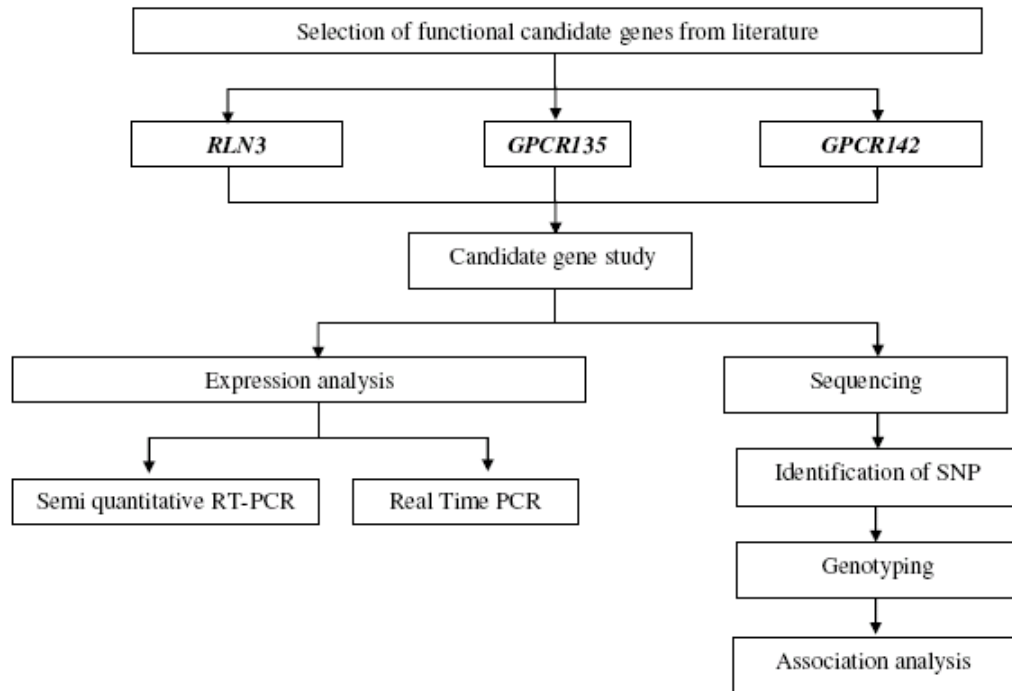
Figure 20:   Overview of the second part of the experiment, the analysis of the selected genes from literature

3.1   Animals

3.1.1   Animal for expression studies

The sows mated for the identification of functional candidate gene for the microarray experiment were kept at the experimental farm Frankenforst of the University Bonn. The dams derived from a commercial pig population of German breeding companies, the samples collected from these animals were used for RNA extraction for the microarray experiments, and the validation study using real-time PCR. Further DNA was isolated for additional genotyping at candidate genes and microsatellite loci for association studies and the validation of possible QTL alleles.

The sows were chosen according to teat characteristic as sow without defect or sow with defect. Two purebred German Landrace (DL) sows were slaughtered two days after weaning in a local slaughterhouse (one was sow without defect another one was

sow with defect). Teats of the sows were dissected into the different tissues of mammary gland, connective tissue and nipple. The samples were snap frozen in liquid nitrogen and stored at -80 °C until RNA isolation. Finally, two normal and one inverted teats from sow with defect and three teats from sow without defect were used for the microarray analysis. Additional teat samples of three young sows from the same breeding population were collected at a commercial slaughterhouse. The samples of one defect and one normal teat of two different sows with defect and two samples of normal teats of sows without defect were used for the experiments. The samples of the whole teat were kept into RNAlater (Sigma-Aldrich), one day later dissected into the three different tissue types before and stored at -80 °C until RNA isolation.

During these experiments, it was essential that the teat samples used for quantitative transcriptions analysis by real-time PCR where the same as the teat sample for the microarray experiments.

3.1.2   Animals for association studies

For SNP detection and the association analysis, samples of animals from an experimental population were used. The experimental population derived from a cross between Berlin miniature (MI) and Duroc (DU) pigs. DU sows were mated with a MI boar to produce the F1 which were further mated to generate the F2 animals. Later the F2 animals were crossed with DU boars; this backcross generation completed the DUMI resource population. In special the pure breed MI and the animals of the first generations showed a high incidence of the inverted teat defect, about 42% of the F1 animals were affected. As this defect was genetically based on the MI animals, the crossbreed with DU animals with relatively less udder problems provides excellent conditions for the detecting of the responsible genetic factors.

To confirm the findings in the experimental population, genetic studies were also performed in animals of the commercially used dam lines. Additional to the animals used for the expression studies, animals of this population were selected for association studies compromising 130 families according the affected sibpair design. For this population, information of the detected QTL regions was available from a previous work (Jonas et al. 2008). Purebreed DL and DE animals and their crossbreeds were used. This would enable the direct application of the finding to possible selection strategies in these commercial populations. For this task, the teat phenotypes of male castrated animals, tested at the performance station were observed after slaughtering

and teat samples collected. From the records of the animals, sow with defect/ sow without defect, fullsibpairs were described from which samples of the parents were collected from either artificial insemination stations or other breeding companies. In total samples of fullsib offspring including sire and dam of 130 families could be used for the association experiments.

Additional to the sampling of tissue of animals of these two populations, random samples of pigs of various breeds were used for the screening of polymorphisms within the candidate genes. This comprised samples (n=10) of animals from the breeds DU, Hampshire, DL, Pietrain and Wild pig. Because of the expected smaller amount of genetic conservation between breeds than within breeds, this strategy was found to implement a higher efficiency to detect polymorphisms in genes.

3.1.3   Phenotypes

Sows for sampling were chosen according their teat characteristics. For each group two sows showing at least one inverted teat were used for the experiment. The teat classification of animals was made based on the condition characterized by the failure of teats to protrude from the udder surface. For the selection of the commercial animals at least two inverted teats were detected in the udder of animals classified as affected.

3.2   Material

3.2.1   Chemicals and Kits

Beckman Coulter (Krefeld):	CEQ™ 8000 Genetic Analysis System, Dye Terminator Cycle Sequencing (DTCS), Glycogen, sample loading solution (SLS)
Biomol (Hamburg):	Phenol
Invitrogen Life Technologies (Karlsruhe):	DTT, SuperScript™ II RNase H⁻ Reverse Transcriptase, 5 × first strand buffer, Random Primers
MBI Fermentas (St. Leon-Rot):	Glycogen

Promega (Mannheim):   BSA, pGEM®-T vector, RQ1 RNase-free DNase,
                             RNasin Ribonuclease inhibitor, 2×rapid ligation
                             buffer, T4 DNA ligase, *Rsa*I endonuclease, *Taq*I
                             endonuclease,   *Msp*I   endonuclease,   *Hae*III
                             endonuclease

Qiagen (Hilden):   RNeasy® Mini kit, GenElute^TM Plasmid
                             Miniprep Kit

Roth (Karlsruhe):   Acetic acid, Agar-Agar, Ampicillin, Bromophenol
                             blue,   Dimethyl   sulfoxide   (DMSO),
                             Ethylenediaminetetraacetic acid (EDTA), Ethanol,
                             Ethidium bromide, Hydrochloric acid, Isopropyl -D-
                             thiogalactoside (IPTG), Nitric acid, Peptone, Sodium
                             acetate, Sodium carbonate, Sodium chloride, Sodium
                             hydroxide,   TrisX-Gal   (5   -bromo-4-chloro-3-
                             indolylbeta-D-galactopyranoside), Yeast extract

Sigma-Aldrich Chemie GmbH   Agarose, Ammonium acetate, Calcium chloride,
(Munich):   Formaldehyde, Glutamine, Isopropanol, Magnesium
                             chloride,   β-Mercaptoethanol,   Oligonucleotide
                             primers, Penicillin, 10 × PCR reaction buffer,
                             Potassium chloride, Sodium dodecyl sulfate (SDS),
                             Taq DNA polymerase, TRIReagent

Stratagene (Amsterdam):   5 c DH *Escherichia coli* competent cells

USB (Ohio):   ExoSAP-IT

3.2.2   Reagents and media

All solutions used in the experiments were prepared with deionized millipore water (ddH_2O). The pH values were adjusted with either sodium hydroxide (NaOH) or hydrochloric acid (HCl).

Agarose loading buffer	Bromophenol blue	0.0625 g
	Xylencyanol	0.0625 g
	Glycerol	7.5 ml
	ddH_2O added to	25 ml
Ampicillin (10 mg/ml)	Ampicillin powder	2 g
	Sterile, distilled water	40 ml
	Filtrate with 0.45 µl filter	
DEPC-treated water	DEPC	1 ml
	ddH_2O	1000 ml
	Incubation at 37 °C and heat inactivated by autoclaving (120 °C for 30 min)	
Digestion buffer	NaCl	100 mM
	Tris-HCl	50 mM
	EDTA pH 8.0	1mM
dNTP solution	dATP (100 mM)	10.0 µl
	dCTP (100 mM)	10.0 µl
	dGTP (100 mM)	10.0 µl
	dTTP (100 mM)	10.0 µl
	ddH_2O added to	400.0 µl
1M EDTA, pH 8.0	EDTA	37.3 g
	ddH_2O added to	1000 ml

10×FA buffer, pH 7.0	MOPS	41.8 g
	Sodium acetate	4.1 g
	EDTA (0.5M)	20.0 ml
	ddH$_2$O added to	1000.0 ml
1.2% FA gel	Agarose	1.2 g
	10×FA buffer	10.0 ml
	DEPC ddH$_2$O	90.0 ml
	Ethidium bromide	2.0 µl
	Formaldehyde (37%)	1.8 ml
IPTG solution	IPTG	1.2 g
	ddH$_2$O added to	10.0 ml
LB-agar plate	Sodium chloride	8.0 g
	Peptone	8.0 g
	Yeast extract	4.0 g
	Agar-Agar	12.0 g
	Sodium hydroxide (40 mg/ml)	480.0 µl
	ddH$_2$O added to	800.0 ml
LB-broth	Sodium chloride	8.0 g
	Peptone	8.0 g
	Yeast extract	4.0 g
	Sodium hydroxide (40 mg/ml)	480.0 µl
	ddH$_2$O added to	800.0 ml
Phenol Chloroform	Phenol : Chloroform	1 : 1 (v/v)
Proteinase K solution	Protein K in 1×TE bufer	2% (w/v)
SDS solution	Sodium dodecylsulfat in ddH$_2$O	10% (w/v)

| 3M Sodium Acetate, pH 5.2 | Sodium Acetate | 123.1 g |
| | ddH_2O added to | 500 ml |

TAE (50×) buffer, pH 8.0	Tris	242.0 mg
	Acetic acid	57.1 ml
	EDTA (0.5 M)	100.0 ml
	ddH_2O added to	1000.0 ml

TBE (10×) buffer	Tris	108.0 g
	Boric acid	55.0 g
	EDTA (0.5 M)	40.0 ml
	ddH_2O added to	1000.0 ml

TE (1×) buffer	Tris (1 M)	10.0 ml
	EDTA (0.5 M)	2.0 ml
	ddH_2O added to	1000.0 ml

| X-gal | X-gal | 50.0 mg |
| | N, N'-dimethylformamide | 1.0 ml |

3.2.3   Used software

Association analysis	Family-based association test (FBAT); http://www.biostat.harvard.edu/~fbat/fbat.htm
Clustering of genes from Microarray	Database for annotation, vizualisation and integrated discovery (DAVID); http://david.abcc.ncifcrf.gov/
Correspondences between human and pig chromosomal segments	Comparative cytogenetic map; http://www2.toulouse.inra.fr/lgc/pig/compare/compare.htm
General sarch, fragment comparision	Basic local alignment search tool (BLAST); http://www.ncbi.nlm.nih.gov/BLAST/

Manipulate and display a DNA sequence	http://www.vivo.colostate.edu/molkit/manip/index.html
Multi amino acid alignment	http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_multalinan.html
Multi sequence alignment	http://prodes.toulouse.inra.fr/multalin/multalin.html
Prediction to affect protein function (SIFT)	http://blocks.fhcrc.org/sift/SIFT.html
Primer design	Primer3; http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi
Restriction enzyme analysis	http://tools.neb.com/NEBcutter2/index.php
Statistical analysis	SAS (version 9.1); SAS Institute Inc., NC, USA
Translation of a base sequence	http://molbiol.ru/eng/scripts/01_13.html

3.2.4   Equipment

ABI PRISM® 7000 SDS	Applied Biosystems, Foster city, USA
Centrifuge	Hermle, Wehingen, Germany
CEQ™ 8000 Genetic Analysis System	Beckman Coulter GmbH, Krefeld, Germany
Electrophoresis (for agarose gels)	BioRad, Munich, Germany
Incubator	Heraeus, Hanau

Millipore apparatus   Millipore corporation, USA

PCR thermocycler (PTC100)   MJ Research, USA & BioRad, Germany

pH Meter   Kohermann

Power supply PAC 3000   Biorad, Munich

Spectrophotometer (DU-62)   Beckman, Unterschleissheim-Lohhof

Spectrophotometer, Ultrospec™ 2100 *pro* Amersham Bioscience, Munich
UV/Visible

Thermalshake Gerhardt   John Morris scientific, Melbourne

Tuttnauer autoclave   Connections unlimited, Wettenberg

Ultra low freezer (-80°C)   Labotect GmbH, Gottingen

UV Transilluminator (Uvi-tec)   Uni Equip, Martinsried, Germany

3.3   Methods for transcriptome analysis

3.3.1   RNA isolation and cDNA synthesis

Samples of different tissues and animals were used for RNA isolation using TRIReagent (Sigma-Aldrich, Munich, Germany). The samples were first grinded in a mortar, then mixed and homogenized with 1 ml TRIReagent. The samples were incubated for 10 min at room temperature to ensure the complete dissociation of the nucleoprotein complexes before adding 0.2 ml of chloroform. The mixtures were shaken and left at room temperature for 10 min and centrifuged at 7500 x *g* for 15 min and 4 °C. The upper aqueous phases were transferred to another sterilized tube and RNA was precipitated using 0.5 ml of isopropanol. After incubation at room temperature for 10 min, the samples were centrifuged at 7500 x *g* for 10 min and 4 °C.

After this step the RNA pellet was visible, and subsequently washed using 75% (v/v) ethanol. Centrifugation was then performed and the RNA pellets were air-dried and resuspended in 40 µl of RNase free water.

In order to remove possible contaminating genomic DNA, the extracted RNA was treated with 5 µl RQ1 DNase buffer, 5 units DNase and 40 units RNase inhibitor in a 40 µl reaction volume. The mixture was incubated at 37 °C for 1 h followed by purification using chemicals and protocols of the RNeasy Mini Kit. The concentrations of the RNA after clean-up were determined spectrophotometrically at 260 and 280 nm; the purity of the RNA was estimated by the ratio A260/A280 with respect to contaminants that absorb the UV. Additional examination of integrity was done in a denaturing agarose gel electrophoresis and ethidium bromide staining. Finally, the purified RNA was stored at -80 °C for further analysis.

The individual RNA was used to synthesize first-strand cDNA using SuperScript II enzyme. First 1 µl of Oligo (dT)$_{12}$ primer (100 µM) and 1 µl random primer were added to 1 µg of total RNA and incubated at 68 °C for 5 min, followed by cooling on ice for 2 min. A transcription mixture including 4 µl first strand 5× buffer, 1 µl 0.1 DTT, 1 µl dNTP mix (10 mM each), 1 µl (200 units) SuperScript II reverse transcriptase, 1 µl (40 units) of RNasin Ribonuclease inhibitor and RNase-free water was prepared for a final volume of 20 µl. The mix was incubated at 25 °C for 5 min followed by 42 °C for 1h and stopped by heating at 70 °C for 15 min. The cDNA was diluted using 80 µl RNase free water. The resulting cDNA was tested using housekeeping gene (*18S*) primers and kept at -20 °C until use.

3.3.2   Microarray analysis

The RNA samples were sent to the Research Institute in Dummerstorf, Germany for hybridisation. The microarray analysis was performed by RZPD (Deutsches Ressourcenzentrum für Genomforschung GmbH) in Berlin, Germany. The relative abundance of the transcript and the p-value which indicate the significance of the detection call was determined from the results. According to the algorithm the information of whether the gene is expressed with a defined confidence level or not (detection call) was providing. The detection call was either present (P), marginal (M) or absent (A). The three categories were made depending on the p-values and the number of stat pairs used. The number of stat pairs used was either 20, 16, 11 or 9. The

default p-values were alpha1 = 0.04 and alpha2 = 0.06 for a number of 16 to 20 stat pairs used and alpha1 = 0.05 and alpha2 = 0.065 for 11 probe sets used. The detection call was used to determine the change in the hybridization intensity levels of two different probe sets on the same array. The house keeping genes Actin and *GAPDH* were used for the normalization of the microarrays. A list of relative expression levels of each transcript on the different chips was provided by RZPD. As a one dye microarray was used, each microarray chip included the information of one sample.

3.3.3   Ligation and transformation

A fragment of genomic DNA derived from a PCR reaction was ligated into a plasmid pGEM®-T vector. In total 5 µl reaction mix containing 2.5 µl 2×ligation buffer, 0.5 µl pGEM®-T (50 µg/µl), 0.5 µl T4 DNA ligase (3 units/µl) and 1.5 µl target template were used. The reaction was incubated at room temperature for 1 h or at 4 °C overnight. It was expected that most of the vectors had inserted a DNA fragment after the incubation period.

The entire ligation reaction was added to 100 µl volume of competent JM109 *E. coli* cells and incubated on ice for 30 min. This mixture was further heat-shocked in a 42 °C water bath for 90 sec and immediately returned to ice for 2 min. Eight hundred microlitres of nutrient medium (LB-broth) were added and the mix was incubated at 37 °C for 90 min in a thermal shaker. At the same time, ampicillin treated LB-agar (50 mg/L LB-agar) plates including 20 µl of X-Gal (50 mg/ml in N, N'-dimethyl-formamide) and 20 µl of IPTG were prepared. At the end of incubation period, each transformation culture was plated on two of the prepared LB-agar plates and incubated at 37 °C overnight.

After the incubation, the colonies were screened by blue white screening test based on the activity of β-galactosidase as white and blue for the presence and absence of inserted DNA fragments. The *lacZ* gene in the pGEM®-T vector produces β-galactosidase which interacts with IPTG to produce a blue color of the bacteria cultures. If the insert is successfully ligated, the *lacZ* gene is disrupted; β-galactosidase can not be produced, the colonies are all white. White bacteria colonies and blue colonies were picked up from the plates and suspended in 30 µl 1×buffer for further testing. Each blue colony was used as a control of the length of the amplified DNA fragment in comparison to the vector without fragment from each plate. The same colonies were

cultured in 5 ml ampicillin/LB-broth (5 mg/100 ml) in a shaking incubator at 37 °C for further plasmid isolation.

To confirm the insertion of the right fragments into the plasmid, a M13 PCR was performed. The bacterial suspensions were boiled at 95 °C for 15 min and these lysed bacterial solutions were used as templates. The M13 PCR was carried out in a 20 μl reaction including 1 μl 10×PCR buffer, 10 μl lysed bacterial solution, 0.5 μl dNTP (10 mM), 0.5 μl (10 μM) of each M13 primer (forward: 5'-TTG-TAA-AAC-GAC-GGC-CAG-T-3'; reverse: 5'-CAG-GAA-ACA-GCT-ATG-ACC-3') and 0.1 U Taq polymerase. The PCR reaction was performed with a thermal cycling program of 95 °C for 5 min followed by 35 cycles of 94 °C for 30 sec, 60 °C for 30 sec, 70 °C for 1 min and an additional extension step for 5 min at 70 °C. An aliquot of 5 μl PCR product was then electrophoresed in 1.5% (w/v) agarose gel with 0.8 μg/ml ethidium bromide (4 μl) in 1×TAE buffer. Under UV-transilluminator, length differentiation of PCR fragments was identified. The M13 PCR products from white colonies were selected for subsequent sequencing while bacterial cultures of these colonies were expanded in a volume of 5 ml and incubated at 37 °C overnight in a shaking incubator for plasmid isolation. The M13 products were used as templates for sequencing according to the Quick Start Kit (BeckmanCoulter, Krefeld, Germany) including DNA polymerase, pyrophosphatase, buffer, dNTP, dye terminator (ddNTP) and either the gene specific forward or reverse primer. After the sequencing PCR, 3 M NaOAc, 100 mM EDTA and glycogen were added to stop the reaction. To each sample, 60 μl of 98% ethanol (Roth) was added and mixed well by vortexing and then centrifuged for 15 min at 12000 x g at 4 °C. All liquid was removed and replaced with 200 μl 70% ethanol without mixing and centrifuged again for 15 min at 12000 x g at 4 °C. The ethanol was then removed and the sample was air dried for 10 min. The sample was then resuspended in 40 μl SLS (Beckman Coulter) then transferred manually to a CEQ sample plate and overlaid with mineral oil. Samples were sequenced using CEQ™ 8000 Genetic Analysis System (Beckman Coulter). The sequence method was based on a chain-reaction method. The sequencing PCR was performed to amplify the target fragment, by addition of the labled base (ddNTP) the amplification reaction stipped after that particular length. Using this method it can be suggested that all possible fragment length are amplified in the product and can further be detected by sequencing, for example using a capillary sequencer.

The results from sequence analysis were compared with published sequences using the program BLAST (http://www.ncbi.nlm.nih.gov/BLAST/). Plasmids from those clones

with identity percentage higher than 90% were considered as the target gene fragment. Plasmids were then isolated for downstream application.

3.3.4   Plasmid DNA isolation

The GenElute^{TM} Plasmid Miniprep Kit was used to isolate plasmid DNA from the bacteria including the insert after the transformation. of the bacterial culture was centrifuged at 12000 x g for 1 min for harvesting cells, the supernatant was discarded. These cells were resuspended in 200 µl of resuspension solution, vortexed, 200 µl of lysis solution were added and subsequently mixed by inversion of tubes until it became a clear and viscous solution. After incubating at room temperature for 4 min 350 µl neutralization/binding buffer was added for cell precipitation, mixed gently and centrifuged at 12000 x g for 10 min. At the same time, the GeneElute Miniprep column was prepared by adding 500 µl preparation solution, centrifuging shortly and discarding the flow-through. After that, the clear supernatant was transferred to this binding column and centrifuged at 12000 x g for 1 min. The flow-through was discarded and the column was washed by adding 750 µl wash solution followed by centrifugation at 12000 x g for 1 min. To elute the DNA, the column was transferred to a fresh collection tube; 50 µl ddH$_2$O was added and centrifuged at 12000 x g for 1 min. The column was discarded, the DNA plasmid was collected in the water in the tube.

For the determination of plasmid size and quality, 5 µl of plasmid DNA was checked together with 2 µl loading buffer using agarose gel electrophoresis. In addition, the quantity of plasmid was also measured by reading the absorbance at 260 nm in a spectrophotometer. An aliquot of the DNA plasmid solution was used to ckeck the fragment using sequencing. The remain part was stored at -20 °C and further used as template for setting up the standard curve in the real-time PCR.

3.3.5   Real-time PCR

An ABI Prism$^®$ 7000 SDS was used for real-time PCR based on the changes in fluorescence proportional to the increase of product. SYBR$^®$ Green, which emits a fluorescent signal upon binding to double stranded DNA, was used as a detector. Fluorescence values were recorded during every cycle representing the amount of product amplified to a point known as threshold cycle (Ct). The higher the initial transcript amount, the sooner accumulated product was detected in the PCR process.

Using the web-based program Molbiol (http://molbiol.ru/eng/scripts/01_07.html) the concentration of the plasmid (ng/µl) was converted into the numbers of molecules used for the plasmid serial dilution preparation. The plasmid concentration was diluted several folds from 10^8 to 10^1 so that the concentration would cover the range of target concentration in the samples. The PCR assay was started to test whether a suitable standard curve could be achieved revealing a high PCR efficiency.

The optimal primer concentration was obtained by testing different primer combinations from 200 nM to 600 nM prior to the quantification. The list of primers used in real-time PCR for quantification transcription analysis is shown in table 3. Results from these primer combinations were compared and the one with the lowest threshold cycle and minimal non-specific amplification was selected for subsequent reactions. After selection of the primer concentration, a final assay consisted of 2 µl cDNA as template, iTag^{TM} SYBR$^®$ Green Supermix with ROX, optimized level of forward and reverse primer and optimized buffer components were performed in 20 µl reaction volume. The amplification reaction started with an initial denaturation at 95 °C for 3 min followed by 40 cycles of 95 °C for 15 sec denaturation and 60 °C for 45 sec annealing and extension. A dissociation curve was generated at the end of the last cycle by collecting the fluorescence data at 60 °C and observing measurements every 7 sec until the temperature reached 95 °C. Final quantification analysis was done by amplifying serial dilutions of target plasmid DNA. The concentration of unknown cDNA was calculated according to the standard curve, and expression levels of transcripts were described relatively to the transcript of the house keeping *RPL32* gene.

Table 3:   List of primers used for real-time PCR with the sequence, the annealing temperature (Ta) and the product size

Gene	Sequence	Ta (C°)	Product size (bp)
RLN3	5' GCAGAGGCTGCTGATTTCAC 3'	61	297
	5' GCAGAGGCTGCTGATTTCAC 3'		
GPCR135	5' CTGCTGGTTCTCTACCTG 3'	59	197
	5' GTTCATGGACGTCACTACGG 3'		
GPCR142	5' CTTTCTGGGTCAATGCGTCT 3'	59	197
	5' AGTGTCAGAAGGTGGGCAAG 3'		
CTGF	5´ TTGCAGACTGGAGAAGCAGA 3´	63	223
	5´ AAGGGTGGTGGTTCTGTGAG 3´		
EGFR	5´ CCTTGGGAACTTGGAGATCA 3´	63	208
	5´ GGTTTTATTGGCCCCGTAGT 3´		
IGF-II	5´ ACACCCTCCAGTTTGTCTGC 3´	55	212
	5´ GGGGTATCTGGGGAAGTTGT 3´		
EGF	5´ TGCCTTCCTCTCAGTCCAGT 3´	55	199
	5´ GGGCCAAAACCATTCCTATT 3´		
GDF8	5´AACAGCGAGCAAAAGGAAAA 3´	56	201
	5´ATCAATCAGTTCCCGGAGTG 3´		

3.3.6   Semi-quantitative RT-PCR

The cDNA from tissues of two different male pigs (muscle, heart, spleen, lymph nodes, skin, brain, teat, lung, testis, tonsil, liver and kidney) and one sow (uterus and inverted teat) was used for semi-quantitative RT-PCR. The forward primer 5'-CGA-GCG-GTC-ATC-TTT-ACC-TG-3' and reverse primer 5'-GCA-GAG-GCT-GCT-GAT-TTC-AC-3' were used for expression analysis of *RLN3* gene tissues following the protocol for semi-quantitative RT-PCR starting with an initial denaturation at 95°C for 5 min followed by 35 cycles with 95 °C for 30 sec, 55 °C for 30 sec and 72 °C for 1 min with a final elongation step at 72 °C for 10 min.

To analyse the expression pattern of the *GPCR135* gene the forward primer 5'-CTG-CTG-GTT-CTC-TAC-CTG-3' and reverse primer 5'-GTT-CAT-GGA-CGT-CAC-TAC-GG-3' were used for semi-quantitative RT-PCR. The PCR protocol was slightly different, starting with an initial denaturation at 95 °C for 5 min followed by 35 cycles

with 95 °C for 45 sec, 59 °C for 45 sec and 72 °C for 1 min with a final elongation step at 72 °C for 10 min.

The forward primer 5'-CTT-TCT-GGG-TCA-ATG-CGT-CT-3' and reverse primer 5'-CTT-CCT-TTA-GGG-CCA-CCT-GT-3' designed wihin the sequence of the *GPCR142* gene were used to analyse the expression pattern of this genes. The protocol for semi-quantitative RT-PCR started with an initial denaturation at 95°C for 5 min followed by 35 cycles with 95 °C for 45 sec, 59 °C for 45 sec and 72 °C for 1 min with a final elongation step at 72 °C for 10 min. For all analysis the expression of *18S* rRNA was used as an internal reference.

3.4   Methods for genome analysis

3.4.1   DNA extraction

For the DNA isolation the tissue sample (approximately 100 mg) was cut into small pieces and placed in a 1.5 ml tube. 700 µl digestion buffer, 70 µl 10% SDS and 18 µl proteinase K were added for protein digestion. The mixture was incubated over night at 37 °C in a thermo shaker at 90 rpm. Completely digested tissue resulted in a viscous homogeneous solution. Seven hundred microlitres of phenol-chloroform was added into each tube and gently mixed by several inversions until an emulsion was formed. After centrifugation at 7000 x *g* for 10 min the mixture was separated into three phases, a lower phenol-chloroform phase, an interphase of precipitated protein and an upper phase containing DNA. The DNA phase was transferred to another 2 ml tube, 700 µl chloroform were added, mixed and centrifuged at 7000 x *g* for 10 min. The aqueous phase was transferred to a fresh tube and mixed with 700 µl isopropanol and 70 µl sodium acetate for DNA precipitation. The solution was centrifuged at 7000 x *g* for 5 min, the DNA pellet was then visiable at the bottom of the tube. The supernatant was removed and the pellet washed with 200 µl 70% ethanol to remove excess salt. After air-drying, the pellet was dissolved in 200 to 500 µl 1×TA buffer. DNA concentration and integrity was evaluated by a spectrophotometer. The working solution of DNA was prepared by diluting stock DNA in 1×TA buffer to the concentration of 50 ng/µl. Stock DNA solution was stored at -20 °C and the working solution was kept at 4 °C.

A modified protocol was used to isolate DNA from agarose gel. Twenty microlitres of the PCR product were separated in a 0.8% (w/v) agarose gel in 1×TAE buffer. The

DNA fragment was visualized under an ultraviolet transilluminator, cutted from the gel, placed in a fresh tube and kept at -20 °C. After at least 30 min, 500 µl 1×TE buffer were added and gel and solution homogenized. An equal volume (500 µl) of phenol:chloroform (1:1) was added, mixed and vortexed well. After centrifugation at 7500 x g for 15 min at 4 °C, the upper aqueous layer was carefully transferred to a new microcentrifuge tube. Five hundred microlitres of chloroform were added, shaken vigorously and centrifuged at 7500 x g for 10 min. The upper phase was transferred to a new tube and DNA was precipitated by adding the double volume of cool 100% ethanol and 1:10 volume of 3 M NaOAc (pH 5.2) and mixed by inversion. The precipitation was performed by incubation at -20 °C overnight or alternatively at -80 °C for 2 hours. The DNA was recovered by centrifugation at 7500 x g for 30 min at 4 °C. To remove residual salt in the sample, the DNA pellet was washed in 75% (v/v) ethanol and centrifugation for 5 min. In a last step, the ethanol was transferred and the pellet left for drying at room temperature before resuspending using 7 µl distilled water. The purified DNA were kept at -20°C.

3.4.2   Polymerase chain reaction

Base on the published sequence of the porcine *RLN3* gene (GeneBank accession number AB076661) and using the software Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) the forward primer 5'-CGA-GCG-GTC-ATC-TTT-ACC-TG-3' and reverse primer 5'-TCT-GCG-TTG-GAA-TCT-GTG-TC -3' were designed. Primers were used to sequence the start and the end of the intron part of the *RLN3* gene. After the first sequences were available, different primers were designed to sequence the complete *RLN3* gene.

The sequence of the genomic DNA of the porcine *GPCR135* gene was obtained starting with heterologous primers (forward primer 5'- CTG-CTG-GTT-CTC-TAC-CTG -3' and reverse primer 5'-GGT-TGA-GGC-AGC-TGT-TGG-AG-3') designed from conserved regions of the human (AY394501), mouse (AY633762) and rat (NM_001008310) and subsequent using homologous primers. The forward primer (5'- CCA-TCT-TCT-CCA-CCA-CCA-TC -3') and the reverse primer (5'- GTT-CAT-GGA-CGT-CAC-TAC-GG-3') were designed from the 859 bp fragment to screen for polymorphisms within the *GPCR135* gene.

The primer pair was designed based on the porcine *GPCR142* gene sequence (GeneBank accession number AY633768) and using the software Primer3. The forward

primer (5'- CTT-TCT-GGG-TCA-ATG-CGT-CT -3') and reverse primer (5'- CTT-CCT-TTA-GGG-CCA-CCT-GT -3') were used for the PCR reaction to screen for polymorphisms within the sequence of the *GPCR142* gene

3.4.3   Sequencing

Comparative sequencing was done to screen for polymorphisms within the PCR fragment amplified using specific primers. The PCR product was checked on 1.5% agarose gel, 5 µl PCR product was incubated with 1 µl ExoSAP-IT at 37 °C for 30 min followed by ExoSAP-IT inactivation at 80 °C for 15 min. After clean up, the PCR product was mixed with 2 µl of either forward or reverse primer, 4 µl DTCS master mix (DNA polymerase, pyrophosphatase, buffer, dNTPs, and dye terminators) and 8 µl ddH$_2$O. The sequence PCR reaction run for 30 cycles, starting with a denaturation step at 96 °C for 20 sec followed by annealing at 50 °C for 20 sec and an extension step at 60 °C for 4 min. A stop solution including 2 µl 3 M NaOAc (pH 5.2), 2 µl of 100 mM EDTA (pH 8.0) and 1 µl glycogen (20 mg/ml) were added to the product after the sequencing PCR. 60 µl cold 95% ethanol were added. The solution was mixed throughly and centrifuged at 12000 x *g* for 15 min at 4 °C. The supernatant was removed and the pellet was rinsed two times with 200 µl cold 70% ethanol with centrifugation steps at 12000 x *g* at 4 °C for 5 min in between. The supernatant was removed and the pellet was air dried or vacuum dried for 10 min. The pellet was resuspended with 40 µl of sample loading solution, transferred to the wells of the sample plate and overlaid with one drop of light mineral oil. The separation buffer was prepared in the buffer plate, both plates were loaded into the CEQ8000 sequencer. A fragment specific sequencing program was started. A open source software program for comparative sequencing was used to polymorphism screening.

3.4.4   Genotyping

3.4.4.1 Relaxin3 (*RLN3*)

Using the DNA samples, animals were genotyped using the PCR-RFLP (restriction fragment length polymorphim) method. The restriction enzymes were selected according to the recognition (http://tools.neb.com/NEBcutter2/index.php) of the polymorphic sites. An 434-bp fragment was amplified using the forward primer 5'-CTA-GGG-TTG-GCT-TCC-TGC-GG-3' and the reverse primer 5'-TCA-CGG-ATA-

CTA-GTT-GGG-TTC-A-3' within the *RLN3* gene to perform the PCR-RFLP analysis using *Rsa*I endonuclease. An 334-bp fragment was amplified using the forward primer 5'-CAG-GCT-GAG-TCA-CAA-GAA-CAG-3' and the reverse primer 5'-GCA-GAG-GCT-GCT-GAT-TTC-AC-3' of *RLN3* gene. *Taq*I endonuclease was used as the enzyme for the PCR-RFLP digestion.

3.4.4.2 G-protein coupled receptor 135 (*GPCR135*)

The *GPCR135* gene was genotyped by PCR-RFLP using the *Msp*I endonuclease restriction enzyme. Digestion of the resulting 859 bp product (forward primer 5'-CTG-CTG-GTT-CTC-TAC-CTG-3' and reverse primer 5'-GGT-TGA-GGC-AGC-TGT-TGG-AG-3') with *Msp*I endonuclease revealed the polymorphism with two alleles as previous detected by sequencing. An 198 bp fragment was amplified using the forward primer 5'-CTG-CTG-GTT-CTC-TA-CCT-G-3' and reverse primer 5'-GTT-CAT-GGA-CGT-CAC-TAC-GG-3' to genotype animals at the loci of the other SNP detected in the *GPCR135* gene using the *Mnl*I endonuclease

3.4.4.3 G-protein coupled receptor 142 (*GPCR142*)

The forward primer (5'- CTT-TCT-GGG-TCA-ATG-CGT-CT -3') and reverse primer (5'- CTT-CCT-TTA-GGG-CCA-CCT-GT -3') generated an 1002 bp fragment in the sequence of the porcine *GPCR142* gene. PCR-RFLP analysis was performed using the *Hae*III endonuclease. All PCR and PCR-RFLP products were run in 1.5% and 3.0% agarose gels respectively, stained with ethidium bromide, and visualized under UV irradiation.

3.5   Statistical analysis

3.5.1   Gene mapping

The gene positions were mapped using linkage and physical mapping. For the linkage mapping, two point and multipoint procedures of the CRI-MAP package version 2.4 were used (Green 1992).

3.5.2   Radiation hybrid (RH) mapping

The RH mapping of the porcine *RLN3* gene was performed using the INRA-Minnesota 7000 rad radiation hybrid panel (IMpRH) (Yerle et al. 1998). Data analysis was performed using software available at IMpRH database (http://imprh.toulouse.inra.fr) for chromosome assignment.

3.5.3   Microarray experiment

A list of all transcriptomes and their relative expression was provided by RZPD for the different sample used for the microarray study. In the first step these data were used to calculate correlations between the different samples using the procedure PROC CORR of the SAS package. Further the different expressed genes were identified comparing the transcript expression between two samples (the less correlated plates were used for this approach). Genes were considered as differentially expressed if the divison of the two expression revealed a value of smaller than 0.5 or larger than 1.5. The genes with the most overlapping results were then considered as most interesting candidate genes. Further the different expressed genes were used for a cluster analysis using the open source program DAVID.

3.5.4   Variance analysis of the expression analysis

The SAS package version 8.02 (SAS Inc., Cary, NC, USA) was applied for statistical analyses. Analysis of variance using PROC GLM of SAS was performed to compare relative abundance of mRNA level. Means of the relative abundance of mRNA level between inverted and normal teat were compared using a *t-test*.
For comparison of the mRNA level from real-time PCR in different tissues only the mRNA level data and the phenotypes (inverted teat trait data) were used.

The general linear model used was:

$$Y = \mu + a + \varepsilon$$

The fixed effect a was the phenotype, the expression Y was tested against the phenotype. No random factor was used, the model implemented the overall mean μ and the random error ε.

3.5.5   Association analysis

The Family-Based Association Test (FBAT) (version 1.4) was used to perform the association analysis of the genotypes with the teat characteristics (Horvath et al. 2001). FBAT perform the testing by two-step procedure. First the test statistic is defined showing the association between the trait locus and the marker locus. In the second step the distribution of the data of genotypes are tested under the hypothesis. The genotypes of the offspring are treated as random (Rabinowitz and Laird 2000).

4   Results

4.1   Microarray study

4.1.1   Microarray analysis

The results of the genome-wide gene expression analysis were compared transcript on the microarray between the pooled results of the three tissues, mammary gland, connective tissue and nipple by calculating the average of the three tissues for each teat sample. To investigate the different expressed genes, the results were compared between the normal teat from the sow without defect, normal teat from the sow with defect and the inverted teat from the sow with defect. In total, the expression of 9273 transcripts was analyzed. The analysis was focused on the 1253 (13.5% of the transcripts assayed) differentially expressed transcripts between normal (from both sow without defect and sow with defect) and inverted teats. In total, 695 transcripts were found being higher expressed in inverted compared to normal teats. A clustering analysis was performed by an internet–based program (DAVID), revealing differentially expressed transcript factor were categorized in 15 different clusters. The results were shown in appendix 9.1. The clusters from up-regulated genes related with the mammary gland development were shown in table 4.

Table 4: The up-regulated genes in inverted compared to normal teats (the divison of
the two expression revealed a value of larger than 1.5) as grouped in the
clusters related with the mammary gland development

Group	AFFY_ID	Gene Acc	Gene name
Development	Ssc.8562.3.A1_at	NM_213833	*CTGF*
	Ssc.148.1.S1_at	NM_214041	*IL10*
	Ssc.20.1.S1_at	NM_213997	*IL18*
Regulation of physiological process	Ssc.148.1.S1_at	NM_214041	*IL10*
	Ssc.8562.3.A1_at	NM_213833	*CTGF*
	Ssc.10015.1.A1_at	NM_214084	*VEGF-A*
	Ssc.20.1.S1_at	NM_213997	*IL18*
	Ssc.16187.1.S1_at	NM_214228	*PTGDS*
Disulfide bond	Ssc.10015.1.A1_at	NM_214084	*VEGF-A*
	Ssc.670.1.S1_at	NM_214392	*LYZ*
	Ssc.148.1.S1_at	NM_214041	*IL10*
	Ssc.20525.1.S1_at	NM_213883	*IGF-II*
	Ssc.11992.1.A1_at	NM_213971	*CLU*
Phosphorylation	Ssc.9075.1.A1_at	NM_213880	*C-JUN*
	Ssc.9819.1.S1_at	NM_213911	*PGRMC1*
Cell proliferation	Ssc.148.1.S1_at	NM_214041	*IL10*
	Ssc.10015.1.A1_at	NM_214084	*VEGF-A*
	Ssc.20.1.S1_at	NM_213997	*IL18*
Cytokine-cytokine receptor interaction	Ssc.148.1.S1_at	NM_214041	*IL10*
	Ssc.55.1.S1_at	NM_214007	*EGFR*
	Ssc.20.1.S1_at	NM_213997	*IL-18*
Membrane-bound organelle	Ssc.16187.1.S1_at	NM_214228	*PTGDS*
	Ssc.3706.1.S1_at,	NM_214127	*SOD2*
Receptor binding	Ssc.148.1.S1_at	NM_214041	*IL10*
	Ssc.20525.1.S1_at	NM_213883	*IGF-II*
	Ssc.10015.1.A1_at	NM_214084	*VEGF-A*
	Ssc.20.1.S1_at	NM_213997	*IL18*
	Ssc.7243.1.A1_at	NM_001009580	*CXCL12*

In total 558 transcripts were found being lower expressed in the samples of the inverted compared to normal teats. A number of genes were also identified being down-regulated in inverted compared to normal teats, genes could be assigned to 11 different clusters. The results were shown in appendix 9.2. The clusters from up-regulated genes related with the mammary gland development were shown in table 5.

Table 5 The up-regulated genes in inverted compared to normal teats (the divison of the two expression revealed a value of smaller than 1.5) as grouped in the clusters related with the mammary gland development

Group	AFFY_ID	Gene Acc	Gene name
EGF-like	Ssc.87.1.S1_at	NM_214020	*EGF*
domain	Ssc.16045.2.A1_at	NM_001001771	*FBN1*
	Ssc.16345.1.S1_at	Y11683	pp47 protein; *MFG-E8*
Glycoprotein	Ssc.16127.1.S1_at	AAC16735	*MC2R*
	Ssc.335.1.S2_at	NM_214435	*GDF8*
	Ssc.136.1.S1_at	NM_214036	*VIPR1*
	Ssc.87.1.S1_at	NM_214020	*EGF*
	Ssc.16045.2.A1_at	NM_001001771	*FBN1*
	Ssc.16345.1.S1_at	Y11683	pp47 protein; *MFG-E8*
Receptor	Ssc.5105.2.S1_a_at	AAG33870	*GPRC5B*
	Ssc.4253.1.S1_at	NM_214272	*TGFBR3*
	Ssc.16006.1.S1_at	NM_214036	*PIGR*
	Ssc.136.1.S1_at	NM_214159	*VIPR1*
Signal	Ssc.335.1.S2_at	NM_214435	*GDF8*
transducer	Ssc.5105.2.S1_a_at	AAG33870	*GPRC5B*
activity	Ssc.4253.1.S1_at	NM_214272	*TGFBR3*
	Ssc.16006.1.S1_at	NM_214036	*PIGR*
	Ssc.29100.1.S1_at	NM_001024695	*CCL28*

Two genes (*EGF* and *GDF8*), shown to be lower expressed in the teat of sow with defect, were selected according to their biological function as growth factor related to development of mammary gland, and *CTGF* and *IGF-II* which were shown to be higher expressed in the teats of sow with defect were selected according to their biological

function as growth factor related to development of mammary gland as well. Additional, *EGFR* which involed the MAPK singnalling pathway as same as *EGF* were seleted from up-regulated gene in the teat of sow with defect. Finally, the results of the expression analysis led to five selected functional candidate genes (*CTGF*, *IGF-II*, *EGFR*, *EGF* and *GDF8*) for the inverted teat defect in pigs.

4.1.2   Validation using real-time PCR

4.1.2.1 Connective tissue growth factor (*CTGF*)

The validation of *CTGF* expression by real-time PCR could confirm the differential expression of inverted and normal teats in lactating sows but not in young sows (results not shown). The *CTGF* gene was higher expressed in samples from sow with defect than sow without defect ($p<0.01$). The expression level in samples of inverted teats was also higher than in normal teat ($p<0.01$). Using samples of the same sow with defect, it was found that the expression level in samples of inverted teats was similar to normal teats but higher than in normal teats from other sow without defect ($p<0.01$) (Figure 21).
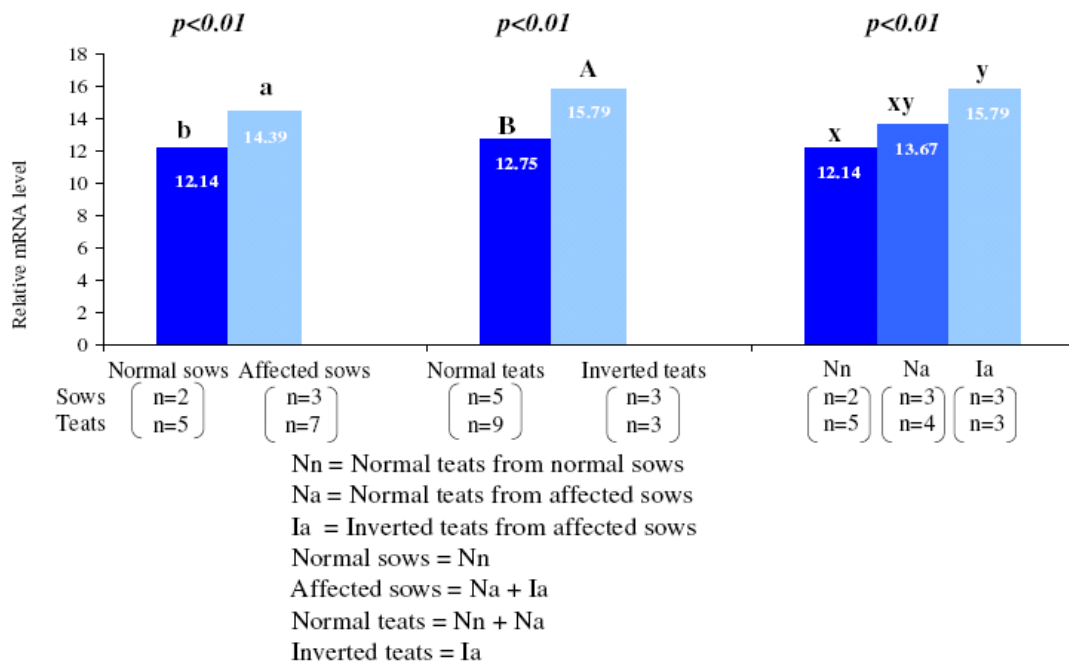
Figure 21: Relative expression levels of the *CTGF* gene determined by real-time PCR using samples of teats from different lactating sows

4.1.2.2 Epidermal growth factor receptor (*EGFR*)

The mRNA of the *EGFR* gene was not differentially expressed in sows with defect compared to sow without defect (Figure 22).

Nn = Normal teats from normal sows
Na = Normal teats from affected sows
Ia = Inverted teats from affected sows
Normal sows = Nn
Affected sows = Na + Ia
Normal teats = Nn + Na
Inverted teats = Ia
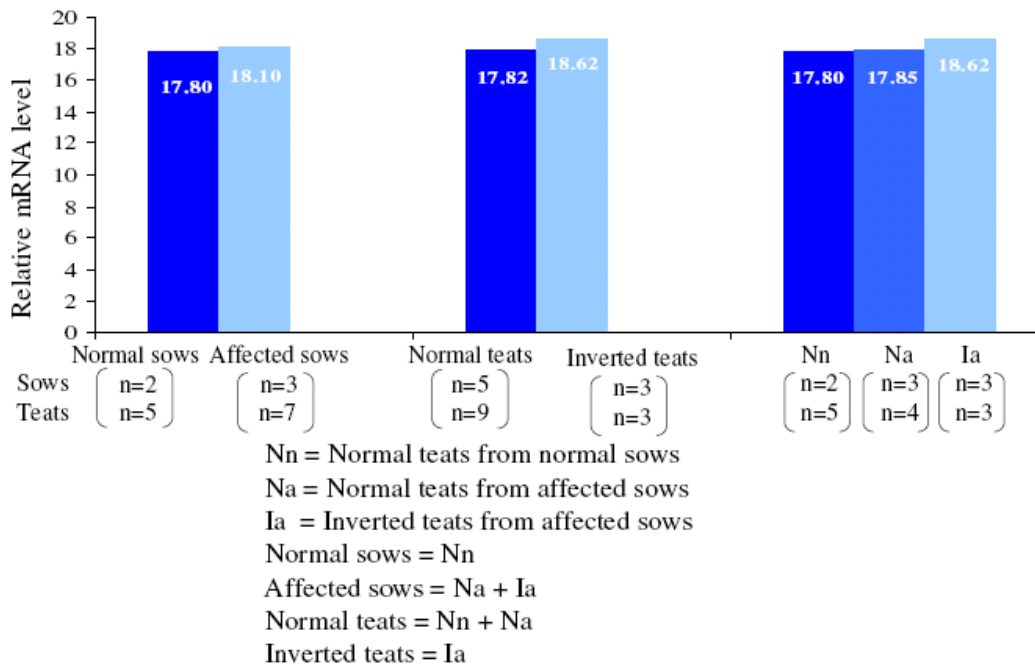
Figure 22: Relative mRNA expression of the *EGFR* gene determined by real-time PCR

4.1.2.3 Insulin-like-growth factor 2 (*IGF-II*)

The *IGF-II* mRNA was higher expressed in tissue samples of sow with defect than sow without defect (*p<0.01*). The expression level in samples of the inverted teats was also higher than in normal teats (*p<0.01*). The expression level found in samples of the inverted teat was similar to the normal teat from the same sow with defect but higher than the expression level in samples of a normal teat from sow without defect (*p<0.01*) (Figure 23A). The differentially expressed of *IGF-II* mRNA was found only in lactating sows (*p<0.05*) (Figure 23B), but there was no difference in the expression level between samples of inverted and normal teats of young sows (results not shown).

(A)

Nn = Normal teats from normal sows
Na = Normal teats from affected sows
Ia = Inverted teats from affected sows
Normal sows = Nn
Affected sows = Na + Ia
Normal teats = Nn + Na
Inverted teats = Ia
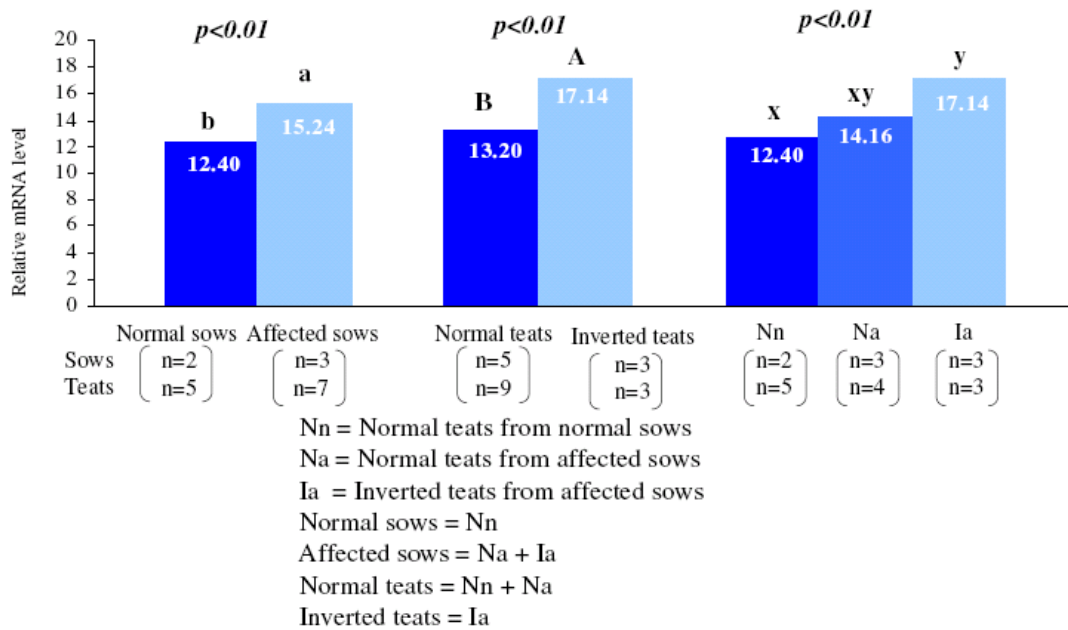
(B)
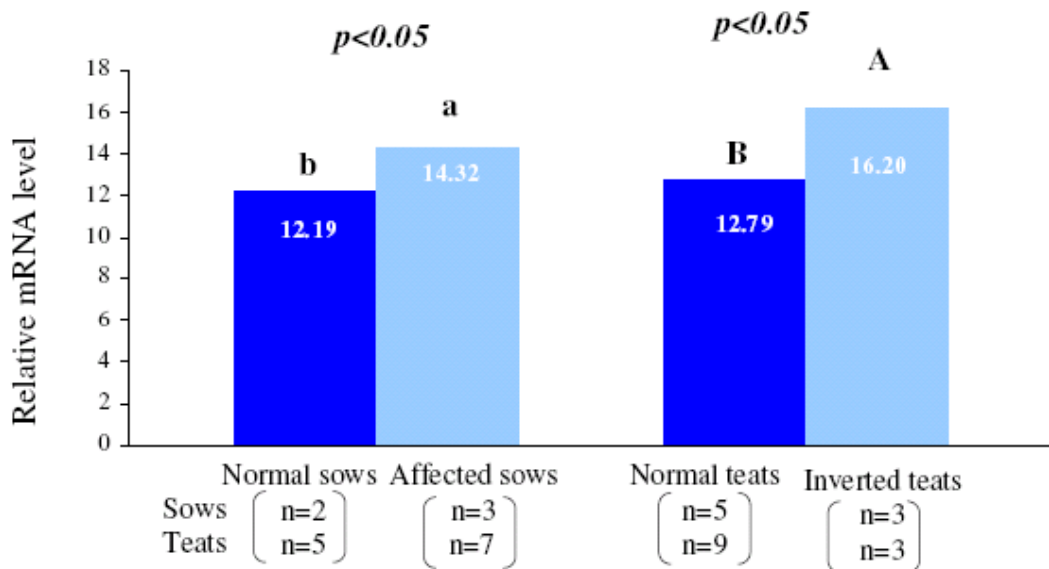
Figure 23: Relative mRNA expression of the *IGF-II* gene determined by real-time PCR comparing samples of lactating sows and young sows (A) and using samples of only lactating sows (B)

4.1.2.4 Epidermal growth factor (*EGF*)

The mRNA of the *EGF* gene was not significantly different expressed in samples from sow with defect compared with samples from sow without defect (Figure 24).

Nn = Normal teats from normal sows
Na = Normal teats from affected sows
Ia = Inverted teats from affected sows
Normal sows = Nn
Affected sows = Na + Ia
Normal teats = Nn + Na
Inverted teats = Ia
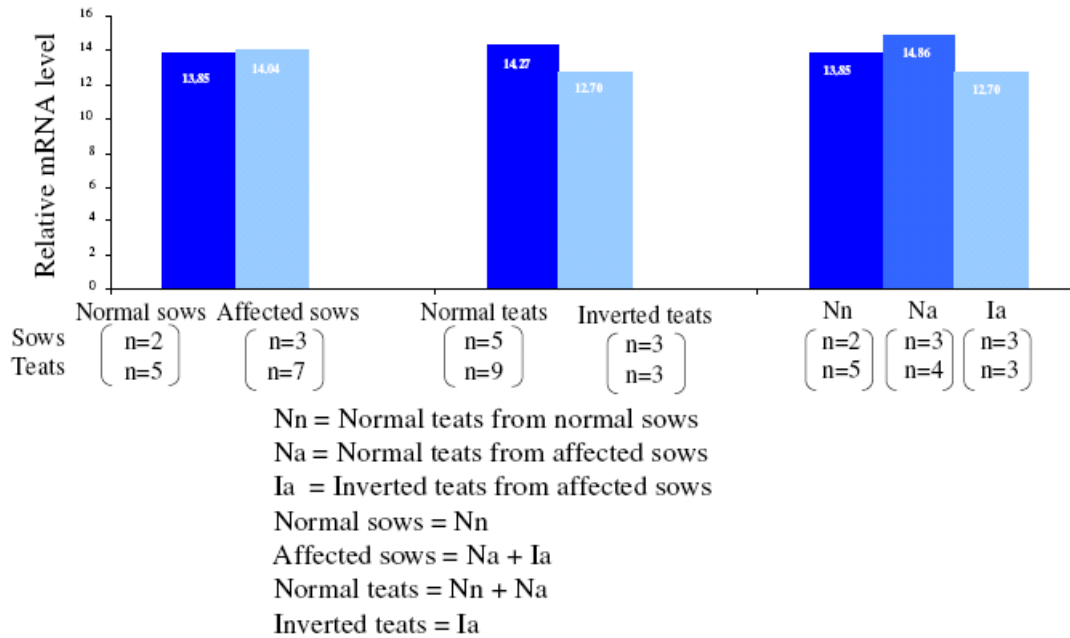
Figure 24:   Relative mRNA expression of the *EGF* gene using different samples analysed by real-time PCR

4.1.2.5 Growth differentiation factor 8 (*GDF8*)

The *GDF8* gene was similarly expressed in samples of the sow with defect compared to the sow without defect, but the expression level of normal teats was higher than inverted teats ($p<0.05$). The expression levels detected in samples of inverted teats was lower than in samples of normal teats either from sow with defect or sow without defect ($p<0.05$) (Figure 25).
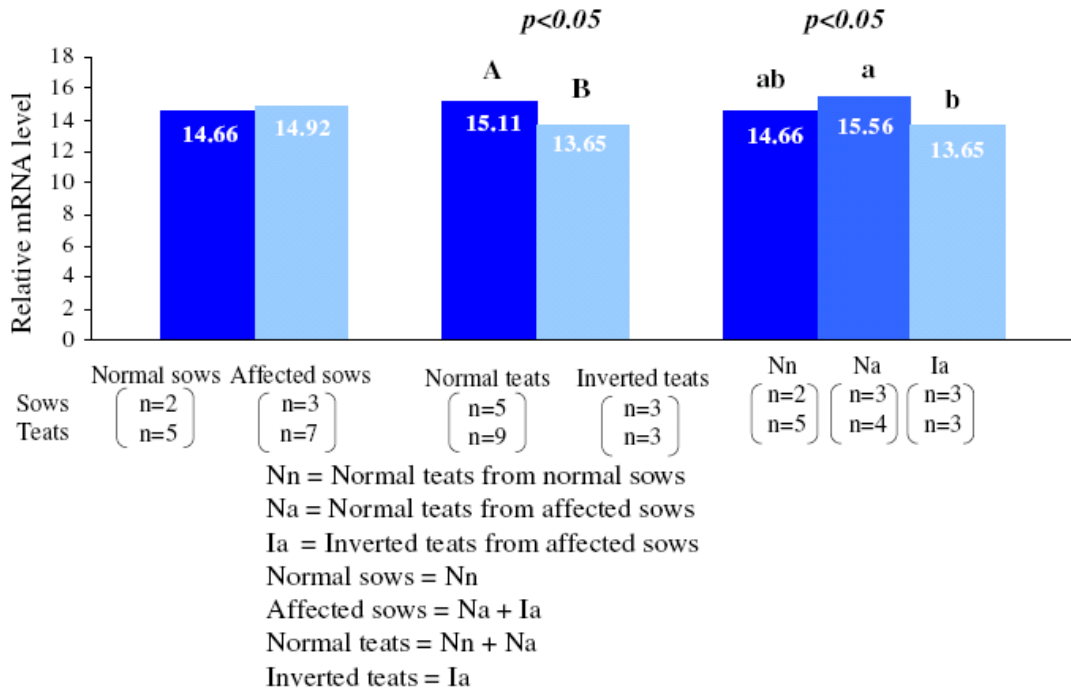
Figure 25: Relative mRNA expression levels of the *GDF8* gene in different samples of lactating sows determined by real-time PCR

4.2 Candidate gene study

4.2.1 Expression analysis by semi-quantitative RT-PCR

The results from the expression profiling of the three functional candidate genes selected from literature revealed that the porcine *RLN3* gene is highly expressed in lung, testis and uterus, moderate in spleen, tonsil, lymph nodes, liver, kidney and skin, lower in muscle, heart brain, teat and inverted teat. For the two investigated receptors of *RLN3*, it was found that *GPCR135* is moderate expressed in muscle, heart, spleen, kidney and uterus while *GPCR142* is highly expressed in spleen, tonsil and lung (Figure 26A and Figure 26B). The expression of the *RLN3* mRNA was observed in all investigated ovarian stages using samples of the uterus of 20 day of estrous cycle sows, sows at day 42 and day 91 of pregnancy as well as two day post partum. The *RLN3* mRNA was highly expressed in the uterus of 20 day of estrous cycle sows, but lower expressed in the uterus of 42 day and 91 day pregnant sows. In uterus samples of sows 2 day post partum, the *RLN3* genes was higher expressed (Figure 26C).

(A)

(B)

(C)
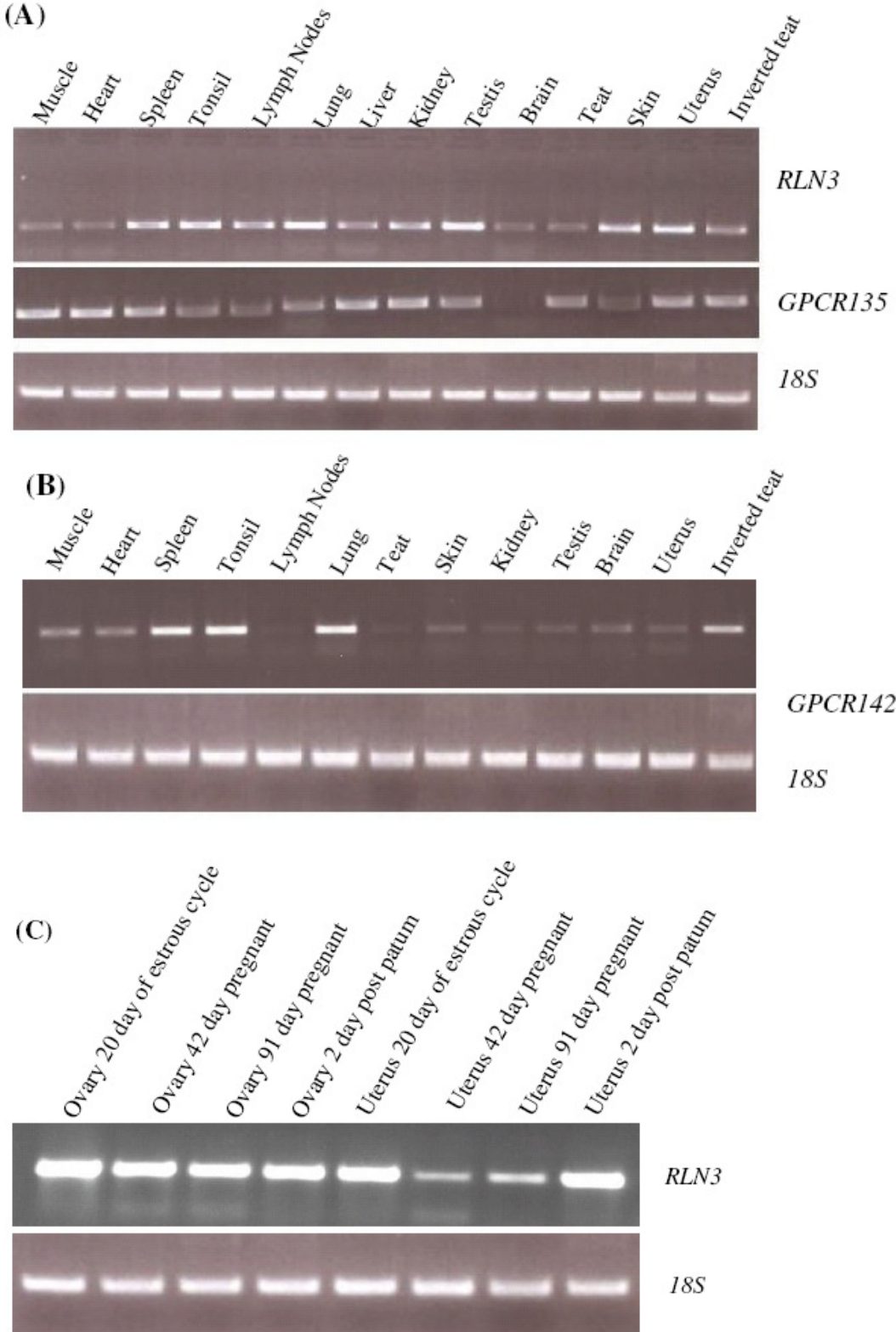
Figure 26:   Results of the semi-quantitative RT-PCR showed the expression levels of
RLN3, GPCR135 (A) and *GPCR142* (B) in different porcine tissues and
the *RLN3* expression in porcine ovary and uterus at different pregnancy
stages (C)

4.2.2   Relative expression using real-time PCR

4.2.2.1 Relaxin 3 (*RLN3*)

The *RLN3* mRNA was higher expressed in teat samples of sows with defect compared to sows without defect (*p<0.01*) (Figure 27), whereas the expression level was similar between inverted and normal teat within the same sow.

(A)
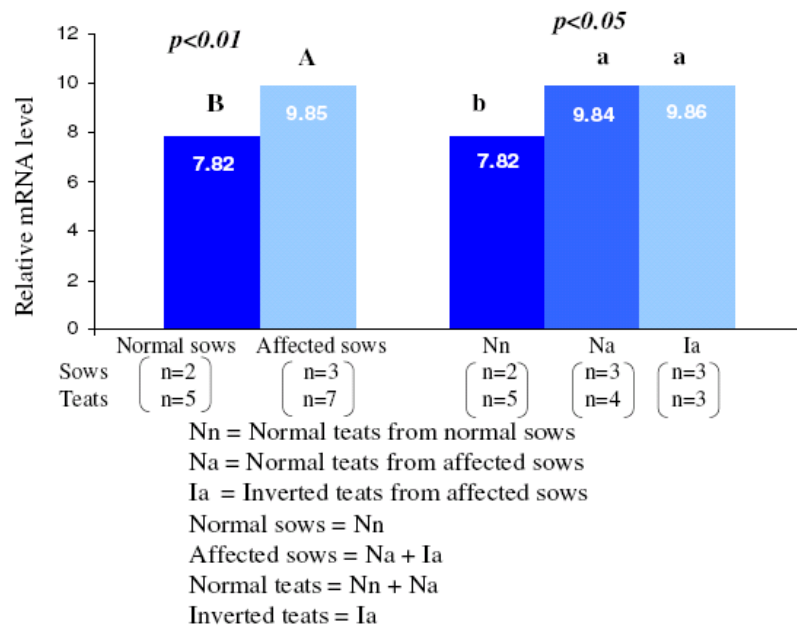
(B)
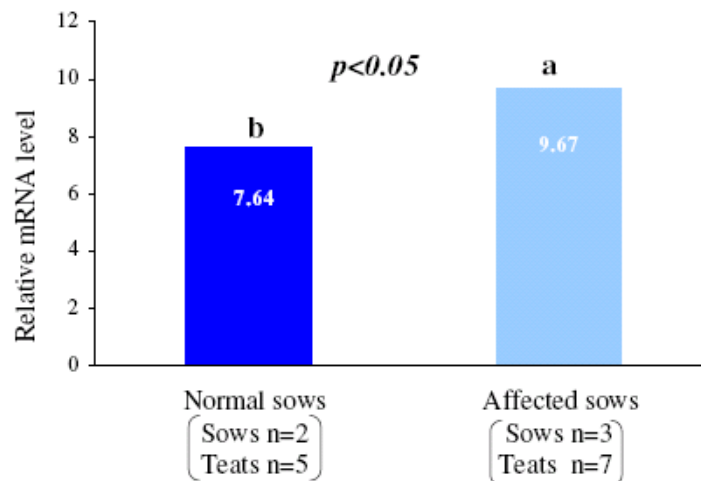
Figure 27: Relative mRNA expression of the *RLN3* gene determined by real-time PCR using teat samples of lactating and young sows (A) and only young sows (B)

It was found that the expression level of teats from the same animal either sows with or without defect were similar whereas the teats with the same phenotype from phenotypic different animals (normal teat from sows without defect and sows with defect) were differentially expressed ($p<0.05$). The *RLN3* mRNA was higher expressed in samples from young sow with defect than young sow without defect but samples from sow with defect and sow without defect was not difference at lactation stage.

4.2.2.2 G-protein coupled receptor 135 (*GPCR135*)

The mRNA of the *GPCR135* gene was higher expressed in teat samples of sow with defect compared to sow without defect ($p<0.05$), whereas the expression level was similar between samples of inverted teats and normal teats within the same sows with defect (Figure 28).
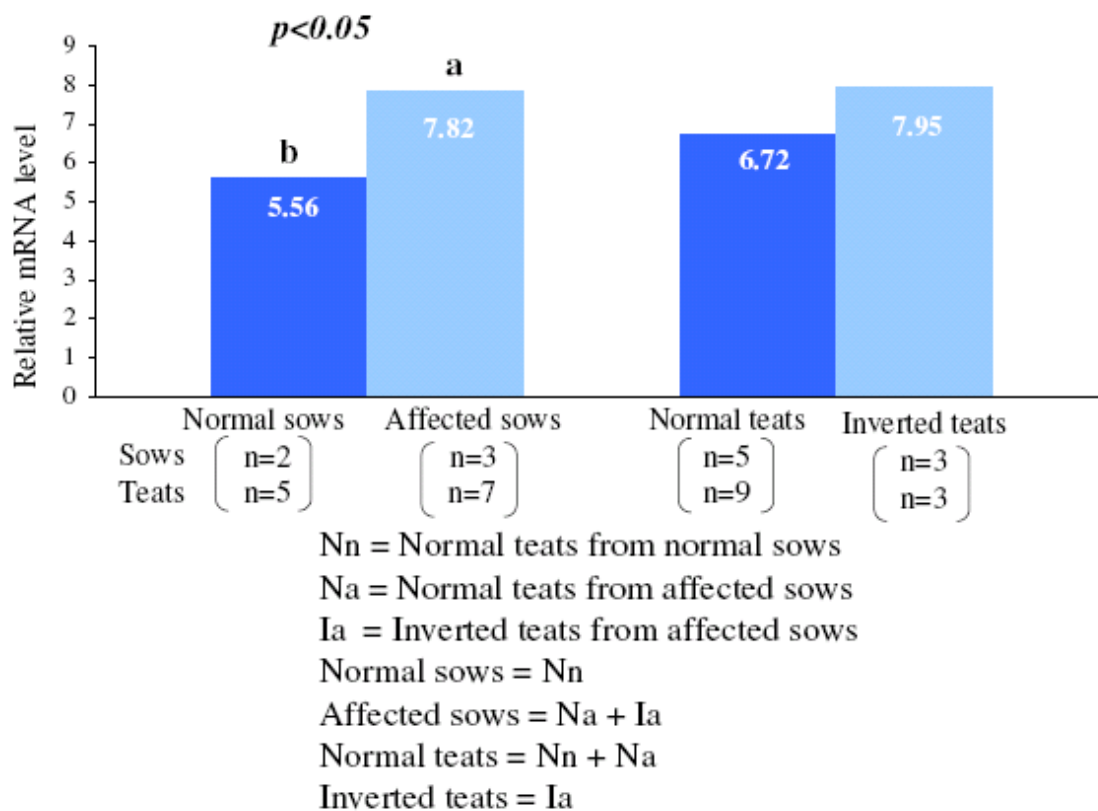
Figure 28:   Relative mRNA expression of *GPCR135* gene determined by quantitative real-time PCR (lactating sows and young sows)

4.2.2.3 G-protein coupled receptor (*GPCR142*)

The mRNA of the *GPCR142* gene of sow with defect was not differentially expressed compared with sow without defect (Figure 29).
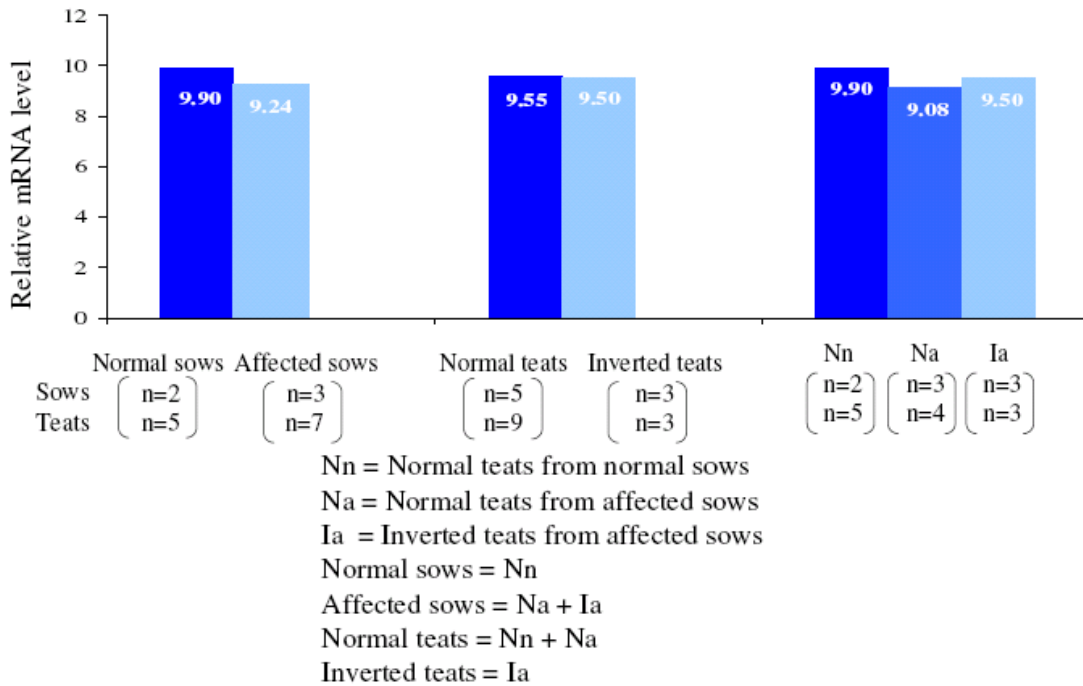
Figure 29: Relative mRNA expression of the *GPCR142* gene determined using real-time PCR

4.2.3   Candidate genes sequencing analysis

4.2.3.1 Relaxin 3 (*RLN3*)

A partial sequence of porcine *RLN3* with 2033 bp length was obtained from the sequencing analysis this identified sequence fragment covering an intron part of the gene. The fragment was deposited with the GenBank accession number DQ974115 (Figure 30).

```
   1 atggccaaac gtccactgct gctgctgctg ctggccgtat gggtgctggc tggggagctg
  61 tggctgagga ctgaggcccg ggcgtcaccc tatggagtga agctttgcgg ccgtgaattc
 121 atccgagcgg tcatctttac ctgcgggggc tcccggtgga gacggtcgga catgctggcc
 181 catgaagctc tgggtgaggc tggggaggta atgtgtgtgg ggaaggggggg tggacgagac
 241 aggacaaggg tccgaggagc caggaaggga ggtgagatga agagcaagag agtcccttcc
 301 ctccacatgc aactggggag acagccaagc cctccaagct gtgccaaggc tggaggagcc
 361 aattagcaga gcagaggagc tgggatgggg tggtgtgtgt gtggggacag caagccattc
 421 accctagcta aaagtcagag aacaggctac tctttctgga cacctgcta tgtgtgagcc
 481 ctgcattagg caaggctgct gtttgcaaaa ccctggtatg tgccgttcag aagcggaaac
 541 taaagtgctt tccaacccct tggggaaggt agtgtgaggg atgtgatgaa tccaaggcta
 601 gatttgccac ctagtaggtg accctcctac caacttgctt tccctccctg atcctcttta
 661 acagcaaata ggagcacagg ccctgtatcc tgcaggtgtt aagtggcaaa agtgaggcaa
 721 gacagagtag catacagttg agtgcttacc atgtgccaga cactgtttta agtgtttttt
 781 tgtttgtttg tttttggggt tttttttgcct tttctagggc cgctcccaca gcatatggaa
 841 gatcccaggc tagggggtcta atcagagctg tggccgctgg cctatgccag agccacagca
 901 acgaaggatc ccagttgtgt ctgcgacctg caccacagct cacagcaatt ccagatcctt
 961 aacgcactga gcatggccag ggatgtgaac cctcaacctc atggtgtcct agtgcagatg
1021 tcgttaatta ctgtgccacg atgggaactc cgttttttttt gttttttgttt ttgttttttgt
1081 ttttgttttt aatgggtatt gagtcttcta ctgggtacga gatcctagaa gataagagtt
1141 attattatcc tcactttgca gacaaggtaa cttaagctca gagaagtaaa ttcacttgct
1201 cagggttatc actcagagag gctggattgg tgcttagtca gttcagctct caaacacatg
1261 ttttttttttt ttttctaggg ttggcttcct gcggcttttta aaggttccca ggctgggggt
1321 ctaatctgag ctacagccac tggcttacac cacagccaca gcaacttggg atccgggcca
1381 catcttcgac ctacaccaca gctcacagca acgctggatc cttaaccctc tgagcaaggc
1441 cagggatcga acccacaacc tcccagttcc tagtcagatt tgctaaccac tgtgccacga
1501 cgggaactcc caaacacatg ttcttaacca tcaggtcctc cacaaataaa tagtagggga
1561 aagaaaatga ttgaagcaga aagccaagct cagtgaggtc gagtaagctg cccatcaaag
1621 ctggaaccca ggtaccagtg gcatcaaaat ctttaactgg gagttcccct tgtggctcag
1681 tgataatgaa cccaactagt atccgtgaag attagggttc aacccctggc cctgttcagt
1741 gggttgggga tccagtgttg ccatgagttg ttgtataggt tgcaaacgca gctcggatct
1801 ggagttgctg tggatgcagc ctaggctggc agctgcagct ccaatttgac tcctagcctg
1861 ggaacttcca tatgccacag gtgctgccct aaaaaaaaaa aaaaaagagt cactgtccta
1921 tagccctgtg ttcccccacc ccccagtcct ccctgaggac cccaaggctg ctaggagcct
1981 ccaggacact gccagggggcc cagctgcacc tgtccttgtc ccacggggcc atgtctattg
2041 gccacaattt ttttttttctttt tagatgtggc tctaaaatga caacaaacaa acaaaaacac
2101 aaaaaaatcat tataactaac aggctgagtc acaagaacag agaagtggct ggaccagcca
2161 ggagccagaa ctctgagtgt ggcctgagct gaatcccaag cactcactct tttcatcttt
2221 ttgcagggga tgtcttctca gacacagatt ccaacgcaga cagcgagttg gacgaggcaa
2281 tggcctccag cgaatggctg gccctgacca agtcccctga gaccttctat ggggttcaac
2341 caggctggca gagaaccccct ggggctctta ggggcagtcg tgatgtcctg gctggcctct
2401 ccagcaactg ctgcaagtgg gggtgcagca agagtgaaat cagcagcctc tgctag
```

Figure 30: Detected nucleotide sequence of the porcine *RLN3* gene published under the accession number DQ974115

The gene structure of porcine *RLN3* comprise 193 bp of first exon, 2033 bp of single intron, and 230 bp of second exon as show in figure 31.

Exon 1 : 193 bp	Intron : 2033 bp	Exon 2 : 230 bp

Figure 31: The gene structure of porcine *RLN3*

The amino acid sequences of the porcine *RLN3* gene showed 78%, 71% and 70% homology compared with human, mouse and rat, respectively. A comparison of the amino acid sequences of the *RLN3* gene among human, pig, mouse and rat showed that all *RLN3* genes and their derived pro-hormone sequences contain a typical signal sequence after the ATG start codon. The porcine *RLN3* comprises a B-chain of 27 amino acids, a C-peptide of 63 amino acids, and an A-chain of 24 amino acids.

The porcine *RLN1* gene has the B-chain motif R*XXX*R*XX*V whereas the B-chain motif of the porcine *RLN3* gene is R*XXX*R*XX*I.

The SNP at the nucleotide position 2338 (A2338G) of the *RLN3* gene was mapped to codon 101 of the C-chain within the second exon. It was found that this SNP is leading to an amino acid change from glutamine (Q) to arginine (R) in the C-peptide region of the protein (Figure 32). The analysis of the prediction for amino acid changes of interest using the SIFT method was performed. SIFT is based one the premise that important amino acids will be conserved among sequences in a protein family, so changes at amino acid conserved in the family should affect protein function. SIFT uses sequence homology to predict whether an amino acid substitution in a protein will affect protein function. SIFT starts with a query protein sequence. Relying on the observation that proteins in the same subfamily have high conservation in conserved regions. The probability (P*ca*) of amino acid *a* appearing at position *c* is estimated (Henikoff and Henikoff 1996). In the most diverse alignment possible, all 20 amino acids might appear in a position with equal probability of 0.05 = 1/ 20, whereas in a conserved position only two amino acids might appear, one with probability 0.05 and the other with 0.95.

If 0.05 were chosen as a cutoff for P*ca* so that substitution to amino acid *a* in position *c* is predicted to be deleterious when P*ca* $\leq$ 0.05, then substitution to any amino acid would be predicted as deleterious in the position (Ng and Henikoff 2001).

In this study, the analysis of the prediction for amino acid changes of interest using the SIFT method revealed that this substitution is predicted to affect the protein function with a score of 0.03.

```
                    10        20        30        40        50        60
                    |         |         |         |         |         |
      A       MAKRPLLLLLLAVWVLAGELWLTEARASPYGVKLCGREFIRAVIFTCGGSRWRRSDMLAH
      G       MAKRPLLLLLLAVWVLAGELWLTEARASPYGVKLCGREFIRAVIFTCGGSRWRRSDMLAH
  Consensus   MAKRPLLLLLLAVWVLAGELWLTEARASPYGVKLCGREFIRAVIFTCGGSRWRRSDMLAH
  Prim.cons.  MAKRPLLLLLLAVWVLAGELWLTEARASPYGVKLCGREFIRAVIFTCGGSRWRRSDMLAH

                    70        80        90        100       110       120
                    |         |         |         |         |         |
      A       EALGDVFSDTDSNADSELDEAMASSEWLALTKSPETFYGVQPGWQRTPGALGSRDVLAGL
      G       EALGDVFSDTDSNADSELDEAMASSEWLALTKSPETFYGVRPGWQRTPGALGSRDVLAGL
  Consensus   EALGDVFSDTDSNADSELDEAMASSEWLALTKSPETFYGVrPGWQRTPGALGSRDVLAGL
  Prim.cons.  EALGDVFSDTDSNADSELDEAMASSEWLALTKSPETFYGV2PGWQRTPGALGSRDVLAGL

                    130
                    |
      A       SSNCCKWGCSKSEISSLC
      G       SSNCCKWGCSKSEISSLC
  Consensus   SSNCCKWGCSKSEISSLC
  Prim.cons.  SSNCCKWGCSKSEISSLC
```

Figure 32: Alignment of the amino acid sequences of the two distinctive nucleotide A and G for the porcine *RLN3* gene

4.2.3.2 G-protein coupled receptor 135 (*GPCR135*)

The sequence of the genomic DNA of the porcine *GPCR135* gene was obtained starting with heterologous primers designed from conserved regions of the published human, mouse and rat gene information and subsequent sequencing using homologous primers. The obtained sequence of 859 bp as shown in figure 33 was published under the accession number EU443643 at the Genebank database. It was further assumed from the comparison with other species that the porcine *GPCR135* genes contains only a single exon.

```
   1 ctgctggttc tctacctgat gaagagtaag cagggctggc gcaagtcctc cataaacctc
  61 ttcgtcacca acctggcgct gacagacttt cagtttgtgc tcaccctgcc cttctgggcg
 121 gtggaaaacg ctctcgactt caaatggccc tttggcaagg ccatgtgtaa gatcgtatcc
 181 gtagtgacgt ccatgaacat gtacgccagc gtgcttcttc ctcacctcca tgagcgtggc
 241 gcgctaccac tcggtggcct cggctcttaa gagtcaccgg acccaagggc acggccgggg
 301 tgactgctgc ggccgaagcc tgggggacag ctgctgcttc tcagccaaag ttctgtgtgt
 361 gttgatctgg gccgcggcca cgctggcctc gctgcccaac gccatcttct ccaccaccat
 421 caaggtgatg ggcgaggagc tgtgcttggt gcgcttccca gacaagttgc tggggcgcga
 481 ccggcagttc tggctgggcc tttaccactt gcagaaggtg ctgctgggct tcctgctgcc
 541 actgggcgtc atcagcctgt gctatctgct gctggtgcgc ttcatctccg accattgcgt
 601 ggctgggacc gaaggaagtg cctcagcagc tgggggaggc ctggccggag ccagcgcccg
 661 aagacgctcc aaggtcacca aatcagtgac catcgtggtc ctatccttct tcctgtgttg
 721 gctgcccaac caggcgctca ccacctggag catcctcatc aagttcaacg tggtgccctt
 781 cagccaagag tacttcttgt gccaggtata cgcattctcg ccgtatgtgc ctggcgcact
 841 ccaacagctg cctcaacc
```

Figure 33: Nucleotide sequence of the porcine *GPCR135* gene as published under the genebank accession number EU443643

4.2.4   Comparative sequencing for polymorphism screening

4.2.4.1 Relaxin 3 (*RLN3*)

Screening for polymorphisms revealed two SNP within the porcine *RLN3* gene (Figure 34), which were located in intron (C1163T) and the second exon (A2338G), respectively.

Intron (C1163T)
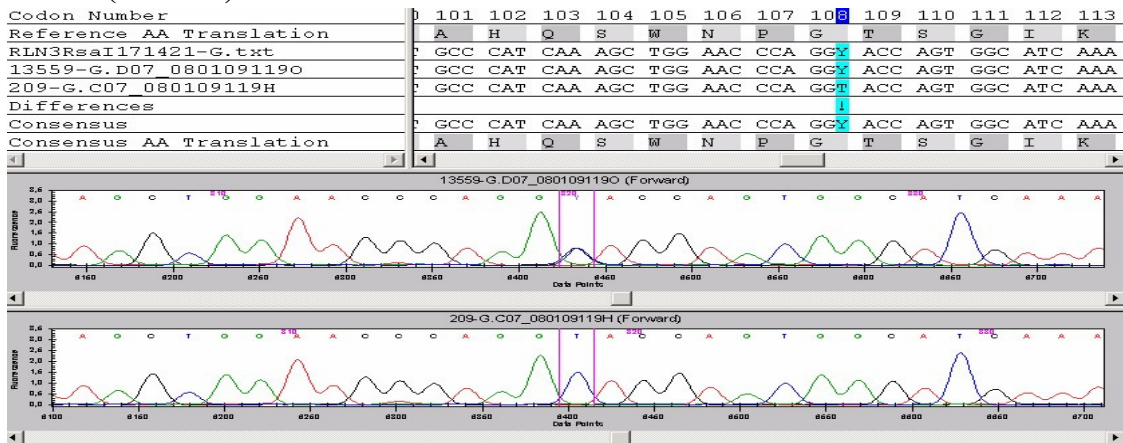
Exon (A2338G)
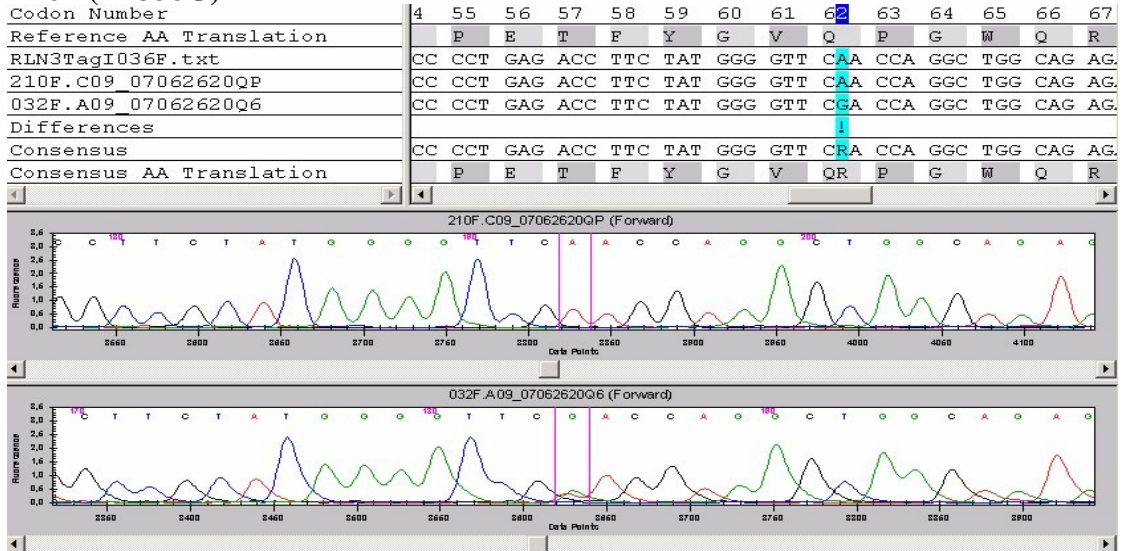
Figure 34: Image showing the SNP within the intron (C1163T) and the second exon
            (A2338G) of the *RLN3* gene using the Investigator program of the CEQ[TM]
            8000

4.2.4.2 G-protein coupled receptor 135 (*GPCR135*)

Two SNP (C to A and C to T) could be detected within the region of the single exon of the porcine *GPCR135* gene (Figure 35).

(base change C to A)
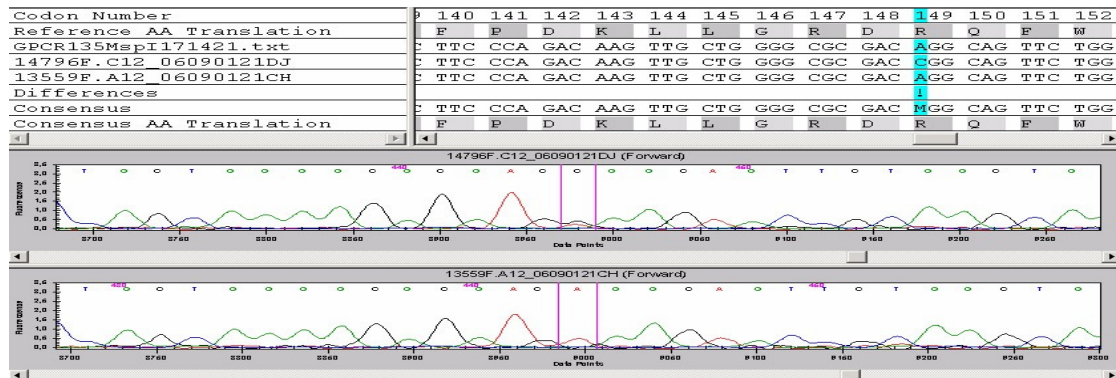
(base change C to T)
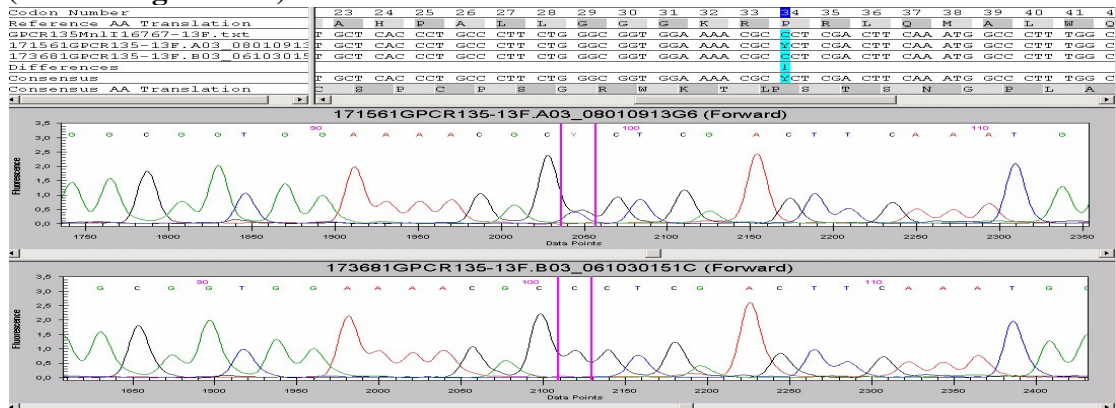
Figure 35: Comparative sequencing revealed two SNPs within the exon of the *GPCR135* gene using the Investigator program of the CEQ^{TM} 8000

4.2.4.3 G-protein coupled receptor 142 (*GPCR142*)

Screening for polymorphism revealed one SNP (G to A) within the single exon of the porcine *GPCR142* gene (Figure 36). It was found that this SNP at nucleotide position 603 (G603A) does not lead to an amino acid change.
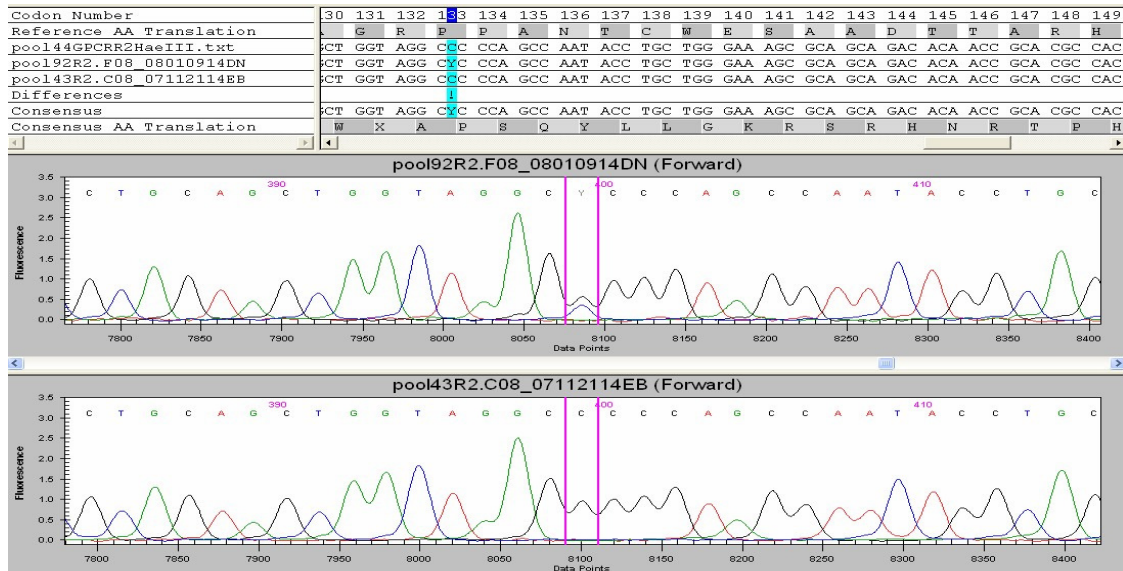
Figure 36:   Image showing the detected SNP within the single exon of the porcine *GPCR142* gene (G603A)

4.2.5   Genotyping and association analysis

4.2.5.1 Relaxin 3 (*RLN3)*

Using the PCR-RFLP method to genotype animals at the two detected SNP loci of the *RLN3* gene, two alternative alleles could be detected at the intron after digestion using the *Rsa*I enzyme (allele C: 434 bp, allele T: 359 and 75 bp). Digesting the amplified 334 bp fragment to genotype animals with the SNP at the second exon, two alleles were generated (allele A: 334 bp, and allele G: 217 and 117 bp) as shown in figure 37.
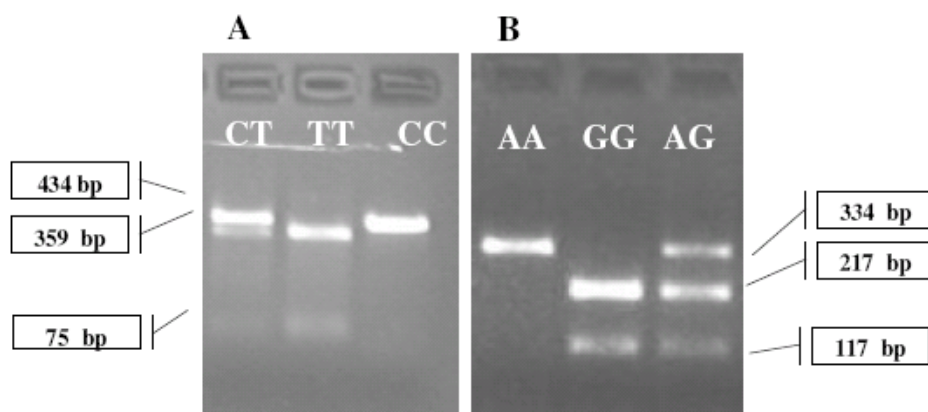
Figure 37: Results of PCR-RFLP for genotyping animals at the locus of the *RLN3* gene
using *Rsa*I (A) and *Taq*I (B) endonuclease. Electrophoresis of the digested
PCR product was performed in a 3% gel containing ethidium bromide and
visualized by a UV transiluminator

In the commercial population, the frequencies of the alleles 'C' and 'T' of the SNP in
the intron region of the *RLN3* gene were 0.15 and 0.85, respectively. The frequencies of
the alleles 'A' and 'G' located in the second exon were 0.74 and 0.26, respectively. The
estimated haplotype frequencies of the haplotypes 'T-A', 'C-G', 'T-G' and 'C-A'
estimated using the FBAT program were 0.79, 0.10, 0.08 and 0.03, respectively (Table
6), whereas the frequencies of the alleles 'C' and 'T' in DUMI population were 0.18
and 0.82, respectively. The frequencies of the alleles 'A' and 'G' located in the second
exon were 0.84 and 0.16, respectively. The estimated haplotype frequencies of the
haplotypes 'T-A', 'C-G', 'T-G' and 'C-A' were 0.73, 0.13, 0.08 and 0.06, respectively
(Table 6).

Table 6:   Allele and   haplotypes frequencies of the two SNP detected within the
candidate gene *RLN3* in the commercial and DUMI population

Population	Number of animal	*Rsa*I		*Taq*I		*RLN3* (Haplotype)			
		C	T	A	G	T-A	C-G	T-G	C-A
Commercial	225	0.15	0.85	0.74	0.26	0.79	0.10	0.08	0.03
DUMI	344	0.18	0.82	0.84	0.16	0.73	0.13	0.08	0.06

Table 7: Result of association analysis of the candidate gene *RLN3* with the inverted teat defect genotyping animals of the commercial population

Trait	Allele	Number of transmitted alleles		Association	
		Observed: S	Expected under null hypothesis: E(S)	Z score	P value
*RLN3/ Rsa*I	C	5.000	2.500	- 0.632	0.527
	T	5.000	2.500	0.632	0.527
*RLN3/Taq*I	A	23.500	6.250	1.400	0.162
	G	23.500	6.250	-1.400	0.162
RLN3 (Haplotype)	T-A	25.417	5.660	2.347	0.019*
	C-G	2.583	1.410	1.193	0.233
	T-G	7.000	3.500	-2.138	0.032*
	C-A	3.000	1.500	-2.449	0.014*

*: significant (p<0.05); Z score < 0: allele more common in not affected animals

Table 8: Result of association analysis of the *RLN3* gene with the inverted teat defect in animals of the DUMI population

Trait	Allele	Number of transmitted alleles		Association	
		Observed: S	Expected under null hypothesis: E(S)	Z score	P value
*RLN3/ Rsa*I	C	30	31	-0.254	0.799
	T	90	89	0.254	0.799
*RLN3/Taq*I	A	105	101.5	0.843	0.399
	G	31	34.5	-0.843	0.399
RLN3 (Haplotype)	T-A	61	54.5	2.030	0.042*
	C-G	5	6.5	-0.832	0.405
	T-G	6	7	-0.535	0.592
	C-A	6	10	-1.789	0.073

* significant (p<0.05); Z score < 0: allele more common in not affected animals

The association analysis with the inverted teats defect showed no significant results in both experimental and the commercial population. Haplotypes constructed using the FBAT program revealed a significant association of *RLN3* gene with the alleles T-A, T-G and C-A in the commercial population and the allele T-A in the DUMI population (Table 7, Table 8).

4.2.5.2 G-protein coupled receptor 135 (*GPCR135)*

An 859 bp fragment of the *GPCR135* locus was amplified and further genotyped. PCR-RFLP leads to the detection of the alternative alleles A (fragments after digestion: 350 bp, 276 bp and 202 bp), and C (fragments after digestion: 276 bp, 202 bp, 186 bp and 164 bp). An 198 bp fragment was amplified including SNP which generated the two possible alleles T (131 and 57 bp bands after cutting), and C (75, 57 and 56 bp bands after cutting) as shown in figure 38.
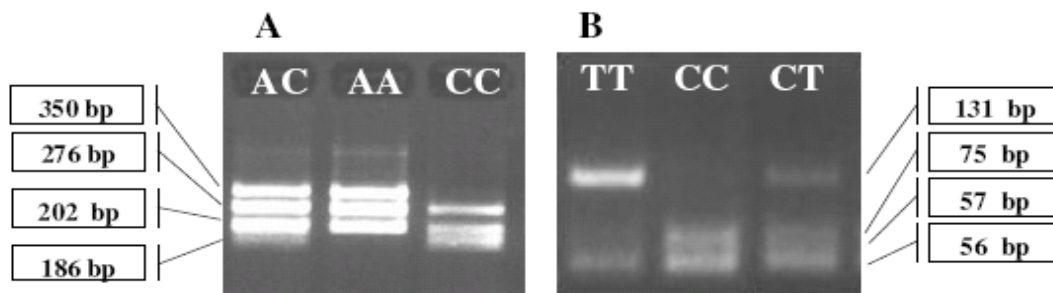
Figure 38: Results of PCR-RFLP for genotyping the animals at the locus of *GPCR135* gene using *Msp*I (A) and *Mnl*I (B) endonuclease. Electrophoresis of the digested PCR product was performed in a 3% agarose gel

In the commercial population, the frequencies of the alleles 'C' and 'A' of the first SNP within the single exon region of the *GPCR135* gene were 0.72 and 0.28 respectively. The frequencies of the alleles 'C' and 'T' of the second SNP were 0.80 and 0.20 respectively. The frequencies of the haplotypes 'A-C', 'A-T', 'C-T' and 'C-C', estimated using the FBAT program, of this locus were 0.61, 0.22, 0.13 and 0.04 respectively (Table 9). In DUMI population, the frequencies of the alleles 'C' and 'A' were 0.76 and 0.24 respectively. The frequencies of the alleles 'C' and 'T' of the second SNP were 0.73 and 0.27 respectively. The frequencies of the haplotypes 'A-C', 'A-T', 'C-T' and 'C-C' of this locus were 0.65, 0.15, 0.15 and 0.05 respectively (Table 9).

Table 9:   Allele and   haplotypes frequencies of the two SNP detected within the candidate gene *GPCR135* in the commercial and DUMI population

Population	Number of animal	*Msp*I		*Mnl*I		*GPCR135* (Haplotype)			
		C	A	C	T	A-C	A-T	C-T	C-C
Commercial	217	0.72	0.28	0.80	0.20	0.61	0.22	0.13	0.04
DUMI	258	0.76	0.24	0.73	0.27	0.65	0.15	0.15	0.05

Haplotypes analysis revealed no significant association between the *GPCR135* gene and the inverted teat trait. The association analysis between each SNP and haplotype of the *GPCR135* gene with different characteristic of the teats showed also no significant association (Table 10).

Table 10: Result of association analysis of the candidate gene *GPCR135* with the inverted teat defect in the commercial population

Trait	Allele	Number of transmitted alleles		Association	
		Observed: S	Expected under null hypothesis: E(S)	Z score	P value
GPCR135/	C	20.000	20.500	-0.229	0.819 [n.s.]
*Msp*I	A	12.000	11.500	0.229	0.819 [n.s.]
GPCR135/	C	44.000	43.500	0.156	0.876 [n.s.]
*Mnl*I	T	20.000	20.500	-0.156	0.876 [n.s.].
GPCR135 (Haplotype)	A-C	40.943	40.943	-0.017	0.986 [n.s.].
	A-T	15.973	15.973	-1.230	0.218 [n.s.]
	C-T	10.905	9.110	0.851	0.394 [n.s.].
	C-C	5.095	3.973	1.136	0.256 [n.s.].

n.s.: non significant (p>0.05); Z score < 0: allele more common in not affected animals

The association analysis using animals of the DUMI population showed also no significant association (Table 11).

Table 11: Result of association analysis of the *GPCR135* gene with the inverted teat defect in the DUMI population

Trait	Allele	Number of transmitted alleles		Association	
		Observed: S	Expected under null hypothesis: E(S)	Z score	P value
GPCR135/	C	370	366.529	0.496	0.62 [n.s.]
*Msp*I	A	126	129.471	-0.496	0.62 [n.s.]
GPCR135/	C	410	395.662	1.756	0.079 [n.s.]
*Mnl*I	T	132	146.338	-1.756	0.079 [n.s.]
GPCR135 (Haplotype)	A-C	238.753	237.55	0.094	0.925 [n.s.]
	A-T	50.247	53.349	-0.408	0.683 [n.s.]
	C-T	38.753	33.371	0.915	0.360 [n.s.]
	C-C	38.247	41.730	-0.492	0.622 [n.s.]

n.s.: non significant; Z score < 0: allele more common in not affected animals

4.2.5.3 G-protein coupled receptor 142 (*GPCR142)*

Genotyping of animals at the *GPCR142* locus was performed using PCR-RFLP. Digestion of the resulting 1002 bp PCR product with *Hae*III revealed one polymorphism. The polymorphism was found segregating in the population with allele A with 13 cutting sites for the enzyme showing an extra 153 bp fragment, whereas allele G yielded one additional cutting side leading to two additional fragments (126 and 27 bp). Monomorphic fragments of 222, 170, 109 and 96 bp were also seen for both alternative alleles. Additional smaller fragments representing the remaining 51 to 8 bp were likely generated by *Hae*III digestion but were not detectable in agarose gels (Figure 39).
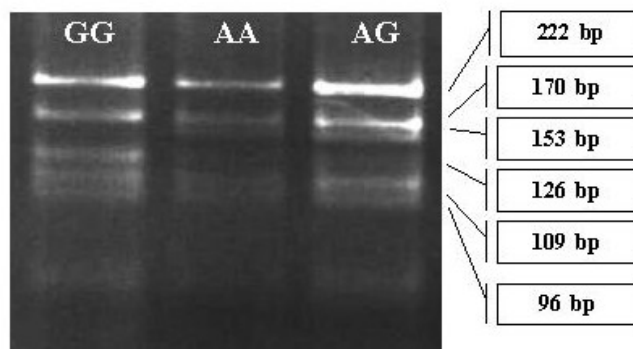
Figure 39: Results of PCR-RFLP for genotyping animals at the locus of the porcine
            GPCR142 gene using *Hae*III

The frequencies of the alleles 'G' and 'A' for SNP in single exon region of the
GPCR142 gene were 0.96 and 0.04 respectively in commercial population, while the
frequencies of the alleles 'G' and 'A' were 0.91 and 0.09 respectively (Table 12).

Table 12: Allele and haplotypes frequency of SNP detected within the candidate gene
           GPCR142 in the commercial and DUMI population

Population	Number of	Allele	
	animal	G	A
Commercial	189	0.96	0.04
DUMI	262	0.91	0.09

The association analysis between the SNP of the *GPCR142* gene with the different
characteristic of the teats showed no significant association in the commercial animals
(Table 13).

Table 13: Result of association analysis of the candidate gene *GPCR142* with the
        inverted teat defect in the commercial population

Trait	Allele	Number of transmitted alleles		Association	
		Observed: S	Expected under null hypothesis: E(S)	Z score	P value
GPCR142/	G	4	4.5	-0.577	0.563 [n.s.]
*Hae*III	A	2	1.5	0.577	0.563 [n.s.]

n.s.: non significant; Z score < 0: allele more common in not affected animals

The association analysis between the genotypes this loci with the inverted teat defect
could neither reveal the importance of this gene for the inverted teat defect using
animals of the DUMI population (Table 14).

Table 14: Result of association analysis of the *GPCR142* gene with the inverted teat
        defect in the DUMI population

Trait	Allele	Number of transmitted alleles		Association	
		Observed: S	Expected under null hypothesis: E(S)	Z score	P value
GPCR142/	G	45	41	1.083	0.278 [n.s.]
*Hae*III	A	21	25	-1.083	0.278 [n.s.]

n.s.: non significant; Z score < 0: allele more common in not affected animals

4.2.5.4 Association analysis for the three genes with the teat number

The haplotype analysis showed a significance of the allele T-G with the teat number on
the left, the right side and the total number of teats in animals of the DUMI population
(Table 15).

Table 15: Summary of the results of the association analysis for the three genes with the teat number in the DUMI population

Trait	Allele	TL		ITL		TR		ITR		TT		TIT	
		Z score	P value	Z score	P value	Z score	P value	Z score	P value	Z score	P value	Z score	P value
RLN3/ RsaI	C	-1.104	0.269	-0.214	0.830	-1.213	0.225	-0.372	0.709	-1.161	0.245	-0.297	0.766
	T	1.104	0.269	0.214	0.830	1.213	0.225	0.372	0.709	1.161	0.245	0.297	0.766
RLN3/TaqI	A	0.245	0.806	0.125	0.900	0.253	0.799	-0.251	0.801	0.250	0.802	-0.066	0.947
	G	-0.245	0.806	-0.125	0.900	-0.253	0.799	0.251	0.801	-0.250	0.802	0.066	0.947
RLN3 (Haplotype) (n=334)	T-A	1.248	0.212	0.172	0.863	1.193	0.232	-0.180	0.856	1.222	0.221	-0.006	0.995
	C-G	0.019	0.984	-0.086	0.931	-0.15	0.885	-0.097	0.922	-0.064	0.948	-0.093	0.926
	T-G	-2.513	0.011*	-0.346	0.729	-2.450	0.014*	-0.491	0.623	-2.486	0.012*	-0.421	0.674
	C-A	-0.291	0.771	0.128	0.898	-0.023	0.981	0.961	0.336	-0.156	0.875	0.553	0.580
GPCR135/ *Mnl*I	C	0.545	0.586	0.687	0.492	0.766	0.443	0.611	0.541	1.756	0.079	0.817	0.413
	T	-0.545	0.586	-0.687	0.492	-0.766	0.443	-0.611	0.541	-1.756	0.079	-0.817	0.413
GPCR135 (Haplotype) (n=258)	A-C	-0.208	0.835	-0.140	0.888	0.163	0.870	-0.258	0.796	0.094	0.925	-0.185	0.853
	A-T	-1.124	0.261	-1.381	0.167	-0.467	0.640	-1.264	0.206	-0.408	0.683	-1.337	0.181
	C-T	1.420	0.155	1.740	0.081	0.982	0.325	1.571	0.116	0.915	0.360	1.650	0.098
	C-C	0.469	0.638	0.284	0.776	-0.660	0.509	0.615	0.538	-0.492	0.622	0.450	0.652
GPCR142/ *Hae*III (n=262)	G	1.318	0.187	1.322	0.186	1.053	0.292	1.090	0.275	1.203	0.228	1.233	0.217
	A	-1.318	0.187	-1.322	0.186	-1.053	0.292	-1.090	0.275	-1.203	0.228	-1.233	0.217

TL: left teat number;   TR: right teat number; TT: total inverted teats; ITL: inverted teats left; ITR: right inverted teat number;

TIT: total inverted teats; * significant (p<0.05)

4.2.6   Linkage and physical mapping

The multipoint linkage map revealed a good evidence of *RLN3* being mapped on SSC2 in relative position of approximately 23 cM (SW2443 -*RLN3* - SW240). The most significantly linked marker (twopoint analysis) using RH mapping was SSC2F05 on SSC2 (30 cR; LOD score 13.91). The multipoint linkage map revealed good evidence *GPCR135* being mapped on SSC16 (S0111 - *GPCR135* - S0026), while the multipoint linkage map revealed good evidence of *GPCR142* being mapped on SSC4 (S0214 - *GPCR142* - S0001).

5   Discussion

5.1   Microarray study

5.1.1   Identification of candidate genes

The regulation mechanism of the mammary gland development is complex. To further understand the molecular mechanisms of the mammary gland development, it might be helpful to obtain a high coverage collection of genes relevant to different physiological stages and construct the physiological stage specific gene expression profile. The lactation performance of sows is correlated with their reproductive performance. Previous studies have shown that the production and composition of milk is critical for the survival of the suckling piglets. Further the milk production is one of the most important factors limiting neonatal pig growth (Su et al. 2006).

In the present report the gene expression profile of the teat from lactating sows was investigated. As a result genes differentially expressed in the mammary gland could be identified, specific gene clusters were identified. These genes may be important for the reproductive performance or play important roles for milk synthesis, secretion and mammary involution. The characterization of the global gene expression profiles may help to elucidate important biological processes in both normal and inverted teat. A large fraction of normal mammary epithelium specific growth factors (*IGF-II*, *VEGF-A*), chemokines (*CXCL12*), cytokines (*IL10, IL-18*, and *CTGF*), and cytokines receptor (*EGFR*) may play a role during the regulation of normal mammary epithelial cell growth, differentiation, and morphogenesis. In the present study, the expression of growth factors (*EGF*, *MFGE8*, *GDF8*) and signal transducer activity genes (*GPRC5B*, *TGFBR3*, *PIGR* and *ccl28*) were significantly decreased in samples of inverted teats. On the basis of these observations and the aforementioned physiologic roles of growth factor in apoptosis induction and potential growth controlling functions in mammary epithelial cells, it can be speculated that the observed significant down regulation could be associated to an escape of inverted structures from the constraints of such growth regulatory mechanisms (Hu et al. 2004).

It was found in human that several classes of genes are down or up regulated in healthy mammary epithelial cells compared with cancer cells. It was considered in that study that the down regulation of a set of genes may be the basic mechanism of cancer

formation, whereas the upregulation may characterize and possibly control the state of evolution of individual inverted teat (Zucchi et al. 2004). This result was consistent with other results suggesting that the mammary gland of gilts was prepared for lactating and its gene expression was active during the late gestation (Su et al. 2006). Genes of ribosome structural constituent were abundantly expressed in this stage and also some proteins involved in the regulation of transcription and translation were abundantly expressed e.g. *GDF8* (Su et al. 2006).

Similar to the gene expression in mouse mammary gland during lactation, the proportion of cell cycle genes had diminished. It was replaced by an increase of the number of genes involved in the development and differentiation, including fatty acid biosynthesis and other metabolic processes. These three clusters should contain genes that are important for terminal differentiation and transition to a secretory phenotype (Clarkson et al. 2004).

The immune related transcripts were expressed in the porcine lactation stage in the present study. This is consistent with previous observations of the immune cell complement in regressing mouse and murine mammary gland (Clarkson et al. 2004, Monks et al. 2002) which showed that a strong statistical relationship exists between involution and immune related genes. The gene expression during involution was found being significant associated with inflammation, the acute phase response, or humoral immunity and innate cellular defence (Clarkson et al. 2004). The pituitary gland plays a key role in the regulation of growth, differentiation and function of all cells in the body, including the immunocytes. Immune reactions are generated through the proliferation of antigen specific lymphocyte clones. In recent years many efforts have been undertaken to elucidate the complex interactions between mediators of the endocrine system and the immune system. The main effector of *GH* is *IGF-I*, an endocrine mediator of growth and development under physiological conditions, *IGF-I* also plays a prominent role in the regulation of immunity and inflammation (Heemskerk et al. 1999). The clinical consequences of the link between the endocrine and the immune system remain to be elucidated (Hansen et al. 2001). The possible impact of the immune related genes on the inverted teat defect can therefore be considered deriving from two different hypotheses, either the impact of reproductive hormones onto the immune system or the relevance of immune related genes during the apoptosis of the mammary gland, which might be of relevance for the inverted teat defect. Further it was found that the inverted teat might lead to a higher risk of mastitis, a further recommendation of the finding of differential

expression of the immune-related gene clusters between normal and defect teats in the present study (Varadin and Filipovic 1975).

5.1.2   Validation using real-time PCR

5.1.2.1   Connective tissue growth factor (*CTGF*)

The *CTGF* is a member of the *FSH*/androgen regulated gene repertoire expressed in the mammalian granulosa cells. This gene is implicated in the regulation of the connective tissue synthesis (Slee et al. 2001). The *CTGF* gene is typically up regulated by serum enrichment or exposure to tissue growth factors such as *TGFB*, platelet-derived growth factor (*PDGF*) or *FGF*. A similar pattern of the expression was observed in pig ovary, where *CTGF* has been hypothesized to promote ovarian cell growth and blood vessel formation during follicular and luteal development (Slee et al. 2001). Based on its expression profile in the ovary and biological properties in other tissue systems, it was suggested that *CTGF* might contribute to the process of thecal cell recruitment, a crucial process in folliculogenesis (Slee et al. 2001).

The *CTGF* plays a role in *TGFB*-mediated formation of granulation tissue. Brigstock (1999) suggested that this gene is involved in the regulation of uterine function as it is present in uterine fluids in pigs and mice, as well as in the uterus of pigs, mouse and human. *CTGF* is mitogenic for connective tissue cells, it is secreted by fibroblasts and endothelial cells, and is selectively induced by fibroblasts after activation with *TGFB*. The high expression of *CTGF* gene in tissue samples of the inverted teat in the present study may support the hypothesis that *CTGF* might be involved in proliferation of mammary gland, further the gene expression profile of the connective tissue may play a important role for the development of the inverted teat. Due to the presence of *TGFB1* in mammary carcinomas, *CTGF* might be involved in stromal proliferation of mammary cancer (Frazier and Grotendorst 1997). This factor has a potential role during embryogenesis and for the uterine function as the *CTGF* is produced by the embryo and in the uterus. The levels of *CTGF* gene expression have been studied in mammary gland tumours, sarcoma cells, chondrosarcoma cells, and various tumours of the nervous and vascular systems, but the data could not strongly support the possible functional relationship of *CTGF* in tumour development to the incidence of desmoplasia in mammary gland carcinomas (Uzumcu et al. 2000).

5.1.2.2   Epidermal growth factor receptor (*EGFR*)

Growth factors such as the *EGF* have also been implicated in the normal mammary gland growth and development (Coleman et al. 1988). *EGF* has been shown to stimulate the proliferation of mammary tissues (Sternlicht et al. 2005, Taketani and Oka 1983, Yang et al. 1980). The *EGF* may act synergistically with estrogen (E) and/or progesterone (P) (Sheffield and Welsch 1987, Sheffield 1998). The normal mouse mammary gland possesses membrane receptors for *EGF* and the concentration of receptors varies at different developmental states. Receptor levels, present in both epithelial and stromal cells, are high in the immature pubertal mammary gland and then decrease with increasing age. Receptor levels increase again with the onset of pregnancy, reaching a peak level at 10 days followed by a rapid decline to very low levels during lactation (Sandra et al. 1992).

EGFR levels increase during the pregnancy at a time when circulating levels of E and P are known to be elevated (Edery et al. 1985). Interestingly, the existing level of *EGFR* in the mature gland was not decreased by ovariectomy, indicating that constitutive levels of *EGFR* are increased only when E and P is added. It was found that progesterone increases *EGFR* levels in human mammary carcinoma cell lines that are estrogen and progesterone receptor positive. Thus it is possible that the progesterone-dependent increase of *EGFR* may be a mammary specific regulatory mechanism (Sandra et al. 1992).

Sandra et al. (1992) showed that the cellular distribution of *EGFR* demonstrated that the receptors are present in epithelial and stromal cells in both pubertal and mature mammary glands. These results agree with previous localization of *EGFR* in stromal and epithelial cells of 5-week-old mammary gland (Sandra et al. 1992). The similarly expression of the *EGFR* gene in tissue samples of normal and inverted teat during the lactation in the present study could not verify the results of the microarray, where this genes was found being differentially expressed between these two tissue types. Anyhow further studies including teat samples of fetus where key stages of the ductal morphogenesis are expected might explain the role of *EGFR* during the ductal morphogenesis by stimulating proliferation in the end bud regarding the possible impact for the inverted teat defect.

Stromal *EGFR* was found to be located in close proximity to the end bud and ductal epithelium where it was present at levels approximately five fold higher than at more distal sites. The stromal *EGFR* are functional when locally-released *EGF* induced both

epithelial and stromal DNA synthesis in the mammary glands of ovariectomized peripubertal and mature female mice (Hovey et al. 1999).

The *EGFR* gene is present in the stromal fibroblasts that separate the ducts from the fatty stroma. Impaired ductal growth in *EGFR* knockout mammary glands was associated with a marked reduction in the density of periductal fibroblasts. Thus, signaling through the *EGFR* may promote fibroblast survival, which in turn induces ductal epithelial cell proliferation. In contrast to the lack of ductal development, alveolar development occurred in grafts of both *EGFR* knockout and wild-type mammary glands in response to prolactin produced by the pituitary gland. Although the alveoli were morphologically normal in grafts of *EGFR* knockout mammary glands, these structures did not penetrate entirely throughout the fat pad of the *EGFR* knockout mammary glands compared to the wild-type glands, presumably due to the underlying defect in ductal development. This defect was not intrinsic to the epithelium, because *EGFR* knockout epithelium transplanted into wild-type cleared pads showed normal ductal development and equivalent alveolar development in response to a pituitary graft. Therefore, signaling through the *EGFR* is dispensable for alveolar development, just as it is indispensable for ductal development. The *EGFR* is not essential in the epithelial component of the mammary gland *in vivo.* Instead, the *EGFR* is absolutely necessary for the stromal component, the fat pad, to induce estrogen-dependent ductal growth and branching morphogenesis as shown by the tissue recombination studies. These results suggest that, under estrogenic conditions, which stimulate the pubertal mammary gland, the stroma responds to estrogen action through an *EGFR*-mediated signaling event that is required for stimulation of epithelial growth and development. In contrast, the epithelial *EGFR* is neither necessary nor sufficient (Wiesen et al. 1999). The *EGFR* axis is an essential mammary signaling system in which *ADAM17* (A disintegrin and metalloproteinase protein) releases epithelial *AREG*, which then activates stromal *EGFR* (Figure 40), thus eliciting reciprocal responses that further orchestrate mammary epithelial development (Sternlicht et al. 2005).
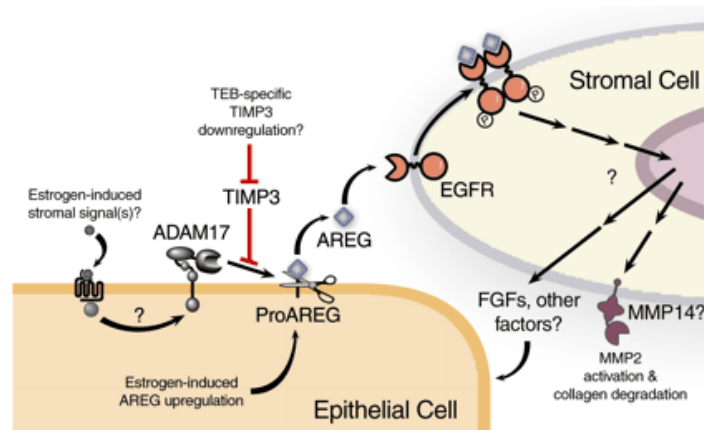
Figure 40: Model depicting epithelial-mesenchymal crosstalk and potential modifiers of *ADAM17-AREG-EGFR* signaling in mammary development (Sternlicht et al. 2005)

In the present study the different expression of the *EGFR* gene was only detected using the genome wide expression study, the expression difference between normal and inverted teats could not be verified using the real-time PCR. Possibly the total amount of this transcript was too low in the tissue samples used for the detection of the transcript abundance as is was found that the *EGFR* transcript abundance in mouse mammary gland decline during the late pregnancy and lactation (Sheffield 1998). Further Sheffield (1998) also found that this gene is lower expressed in mammary tissue of lactating cattle compared to the tissue of nonlactating cattle. This might be one possible explaination that in the present study no different expression could be detected using the more sensitive real-time PCR method.

5.1.2.3 Insulin-like-growth factor 2 (*IGF-II*)

The insulin-like growth factors, *IGF-I* and *IGF-II*, are additional peptide growth factors with proposed roles in postnatal mammary growth. These peptides and their primary signaling receptor, the *IGF* type I receptor (*IGF-IR*), are essential for normal embryogenesis and have mitogenic, survival and differentiative actions on a variety of developing tissues and cell types. The distinct patterns of expression for *IGF-I* and *IGF-II* suggest that the two *IGF* ligands are distinctly regulated. While this observation may simply represent a mechanism to finely regulate or restrict autocrine/paracrine *IGF* actions at specific stages of growth, it is also possible that the two ligands have distinct functions in the developing gland. In addition to *IGFBP* specificity, *IGF-II* also can

mediate fetal growth and cell proliferation through the insulin receptor as well as through the *IGF-IR*. Actions of *IGF-II* in growth of the mammary epithelium are likely subtle since mice carrying a deletion of *IGF-II* were found to be viable and fertile. Female *IGF-II* mutant mice support litters and thus develop sufficient alveoli for lactation, however, careful analyses will be required to determine whether glands from these mice have subtle growth deficits (Wood et al. 2000). In the present study using samples of pigs, it was found that the *IGF-II* gene was higher expressed in samples of inverted teats compared to normal teats, which was congruent using the microarrays and the real-time study. This might promote the impact of this gene for the inverted teat defect due to a sufficient alveoli development.

The *IGFs* would promote proliferation of the mammary epithelium. This mechanism of *IGF* action is consistent with proposed actions of other growth factors through extracellular matrix binding. The extracellular matrix is known to sequester a number of growth factors including *TGF-b* and the *FGFs*. Finally, since the *IGFs* and *IGF-IR* are well-characterized survival factors for many cell types, it is possible that *IGF*-mediated epithelial growth is due predominantly to promotion of cell survival. Maintenance of mammary epithelial and stromal cells is essential for normal development. Moreover, it is known that apoptotic cell death occurs at specific locations and times during postnatal mammary growth including in the terminal end buds (TEBs) during formation of ductal structures and throughout the alveoli during involution (Loladze et al. 2006, Streuli and Haslam 1998). Thus, it is possible that *IGFBP* regulation of *IGF* availability and localization is critical for promoting or inhibiting *IGF*-mediated cell survival (Wood et al. 2000).

5.1.2.4   Epidermal growth factor (*EGF*)

The *EGF* is a polypeptide which, acting systemically or locally, may effect responses in the mammary gland through the fat pad. *EGF* may stimulate epithelial synthesis of type IV collagen, a component of the basal lamina required for epithelial attachment and proliferation and required for growth of cultured mammary gland. *EGF* causes the reappearance and growth of involuted mammary end buds in ovariectomized mice. Administering *EGF* to mice may reelevate lowered mammary and intestinal rates of epithelial mitosis, expedite mammary development, and reelevate mammary tumor risk. The similar expression of *EGF* in normal and inverted teat indicated that the inverted teat defect is not afftected by *EGF* during lactation because *EGF* required for early

ductal outgrowth. These results are in accordance with report in mice which show that *EGF* specifically binds and stimulates the proliferation of primary mammary epithelial cells derived from virgin or pregnant mice, it is reasonable to postulate that ligand activation of the *EGFR* is an integral component of the normal growth regulation of these cells (DiAugustine et al. 1997).

EGF is able to inhibit the apoptosis of apoptotic mouse mammary epithelial cells. During the mammary gland development, the growth and development of the ducts are directed by the terminal end buds. *EGF* appears to function as a survival factor, *EGF* receptors are important for the survival of epithelial cells in the mammary gland (Rosfjord and Dickson 1999).

An analysis of *in vivo* effects of *EGF* in 5-week-old ovariectomized mice has shown that *EGF* promotes the normal ductal morphogenesis by stimulating proliferation in the end buds (Sandra et al. 1992).

EGF may affect responses in the mammary gland through the fat pad. In addition to its presence in epithelial cells, receptors for *EGF* were detected within the mouse mammary fat pad by a ligand binding assay. *EGF* signals through the stroma and controls early ductal outgrowth. Together with E, it also controls ductal elongation and branching during puberty. Progesterone, placental lactogens, prolactin, and the osteoclast differentiation factor signal alveolar proliferation and differentiation during pregnancy and possibly lactation. The signals inducing tissue remodeling during involution have not been defined. The genes that control distinct stages of mammary development are shown in figure 41. Ductal elongation and branching during puberty is controlled by inhibin B, colony-stimulating factor 1 (*CSF-1*), progesterone receptor, and *Wnt*. Proliferation and differentiation of mammary alveolar cells is controlled by the *PRLR*, signal transducer and activator of signalling-5 (*STAT5*), receptor activator for nuclear factor κ B ligand *(RANKL)*, *CCND1*, pelargonium flower break virus (*p27*), *Id2*, and CCAAT/enhancer binding protein (*C/EBP*). Mammary function during lactation is controlled by *PRL* through *STAT5*, and tissue remodeling and cell death during involution by *STAT3* (Hennighausen and Robinson 2001).
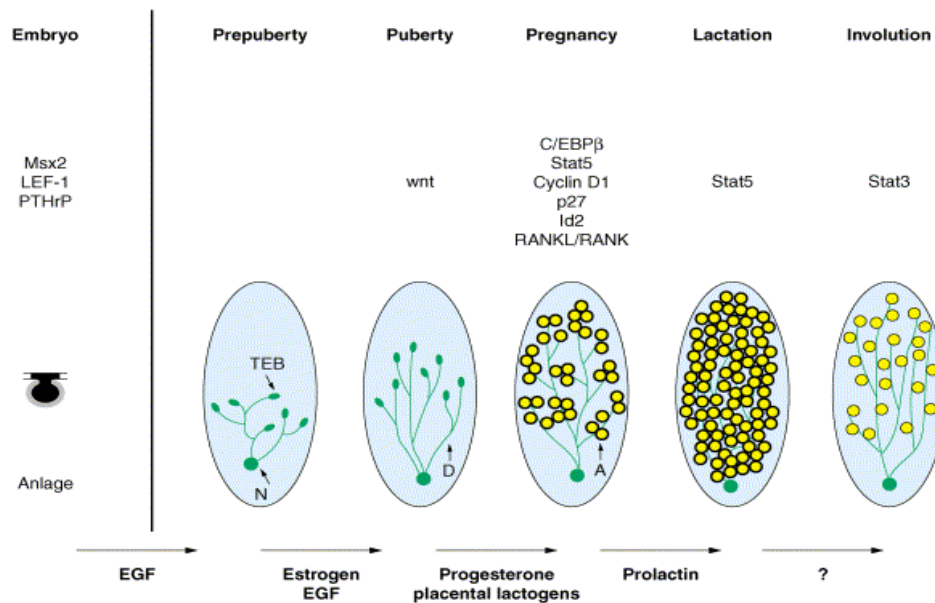
Figure 41:   The schematic drawing represents different developmental stages. The anlage is composed of epithelium (dark knob) and stroma (gray surrounding). The oval shown in the postnatal stages depicts the mammary fat pad (stroma). The solid green circle represents the nipple (N) from which the ducts originate. The ends of growing ducts form TEB during puberty (Hennighausen and Robinson 2001)

5.1.2.5   Growth differentiation factor 8 (*GDF8*)

*GDF*8 is predominantly expressed in the skeletal muscle and may be a key regulator of the development and growth of skeletal muscle (Ji et al. 1998, Wehling et al. 2000). In muscle, the physiological role of *GDF*8 is largely associated with the prenatal period of muscle growth in which myoblasts are proliferating, differentiating, and fusing to form multinucleated myofibers. The *GDF8* performs a regulatory role pertaining to gestational or lactational mammary gland growth and development and/or metabolism. It seems possible that *GDF*8 is secreted from the mammary gland into the milk and serves a regulatory role in neonatal pigs (Ji et al. 1998). The lower expression of *GDF8* in tissue samples of the inverted teat in this study indicated that the inverted teat may be caused by an insufficient of cell proliferation and differentiation due to the lower expression of *GDF8*. It is well established that milk production is positively related to milking frequency (Bar-Pelled et al. 1995, Erdman and Varner 1995). The increased milking frequency in dairy cattle is associated with alteration of mammary cell- ECM interactions and signalling that support milk synthesis (Connor et al. 2008). It is demonstrated that milk production of the inverted teats is lower than the normal teats.

Because teat canal of inverted teats is held inward and forming a small crater therefore the milk can not flow out. As a consequence, the teat stimulation of inverted teat is lower than normal teat. It might be also possible that the lower expression level of this gene is related to the decreased amount of milk which can be found in the inverted teats shortly after the birth as less transcript can also be secreted from the mammary glands into the milk which was shown by Ji et al. (1998) to play a regulatory role in neonatal pigs.

5.2   Candidate gene study

5.2.1   Expression analysis by semi-quantitative RT-PCR

5.2.1.1 Relaxin 3 (*RLN3*)

The expression profile of porcine *RLN3* in this study was similar to the *RLN3* expression published in mouse (Bathgate et al. 2002). The murine *RLN3* mRNA is present in several tissues including the brain, thymus, spleen, lung, testis, ovary, and mammary gland. It was also shown being weakly expressed in the heart, liver, epidermis, prostate, and uterus. The findings that *RLN3* gene is expressed in various tissues, confirming that relaxin is more than a hormone of pregnancy and acts on cells and tissues other than those of the female reproductive system (Bathgate et al. 2002). Relaxin causes a widening of blood vessels (vasodilatation) in the kidney, mesocecum, lung, and peripheral vasculature, which leads to increased blood flow or perfusion rates in these tissues. It also stimulates an increase in heart rate and coronary blood flow and increases both glomerular filtration rate and renal plasma flow. In addition to airway fibrosis, relaxin deficiency results in airway structural changes (epithelial thickening) and increased lung recoil, suggesting that relaxin may impact other aspects of airway/lung structure and function beyond its ability to regulate collagen turnover (Samuel et al. 2007). The brain is another target tissue for relaxin, where the peptide has been shown to bind to receptors in the circumventricular organs to affect blood pressure and drinking (Bathgate et al. 2002, Sherwood 2004).

The functional role of *RLN3* also includes the ability to inhibit myometrial contractions, to stimulate remodeling of the connective tissue, and to induce softening of the tissues of the birth canal. Additionally, relaxin increases growth and differentiation of the

mammary gland and nipple and induces the breakdown of collagen, one of the main components of the connective tissue (Sherwood 2004). It was found in female guinea pig that the mammary gland is a minor source of relaxin production (Bathgate et al. 2002). Within the female reproductive tract, relaxin is primarily produced by the corpus luteum in pregnant and nonpregnant mammalian species (Knox et al. 1994). Significant expression of mouse *RLN3* mRNA was observed in different ovarian stages which supports the detection of *RLN3* mRNA in the ovary of nonpregnant, pregnant and lactating sows in the present study in pigs (Bathgate et al. 2002). These results are in accordance with the findings that the *RLN3* gene is expressed in the luteal tissue during pregnancy, ovarian cycle and early lactation as revealed using immunohistochemistry (Bagnell et al. 1990).

The porcine *RLN3* was in the present study higher expressed in the uterus of non pregnant sows compared to the lower expression during pregnancy and again higher expression during lactation. Similar expression profiles of the relaxin gene were also reported during the early pregnancy in pigs and guinea pigs in the uterine endometrium (Knox et al. 1994, Larkin and Renegar 1986). The relaxin in the uteri of nonpregnant gilts is of luteal origin (Knox et al. 1994). Its level was minimal during the early pregnancy and increased by day 16. It seems likely that the presence of conceptus tissue during the early pregnancy may enhance the production of relaxin by the luminal epithelium in the uterus in pigs. The corpora lutea of pigs and rats are also the major source of circulating relaxin during late pregnancy (Crish et al. 1986). In pregnant pigs, relaxin can be detected in the corpus luteum where the amount of relaxin increases dramatically from days 17-106 of gestation (Anderson et al. 1973). This was in accordance with the findings in the present study that the *RLN3* gene expression is increasing from day 42 of the pregnancy until the lactation using the semi quantitative RT PCR.

5.2.1.2   G-protein coupled receptor 135 and 142 (*GPCR*135 and *GPCR142*)

The human *RLN3* gene has recently been identified as a ligand for two structurally related GPCRs *GPCR135* and *GPCR142* (Kuei et al. 2007, Liu et al. 2003a, Liu et al. 2003b, Liu et al. 2005). It could be shown in the present study that the porcine *GPCR142* mRNA expression pattern is very different from that of *GPCR135*. The porcine *GPCR135* gene is high expressed in reproductive tissue while the *GPCR142* gene was found in a broader range of peripheral tissues. These results are in accordance

with the previous reports describing the inconsistent mRNA expression patterns between *GPCR142* and *GPCR135* (Liu et al. 2003a, Matsumoto et al. 2000). The *GPCR135* gene is expressed in restricted tissues with the predominant expression in the brain (Matsumoto et al. 2000), whereas the *GPCR142* gene is expressed in a broader range of peripheral tissues in rats (Liu et al. 2003a). These results of Liu et al. (2003a) suggest that *GPCR135* and *GPCR142* may share a ligand but seem to have different signal pathways. The two receptors exert different physiological functions but probably orchestrated by the common ligand *RLN3*. *GPCR135* may be involved in the regulation of feeding, energy expenditure, metabolism, and other related central functions supplementing the reproductive needs of the body. The relative abundant expression of *GPCR142* suggests the possible existence of an additional ligand expressed in the peripheral tissue (Liu et al. 2003a).

5.2.2   Relative expression analysis using real-time PCR

5.2.2.1   Relaxin 3 (*RLN3*)

The *RLN3* gene was higher expressed in the tissue samples of the mammary gland of inverted teats from sows with defect compared to the tissue of normal glands from sows with defect. There is only limited evidence that *RLN3* gene may act locally on reproductive tissue in pigs but mouse *RLN3* gene is expressed in the ovary. Both *RLN1* and *RLN3* were also detected in the mammary gland, ovaries of non pregnant, pregnant, and lactating mice, and the endometrium and myometrium of pregnant mice.

5.2.2.2   G-protein coupled receptor 135 and 142 (*GPCR135* and *GPCR142*)

The mRNA of *GPCR135* gene in the teat samples of sows with inverted teats was higher expressed compared to the tissue of sows without defect. This is supported by the regulation of cell growth and differentiation by G-proteins. The *GPCR142* gene expression was similar between tissues of inverted and normal teats.

The difference effect of gene expression between *GPCR135* and *GPCR142* can be explained in terms of signalling pathways activated by the *GPCR*s (Figure 42). *GPCR*s interact with heterotrimeric G proteins composed of α, β and γ subunits. The α subunits of G proteins are divided into four subfamilies ($G\alpha_s$, $G\alpha_i$, $G\alpha_q$ and $G\alpha_{12}$). *GPCR142*, is linked to $G\alpha_i$ subunits, inhibits adenylyl cyclase and lowers cAMP levels whereas

GPCR135, is linked to Gα_q subunits, bind to and activate phospholipase C (PLC), and can also control the activity of key intracellular signal-transducing molecules, including small GTP-binding proteins of the Ras and Rho families and members of the mitogen-activated protein kinase (MAPK) family and mediated the cell proliferation by the transactivation of *EGFR* (Halls et al. 2007, Liu et al. 2003b, Liu et al. 2003a). Because *RLN3* is primarily expressed in the brain, the relative abundant expression of *GPCR142* in the peripheral tissues also suggests the possible existence of an additional ligand expressed in peripheral tissues (Liu et al. 2003a).
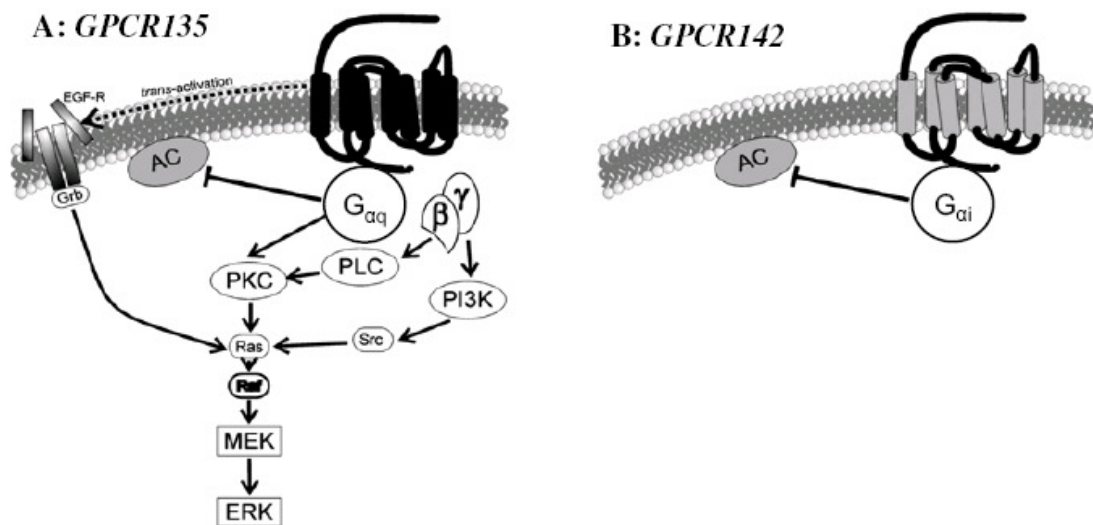
Figure 42: Signalling pathways activated by the *GPCR*s. (A) *GPCR135*, is linked to Gα_q subunits, bind to and activate phospholipase C (PLC) (B) *GPCR142* is linked to Gα_i subunits (adapted from the review of Halls et al. 2007))

5.2.3 Candidate gene sequencing analysis

5.2.3.1 Relaxin 3 (*RLN3*)

In recent years many new members of the relaxin/insulin/*IGF* superfamily have been identified. The derived amino acid sequences of these genes reflect the characteristics of members of the relaxin/insulin family of peptide hormones. The sequences show the highest amino acid homology to *INSL5*, but the presence of an amino acid motif in the B-chain (R*XXX*R*XX*I), which is essential for relaxin receptor binding, indicates that they are more relaxin than insulin like. Furthermore, the structure of the A-chain of *RLN3* gene is more typical to other members of the relaxin peptide family, wherein the terminal A-chain cysteine residue is separated from the intra-disulfide bond-linked

cysteine by eight amino acids. The human and mouse *RLN3* share 93% homology within the A- and B-domains. The homology of the C-peptide chain between relaxins is also generally quite low. However, human *RLN3* and mouse *RLN3* show very high homology in this region and are also much shorter than other relaxins. Interestingly, the human *RLN3* B-chain sequence shares 59% sequence homology to a newly discovered frog relaxin sequence, although it only shares 38% homology in the A-chain and 14% homology in the C-chain. Considering the high homology between the human *RLN3*, mouse *RLN3*, and rat *RLN3* in the A-chain sequence, it is unlikely that this frog sequence is an human *RLN3* homologue (Bathgate et al. 2002). The porcine *RLN3* gene amino acid sequence (NP_001039144) includes 140 AA showing 78%, 71% and 70% homology compared with human (NP_543140), mouse (NP_775276), and rat (NP_733767), respectively. A comparison of the amino acid sequences of *RLN3* gene among human, pig, mouse, and rat showed that all genes and derived pro-hormone sequences contain a typical signal sequence after the ATG start codon (Figure 43).

Signal peptide

```
Human     -MARYMLLLLLAVWVLTGELWPGAEA
Pig       MAKRPLLLLLAVWVLAGELWLRTEA
Mouse     --MAMLGLLLLASWALLGALGLQAEA
Rat       --MATRGLLL-ASWALLGALVLQAEA
```

B chain

```
Human     RAAPYGVRLCGREFIRAVIFTCGGSRW
Pig       RASPYGVKLCGREFIRAVIFTCGGSRW
Mouse     RPAPYGVKLCGREFIRAVIFTCGGSRW
Rat       RPAPYGVKLCGREFIRAVIFTCGGSRW
```

C chain

```
Human     RRSDILAHEAMGDTFPDADADEDSLAGELDEAMGSSEWLALTKSPQAFYRGRPSWQGTPGVLRGSR
Pig       RRSDMLAHEALGDVFSDTDSNADS---ELDEAMASSEWLALTKSPETFYGVQPGWQRTPGALRGSR
Mouse     RRADILAHESLGDFFADGEANTDHLASELDEAVGSSEWLALTKSPQAFYGGRASWQGSPGVVRGSR
Rat       RRADILAHDPLGEFFADGEANTDHLASELDEAVGSSEWLALTKSPQVFYGGRSSWQGSPGVVRGSR
```

A chain

```
Human     DVLAGLSSSCCKWGCSKSEISSLC
Pig       DVLAGLSSNCCKWGCSKSEISSLC
Mouse     DVLAGLSSSCCEWGCSKSQISSLC
Rat       DVLAGLSSSCCEWGCSKSQISSLC
```

Figure 43: Multi-alignment of the full length amino acid sequences of the three chains of the porcine *RLN3* gene to human, mouse and rat *RLN3* gene including the consensus sequences. The position of the detected SNP in the porcine *RLN3* gene is underlined. Positions where the same amino acids were detected are highlighted

The arginine-arginine pair of basic amino acids (RR) at the beginning of the C-peptide (B/C junction) found with all species cleavage between tryptophan and arginine (WR) at the end of the B-chain. Similarly, cleavage at the C/A junction is most likely to occur between the arginine (R) and aspartic acid (D) (Bathgate et al. 2002). The porcine *RLN3* comprise a B-chain of 27 amino acids, a C-peptide of 63 amino acids, and an A-chain of 24 amino acids. The homology between the *RLN1* and the *RLN3* gene is only 38% and 28% in the A- and B-chains, respectively (Figure 44).

```
B chain

RLN1            QSTNDFIKACGRELVRLWVEICGSVSWGRTAL
RLN3            RASPYGVKLCGREFIRAVIFTCGGSRW

A chain

RLN1                RMTLSEKCCQVGCIRKDIARLC
RLN3                DVLAGLSSNCCKWGCSKSEISSLC
```

Figure 44: The alignment of partial amino acid sequences of the A and B chain of the porcine *RLN1* and the porcine *RLN3* gene. All cystein residues are highlighted in bold letters. Underlined amino acids show the amino acids that do not differ between *RLN1* and *RLN3*

The cysteine (C) residues are retained in the correct positions. They are together with conserved glycine (G) residues necessary for flexibility around the cysteine linkages (Bathgate et al. 2002). Interestingly, there is the difference of the motif in the *RLN* receptor binding in the core of the B-chain (R*XXX*R*XX*I/V) between porcine *RLN1* and *RLN3* gene. The porcine *RLN1* gene has the motif B-chain as (R*XXX*R*XX*V) whereas the motif B-chain of porcine *RLN3* gene is (R*XXX*R*XX*I).

The phylogeny of the *RLN* peptides was investigated showing the relationships between the genes. Amino acid sequences are from GenBank (accession no. in parenthesis): human *RLN2* (BC126419), human *RLN1* (NM_006911), porcine *RLN1* (NM_213872), mouse *RLN1* (NM_011272), porcine *RLN3* (NM_001045679), human *RLN3* (NM_080864), rat *RLN3* (NM_170667), mouse *RLN3* (NM_173184). The phylogenetic analysis of the known *RLN* peptides of pigs and of human are shown in the figure 45.
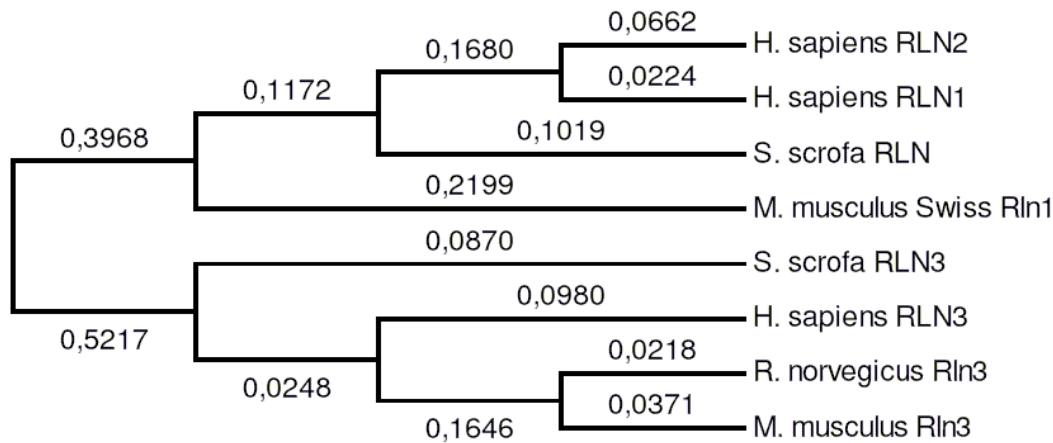
Figure 45: Unrooted NJ tree showing phylogenetic relationships among the *RLN* family. In total eight predicted amino acid sequences were used for constructing the tree. The numbers of the nodes are the bootstrap scores (of 1000 replicates)

The result indicates a closer relationship of the *RLN1* to *RLN2* than to *RLN3*, similar to previously published results (Wilkinson et al. 2005). The phylogenetic analysis of the known insulin-relaxin superfamily shows homologies of both the *RLN3* and *INSL5* genes (results not shown). This might implement that the two *RLN3*-like genes existed prior to the genome duplication event proposed to have occurred in the teleost ancestor. The human *RLN1* and *RLN2*, appear to be derived from recent gene duplication in mammals. Therefore, *RLN*-family peptides could be important for the evolution and adaptation to lineage-specific physiological processes during evolution (Hsu et al. 2005).

5.2.3.2   G-protein coupled receptor 135 (*GPCR135*)

GPCR135 is highly conserved among different species. The porcine *GPCR135* gene sequence showed 90%, 85% and 85% homology with the human, rat and mouse, respectively. *GPCR135* shares many features common to *GPCR* proteins (Chen et al. 2005), including seven hydrophobic TM domains, potential *N*-linked glycosylation sites in the amino terminal extracel extracellular loop which may form a disulfide bridge. Also included are asparagine in TM I, aspartate in TM II, tryptophans in TM IV and TM VI, prolines in TMs IV–VII and the conserved NPXXY motif in TM VII. However, it does possess unique characteristics threonine replaces the aspartate or glutamate usually found in the (D/E)RY motif in the amino-terminus of the second cytoplasmic loop.

5.2.4 Association analysis

5.2.4.1 Relaxin3 (*RLN3*)

The association analysis revealed some evidence of the possible role of the porcine *RLN3* gene during the teat development. In previous studies the related *RLN1* gene was described with effects on the inverted teat defect in pigs (Chomdej 2005). Studies using *RLN1* knock-out mice, showed furthermore that these animals could not deliver milk to their offspring (Kuenzi et al. 1995, Zhao et al. 1999). The present study should further investigate the relevance of the porcine *RLN3* gene for the development of the mammary gland in pigs. Even though the *RLN3* gene was described being expressed in mammary gland in mice, there is no clear evidence about its role during the mammary gland development (Bathgate et al. 2002). In the present study significant association was only found for the estimated haplotypes with the inverted teat defect in pigs. Endogenous relaxin is known to have dramatic effects on the development of the mammary apparatus in pregnant pigs, rats, and mice (Kass et al. 2001, Kuenzi et al. 1995, Zaleski et al. 1996, Zhao et al. 1999). It has profound effects on the development of the mammary gland in pigs, the effects of the hormone on the development of the mammary nipples are most dramatic and vital in rodents (Sherwood 2004). In primiparous pigs, mammary lobulo-alveolar development begins around day 80 of pregnancy, and it continues until term, a period that coincides with rising levels of estrogen, relaxin but also relaxin-dependent growth and softening of the cervix. Endogenous relaxin plays a major role in promoting the development of the mammary gland parenchyma. A subsequent study in ovariectomized nonpregnant gilts provided evidence that the effects of relaxin on mammary parenchymal development are estrogen dependent in pigs, and that they are accompanied by a reduction in the organization of collagen fiber bundles in the stroma (Winn et al. 1994). Endogenous relaxin was found to have no apparent effect on the weight of the mammary glands during the second half of the pregnancy in rats and mice (Hwang et al. 1991, Winn et al. 1994, Zhao et al. 2000). Nevertheless, relaxin influences the mammary gland differentiation. In rats, relaxin reduces the density and organization of collagen fiber bundles, reduces the length of elastin fibers, and increases the cross-sectional area of arterioles. Moreover, in both rats and mice, relaxin increases alveolar development. Relaxin likely mediates the mammary development through direct effects on the mammary glands. Relaxin-binding sites have been reported in epithelial cells associated with lactiferous ducts and lobulo-

alveolar structures in rats, pigs, and humans (Kohsaka et al. 1998, Kuenzi and Sherwood 1995, Min and Sherwood 1996). The endogenous relaxin is required for the development of the mammary nipples that occurs during the second half of pregnancy in rats. The nipples in relaxin-deficient rats are so small at term that the pups cannot grasp them to obtain milk. Analysis of rat and mouse nipple histology demonstrated that relaxin promotes a reduction in the density of collagen fiber bundles that is similar to that which occurs in the cervix, vagina, and mammary glands (Hwang et al. 1991, Kuenzi and Sherwood 1992).

The results of QTL analysis revealed significant QTL for the inverted teat defect on SSC2 (Jonas et al. 2008, Oltmanns 2003). These results led also to the suggestion that *RLN3* may be a positional candidate gene for the inverted teat defect.

There is no good evidence that the identified variants of the *RLN3* genotypes are the actual causative mutations. In fact, the location of the site makes it seem unlikely that it affects protein structure or expression. The SNP at nucleotide position 2338 (A2338G) of the *RLN3* gene leads to an amino acid change from glutamine (Q) to arginine (R) in the C peptide region of the protein. Using the prediction of the SIFT analysis it was found that for the *RLN3* gene the substitution at position 101 from R to Q is predicted to affect the protein function with a score of only 0.03.

5.2.4.2   G-protein coupled receptor 135 and 142 (*GPCR135* and *GPCR142*)

The association analysis between *GPCR135* and *GPCR142* gene with the inverted teat defect showed no significant in both the experimental and the commercial pig population. It was suggested that this receptor may be not involved in reproductive tissue; *RLN3* may influence the teat characteristic via other receptors. *GPCR135* and *LGR7* share a ligand but have a vastly different tissue expression pattern, these two receptors exert different physiological functions. The physiological role of *GPCR135* may involve feeding, energy expenditure, metabolism, or other related central functions with impact on the energy metabolism also regarding the reproduction. The results from *LGR7* knockout mice indicate an essential role for the *LGR7* receptor in nipple development during pregnancy. The timed expression of *LGR7* during pregnancy in the nipple and pituitary gland suggests that pregnancy hormones such as E or P might be involved in regulating *LGR7* gene expression in these tissues (Krajnc-Franken et al. 2004).

5.2.5   Physical and genetic mapping

5.2.5.1   Relaxin3 (*RLN3*)

The positions of the porcine *RLN3* genes detected using linkage mapping are in good agreement with the published assignment to HSA19p13.2 (Bathgate et al. 2002, Matsumoto et al. 2000). By comparative mapping of pig chromosome to human, an evidence was found that the *RLN3* gene mapped on SSC2 (Meyers et al. 2005) (Figure 46).
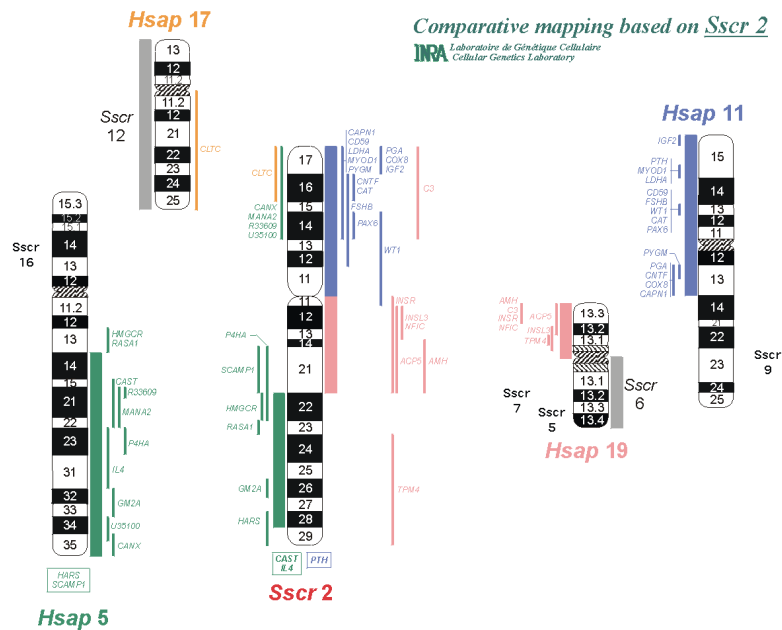
Figure 46: A comparative map of porcine chromosome 2 to the human chromosomes, whereas the location of *RLN3* is most likely on SSC2 compared to its assignement to HSA19

5.2.5.2   G-protein coupled receptor 135 (*GPCR135*)

Also the position of the porcine *GPCR135* gene is in good agreement with the published assignment to HSA5p15.1-p14 (Bathgate et al. 2002, Matsumoto et al. 2000). By comparative mapping of pig chromosome to human it was suggested that the porcine, *GPCR135* gene is mapped on SSC16 (Meyers et al. 2005) (Figure 47).
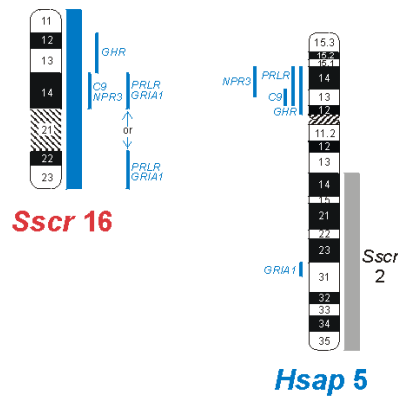
Figure 47: A comparative map of the porcine chromosome 16 and the human chromosome 5 revealed the assignment of *GPCR135* to SSC16

5.2.5.3 G-protein coupled receptor 142 (*GPCR142*)

The porcine *GPCR142* gene could be mapped on SSC4 using genetic mapping. These findings are in good agreement with the assignment to the comparative human chromosome 1q22 (Fredriksson et al. 2003). Comparative mapping of the pig chromosome to human revealed the assignment of the porcine *GPCR142* gene to SSC4 (Meyers et al. 2005) (Figure 48).
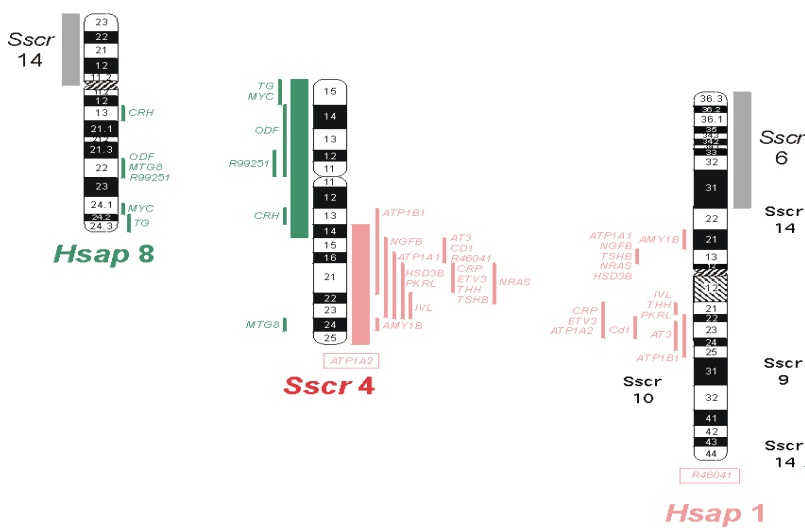
Figure 48: The comparative map of the porcine chromosome 4 and the human chromosomes 1 and 8 revealed the assignment of *GPCR142* to SSC4

5.3   Future prospects

Lifestock health and welfare traits are mostly complex and multifactorial. The identification of functional candidate genes and causative mutations is therefore a challenge. The detection of polymorphisms as well as the analysis of association and the trait associated alleles would lead to the functional characterization of genes (Hiendleder et al. 2005).

In the present study three related genes, *RLN3* and its receptors (*GPCR135* and *GPCR 142*), were molecularly characterized, the association of haplotypes of the *RLN3* with the inverted teat defect could be demonstrated. As *RLN3* and *GPCR135* were found to be differentially expressed between normal and inverted teats, future studies should focus on the additional screening for potential SNPs in the promoter regions of these genes.

The term of functional genomics refers to the development and application of holistic experimental approaches to assess gene function by using the information and reagents provided by structural genomics. It could be demonstrated in the present study that the application of whole genome microarrays is an useful tool to compare the expression patterns within relevant tissues. The results of the expression analysis delivered a list of differentially regulated genes that represent functional candidate genes for the inverted teat defect in pigs. Future investigations are ongoing; they will include the analysis of complex changes in the transcriptome during the development of the mammary gland in pigs. Further most promosing genes will be used for furher investigation of their protein profile in the target tissues. A proteomics strategy involves the localization of proteins by immunolocalization in cells as a necessary first step towards understanding protein function in complex cellular networks (Phizicky et al. 2003). A number of techniques including Western blot, immunohistochemical staining, or enzyme linked immunosorbent assay (ELISA) allow to test for proteins produced during a particular disease, which helps to understanding the proteome, the structure and function of each protein and the complexities of protein-protein interactions (Cheng et al. 2007, Shakhnovich et al. 2003).

6   Summary

This study was carried out to detect and to analyze functional candidate genes for the inverted teat defect in pigs. Using two different strategies genes were selected due to (1) their expression in tissues of normal and inverted teats of phenotypically different sows using microarray experiments, and (2) their described function in the literature. The microarray analysis led to the selection of the five functional candidate genes *CTGF*, *EGFR*, *IGF-II*, *EGF* and *GDF8*. The *RLN3* gene and its two receptors, *GPCR135* and *GPCR142* were selected due to their described physiology in literature.

In total twelve commercial DL and crossbred DL×DE sows were selected for the collection of samples for the expression study. Ten sows with at least one inverted teat, and two animals without any defect teat were mated to different boars for which a high number of affected offspring were observed during the performance tests. For this study two teats from a sow with defect and one teat of a sow without defect were analyzed using eleven genome wide pig specific microarrays. In total 695 genes were detected being up and 558 genes being down regulated in inverted compared to normal teats. The combination with clustering analysis, the results of the expression analysis led to the selection of three genes which were shown to be higher expressed in the samples of animals with inverted teats (*CTGF*, *EGFR* and *IGF-II*), and two genes were shown to be lower expressed in animals with inverted teats (*EGF* and *GDF8*).

Using real-time PCR to validate the expression results, *CTGF* and *IGF-II* genes showed higher transcript abundances in tissue of inverted teats compared to normal teats whereas the *GDF8* gene was found to have lower transcript abundance in inverted teats compared to normal teats. For the two genes *EGF* and *EGFR* no significant different transcript abundances could be shown between inverted and normal teats. In summary these results suggest that the genes *CTGF*, *IGF-II* and *GDF8* play a potential role for the development of the inverted teat defect in pigs.

For the second approach the three functional candidate genes *RLN3*, *GPCR135* and *GPCR142* were selected due to their described function in literature. Based on the porcine cDNA *RLN3* gene sequence available from NCBI (accession number AB076661) primers were selected at the end of frist exon and within the first part of second exon to sequence the intron part of this gene. Different primers were used for further sequencing of the 2033 bp long intron of the *RLN3* gene. In accordance with previous publications an 193 bp first exon and an 230 bp second exon were identified. Screening for polymorphism revealed two SNP within the *RLN3* gene located in the

intron (C1163T) and exon 2 (A2338G). Both base changes lead to a change of the cutting side of a restriction enzyme. To genotype experimental and commercial pigs at the first locus, an 434 bp fragment was amplified and further digested using *Rsa*I endonuclease. To genotype animals at the second SNP, PCR-RFLP was used to detect the different alleles within an 334 bp fragment after digestion using *Taq*I endonuclease. The allele variante within the intron including the base C did not had any cutting side and due to that led to an 434 bp fragment, whereas alleles with the base T led to two, 359 and 75 bp fragments. By digestion the fragment within the second exon, two possible alleles were generated including either the base A (no cutting side: 334 bp) or the base G (one cutting side: 217 and 117 bp).

Different from the porcine *RLN3* no prior information of the sequence of the porcine *GPCR135* gene was available. The sequence of the porcine *GPCR135* gene was therefore obtained starting with heterologous primers designed from conserved regions of the human, mouse and rat *GPCR135* genes and subsequent sequencing using homologous primers. It was found that *GPCR135* is highly conserved among different species, the porcine *GPCR135* gene showed 90%, 85% and 85% homology with human, rat and mouse genes, respectively. A primer pair was designed to screen for polymorphism within an 859 bp fragment of the *GPCR135* gene. Two SNP (C to A and C to T) were detected in the exon part of the *GPCR135* gene and genotyped using PCR-RFLP. The digestion using the *Msp*I restriction enzyme validated the detected polymorphism and led to the detection of two alleles A (350 bp, 276 bp and 202 bp fragment), and C (276 bp, 202 bp, 186 bp, and 164 bp). To genotype animals for the other SNP, PCR-RFLP using *Mnl*I endonuclease was performed. An 859 bp product was amplified showing the alternative alleles T (131 and 57 bp) and C (75, 57 and 56 bp).

Primers pairs were designed from published sequences (accession number AY633768) to screen for polymorphisms within an 1002 bp fragment of the porcine *GPCR142* gene. The screening for polymorphism revealed a SNP (G to A) within the single exon of the porcine *GPCR142* gene. Genotyping was performed by PCR-RFLP using the *Hae*III endonuclease to verify the detected polymorphism. Two alleles could be detected with allele A leading to an additional 153 bp fragment, and allele A resulted in two different fragments after digestion (126 and 27 bp). Monomorphic fragments of 222, 170, 109 and 96 bp were found in both allele variants.

Animals of an experimental pig population derived from a cross between Duroc and Berlin miniature pigs (DUMI) showing a high incidence of the inverted teat defect were used for genotyping. Additional association studies were also performed using samples of animals of the commercially used dam lines German Landrace (DL) and German Large White (DE). The Family-Based Association Test (FBAT) was used to analyse the association of the genotypes with different productive and reproductive traits. Association analysis using haplotypes revealed only a significant association of *RLN3* with the inverted teat defect. The association analysis between each SNP of *RLN3*, *GPCR135* and *GPCR142* gene with the appearance of the inverted teat defect showed no significant association in both pig populations. Haplotypes revealed no significant association between the *GPCR135* and *GPCR 142* gene and the inverted teat defect.

The position of the *RLN3* gene was mapped using linkage and physical mapping; the two receptors were mapped using only the genetic mapping. For the genetic mapping, two point and multipoint procedures of the CRI-MAP package version 2.4 were used. The multipoint linkage map revealed a good evidence of *RLN3* being mapped to SSC2. This result could be confirmed by physical mapping using a RH mapping. The results are in accordance with the comparative mapping to human, additionally it was found that *RLN3* might be located in the region of a QTL detected for the inverted teat defect on SSC2. The multipoint linkage map revealed that *GPCR135* and *GPCR142* are located on SSC16 and SSC4, respectively.

Additional to the sequencing and association analysis, the expression of these three genes was tested. Total RNA was isolated from tissues of two different male pigs (muscle, heart, spleen, lymph nodes, skin, brain, teat, lung, testis, tonsil, liver and kidney) and one sow (uterus and inverted teat). First strand cDNA was synthesised to analyse the expression pattern of the genes in different tissues following the protocol for semi-quantitative RT-PCR. The expression of *18S* rRNA was used as an internal reference. The expression profile of porcine *RLN3* was similar to the *RLN3* expression in mouse which is present in several tissues including the brain, thymus, spleen, lung, testis, ovary, and mammary gland.

It was found that the mRNA expression pattern of *GPCR142* and *GPCR135* are different, the *GPCR135* mRNA was higher expressed in reproductive tissue while the *GPCR142* mRNA was found in a broader range of peripheral tissues.

Both, the *RLN3* and *GPCR135* gene were also investigated using real-time PCR. Both genes were higher expressed in teats from sows with defect compared to sows without

defect, whereas the expression levels of inverted teat and normal teat from the same sows with defect were similar. The mRNA expression of the *GPCR142* gene was similar in teats of sows with defect and the normal teats of sows without defect.

In conclusion a number of functional candidate genes was identified and further analyzed to detect their impact on the inverted teat defect in pigs. A whole genome expression analysis showed for the first time the expression patterns of genes in relevant tissues of inverted compared to normal teats in phenotypic different sows. A list of trait dependent regulated genes that represent functional candidate genes could be validated using real-time PCR. Further this was the first study characterizing the porcine *RLN3* gene and its receptors *GPCR135* and *GPCR142*. The association between haplotypes of the *RLN3* gene and the inverted teat defect could be demonstrated. Further expression studies revealed impact of these genes on the heritable inverted teat defect in pigs.

7 Zusammenfassung

Diese Untersuchung wurde zur Identifizierung von funktionellen Kandidatengenen für den Stülpzitzendefekt beim Schwein durchgeführt. Mit Hilfe von zwei unterschiedlichen Strategien wurden Gene aufgrund (1) ihrer Expression in Geweben von normalen und Stülpzitzen von phänotypisch unterschiedlichen Sauen mit einer Microarray Analyse und (2) ihrer in der Literatur beschriebenen Funktion ausgewählt. Aufgrund der Microarray Analyse wurden die fünf funktionellen Kandidatengene *CTGF*, *EGFR*, *IGF-II*, *EGF* und *GDF8* ausgewählt. Das *RLN3* Gen und die beiden Rezeptoren *GPCR135* und *GPCR142* wurden aufgrund der beschriebenen Physiologie in der Literatur ausgewählt.

Insgesamt wurden zwölf kommerzielle reine DL und Hybridsauen (DL×DE) für die Sammlung der Proben für die Expressionsanalysen selektiert. Zehn Sauen mit mindestens einer Stülpzitze und zwei Sauen ohne Stülpzitzen wurden an Eber angepaart für die eine hohe Anzahl an betroffenen Nachkommen in der Nachkommenprüfung beobachtet wurde. Im Rahmen dieser Untersuchung wurden zwei Zitzen von einer Sau mit und eine Zitze von einer Sau ohne Stülpzitze mit elf genomweiten schweinespezifischen Microarray analysiert. Insgesamt wurden 695 Gene identifiziert, die in Stülpzitzen im Vergleich zu normalen Zitzen stärker und 558 Gene die geringer exprimiert waren. Die unterschiedlich exprimierten Gene konnten spezifischen Genclustern zugeordnet werden, die relevant für verschiedene reproduktive Mechanismen wie Reproduktionsfähigkeit, Milchsynthese, Sekretion oder Milchdrüseninvolution sind. Die Ergebnisse der Expressions- und Clusteranalyse führte zu der Auswahl von drei Genen (*CTGF*, *EGFR* und *IGF-II*), die in Proben von Stülpzitzen stärker exprimiert waren und zwei Gene (*EGF* und *GDF8*), die in diesen Geweben geringer exprimiert waren. Die Gene wurden anschließend mit einer real-time PCR untersucht, um die Ergebnisse zu bestätigen. Dabei zeigten die *CTGF* und *IGF-II* Gene eine höhere Transkriptabundanz in Geweben von Stülpzitzen im Vergleich zu normalen Zitzen und das *GDF8* Gen war geringer exprimiert in diesen Geweben. Für die beiden Gene *EGF* und *EGFR* konnte keine unterschiedliche Expression nachgewiesen werden. Zusammenfassend konnte an diesen Ergebnissen gezeigt werden, dass die Gene *CTGF*, *IGF-II* und *GDF8* eine potentielle Rolle für die Entwicklung der Stülpzitze beim Schwein spielen.

Für den zweiten Ansatz der Untersuchung wurden die drei funktionellen Kandidatengene *RLN3*, *GPCR135* und *GPCR142* aufgrund ihrer beschriebenen

Funktion in der Literatur ausgewählt. Basierend auf der porcinen cDNA *RLN3* Gensequenz der NCBI Datenbank (Nummer AB076661) wurden Primer am Ende des ersten Exon und am Beginn des zweiten Exons abgeleilet, um die Intronregion zu sequenzieren. Verschiedene Primer wurden für die weitere Sequenzierung des 2033 bp langen Introns des *RLN3* Gens gebildet. In Übereinstimmung mit vorangegangenen Veröffentlichungen konnte ein 193 bp langes Exon 1 und 230 bp langes Exon 2 identifiziert werden. Die Suche nach Polymorphismen ergab zwei SNP innerhalb der *RLN3* Gensequenz, welche im Intron (C1163T) und im Exon 2 (A2338G) lokalisiert waren. Beide Polymorphismen führen zu eine Änderung der Schnittstelle für ein Restriktionsenzym. Zur Genotypisierung von experimentellen und kommerziellen Schweinen am ersten Genort wurde ein 434 bp langes Fragment vervielfältigt und mit der *Rsa*I Endonuklease verdaut. Zur Genotypisierung von Tieren an dem Genort des zweiten SNP wurde ebenfalls eine PCR-RFLP zur Detektion der verschiedenen Allele innerhalb eines 334 bp langen Fragmentes nach dem Verdau mit *Taq*I Endonuklease durchgeführt. Die Allelvariante innerhalb des Introns mit einer C Base hatte keine Schnittstelle für das Enzym, während die Allele mit der Base T zu zwei 359 und 75 bp großen Fragmenten führte. Nach dem Verdau des Fragmentes innerhalb des zweiten Exons wurden zwei Allele produziert, welche entweder die Base A (keine Schnittstelle, 334 bp langes Fragment) oder die Base G (eine Schnittstelle: 217 und 117 bp Fragmente) enthielten.

Anders als für das porcine *RLN3* Gen war keine Information für die Sequenz des porcinen *GPCR135* Gens vorhanden. Die Sequenz dieses Gens wurde daher mit Hilfe von heterologen Primern analysiert, welche innerhalb von konservierten Regionen dieses Gens von Mensch, Maus und Ratte abgeleilet wurden. Weiteres Sequenzieren mit homologen Primern führte zur Identifizierung der Sequenz dieses Gens. Es wurde festgestellt, dass das *GPCR135* Gen unter verschiedenen Spezies sehr stark konserviert ist, die Sequenz des porcinen *GPCR135* Gens zeigt 90%, 85% und 85% Homologie mit den Sequenzen von Mensch, Ratte und Maus. Ein Primerpaar wurde zur Suche von Polymorphismen innerhalb eines 859 bp langes Fragmentes designed. Zwei SNP (C zu A und C zu T) wurden im Exon des *GPCR135* Gens detektiert und mit einer PCR-RFLP genotypisiert. Der Verdau mit dem Restriktionsenzym *Msp*I bestätigte einen Polymorphismus und führte zu den zwei Allelen A (350 bp, 276 bp und 202 bp Fragment) und C (276 bp, 202 bp, 186 bp und 164 bp). Zur Genotypisierung von Tieren an dem zweiten Genort wurde eine PCR-RFLP mit der *Mnl*I Endonuklease

durchgeführt. Ein 859 bp langes Produkt wurde amplifiziert welches die alternativen Allele T (131 und 57 bp) und C (75, 57 und 56 bp) beinhaltete.

Primerpaare wurden ausgehend von der publizierten Sequenz (Genbanknummer AY633768) des porcinen *GPCR142* Gens zur Detektion von Polymorphismen innerhalb eines 1002 bp langen Fragmentes abgeleilet. Die Suche nach Polymorphismen führte zu einem SNP (G zu A) innerhalb des Exons. Die Genotypisierung wurde mit einer PCR-RFLP mit der *Hae*III Endonklease durchgeführt. Der SNP konnte verifiziert werden, die Restriktion führte zu einem Allel A mit einem zusätzlichen 153 bp langen Fragment und einem Allel G mit zwei zusätzlichen Allelen (126 and 27 bp). Weiterhin wurden die monomorphen Fragmente der Längen 222, 170, 109 und 96 bp in beiden Allelen identifiziert.

Tiere einer experimentellen Schweinepopulation, einer Kreuzung zwischen Duroc und Berliner Miniaturschwein (DUMI), mit einer hohen Frequenz an Stülpzitzen wurden für die Genotypisierung verwendet. Zusätzlich wurden Proben der kommerziell genutzen Sauenlinien Deutsche Landrasse (DL) und Deutsches Edelschwein (DE) verwendet. Der Familien-basierte Assoziationstest wurde zur Analyse der Assoziation der Genotypen zu verschiedenen Reproduktions- und Produktionsmerkmalen verwendet. Die Assoziationsanalyse ergab eine signifikante Assoziation der Haplotypen des *RLN3* Gens zu dem Stülpzitzendefekt. Die Assoziationsanalyse zwischen den SNPs der *RLN3*, *GPCR135* und *GPCR142* Gene und dem Stülpzitzendefekt ergab keine signifikante Assoziation in beiden Populationen. Haplotypen, welche mit dem FBAT Programm gebildet wurden, zeigten ebenfalls keine signifikante Assoziation zu dem Stülpzitzendefekt.

Die Position des *RLN3* Gens konnte mit einer genetischen und physischen Kartierung festgestellt werden; die relativen Positionen der beiden Rezeptoren konnten ebenfalls mit einer genetischen Kartierung identifiziert werden. Die genetische Kartierung des *RLN3* Gens mit dem CRI-MAP Paket, Version 2.4 ergab die Zuordnung zu SSC2. Dieses Ergebnis konnte mit der physischen Kartierung mit einem RH Panel bestätigt werden. Die Ergebnisse stimmen ebenfalls mit der vergleichenden Kartierung zum Menschen überein. Es konnte darüber hinaus festgestellt werden, dass das porcine *RLN3* Gen in einer QTL Region für den Stülpzitzendefekt liegt. Die Kopplungskarte ordnete die Gene *GPCR135* und *GPCR142* den Chromosomen SSC16 und SSC4 zu.

Zusätzlich zu der Sequenzierung und der Assoziationsanalyse wurde ebenfalls die Expression dieser drei Gene untersucht. Die RNA wurde aus verschiedenen Geweben

von zwei männlichen Schweinen (Muskel, Herz, Milz, Lymphknoten, Haut, Gehirn, Zitzen, Lunge, Hoden, Mandel, Leber und Niere) und einer Sau (Uterus und Stülpzitze) isoliert. Einzelstrang cDNA wurde zur Analyse der Expressionsprofile der Gene in den verschiedenen Geweben synthetisiert. Die Expressionprofile wurden mit einer semi-quantitativen RT-PCR erstellt, die Expression des *18S* rRNA Gens wurde als Kontrolle verwendet. Die Expressionprofile des porcinen *RLN3* Gens waren ähnlich der Expression in Mäusen, wo dieses Gen in verschiedenen Geweben wie Gehirn, Thymus, Milz, Lunge, hoden, Ovar und Zitze, detektiert wurde. Es wurde ebenfalls festgestellt, das die Expressionsmuster der beiden Gene *GPCR142* und *GPCR135* sich unterschieden, die *GPCR135* mRNA war stärker in reproduktiven Geweben exprimiert während die *GPCR142* mRNA in mehr peripheralen Geweben detektiert wurde.

Die beiden Gene *RLN3* und *GPCR135* wurden ebenfalls mit einer real-time PCR untersucht. Beide Gene waren stärker in Geweben von betroffenen Sauen exprimiert im Vergleich zu Geweben von normalen Sauen, während sich die Expression in Geweben von normalen und Stülpzitzen der gleichen Sau nicht unterschied. Die mRNA Expression des *GPCR142* Gens unterschied sich in den untersuchten Geweben kaum.

Zusammenfassend wurden verschiedene funktionelle Kandidatengene identifiziert und weiter analysiert, um ihren Einfluss auf den Stülpzitzendefekt beim Schwein zu untersuchen. Eine genomweite Expressionsanalyse zeigte zum ersten Mal die Expressionsmuster von Genen in den wichtigsten Geweben von normalen und Stülpzitzen in phänoypisch unterschiedlichen Sauen. Die merkmalsabhängige Regulation von verschiedenen Gene konnte mit einer real-time PCR bestätigt werden. Diese Untersuchung war weiterhin die erste zur Charakterisierung des porcinen *RLN3* Gens und seiner Rezeptoren *GPCR135* und *GPCR142*. Die Assoziation zwischen den Haplotypen des *RLN3* Gens und dem Stülpzitzendefekt konnte gezeigt werden. Weiter zeigten Expressionanalysen den Einfluss dieser Gene auf den Erbfehler Stülpzitze beim Schwein.

8   Reference

Anderson LL, Ford JJ, Melampy RM, Cox DF (1973): Relaxin in porcine corpora lutea during pregnancy and after hysterectomy. Am J Physiol 225, 1215-1219

Andronowska A, Postek A, Doboszyńska T (2006): Epidermal growth factor and epidermal growth factor receptor immunoreactivity in the endothelial cells of the uterine artery and its branches during different stages of the estrous cycle in the pig. Pol J Vet Sci 9, 165-170

Ashworth MD, Ross JW, Stein DR, Allen DT, Spicer LJ, Geisert RD (2005): Endocrine disruption of uterine insulin-like growth factor expression in the pregnant gilt. Reproduction 130, 545-551

Bagnell CA, Tashima L, Tsark W, Ali SM, McMurtry JP (1990): Relaxin gene expression in the sow corpus luteum during the cycle, pregnancy, and lactation. Endocrinology 126, 2514-2520

Bar-Pelled U, Maltz E, Bruckental I, Folman Y, Kali Y, Gacitua H, Lehrer AR, Knight CH, Robinson B, Voet H, Tagari H (1995): Relationship between frequent milking or suckling in early lactation and milk Production of high producing dairy cows. J. Dairy Sci. 78, 2726-2736

Bathgate RAD, Samuel CS, Burazin TCD, Layfield S, Claasz AA, Reytomas IGT, Dawson NF, Zhao C, Bond C, Summers RJ, Parry LJ, Wade JD, Tregear GW (2002): Human relaxin gene 3 (H3) and the equivalent mouse relaxin (M3) gene. novel members of the relaxin peptide family. J Biol Chem 277, 1148-1157

Bazley LA, Gullick WJ (2005): The epidermal growth factor receptor family. Endocr Relat Cancer 12, S17-27

Beilage E, Steffens S, Schoon H, Bollwahn W (1996): Mammary gland hypoplasia and aplasia (inverted nipples) in female and male swine. 1. Clinical-morphological investigations on the occurrence and development of inverted nipples in female and male swine of various age and production groups. Tierarztl Prax 24, 31-35

Bockaert J, Pin JP (1999): Molecular tinkering of G protein-coupled receptors: an evolutionary success. J EMBO 18, 1723–1729

Boels K, Chica Schaller H (2003): Identification and characterisation of GPR100 as a novel human G-protein-coupled bradykinin receptor. Br J Pharmacol 140, 932-938

Borchers N, Reinsch N, Kalm E (2002): Teat number, hairiness and set of ears in a Piétrain cross:variation and effects on performance traits. Arch. Tierz., Dummerstorf 45, 465-480

Bradham DM, Igarashi A, Potter RL, Grotendorst GR (1991): Connective tissue growth factor: a cysteine-rich mitogen secreted by human vascular endothelial cells is related to the SRC-induced immediate early gene product CEF-10. J. Cell Biol. 114, 1285-1294

Brigstock DR (1999): The connective tissue growth factor/cysteine- rich 61/nephroblastoma overexpressed (CCN) family. Endocr Rev 20, 189-206

Brigstock DR (2002): Regulation of angiogenesis and endothelial cell function by connective tissue growth factor (CTGF) and cysteine-rich 61 (CYR61). Angiogenesis 5, 153-165

Brisken C, Ayyannan A, Nguyen C, Heineman A, Reinhardt F, Jan T, Dey SK, Dotto GP, Weinberg RA (2002): IGF-2 Is a mediator of prolactin-induced morphogenesis in the breast. Dev Cell 3, 877-887

Brissenden JE, Ullrich A, Francke U (1984): Human chromosomal mapping of genes for insulin-like growth factors I and II and epidermal growth factor. Nature 310, 781-784

Campbell ID, Baron M, Cooke RM, Dudgeon TJ, Fallon A, Harvey TS, Tappin MJ (1990): Structure-function relationships in epidermal growth factor (egf) and transforming growth factor-alpha (TGF-[alpha]). Biochem Pharmacol 40, 35-40

Carlin CR, Knowles BB (1982): Identity of Human Epidermal Growth Factor (EGF) Receptor with Glycoprotein SA-7: Evidence for differential phosphorylation of the two components of the EGF Receptor from A431 Cells. Proc Nat Acad Sci 79, 5026-5030

Carpenter G, Cohen S (1979): Epidermal growth factor. Annu Rev Biochem 48, 193-216

Chen J, Kuei C, Sutton SW, Bonaventure P, Nepomuceno D, Eriste E, Sillard R, Lovenberg TW, Liu C (2005): Pharmacological characterization of relaxin-3/INSL7 receptors GPCR135 and GPCR142 from different mammalian species. J Pharmacol Exp Ther 312, 83-95

Cheng Y, Zhang J, Li Y, Wang Y, Gong J (2007): Proteome analysis of human gastric cardia adenocarcinoma by laser capture microdissection. BMC Cancer 7, 191

Cho IC, Choi YL, Ko MS, Kim HS, Lee JG, Jeon JT, Han SH (2005): Association between MSTN gene polymorphism and growth traits in Landrace pigs. J of Ani Sci and Tech 47, 159-166

Chomdej S (2005): Molecular genetic analysis of positional candidate genes for mammary gland characteristics in pigs. Dissertation, Rheinische Friedrich-Wilhelms-Universität

Clarkson R, Wayland M, Lee J, Freeman T, Watson C (2004): Gene expression profiling of mammary gland development reveals putative roles for death receptors and immune mediators in post-lactational regression. Breast Cancer Res 6, R92 - R109

Clayton G, Powell J, Hiley P (1981): Inheritance of teat number and teat inversion in pigs. Anim Prod 33, 299-304

Coleman S, Silberstein GB, Daniel CW (1988): Ductal morphogenesis in the mouse mammary gland: evidence supporting a role for epidermal growth factor. Dev Biol 127, 304-315

Connor EE, Meyer MJ, Li RW, Van Amburgh ME, Boisclair YR, Capuco AV (2007): Regulation of gene expression in the bovine mammary gland by ovarian steroids. J. Dairy Sci. 90, E55-65

Connor E, Siferd S, Elsasser T, Evock-Clover C, Van Tassell C, Sonstegard T, Fernandes V, Capuco A (2008): Effects of increased milking frequency on gene expression in the bovine mammary gland. BMC Genomics 9, 362

Crish JF, Soloff MS, Shaw AR (1986): Changes in relaxin precursor mRNA levels in the rat ovary during pregnancy. J. Biol. Chem. 261, 1909-1913

DiAugustine RP, Richards RG, Sebastian J (1997): EGF-related peptides and their receptors in mammary gland development. J Mamm Gland Biol Neoplasia 2, 109-117

Dorsam RT, Gutkind JS (2007): G-protein-coupled receptors and cancer. Nat Rev Cancer 7, 79-94

Edery M, Pang K, Larson L, Colosi T, Nandi S (1985): Epidermal growth factor receptor levels in mouse mammary glands in various physiological states. Endocrinology 117, 405-411

Enfield FD, Rempel WE (1961): Inheritance of teat number and relationship of teat number to various maternal traits in swine. J. Anim Sci. 20, 876-879

Engelman RW, Owens UE, Bradley WG, Day NK, Good RA (1995): Mammary and submandibular gland epidermal growth factor expression is reduced by calorie restriction. Cancer Res 55, 1289-1295

Erdman RA, Varner M (1995): Fixed yield responses to increased milking frequency. J. Dairy Sci. 78, 1199-1203

Fenton SE (2006): Endocrine-disrupting compounds and mammary gland development: early exposure and later life consequences. Endocrinology 147, s18-24

Fenton SE, Sheffield LG (1993): Prolactin inhibits epidermal growth factor (EGF)-stimulated signaling events in mouse mammary epithelial cells by altering EGF receptor function. Mol. Biol. Cell 4, 773-780

Fernandez A, Toro M, Rodriguez C, Silio L (2004): Heterosis and epistasis for teat number and fluctuating asymmetry in crosses between Jiaxing and Iberian pigs. Heredity 93, 222-227

Fonseca C, Lindahl GE, Ponticos M, Sestini P, Renzoni EA, Holmes AM, Spagnolo P, Pantelidis P, Leoni P, McHugh N, Stock CJ, Shi-Wen X, Denton CP, Black CM, Welsh KI, du Bois RM, Abraham DJ (2007): A Polymorphism in the CTGF promoter region associated with systemic sclerosis. N Engl J Med 357, 1210-1220

Ford JA, Jr., Kim SW, Rodriguez-Zas SL, Hurley WL (2003): Quantification of mammary gland tissue size and composition changes after weaning in sows. J. Anim Sci. 81, 2583-2589

Frazier KS, Grotendorst GR (1997): Expression of connective tissue growth factor mRNA in the fibrous stroma of mammary tumors. The International J of Biochem Cell Biol 29, 153-161

Fredriksson R, Hoglund PJ, Gloriam DEI, Lagerstrom MC, Schioth HB (2003): Seven evolutionarily conserved human rhodopsin G protein-coupled receptors lacking close relatives. FEBS Letters 554, 381-388

Garrett TPJ, McKern NM, Lou M, Elleman TC, Adams TE, Lovrecz GO, Zhu HJ, Walker F, Frenkel MJ, Hoyne PA, Jorissen RN, Nice EC, Burgess AW, Ward CW (2002): Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor alpha. Cell 110, 763-773

Gary BS (2001): Postnatal mammary gland morphogenesis. Micros Res Tech 52, 155-162

Gaunt TR, Cooper JA, Miller GJ, Day INM, O'Dell SD (2001): Positive associations between single nucleotide polymorphisms in the IGF2 gene region and body mass index in adult males. Hum. Mol. Genet. 10, 1491-1501

Gonzalez-Cadavid NF, Taylor WE, Yarasheski K, Sinha-Hikim I, Ma K, Ezzat S, Shen R, Lalani R, Asa S, Mamita M, Nair G, Arver S, Bhasin S (1998): Organization of the human myostatin gene and expression in healthy men and HIV-infected men with muscle wasting. Proc Nat Acad Sci 95, 14938-14943

Green P (1992): Construction and Comparison of Chromosome-21 Radiation Hybrid and Linkage Maps Using Cri-Map. Cytogenet Cell Gene 59, 122-124

Große EB, Steffens S, Schoon H, Bollwahn W (1996): Zitzenkörperhypoplasien und -aplasien (Stülpzitzen) bei weiblichen und männlichenSchweinen. Tierärztliche Praxis 24, 31-35

Grotendorst GR, Lau LF, Perbal B (2000): CCN proteins are distinct from and should not be considered members of the insulin-like growth factor-binding protein superfamily. Endocrinology 141, 2254-2256

Günther C (1984): Morphologie der sogenannten "Stülpzitzen" beim Schwein im Vergleich zum histologischen Bild der normalen Zitzen. Dissertation, Freie Universität

Hackel PO, Zwick E, Prenzel N, Ullrich A (1999): Epidermal growth factor receptors: critical mediators of multiple receptor pathways. Curr Opin Cell Biol 11, 184-189

Haley J, Crawford R, Hudson P, Scanlon D, Tregear G, Shine J, Niall H (1987): Porcine relaxin. Gene structure and expression. J. Biol. Chem. 262, 11940-11946

Halls ML, van der Westhuizen ET, Bathgate RAD, Summers RJ (2007): Relaxin family peptide receptors - former orphans reunite with their parent ligands to activate multiple signalling pathways. Br J Pharmacol 150, 677-691

Hamai Y, Matsumura S, Matsusaki K, Kitadai Y, Yoshida K, Yamaguchi Y, Imai K, Nakachi K, Toge T, Yasui W (2005): A single nucleotide polymorphism in the 5' untranslated region of the EGF gene is associated with occurrence and malignant progression of gastric cancer. Pathobiology 72, 133-138

Hansen TK, Thiel S, Dall R, Rosenfalck AM, Trainer P, Flyvbjerg A, Jorgensen JOL, Christiansen JS (2001): GH strongly affects serum concentrations of mannan-binding lectin: evidence for a new IGF-I independent immunomodulatory effect of GH. J Clin Endocrinol Metab 86, 5383-5388

Harris RC, Chung E, Coffey RJ (2003): EGF receptor ligands. Ex Cell Res 284, 2-13

Heemskerk VH, Daemen M, Buurman WA (1999): Insulin-like growth factor-1 (IGF-1) and growth hormone (GH) in immunity and inflammation. Cytokine & Growth Factor Rev 10, 5-14

Henikoff JG, Henikoff S (1996): Using substitution probabilities to improve position-specific scoring matrices. Comput. Appl. Biosci. 12, 135-143

Hennighausen L, Robinson GW (2001): Signaling pathways in mammary gland development. Dev Cell 1, 467-475

Hiendleder S, Bauersachs S, Boulesteix A, Blum H, Arnold GJ, Fröhlich T, Wolf E (2005): Functional genomics: tools for improving farm animal health and welfare Rev. sci. tech. Off. int. Epiz 24 355-377

Hirooka H, de Koning DJ, Harlizius B, van Arendonk JA, Rattink AP, Groenen MA, Brascamp EW, Bovenhuis H (2001): A whole-genome scan for quantitative trait loci affecting teat number in pigs. J. Anim Sci. 79, 2320-2326

Horvath S, Xu X, Laird NM (2001): The family based association test method: strategies for studying general genotype-phenotype associations. Eu J Hum Genet 9, 301-306

Hovey RC, McFadden TB, Akers RM (1999): Regulation of mammary gland growth and morphogenesis by the mammary fat pad: a species comparison. J Mamm Gland Biol Neoplasia 4, 53-68

Hsu SY, Nakabayashi K, Nishi S, Kumagai J, Kudo M, Sherwood OD, Hsueh AJW (2002): Activation of orphan receptors by the hormone relaxin. Science 295, 671-674

Hsu SYT, Semyonov J, Park J-I, Chang CL (2005): Evolution of the signaling system in relaxin-family peptides. Ann N Y Acad Sci 1041, 520-529

Hu Y, Sun H, Drake J, Kittrell F, Abba MC, Deng L, Gaddis S, Sahin A, Baggerly K, Medina D, Aldaz CM (2004): From mice to humans: identification of commonly deregulated genes in mammary cancer via comparative SAGE Studies. Cancer Res 64, 7748-7755

Hurvitz JR, Suwairi WM, Van Hul W, El-Shanti H, Superti-Furga A, Roudier J, Holderbaum D, Pauli RM, Herd JK, Hul EV, Rezai-Delui H, Legius E, Le Merrer M, Al-Alami J, Bahabri SA, Warman ML (1999): Mutations in the CCN gene family member WISP3 cause progressive pseudorheumatoid dysplasia. Nat Genet 23, 94-98

Hwang JJ, Lee AB, Fields PA, Haab LM, Mojonnier LE, Sherwood OD (1991): Monoclonal antibodies specific for rat relaxin. V. Passive immunization with monoclonal antibodies throughout the second half of pregnancy disrupts development of the mammary apparatus and, hence, lactational performance in rats. Endocrinology 129, 3034-3042

Jackson L, Qiu T, Sunnarborg S, Chang A, Zhang C, Patterson C, Lee D (2003): Defective valvulogenesis in HB-EGF and TACE-null mice is associated with aberrant BMP signaling. J EMBO 22, 2704-2716

Jamnongjit M, Gill A, Hammes SR (2005): Epidermal growth factor receptor signaling is required for normal ovarian steroidogenesis and oocyte maturation. Proc Natl Acad Sci 102, 16257-16262

Jeon J-T, Carlborg O, Tornsten A, Giuffra E, Amarger V, Chardon P, Andersson-Eklund L, Andersson K, Hansson I, Lundstrom K, Andersson L (1999): A paternally expressed QTL affecting skeletal and cardiac muscle mass in pigs maps to the IGF2 locus. Nat Genet 21, 157-158

Ji S, Losinski RL, Cornelius SG, Frank GR, Willis GM, Gerrard DE, Depreux FFS, Spurlock ME (1998): Myostatin expression in porcine tissues: tissue specificity and developmental and postnatal regulation. Am J Physiol Regul Integr Comp Physiol 275, R1265-1273

Jiang Y, Li N, Wu C, Plastow G, Liu Z, Hu X, Wu C (2002): Identification of three SNPs in the porcine myostatin gene (MSTN). Anim Biotechnol 13, 173-178

Jonas E (2006): Ansätze zur Untersuchung der genetischen Ursachen für den Erbfehler Stülpzitze beim Schwein. Dissertation, Rheinische Friedrich-Wilhelms-Universität

Jonas E, Chomdej S, Yammuen-art S, Phatsara C, Schreinemachers H-J, Jennen D, Tesfaye D, Ponsuksili S, Wimmers K, E ET, Schellander K (2007): Verification of chromosomal regions affecting the inverted teat development and their derivable candidate genes in pigs, proceeding of 58th meeting of the european association for animal production (EAAP),26.-29.8.2007, Dublin, Ireland, 348

Jonas E, Schreinemachers H-J, Kleinwächter T, Ün C, Oltmanns I, Tetzlaff S, Jennen D, Tesfaye D, Ponsuksili S, Murani E, Juengst H, Tholen E, Schellander K, Wimmers K (2008): QTL for the heritable inverted teat defect in pigs. Mamm Genome 19, 127-138

Kass L, Ramos J, Ortega H, Montes G, Bussmann L, Luque E, de Toro M (2001): Relaxin has a minor role in rat mammary gland growth and differentiation during pregnancy. Endocrine 15, 263-269

Kim H-S, Nagalla SR, Oh Y, Wilson E, Roberts CT, Jr., Rosenfeld RG (1997): Identification of a family of low-affinity insulin-like growth factor binding proteins (IGFBPs): Characterization of connective tissue growth factor as a member of the IGFBP superfamily. Proc Natl Acad Sci 94, 12981-12986

Knox RV, Zhang Z, Day BN, Anthony RV (1994): Identification of relaxin gene expression and protein localization in the uterine endometrium during early pregnancy in the pig. Endocrinology 135, 2517-2525

Kohsaka T, Min G, Lukas G, Trupin S, Campbell ET, Sherwood OD (1998): Identification of specific relaxin-binding cells in the human female. Biol Reprod 59, 991-999

Krajnc-Franken MAM, van Disseldorp AJM, Koenders JE, Mosselman S, van Duin M, Gossen JA (2004): Impaired nipple development and parturition in LGR7 knockout mice. Mol. Cell. Biol. 24, 687-696

Kuei C, Sutton S, Bonaventure P, Pudiak C, Shelton J, Zhu J, Nepomuceno D, Wu J, Chen J, Kamme F, Seierstad M, Hack MD, Bathgate RAD, Hossain MA, Wade JD, Atack J, Lovenberg TW, Liu C (2007): R3(B{Delta}23 27)R/I5 chimeric peptide, a selective antagonist for GPCR135 and GPCR142 over relaxin receptor LGR7: in vitro and in vivo characterization. J. Biol. Chem. 282, 25425-25435

Kuenzi MJ, Connolly BA, Sherwood OD (1995): Relaxin acts directly on rat mammary nipples to stimulate their growth. Endocrinology 136, 2943-2947

Kuenzi MJ, Sherwood OD (1995): Immunohistochemical localization of specific relaxin-binding cells in the cervix, mammary glands, and nipples of pregnant rats. Endocrinology 136, 1367-1373

Kuenzi MJ, Sherwood OD (1992): Monoclonal antibodies specific for rat relaxin. VII. Passive immunization with monoclonal antibodies throughout the second half of pregnancy prevents development of normal mammary nipple morphology and function in rats. Endocrinology 131, 1841-1847

Kumagai J, Hsu SY, Matsumi H, Roh J-S, Fu P, Wade JD, Bathgate RAD, Hsueh AJW (2002): INSL3/Leydig insulin-like peptide activates the LGR8 receptor important in testis descent. J. Biol. Chem. 277, 31283-31286

Kwon MJ, Goate AM (2000): The candidate gene approach. Alcohol research & health 24, 164-168

Larkin HL, Renegar RH (1986): Immunochemical and cytochemical studies of relaxin-containing cells in the guinea pig uterus. Am J Anat 176, 353-365

Lau LF, Lam SCT (1999): The CCN family of angiogenic regulators: the integrin connection. Exp Cell Res 248, 44-57

Lax I, Bellot F, Howk R, Ullrich A, Givol D, Schlessinger J (1989): Functional-analysis of the ligand-binding site of EGF-receptor utilizing chimeric chicken human receptor molecules. Embo J 8, 421-427

Lee C, Wang CD (2001): Bayesian inference on variance components using Gibbs sampling with various priors. Asian Australasian J. Anim. Sci 14, 1051-1056

Li S, Xiong Y, Zheng R, Li A, Deng C, Jiang S, Lei M, Wen Y, Cao G (2002): Polymorphism of porcine myostatin gene. Yi Chuan Xue Bao. 29, 326-331

Liu C, Chen J, Kuei C, Sutton S, Nepomuceno D, Bonaventure P, Lovenberg TW (2005): Relaxin-3/insulin-like peptide 5 chimeric peptide, a selective ligand for G Protein-Coupled Receptor (GPCR)135 and GPCR142 over leucine-rich repeat-containing G protein-poupled receptor 7. Mol Pharmacol 67, 231-240

Liu C, Chen J, Sutton S, Roland B, Kuei C, Farmer N, Sillard R, Lovenberg TW (2003a): Identification of relaxin-3/INSL7 as a ligand for GPCR142. J. Biol. Chem. 278, 50765-50770

Liu C, Eriste E, Sutton S, Chen J, Roland B, Kuei C, Farmer N, Jornvall H, Sillard R, Lovenberg TW (2003b): Identification of relaxin-3/INSL7 as an endogenous ligand for the orphan G-protein-coupled receptor GPCR135. J. Biol. Chem. 278, 50754-50764

Loladze AV, Stull MA, Rowzee AM, DeMarco J, Lantry JH, III, Rosen CJ, LeRoith D, Wagner K-U, Hennighausen L, Wood TL (2006): Epithelial-specific and stage-specific functions of insulin-like growth factor-I during postnatal mammary development. Endocrinology 147, 5412-5423

Luetteke NC, Qiu TH, Fenton SE, Troyer KL, Riedel RF, Chang A, Lee DC (1999): Targeted inactivation of the EGF and amphiregulin genes reveals distinct roles for EGF receptor ligands in mouse mammary gland development. Development 126, 2739-2750

Martinerie C, Viegas-Pequignot E, Guenard I, Dutrillaux B, Nguyen VC, Bernheim A, Perbal B (1992): Physical mapping of human loci homologous to the chicken nov proto-oncogene. Oncogene 7, 2529-2534

Matsumoto M, Kamohara M, Sugimoto T, Hidaka K, Takasaki J, Saito T, Okada M, Yamaguchi T, Furuichi K (2000): The novel G-protein coupled receptor SALPR shares sequence similarity with somatostatin and angiotensin receptors. Gene 248, 183-189

McPherron AC, Lawler AM, Lee S-J (1997): Regulation of skeletal muscle mass in mice by a new TGF-p superfamily member. Nature 387, 83-90

Mendez EA, Messer LA, Larsen NJ, Robic A, Rothschild MF (1999): Epidermal growth factor maps to pig chromosome 8. J. Anim Sci. 77, 494-495

Meyers SN, Rogatcheva MB, Larkin DM, Yerle M, Milan D, Hawken RJ, Schook LB, Beever JE (2005): Piggy-BACing the human genome: II. A high-resolution, physically anchored, comparative map of the porcine autosomes. Genomics 86, 739-752

Min G, Sherwood OD (1996): Identification of specific relaxin-binding cells in the cervix, mammary glands, nipples, small intestine, and skin of pregnant pigs. Biol Reprod 55, 1243-1252

Molenat M, Thibault B (1977): Heritability of number of false teats in gilts. J Rech Porce 69-73

Monks J, Geske FJ, Lehman L, Fadok VA (2002): Do inflammatory cells participate in mammary gland involution? J Mamm Gland Biol Neoplasia 7, 163-176

Moore K, Kramer JM, Rodriguez-Sallaberry CJ, Yelich JV, Drost M (2007): Insulin-like growth factor (IGF) family genes are aberrantly expressed in bovine conceptuses produced in vitro or by nuclear transfer. Theriogenol, 68 717-727

Morton C, Byers M, Nakai H, Bell G, Shows T (1986): Human genes for insulin-like growth factors I and II and epidermal growth factor are located on 12q22-q24.1, 11p15, and 4q25-q27, respectively. Cytogenet Cell Genet. 41, 245-249

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

Ng PC, Henikoff S (2001): Predicting Deleterious Amino Acid Substitutions. Genome Res 11, 863-874

Nygard A-B, Jorgensen C, Cirera S, Fredholm M (2007): Selection of reference genes for gene expression studies in pig tissues using SYBR green qPCR. BMC Mol Biol 8, 67

O'Dell SD, Day INM (1998): Molecules in focus insulin-like growth factor II (IGF-II). The International J Biochem Cell Biol 30, 767-771

Ogiso H, Ishitani R, Nureki O, Fukai S, Yamanaka M, Kim J-H, Saito K, Sakamoto A, Inoue M, Shirouzu M, Yokoyama S (2002): Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. Cell 110, 775-787

Oltmanns I (2003): QTL-Analyse an den Chromosomen 2, 3, 5, 7, 9, 11, 13, 15 und 17 in einer porcinen Ressourcenpopulation. Dissertation, Rheinische Friedrich – Wilhelms – Universität zu Bonn

Phizicky E, Bastiaens PIH, Zhu H, Snyder M, Fields S (2003): Protein analysis on a proteomic scale. Nature 422, 208-215

Rabinowitz D, Laird N (2000): A Unified approach to adjusting association tests for opulation admixture with arbitrary pedigree structure and arbitrary missing marker information. Hum Hered 50, 211

Radonic A, Thulke S, Mackay IM, Landt O, Siegert W, Nitsche A (2004): Guideline to reference gene selection for quantitative real-time PCR. Biochem Biophysi Res Commun 313, 856-862

Robinson G (2004): Identification of signaling pathways in early mammary gland development by mouse genetics. Breast Cancer Res 6, 105 - 108

Rosengren KJ, Lin F, Bathgate RAD, Tregear GW, Daly NL, Wade JD, Craik DJ (2006): Solution structure and novel insights into the determinants of the receptor specificity of human relaxin-3. J. Biol. Chem. 281, 5845-5851

Rosfjord EC, Dickson RB (1999): Growth factors, apoptosis, and survival of mammary epithelial cells. J Mamm Gland Biol Neoplasia 4, 229-237

Samuel CS, Royce SG, Burton MD, Zhao C, Tregear GW, Tang MLK (2007): Relaxin plays an important role in the regulation of airway structure and function. Endocrinology 148, 4259-4266

Sandra ZH, Laura JC, Katherine AN (1992): EGF receptor regulation in normal mouse mammary gland. J Cell Physiol 152, 553-557

Schmidt GH (1971): Biology of lactation. W. H. Freeman and Company San Francisco, CA.

Schroeder JA, Lee DC (1998): Dynamic expression and activation of ERBB receptors in the developing mouse mammary gland. Cell Growth Differ 9, 451-464

Sebastian J, Richards RG, Walker MP, Wiesen JF, Werb Z, Derynck R, Hom YK, Cunha GR, DiAugustine RP (1998): Activation and function of the epidermal growth factor receptor and erbB-2 during mammary gland morphogenesis. Cell Growth Differ 9, 777-785

Seo KS, Kim SH, Park YI (1996): Estimation of genetic parameters for teat number in pigs. Korean J. Anim. Sci 38 133-138

Shahbazi M, Pravica V, Nasreen N, Fakhoury H, Fryer AA, Strange RC, Hutchinson PE, Osborne JE, Lear JT, Smith AG, Hutchinson IV (2002): Association between functional polymorphism in EGF gene and malignant melanoma. Lancet 359, 397-401

Shakhnovich B, Harvey J, Comeau S, Lorenz D, DeLisi C, Shakhnovich E (2003): ELISA: Structure-function inferences based on statistically significant and evolutionarily inspired observations. BMC Bioinformatics 4, 34

Sheffield LG (1998): Hormonal regulation of epidermal growth factor receptor content and signaling in bovine mammary tissue. Endocrinology 139, 4568-4575

Sheffield LG, Welsch CW (1987): Influence of submandibular salivary glands on hormone responsiveness of mouse mammary glands. Proc Soc Exp Biol Med 186, 368-377

Sherwood OD (2004): Relaxin's physiological roles and other diverse actions. Endoc Re 25, 205-234

Slee RB, Hillier SG, Largue P, Harlow CR, Miele G, Clinton M (2001): Differentiation-dependent expression of connective tissue growth factor and lysyl oxidase messenger ribonucleic acids in rat granulosa cells. Endocrinology 142, 1082-1089

Sonstegard TS, Rohrer GA, Smith TPL (1998): Myostatin maps to porcine chromosome 15 by linkage and physical analyses. Animal Genetics 29, 19-22

Sood R, Zehnder JL, Druzin ML, Brown PO (2006): Gene expression patterns in human placenta. Proc Natl Acad Sci 103, 5478-5483

Steffens S (1993): Stülpzitzen-ein Problem in der Schweinezucht. Schweinezucht und Schweine mast 3, 8-11

Sternlicht MD, Sunnarborg SW, Kouros-Mehr H, Yu Y, Lee DC, Werb Z (2005): Mammary ductal morphogenesis requires paracrine activation of stromal EGFR via ADAM17-dependent shedding of epithelial amphiregulin. Development 132, 3923-3933

Stratil A, Geldermann H (2004): Analysis of porcine candidate genes from selected QTL regions affecting production traits. Anim Sci Pap Rep 22, 123-125

Stratil A, Kopecny M (1999): Genomic organization, sequence and polymorphism of the porcine myostatin (GDF8; MSTN) gene. Anim Gen 30, 462-478

Streuli CH, Haslam SZ (1998): Control of mammary gland development and neoplasia by stromal-epithelial interactions and extracellular matrix. J Mamm Gland Biol Neoplasia 3, 107-108

Su Z, Dong X, Zhang B, Zeng Y, Fu Y, Yu J, Hu S (2006): Gene expression profiling in porcine mammary gland during lactation and identification of breed-and developmental-stage-specific genes. Science in China Series C: Life Sci 49, 26-36

Sudo S, Kumagai J, Nishi S, Layfield S, Ferraro T, Bathgate RAD, Hsueh AJW (2003): H3 relaxin is a specific ligand for LGR7 and activates the receptor by interacting with both the ectodomain and the exoloop 2. J Biol Chem 278, 7855-7862

Tabor HK, Risch NJ, Myers RM (2002): Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat Rev Genet 3, 391-397

Taketani Y, Oka T (1983): Epidermal growth factor stimulates cell proliferation and inhibits functional differentiation of mouse mammary epithelial cells in culture. Endocrinology 113, 871-877

Trakooljul N (2004): Molecular and association analyses of the androgen receptor gene as a candidate for production and reproduction traits in pigs. Dissertation, Rheinische Friedrich-Wilhelms-Universität Bonn.

Trakooljul N, Ponsuksili S, Schellander K, Wimmers K (2004): Polymorphisms of the porcine androgen receptor gene affecting its amino acid sequence and expression level. Biochimi Biophys Acta (BBA) - Gene Structure and Expression 1678, 94-101

Tricoli JV, Rall LB, Scott J, Bell GI, Shows TB (1984): Localization of insulin-like growth factor genes to human chromosomes 11 and 12. Nature 310, 784-786

Ün C (2002): QTL-Analyse an Chromosomen 1, 4, 6, 8, 10, 12, 14, 16 und 18 in einer porcinen Ressourcenfamilie. Dissertation, Rheinische Friedrich - Wilhelms - Universität zu Bonn

Uzumcu M, Homsi MFA, Ball DK, Coskun S, Jaroudi K, Hollanders JMG, Brigstock DR (2000): Localization of connective tissue growth factor in human uterine tissues. Mol. Hum. Reprod. 6, 1093-1098

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

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F (2002): Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biology 3, 1-11

Varadin M, Filipovic M (1975): Klinisches Bild der Milchdrüse bei Sauen und einige Besonderheiten ihres Sekretes. Zuchthygiene 10, 109-118

Wandji SA, Gadsby JE, Barber JA, Hammond JM (2000): Messenger ribonucleic acids for MAC25 and connective tissue growth factor (CTGF) are inversely regulated during folliculogenesis and early luteogenesis. Endocrinology 141, 2648-2657

Wang CD, Chen QM, Zhou WL, Wang ZQ, Han SY, Li ZK (2000): Bayesian inference of teat number in Landrace using Gibbs sampling. J. China Agric. Univ 5 92-95

Wehling M, Cai B, Tidball JG (2000): Modulation of myostatin expression during modified muscle use. FASEB J. 14, 103-110

Wetzker R, Bohmer F-D (2003): Transactivation joins multiple tracks to the ERK/MAPK cascade. Nat Rev Mol Cell Biol 4, 651-657

Wiesen JF, Young P, Werb Z, Cunha GR (1999): Signaling through the stromal epidermal growth factor receptor is necessary for mammary ductal development. Development 126, 335-344

Wilkinson TN, Speed TP, Tregear GW, Bathgate RA (2005): Evolution of the relaxin-like peptide family. BMC Evolutionary Biology 5, 14

Willham RL, Whatley JA (1963): Genetic variation in nipple number in swine. Z f Tierzüchtung und Züchtungsbiologie 78, 350-363.

Winn RJ, Baker MD, Merle CA, Sherwood OD (1994): Individual and combined effects of relaxin, estrogen, and progesterone in ovariectomized gilts. II. Effects on mammary development. Endocrinology 135, 1250-1255

Wood TL, Richert MM, Stull MA, Allar MA (2000): The Insulin-like growth factors (IGFs) and IGF binding proteins in postnatal development of murine mammary glands. J Mamm Gland Biol Neoplasia 5, 31-42

Yammuen-art S, Jennen D, Jonas E, Phatsara C, Ponsuksili S, Tesfaye D, Tholen E, Wimmers K, Schellander K (2007a): Characterization of relaxin-3/INSL7 and its receptor being candidate genes for the inverted teat defect in pigs, proceeding of 58th meeting of the european sssociation for animal production (EAAP), 26.-29.8.2007, Dublin, Ireland 350

Yammuen-art S, Phatsara C, Wimmers K, Ponsuksili S, Tholen E, Schellander K, Jonas E (2007b): Association of the pituitary-specific transcription factor-1 and leptin genes with traits related to the teat development in pigs, proceeding of 13th international conference on production diseases in farm animals (ICPD), 29.7.-4.8.2007, Leipzig, Germany (Abstr)

Yang J, Guzman R, Richards J, Imagawa W, McCormick K, Nandi S (1980): Growth factor- and cyclic nucleotide-induced proliferation of normal and malignant mammary epithelial cells in primary culture. Endocrinology 107, 35-41

Yerle M, Pinton P, Robic A, Alfonso A, Palvadeau Y, Delcros C, Hawken R, Alexander L, Beattie C, Schook L, Milan D, Gellin J (1998): Construction of a whole-genome radiation hybrid panel for high-resolution gene mapping in pigs. Cytogenet Cell Genet 82, 182-188

Zaleski HM, Winn RJ, Jennings RL, Sherwood OD (1996): Effects of relaxin on lactational performance in ovariectomized gilts. Biol Reprod 55, 671-675

Zhang S, Bidanel J, Burlot T, Legault C, Naveau J (2000): Genetic parameters and genetic trends in the Chinese x European Tiameslan I. Genetic parameters Genet Sel Evol 32,

Zhao L, Roche PJ, Gunnersen JM, Hammond VE, Tregear GW, Wintour EM, Beck F (1999): Mice without a functional relaxin gene are unable to deliver milk to their pups. Endocrinology 140, 445-453

Zhao L, Samuel CS, Tregear GW, Beck F, Wintour EM (2000): Collagen studies in late pregnant relaxin null mice. Biol Reprod 63, 697-703

Zhi-Hua J, John PG (1999): Genetic polymorphisms in the leptin gene and their association with fatness in four pig breeds. Mamm Genome 10, 191-193

Zhu M, Zhao S (2007): Candidate gene identification approach: progress and challenges. Int J Biol Sci 3, 420–427

Zimmers TA, Davies MV, Koniaris LG, Haynes P, Esquela AF, Tomkinson KN, McPherron AC, Wolfman NM, Lee S-J (2002): Induction of cachexia in mice by systemically administered myostatin. Science 296, 1486-1488

Zucchi I, Mento E, Kuznetsov VA, Scotti M, Valsecchi V, Simionati B, Vicinanza E, Valle G, Pilotti S, Reinbold R, Vezzoni P, Albertini A, Dulbecco R (2004): Gene expression profiles of epithelial cells microscopically isolated from a breast-invasive ductal carcinoma and a nodal metastasis. Proc Natl Acad Sci 101, 18147-18152

9 Appendixes

9.1 The functional annotations of up-regulated genes in inverted teat

Table 9.1: The functional annotations among 695 up-regulated genes in inverted teat

Annotation Cluster 1	Enrichment Score: 30.76	Number of genes	P-Value
GOTERM_BP_ALL	Antigen presentation	76	2.5E-59
GOTERM_BP_ALL	Immune response	81	3.3E-49
GOTERM_BP_ALL	Antigen processing	63	6.3E-48
SP_PIR_KEYWORDS	MHC II	48	7.0E-46
GOTERM_BP_ALL	Defense response	81	2.7E-45
GOTERM_BP_ALL	Response to biotic stimulus	81	8.5E-45
GOTERM_BP_ALL	Organismal physiological process	85	1.1E-44
GOTERM_MF_ALL	MHC class II receptor activity	50	8.6E-40
GOTERM_BP_ALL	Response to stimulus	85	1.6E-39
INTERPRO_NAME	MHC class II, beta chain, N-terminal	46	4.3E-39
GOTERM_BP_ALL	Antigen presentation, exogenous antigen	49	1.2E-36
GOTERM_BP_ALL	Antigen processing, exogenous antigen via MHC class II	49	1.2E-36
GOTERM_MF_ALL	Transmembrane receptor activity	78	1.3E-35
GOTERM_MF_ALL	Receptor activity	81	6.0E-24
GOTERM_MF_ALL	Signal transducer activity	88	1.2E-18
SP_PIR_KEYWORDS	Transmembrane	56	6.7E-14
GOTERM_CC_ALL	Membrane	89	1.5E-11
GOTERM_CC_ALL	Intrinsic to membrane	64	1.2E-9
GOTERM_CC_ALL	Integral to membrane	64	1.2E-9
GOTERM_BP_ALL	Physiological process	105	4.3E-7
GOTERM_CC_ALL	Cell	99	2.4E-2

Continued Table 9.1

Annotation Cluster 2	Enrichment Score: 15.5	Number of genes	P-Value
INTERPRO_NAME	Immunoglobulin C1 type	43	8.8E-33
INTERPRO_NAME	Immunoglobulin/major histocompatibility complex	39	6.3E-30
SMART_NAME	IGc1	36	1.8E-25
INTERPRO_NAME	Immunoglobulin-like	43	7.7E-24
INTERPRO_NAME	MHC class I, alpha chain, alpha1 and alpha2	25	1.8E-16
GOTERM_MF_ALL	MHC class I receptor activity	25	2.0E-15
GOTERM_CC_ALL	Immunological synapse	27	1.2E-13
GOTERM_CC_ALL	MHC class I protein complex	27	1.2E-13
GOTERM_CC_ALL	MHC protein complex	27	1.2E-13
INTERPRO_NAME	MHC class I, alpha chain, C-terminal	15	5.2E-12
GOTERM_CC_ALL	Plasma membrane	30	9.3E-10
GOTERM_BP_ALL	Antigen processing, endogenous antigen via MHC class I	14	2.4E-8
GOTERM_BP_ALL	Antigen presentation, endogenous antigen	14	2.4E-8
GOTERM_CC_ALL	Protein complex	32	6.7E-4
Annotation Cluster 3	**Enrichment Score: 2.46**		
SP_PIR_KEYWORDS	Immune response	5	2.7E-4
SP_PIR_KEYWORDS	Immunoglobulin domain	6	2.2E-3
INTERPRO_NAME	MHC class II, alpha chain, N-terminal	5	4.8E-3
SP_PIR_KEYWORDS	MHC II	3	5.0E-2
Annotation Cluster 4	**Enrichment Score: 1.12**		
INTERPRO_NAME	Peptidase C1A, papain C-terminal	4	3.2E-3
INTERPRO_NAME	Peptidase C1A, papain	4	3.2E-3

Continued Table 9.1

Annotation Cluster 4	Enrichment Score: 1.12	Number of genes	P-Value
SP_PIR_KEYWORDS	Lysosome	4	6.7E-3
INTERPRO_NAME	Peptidase, cysteine peptidase active site	4	8.2E-3
SMART_NAME	Pept_C1	3	1.4E-2
INTERPRO_NAME	Proteinase inhibitor I29, cathepsin propeptide	3	1.8E-2
GOTERM_MF_ALL	Cysteine-type peptidase activity	4	7.7E-2
GOTERM_MF_ALL	Cysteine-type endopeptidase activity	4	7.7E-2
SP_PIR_KEYWORDS	Thiol protease	3	9.0E-2
SP_PIR_KEYWORDS	Zymogen	3	4.6E-1
GOTERM_MF_ALL	Endopeptidase activity	5	8.0E-1
SP_PIR_KEYWORDS	Protease	3	8.2E-1
SP_PIR_KEYWORDS	Hydrolase	8	8.7E-1
GOTERM_MF_ALL	Peptidase activity	5	9.2E-1
GOTERM_MF_ALL	Hydrolase activity	11	9.9E-1
Annotation Cluster 5	**Enrichment Score: 0.77**		
INTERPRO_NAME	Immunoglobulin subtype	6	6.7E-2
SMART_NAME	IG	5	2.4E-1
INTERPRO_NAME	Immunoglobulin V-set	3	3.2E-1
Annotation Cluster 6	**Enrichment Score: 0.37**		
GOTERM_BP_ALL	Response to wounding	5	8.4E-2
GOTERM_BP_ALL	Response to external stimulus	5	2.2E-1
GOTERM_BP_ALL	Development	5	6.6E-1
GOTERM_BP_ALL	Response to stress	6	7.2E-1
GOTERM_BP_ALL	Response to pest, pathogen or parasite	3	8.0E-1
GOTERM_BP_ALL	Response to other organism	3	8.2E-1

Continued Table 9.1

Annotation Cluster 7	Enrichment Score: 0.28	Number of genes	P-Value
UP_SEQ_FEATURE	Glycosylation site:N-linked (GlcNAc...)	14	3.9E-1
SP_PIR_KEYWORDS	Glycoprotein	16	5.3E-1
UP_SEQ_FEATURE	Disulfide bond	14	7.1E-1
Annotation Cluster 8	**Enrichment Score: 0.17**		
GOTERM_BP_ALL	Regulation of organismal physiological process	3	1.7E-1
GOTERM_BP_ALL	Cellular biosynthesis	6	9.1E-1
GOTERM_BP_ALL	Biosynthesis	6	9.6E-1
GOTERM_BP_ALL	Regulation of physiological process	8	9.8E-1
Annotation Cluster 8	**Enrichment Score: 0.17**		
GOTERM_BP_ALL	Regulation of biological process	8	9.9E-1
Annotation Cluster 9	**Enrichment Score: 0.12**		
UP_SEQ_FEATURE	Transmembrane region	10	6.3E-1
SP_PIR_KEYWORDS	Membrane	11	8.0E-1
SP_PIR_KEYWORDS	Phosphorylation	6	8.7E-1
Annotation Cluster 10	**Enrichment Score: 0.1**		
GOTERM_BP_ALL	Cell proliferation	3	2.5E-1
GOTERM_BP_ALL	Cell cycle	3	4.2E-1
GOTERM_BP_ALL	Development	5	6.6E-1
GOTERM_MF_ALL	Growth factor activity	3	8.5E-1
GOTERM_BP_ALL	Regulation of physiological process	8	9.8E-1
GOTERM_CC_ALL	Extracellular region	9	9.9E-1
GOTERM_BP_ALL	Regulation of biological process	8	9.9E-1
GOTERM_MF_ALL	Receptor binding	5	9.9E-1

Continued Table 9.1

Annotation Cluster 11	Enrichment Score: 0.08	Number of genes	P-Value
GOTERM_BP_ALL	Regulation of cellular physiological process	6	1.0E0
GOTERM_MF_ALL	Protein binding	9	1.0E0
GOTERM_BP_ALL	Regulation of cellular process	6	1.0E0
GOTERM_BP_ALL	Regulation of metabolism	3	1.0E0
INTERPRO_NAME	Calcium-binding EF-hand	4	5.5E-1
GOTERM_MF_ALL	calcium ion binding	5	7.1E-1
INTERPRO_NAME	EF-Hand type	3	8.1E-1
GOTERM_MF_ALL	Ion binding	7	1.0E0
GOTERM_MF_ALL	Metal ion binding	7	1.0E0
GOTERM_MF_ALL	Cation binding	6	1.0E0
Annotation Cluster 12	**Enrichment Score: 0.07**		
GOTERM_BP_ALL	Translation	3	2.2E-1
GOTERM_BP_ALL	Protein biosynthesis	5	6.4E-1
GOTERM_BP_ALL	Macromolecule biosynthesis	5	7.9E-1
KEGG_PATHWAY	Cytokine-cytokine receptor interaction	3	8.9E-1
GOTERM_BP_ALL	Cellular biosynthesis	6	9.1E-1
GOTERM_BP_ALL	Biosynthesis	6	9.6E-1
GOTERM_BP_ALL	Protein metabolism	10	9.9E-1
GOTERM_BP_ALL	Biopolymer metabolism	4	1.0E0
GOTERM_BP_ALL	Cellular protein metabolism	8	1.0E0
GOTERM_BP_ALL	Cellular macromolecule metabolism	8	1.0E0
GOTERM_BP_ALL	Macromolecule metabolism	10	1.0E0
GOTERM_BP_ALL	Primary metabolism	12	1.0E0
GOTERM_BP_ALL	Cellular metabolism	13	1.0E0
GOTERM_BP_ALL	Metabolism	16	1.0E0
GOTERM_BP_ALL	Cellular physiological process	21	1.0E0
GOTERM_BP_ALL	Cellular process	23	1.0E0

Continued Table 9.1

Annotation Cluster 13	Enrichment Score: 0.01	Number of genes	P-Value
GOTERM_MF_ALL	Structural molecule activity	3	9.3E-1
GOTERM_CC_ALL	Non-membrane-bound organelle	3	1.0E0
GOTERM_CC_ALL	Intracellular non-membrane-bound organelle	3	1.0E0
Annotation Cluster 14	**Enrichment Score: 0**		
GOTERM_BP_ALL	Regulation of physiological process	8	9.8E-1
GOTERM_BP_ALL	Regulation of biological process	8	9.9E-1
GOTERM_CC_ALL	Cytoplasm	8	1.0E0
GOTERM_CC_ALL	Nucleus	4	1.0E0
GOTERM_CC_ALL	Intracellular membrane-bound organelle	6	1.0E0
GOTERM_CC_ALL	Membrane-bound organelle	6	1.0E0
GOTERM_CC_ALL	Intracellular organelle	9	1.0E0
GOTERM_CC_ALL	Organelle	9	1.0E0
GOTERM_CC_ALL	Intracellular	13	1.0E0
Annotation Cluster 15	**Enrichment Score: 0**		
GOTERM_BP_ALL	Ion transport	4	9.9E-1
SP_PIR_KEYWORDS	Transport	3	1.0E0
GOTERM_MF_ALL	Transporter activity	3	1.0E0
GOTERM_BP_ALL	Establishment of localization	8	1.0E0
GOTERM_BP_ALL	Localization	8	1.0E0
GOTERM_BP_ALL	Transport	7	1.0E0

9.2 The functional annotations of down-regulated genes in inverted teat

Table 9.2 The functional annotations among 558 down-regulated genes in inverted teat

Annotation Cluster 1	Enrichment Score: 10.06	Number of genes	P-Value
GOTERM_CC_ALL	MHC class I protein complex	27	7.3E-23
GOTERM_CC_ALL	MHC protein complex	27	7.3E-23
GOTERM_CC_ALL	Immunological synapse	27	7.3E-23
INTERPRO_NAME	MHC class I, alpha chain, alpha1 and alpha2	25	1.8E-21
GOTERM_MF_ALL	MHC class I receptor activity	25	4.8E-21
INTERPRO_NAME	Immunoglobulin C1 type	27	2.7E-19
GOTERM_CC_ALL	Plasma membrane	29	2.0E-18
INTERPRO_NAME	Immunoglobulin/major histocompatibility complex	24	5.2E-17
INTERPRO_NAME	MHC class I, alpha chain, C-terminal	15	7.5E-15
GOTERM_CC_ALL	Protein complex	32	1.2E-12
GOTERM_BP_ALL	Antigen presentation	27	4.3E-12
GOTERM_BP_ALL	Antigen presentation, endogenous antigen	14	1.3E-11
GOTERM_BP_ALL	Antigen processing, endogenous antigen via MHC class I	14	1.3E-11
INTERPRO_NAME	Immunoglobulin-like	24	1.5E-11
SMART_NAME	IGc1	20	1.9E-11
GOTERM_BP_ALL	Immune response	28	1.5E-8
GOTERM_BP_ALL	Defense response	28	1.3E-7
GOTERM_BP_ALL	Response to biotic stimulus	28	1.8E-7
GOTERM_BP_ALL	Organismal physiological process	28	3.2E-6
GOTERM_MF_ALL	Transmembrane receptor activity	27	1.4E-5
GOTERM_BP_ALL	Response to stimulus	28	4.2E-5
GOTERM_BP_ALL	Antigen processing	14	5.5E-4
GOTERM_CC_ALL	Membrane	40	7.9E-4

Continued Table 9.2

Annotation Cluster 1	Enrichment Score: 10.06	Number of genes	P-Value
GOTERM_MF_ALL	Receptor activity	30	1.7E-3
GOTERM_BP_ALL	Physiological process	57	1.7E-2
GOTERM_CC_ALL	Cell	50	4.9E-2
GOTERM_MF_ALL	Signal transducer activity	32	6.1E-2
GOTERM_CC_ALL	Integral to membrane	21	2.0E-1
GOTERM_CC_ALL	Intrinsic to membrane	21	2.0E-1
Annotation Cluster 2	**Enrichment Score: 0.88**		
UP_SEQ_FEATURE	Domain:EGF-like 1	3	1.1E-2
SP_PIR_KEYWORDS	EGF-like domain	3	7.9E-2
SMART_NAME	EGF	3	1.4E-1
INTERPRO_NAME	EGF	3	1.9E-1
INTERPRO_NAME	EGF-like, type 3	3	2.5E-1
INTERPRO_NAME	EGF-like	3	3.5E-1
INTERPRO_NAME	EGF-like region	3	3.6E-1
Annotation Cluster 3	**Enrichment Score: 0.55**		
GOTERM_BP_ALL	Intracellular protein transport	3	1.6E-1
GOTERM_BP_ALL	Protein transport	3	2.3E-1
GOTERM_BP_ALL	Establishment of protein localization	3	2.5E-1
GOTERM_BP_ALL	Protein localization	3	2.7E-1
GOTERM_BP_ALL	Intracellular transport	3	2.8E-1
GOTERM_BP_ALL	Cellular localization	3	2.8E-1
GOTERM_BP_ALL	Establishment of cellular localization	3	2.8E-1
GOTERM_BP_ALL	Cell organization and biogenesis	3	7.6E-1
Annotation Cluster 4	**Enrichment Score: 0.54**		
SP_PIR_KEYWORDS	Glycoprotein	13	8.8E-2
SP_PIR_KEYWORDS	Membrane	11	1.1E-1
UP_SEQ_FEATURE	Glycosylation site:N-linked (GlcNAc...)	12	2.8E-1
UP_SEQ_FEATURE	Transmembrane region	9	4.5E-1

Continued Table 9.2

Annotation Cluster 5	Enrichment Score: 0.46	Number of genes	P-Value
UP_SEQ_FEATURE	Disulfide bond	11	7.1E-1
SP_PIR_KEYWORDS	Transmembrane	11	7.2E-1
GOTERM_MF_ALL	Solute	3	2.7E-2
GOTERM_MF_ALL	Symporter activity	3	7.6E-2
GOTERM_MF_ALL	Electrochemical potential-driven transporter activity	3	1.9E-1
GOTERM_MF_ALL	Porter activity	3	1.9E-1
GOTERM_MF_ALL	Transporter activity	11	2.5E-1
GOTERM_MF_ALL	Cation transporter activity	7	3.0E-1
GOTERM_BP_ALL	Metal ion transport	4	3.7E-1
GOTERM_MF_ALL	Ion transporter activity	7	5.0E-1
GOTERM_BP_ALL	Transport	14	5.7E-1
GOTERM_BP_ALL	Establishment of localization	14	6.0E-1
GOTERM_BP_ALL	Localization	14	6.1E-1
GOTERM_BP_ALL	Cation transport	4	6.4E-1
GOTERM_BP_ALL	Ion transport	5	7.3E-1
GOTERM_MF_ALL	Carrier activity	4	7.8E-1
GOTERM_BP_ALL	Cellular physiological process	26	1.0E0
GOTERM_BP_ALL	Cellular metabolism	14	1.0E0
GOTERM_BP_ALL	Metabolism	17	1.0E0
GOTERM_BP_ALL	Cellular process	29	1.0E0
Annotation Cluster 6	**Enrichment Score: 0.41**		
UP_SEQ_FEATURE	Nucleotide phosphate-binding region:ATP	3	2.5E-1
SP_PIR_KEYWORDS	Nucleotide-binding	5	3.3E-1
SP_PIR_KEYWORDS	ATP-binding	3	6.9E-1
Annotation Cluster 7	**Enrichment Score: 0.36**		
SP_PIR_KEYWORDS	Calcium	6	4.2E-2
INTERPRO_NAME	Calcium-binding EF-hand	4	2.7E-1
SMART_NAME	EFh	3	3.0E-1
GOTERM_MF_ALL	Calcium ion binding	4	5.5E-1

Continued Table 9.2

Annotation Cluster 7	Enrichment Score: 0.36	Number of genes	P-Value
INTERPRO_NAME	EF-Hand type	3	5.8E-1
GOTERM_MF_ALL	Iron ion binding	3	6.2E-1
GOTERM_MF_ALL	Cation binding	8	7.0E-1
GOTERM_MF_ALL	Metal ion binding	8	7.6E-1
GOTERM_MF_ALL	Ion binding	8	7.6E-1
GOTERM_MF_ALL	Transition metal ion binding	4	8.7E-1
Annotation Cluster 8	**Enrichment Score: 0.33**		
SP_PIR_KEYWORDS	Metal-binding	7	3.4E-1
GOTERM_MF_ALL	Heme binding	3	4.5E-1
GOTERM_MF_ALL	Tetrapyrrole binding	3	4.5E-1
SP_PIR_KEYWORDS	Heme	3	5.2E-1
SP_PIR_KEYWORDS	Iron	3	6.1E-1
Annotation Cluster 9	**Enrichment Score: 0.33**		
GOTERM_BP_ALL	Translation	4	1.2E-2
GOTERM_BP_ALL	Protein biosynthesis	5	2.2E-1
GOTERM_BP_ALL	Biosynthesis	8	2.4E-1
GOTERM_BP_ALL	Nitrogen compound metabolism	3	2.7E-1
GOTERM_BP_ALL	Cellular biosynthesis	7	2.8E-1
GOTERM_BP_ALL	Macromolecule biosynthesis	5	3.5E-1
GOTERM_BP_ALL	Organic acid metabolism	3	4.6E-1
GOTERM_BP_ALL	Carboxylic acid metabolism	3	4.6E-1
GOTERM_BP_ALL	Biopolymer metabolism	5	8.4E-1
GOTERM_BP_ALL	Cellular protein metabolism	7	9.2E-1
GOTERM_BP_ALL	Cellular macromolecule metabolism	7	9.3E-1
GOTERM_BP_ALL	Protein metabolism	7	9.4E-1
GOTERM_BP_ALL	Macromolecule metabolism	9	9.8E-1
GOTERM_BP_ALL	Primary metabolism	14	1.0E0
GOTERM_MF_ALL	Catalytic activity	22	1.0E0
GOTERM_BP_ALL	Cellular metabolism	14	1.0E0
GOTERM_BP_ALL	Metabolism	17	1.0E0

Continued Table 9.2

Annotation Cluster 10	Enrichment Score: 0.19	Number of genes	P-Value
GOTERM_MF_ALL	GTP binding	3	4.2E-1
GOTERM_MF_ALL	Guanyl nucleotide binding	3	5.0E-1
GOTERM_MF_ALL	Nucleotide binding	6	8.8E-1
GOTERM_MF_ALL	Purine nucleotide binding	5	9.3E-1
Annotation Cluster 11	Enrichment Score: 0.05		
GOTERM_CC_ALL	Organelle membrane	3	5.5E-1
GOTERM_CC_ALL	Cytoplasm	9	8.2E-1
GOTERM_CC_ALL	Nucleus	4	9.9E-1
GOTERM_CC_ALL	Intracellular	14	9.9E-1
GOTERM_CC_ALL	Intracellular membrane-bound organelle	8	9.9E-1
GOTERM_CC_ALL	Membrane-bound organelle	8	9.9E-1
GOTERM_CC_ALL	Intracellular organelle	10	1.0E0
GOTERM_CC_ALL	Organelle	10	1.0E0

Curriculum vitae

1. Personal information

Full name	: Saowaluck Yammuen-Art
Date of birth	: August 14th, 1973
Place of birth	: Tak
Nationality	: Thai
Marital status	: Single

2. Educational background

1992 to 1995:   Chiangmai University, Bachelor of Science (Agriculture)

1996 to1999:   Chiangmai University, Master of Science (Agriculture, Animal Science);
Title of Special Problem: Supplementation Effect of Leucaena Leaves to Rice Straw and Ruzi Grass on Rumen Degradation Determined by Nylon Bag Technique
Title of Thesis: Evaluation of Net Energy and Digestibility of Rice Straw in Dairy Cattle and Sheep

Since 2005:   Scientific Associate at the Institute of Animal Science, University Bonn (Agriculture, Animal Science)

3. Working Experience

April to June 1999   Assistant Researcher; Project "Production and Utilization of good quality corn silage in total mixed ration and the nutrient requirement of high producing cows" Department of Animal Science Faculty of Agriculture, Chiangmai University Financial Support by Thailand Research Fund (TRF)

July 1999 to August 2001	Researcher; Project "Corn silage and total mixed ration (TMR) production for cooperative" SME project Department of Animal Science Faculty of Agriculture, Chiangmai University Financial Support by Thailand Research Fund (TRF) and Chaiprakarn Agriculture Cooperative
September 2001 to August 2005	Instructor at the Rajamangala University of Technology Thunyaburi, Patumthani Campus
September 2005 to present	Scientific Associate at the Institute of Animal Science, University of Bonn, Germany

4. Publications

Cheva-Isarakul, B., **S. Yammuen-art**, C. Nuammuan and K Vasupen. 1997. The prediction of nutrition value of rice straw, ruzi grass and leucaena leaves, common feed for dairy cattle, using nylon bag technique.pp 1-15 *In*: Recent research for the development of nutrient requirements of Thai dairy cattle. Department of animal science Faculty of agriculture Chiangmai University, Chiangmai, THAILAND.

Cheva-Isarakul, Bl., **S. Yammuen-art**, Bs. Cheva-Isarakul and S. Promma. 1999. Energy values *in vivo* digestibility and dry matter intake of rice straw determined by regression method in dairy cows and sheep. pp. 93-101. *In*: The 37th Kasetsart University annual conference. Kasetsart University, Bangkok, THAILAND.

Yammuen-art, S., Bl. Cheva-Isarakul, Bs. Cheva-Isarakul and S. Promma. 1999. Digestibility and energy of rice straw evaluated by *in sacco* and *in vitro* gas production. pp. 76-85. *In*: The 37th Kasetsart University annual conference. Kasetsart University, Bangkok, THAILAND.

Cheva-Isarakul, Bl., Bs. Cheva-Isarakul, S. Promma, K Vasupen, I. Paopaisan and **S. Yammuen-art**. 1999. Evaluation of energy content in dry roughage for dairy cow I Soybean plant and pod husk sugar cane and rice straw. pp. 25-28. The 3th research and development for dairy cow and product conference (Abstract). Thailand research fund and national center for genetic engineering and biotechnology, Bangkok THAILAND.

Yammuen-art, S., Bl. Cheva-Isarakul, Bs. Cheva-Isarakul and S. Promma. 2000. Digestibility and calculated net energy of sweet corn stover silage in dairy cows. pp. 163-174. *In*: The 38th Kasetsart University annual conference. Kasetsart University, Bangkok THAILAND.

Promma, S., Bl. Cheva-Isarakul, **S. Yammuen-art** and Bs. Cheva-Isarakul. 2000. The production of total mixed ration (TMR) containing corn silage for dairy on commercial. *In*: Proceeding of animal science conference, Indochina agriculture fair. Ubonrachathni University, Ubonrachathni, THAILAND.

Cheva-Isarakul, Bl., S. Promma, Bs. Cheva-Isarakul and **S. Yammuen-art**. 2000. The Production of good quality corn silage for commercial of cooperative. *In*: Proceeding of animal science conference, Indochina agriculture fair. Ubonrachathni University, Ubonrachathni, THAILAND.

Cheva-Isarakul, Bs., S. Promma**, S. Yammuen-art** and Bl. Cheva-Isarakul. 2000. The production of corn silage and total mixed ration (TMR) for dairy cows on commercial. *In*: The 3th annual conference. Maejo University, Chiangmai, THAILAND.

Cheva-Isarakul, Bl., S. Promma , Bl. Cheva-Isarakul and **S. Yammuen-art**. 2000. The effect of formalin supplement on the quality of corn silage and total mixed ration (TMR). *In*: The 3th annual conference. Maejo University, Chiangmai, THAILAND.

Yammuen-art, S., S. Promma, Bs. Cheva-Isarakul and Bl. Cheva-Isarakul. 2000. The effect of formic and propionic acid on the secondary fermentation of corn silage and total mixed ration (TMR). In: The 3th annual conference. Maejo University, Chiangmai, THAILAND.

Yammuen-Art S, Phatsara C, Wimmers K, Ponsuksili S, Tholen E, Schellander K, Jonas E (2007): Association of the Pituitary-Specific Transcription Factor-1 and Leptin genes with traits related to the teat development in pigs. 13th International Conference on Production Diseases in Farm Animals (ICPD), 29.7.-4.8.2007, Leipzig, Germany (Abstr)

Yammuen-Art S, Jennen D, Jonas E, Phatsara C, Ponsuksili S, Tesfaye D, Tholen E, Wimmers K, Schellander K (2007): Characterization of Relaxin-3/INSL7 and its receptor being candidate genes for the inverted teat defect in pigs. 58th Meeting of the European Association for Animal Production (EAAP), 26.-29.8.2007, Dublin, Ireland (Proc, Abstr)

Jonas E, Chomdej S, **Yammuen-art S**, Phatsara C, Schreinemachers H-J, Jennen D, Tesfaye D, Ponsuksili S, Wimmers K, Tholen E, Schellander K (2007): Verification of chromosomal regions affecting the inverted teat development and their derivable candidate genes in pigs. 58th Meeting of the European Association for Animal Production (EAAP), 26.-29.8.2007, Dublin, Ireland (Proc, Abstr)

Yammuen-art S, Ponsuksili S, Jonas E, Phatsara C, Wimmers K, Tesfaye D, Tholen E, Jüngst H, Schellander K (2007): Analysis of functional candidate genes for the mammary gland defect in pigs. Vortragstagung der DGfZ und der GfT, 26./27.9.2007, Stuttgart Hohenheim, Deutschland (Proc)

Phatsara C, Jonas E, **Yammuen-art S**, Buschbell H, Tesfaye D, Ponsuksili S, Tholen E, Jüngst H, Schellander K, Wimmers K (2007): Physiological interactions between the endocrine and immune systems shown in gene analysis in pigs. In: Proceeding of the International Symposium of Animal Genomics for Animal Health, 23.-25.10.2007, Paris, France (Poster/Abstr)

Yammuen-art S, Jonas E, Wimmers K, Ponsuksili S, Tesfaye D, Tholen E, Jüngst H, Phatsara C, Schellander K (2008): Comparison of the gene expression profile in normal and inverted teats in lactating sows. Third International Symposium on Animal Functional Genomics (ISAFG), 7.-9.4.2008, Edinburgh, Scotland, Proc, 49 (Abstr)

Yammuen-art S, , E. Jonas, K. Wimmers, S. Ponsuksili, D. Tesfaye, M. Hölker, E. Tholen, H. Jüngst, K. Schellander, C. Phatsara (2008): Differential expression of porcine *EGF* and *EGFR* mRNA in normal and inverted teats at different stages of mammary gland development. Vortragstagung der DGfZ und der GfT, 17./18.9.2008, Bonn, Deutschland (Proc)

5. Research topics

Rajamangala University of Technology Thunyaburi Patumthani Campus, Patumthani THAILAND

'Comparison on Cork Tree Leaves (*Millingtonia hortensis*), *Alpinia galanga (Linn.)Swartz*, Turmeric (*Curcuma longa* Linn.) and Holy basil Leaves (*Ocimum sanctum* Linn.) on Production Performance in Broiler' (2003)

'Effect of Tomato Pomace as an Antioxidation on Growth Performance of Broiler' (2004)

FUGATO – **F**unktionelle **G**enom **A**nalyse im **T**ierischen **O**rganismus / Functional Genome Analysis in Animal Organisms Supported project HeDiPig (2004 to 2008).

'Identifizierung der ursächlich an Erbdefekten beteiligten Gene beim Schwein durch Integration struktureller und funktioneller Genomik / Elucidating the aetiology of hereditary diseases in the pig by integrating structural and functional genomics ; heritable defects in pigs'