From QTL to QTN

Identification of a Quantitative Trait Nucleotide Influencing Muscle Development and Fat Deposition in Pig

Anne-Sophie Van Laere
Veterinary Faculty
Department of Animal Breeding and Genetics
Uppsala

Doctoral thesis
Swedish University of Agricultural Sciences
Uppsala 2005
Abstract


Most traits of economical importance in animal production are quantitative i.e. they are characterized by a continuous variation of phenotypic values. Examples for such traits are carcass weight, milk production and lean meat content. The phenotype of an animal for a quantitative trait depends on its genotype at several loci (called quantitative trait loci, QTL) as well as on environmental factors. Up to date, a large number of QTLs have been identified in farm animals by segregation analysis either within commercial populations or in crossbreed populations. Animal geneticists face now the challenge to identify the causative mutations lying behind these QTLs.

In this thesis, we report the identification of the causative mutation for a major QTL influencing muscle development, fat deposition and heart size in pig. Previous studies have mapped this locus to the distal end of pig chromosome 2p. Furthermore, they have hypothesized that the causative mutation(s) may lie in an element regulating the expression of insulin-like growth factor 2 (IGF2). Firstly, we sequenced the IGF2 region in the pig and made comparative sequence analysis with available human and mouse sequences. We then used an identity-by-descent approach and managed to pinpoint the causative mutation to a G⇒A transition located in an evolutionary conserved CpG island in IGF2 intron 3 (IGF2-intron3-G3070A). Subsequently, we used electrophoretic mobility shift assay and transient transfection experiments and showed that the QTN (quantitative trait nucleotide) abrogates the binding of a putative repressor. We completed our study by determining the core binding site of this trans-acting factor and by performing DNase I footprinting of the CpG island containing the QTN. In addition, we identified an IGF2 antisense transcript (IGF2-AS) and showed that its expression was also influenced by the QTN.

The discovery of mutations causing QTLs in farm animals opens great future prospects. Besides evident practical breeding interests there are also major scientific interests, as understanding the mechanism causing the QTL effects will broaden our general knowledge on how the genome operates.

Keywords: antisense transcript, CpG island, quantitative trait locus, quantitative trait nucleotide, repressor, Sus scrofa.

Author’s address: Anne-Sophie Van Laere, Department of Animal Breeding and Genetics, SLU, BMC, Box 597, S-751 24 UPPSALA, Sweden. Email: Anne-Sophie.Van-Laere@bmc.uu.se
Contents

Introduction, 7
I. Chromosomal location of IGF2, 8
II. Effects of IGF2, 8
   Receptors, 8
   Binding proteins, 9
   Biological actions, 9
III. Imprinting, 10
   Definition, 10
   Transmission of imprints, 10
   (Epi)genetic characteristics of imprinted genes, 11
   Regulation of expression at the IGF2 locus, 13

Aims of the thesis, 15

Methods, 16
I. Transient transfection, 16
II. Electrophoretic mobility shift assay, 16
III. DNase I footprinting, 17

Results and discussion, 19
I. Comparative sequence analysis of the INS-IGF2-H19 gene cluster in pigs (Paper 1), 19
II. A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig (Paper 2), 20
III. IGF2 antisense transcript expression in porcine postnatal muscle is affected by a quantitative trait nucleotide in intron 3 (Paper 3), 22
IV. Molecular Characterization of a Region in IGF2 Intron 3 harbouring a Quantitative Trait Nucleotide affecting Muscle Growth in the Pig (Paper 4), 24

Future prospects, 26

References, 27

Acknowledgements, 31
Appendix

Papers I-IV

The present thesis is based on the following papers, which will be referred to by their Roman numerals:


IV. Van Laere, A.S., Andersson, G., Kindmark, A. & Andersson, L. Molecular Characterization of a Region in IGF2 Intron 3 harbouring a Quantitative Trait Nucleotide affecting Muscle Growth in the Pig. (Manuscript).

Papers I-III are reproduced by permission of the journals concerned.
Introduction

Most traits of economical importance in animal production are quantitative *i.e.* they are characterized by a continuous variation of phenotypic values. Examples for such traits are carcass weight, milk production and lean meat content. The phenotype of an animal for a quantitative trait depends on its genotype at several loci (called quantitative trait loci, QTL) as well as on environmental factors (Andersson, 2001). Up to date, a large number of QTLs have been identified in farm animals by segregation analysis either within commercial populations or in crossbreed populations. Animal geneticists face now the challenge to identify the causative mutations lying behind these QTLs. The major obstacle is the poor precision in the location of those loci. Indeed, the complex relation between genotype and phenotype complicates the detection of recombinants between markers and QTL as the genotype of an individual can only be determined by progeny testing. In addition, the nature of QTL mutations might complicate their identification. QTLs are not responsible for disorders but only for mild variation in phenotypic value and are therefore expected to be caused by a variant gene product or an altered gene expression rather than by a defect in gene product or in gene expression. The causative mutation(s) can hence be regulatory or structural and might be extremely difficult to distinguish from neutral linked mutations (Georges & Andersson, 1996).

During the last decades, overweight and metabolic disorders have been increasing in western countries. As a consequence, the demand on “lighter” more healthy products has also increased. This led the pig industry to select for animals with higher lean muscle and reduced fat deposition. This selection goes in opposite direction to the one occurring on wild boars. Indeed, natural selection favours animals that can store energy (*i.e.* fat), as those will be able to survive periods of starvation. A three-generation intercross was made between wild boars and Large White domestic pigs in an attempt to discover QTLs responsible for the differences in growth and fat deposition observed between those animals (Andersson *et al.*, 1994; Andersson-Eklund *et al.*, 1998). This successful approach led to the discovery of several QTLs, including one influencing muscle development, fat deposition and heart size. This locus maps to the distal end of pig chromosome 2p (SSC2p) and has the particularity of being imprinted (maternally silenced). Early studies (Jeon *et al.*, 1999; Nezer *et al.*, 1999) have suggested *IGF2* (insulin-like growth factor 2) as a candidate gene for this QTL because of:
- its chromosomal location,
- its paternal-specific expression,
- its effect on myogenesis.
I. Chromosomal location of IGF2

The chromosomal location of IGF2 in pigs was unknown at the time the QTL was discovered. However, it was suspected to co-localize with the QTL because of its position in the human genome. Indeed, IGF2 was known to map to human chromosome 11p15.5 and bidirectional chromosome painting had shown that HSA11pter-q13 corresponds to SSC2p (Goureau et al., 1996). A FISH (fluorescent in situ hybridisation) experiment was consequently set up to confirm the assignment of IGF2 to SSC2p. In this experiment, a porcine BAC clone containing IGF2 was hybridized to porcine metaphase chromosomes and gave a consistent signal on the distal end of chromosome 2p (band 2p1.7) (Jeon et al., 1999). This confirmed that IGF2 and the QTL both mapped to the distal end of pig chromosome 2.

II. Effects of IGF-II

Insulin-like growth factor II (IGF-II) is a 67 amino acid-long, single chain polypeptide belonging to the insulin family. This family also includes insulin and insulin-like growth factor I (IGF-I). The genes coding for these three proteins are orthologs (i.e. they have evolved from a common ancestral gene) and are the result of two duplication events. The first duplication occurred approximately 600 million years ago and gave rise to insulin and a common ancestor for the two insulin-like growth factors. This ancestor then led to the genes coding for IGF-I and IGF-II after an additional duplication event that took place around 300 million years ago (Froesch et al., 1985).

IGF-II and insulin show 47% sequence identity at the amino acid level. Furthermore, they have the same three-dimensional structure since they have the same three interchain disulphide bridges and hydrophobic core (O’Dell & Day, 1998).

Receptors

IGF-II exerts its biological effects through three receptors:

IGF-I receptor

The IGF-I receptor binds IGF-I with highest affinity but it binds also IGF-II (with 2-15 times lower affinity) and insulin (with 100-500 lower affinity). It is present in a large variety of tissues where it mediates most of the effects of both IGF-I and IGF-II (Cohick & Clemmons, 1993).

IGF-II receptor

The IGF-II receptor has a high affinity for IGF-II. It can also bind IGF-I but with a 100 to 500 times lower affinity and it does not bind insulin at all. This receptor is mainly known for its clearance role; hence, it internalizes IGF-II upon binding and transports it to the lysosomes for degradation (Jones & Clemmons, 1995). In addition, it has been demonstrated to mediate part of the physiological actions of IGF-II e.g. on myosarcoma cell motility (Minniti et al., 1992) and on extravillous
trophoblast cell migration (McKinnon et al., 2001). The IGF-II receptor possesses two binding sites for mannose-6-phosphate (Man-6-P) in addition to its IGF-II binding site and is therefore also known as cation-independent mannose-6-phosphate receptor. These Man-6-P sites mediate the transport of lysosomal enzymes from the Golgi apparatus to the pre-lysosomes and the endocytosis of ligands containing Man-6-P e.g. thyroglobulin.

Insulin receptor
The insulin receptor binds both insulin and IGF-II, but it has a ten times lower affinity for IGF-II compared to insulin. It was shown to mediate part of the growth promoting function of IGF-II in human and mouse fetus (Louvi, Accili & Efstratiadis, 1997). The insulin and IGF-I receptors are structurally highly similar heterotetrameric glycoproteins composed of two alpha and two beta subunits (αβ). Hybrid insulin/IGF-I receptors composed of one αβ IGF-I half receptor and one αβ insulin half receptor have even been found on cells expressing both types of receptors (Jones & Clemmons, 1995).

Binding Proteins
More than 99% of circulating IGFs are bound by Insulin-like Growth Factor Binding Proteins (IGFBP) (Dupont et al., 2003). Up till now, six IGFBP (IGFBP1-6) have been described. They are characterized by conserved amino- and carboxy-terminal but each of them has a unique central domain. Their main role is to modulate the biological effects of the IGFs by (1) maintaining a reservoir of IGFs in circulation, (2) transporting IGFs across the capillary membrane, (3) localizing the IGFs to specific tissues, (4) modulating binding of the IGFs to their receptors and (5) prolonging the half-life of the IGFs (Wood, 1995). In addition, they have also been shown to have various IGF-independent actions e.g. as growth modulators (Mohan & Baylink, 2002).

Biological actions
IGF-II acts both through endocrine and autocrine / paracrine pathways and has been shown to:
- Promote feto-placental growth: IGF-II has metabolic, mitogenic and differentiative actions on a wide range of fetal tissues and on the placenta (Jones & Clemmons, 1995). Experiments using transgenic mice have proven that IGF-II is a potent fetal growth factor. DeChiara, Robertson & Efstratiadis (1990) showed, for example, that knockout Igf2 mice weigh only 60% of the normal weight at birth.
- Promote both cell proliferation and cell differentiation. Hence, Florini and co-workers (1991) demonstrated that autocrine secretion of IGF-II plays a major role in skeletal muscle cell differentiation. Oksbjerg, Gondret & Vestergaard (2004) reported that, in muscle cells, the stimulation of proliferation and differentiation by the IGFs is concentration- and time-dependant.
- Prevent apoptosis: this has been shown e.g. in cultures of myoblast, neurons and oligodendrocytes (Jones & Clemmons, 1995)
- Mediate insulin-like effects e.g. on glucose and fat metabolism (Jones & Clemmons, 1995).
- Increase cell motility (and hence malignancy) in myosarcoma (Minniti et al., 1992), to increase migration of extravillous trophoblastic cells (McKinnon et al., 2001).

In addition, over-expression of IGF-II has been shown to cause cell hyperproliferation associated with tumour formation (Wood, 1995).

### III. Imprinting

#### Definition

Genomic imprinting has been defined as “an epigenetic modification that is parental-origin specific, and/or preferential expression of a specific parental allele in somatic cells of the offspring” (Feinberg, Cui & Ohlsson, 2002). The term “epigenetic” literally means outside conventional genetics (Jaenisch & Bird, 2003); thus, epigenetic modifications are modifications of the chromatin (e.g. histone acetylation, DNA methylation) without modification of the DNA sequence (Wilkins & Haig, 2003). Those modifications are heritable through many cell divisions but can also be reset (at least in germline).

#### Transmission of imprints

The exact nature of the primary epigenetic modification(s) responsible for the establishment of imprinting is still unknown. Nevertheless, Li and co-worker’s (1993) study on knockout mice has proven that methylation is necessary at least for maintaining imprinting. The imprints causing parent-of-origin specific expression have to be reset at each generation in order to correspond to the germline of the new individual. Hence, imprints go through a three-step life cycle (Reik & Walter, 2001a):

**Erasure**

This first step occurs in the primordial germ cells. Imprints inherited from the parents are removed and DNA is totally unmethylated. Nuclear transplantation experiments in mouse have shown that Igf2 is silenced at this stage whereas H19 is expressed (Labosky et al., 1994).

**Establishment**

The new imprints specific to the germline (oocyte or sperm) are set up at a late fetal stage in males and after birth in females. De novo methylation takes place and results in overall higher methylation in male germ cells than in oocytes.

**Maintenance**

The new imprints have to be transmitted to both daughter cells at mitoses. This is more challenging than it first appears as the new imprints have to resist the genome-wide demethylation occurring after fertilization and the de novo methylation taking place after implantation.
(Epi)genetic characteristics of imprinted genes

Clusters
Eighty percents of imprinted genes are found in clusters. Genes linked in the same cluster are believed to be co-regulated (Reik & Walter, 2001a), notably through Imprinting Control Regions (ICR). These are CpG-rich cis-acting elements that are found associated to approximately 50% of the known imprinted genes and are essential for the correct imprinting to occur (Fergusson-Smith & Surani, 2001). ICRs can be up to several kilobases long and are differentially methylated (usually the maternally-derived ICR is methylated) (Delaval & Feil, 2004).

CpG islands
CpG islands are not an exclusive characteristic of imprinted genes but imprinted genes are much more often associated with CpG islands than non-imprinted genes (88% versus 47% in mouse) (Reik & Walter, 2001a). Gardiner-Garden and Frommer (1987) defined CpG islands as DNA stretches fulfilling the three following criteria:
- length > 200 bp,
- G + C content > 50%,
- Observed CpG / Expected CpG > 0.6.

Direct repeats
The presence of tandem direct repeats associated to GC-rich sequences is a common characteristic among many imprinted genes. These repeats have been suggested to attract the methylation machinery by mimicking foreign DNA structure. Indeed, DNA methylation has been proposed to have evolved to protect the host against the spreading of transposons and endogenous retroviruses. The methylation and subsequent heterochromatization of the tandem repeats could lead to spreading of methylation to the nearby GC-rich region. Nevertheless, deletion experiments realized with H19 transgenes have shown that the tandem repeats alone are not sufficient to cause allele-specific methylation (Reik & Walter, 2001a).

Except for their association with imprinted genes, the repeats themselves do not have much in common. Their sequence, number of repetitions, length, position relatively to the gene, position relatively to the CpG island or DMR (differentially methylated region) varies. Consequently, if they are involved in the acquisition and/or the maintenance of differential methylation they would probably act through their organization. This could be done by:
- Influencing the DNA secondary structure
- Being recognized by protein complexes, e.g. methyltransferase

DNA methylation
DNA methylation has been shown to be a key element to maintain imprinting (Brannan & Bartolomei, 1999; Tilghman, 1999). In eukaryotes, methylation occurs on the carbon at position 5 of cytosines found in CpG dinucleotides (and much more rarely in CpGpG trimucleotides) (Strachan & Read, 1999).
The resulting 5-methylcytosines are unstable and tend to deaminate into thymines. This phenomenon has resulted in a decrease of the frequency of CpG dinucleotides in the genome over time so that the actual observed frequency only corresponds to 23% and 19% of the expected frequency in human and mouse, respectively (Fazzari & Greally, 2004).

The majority of known imprinted genes have been shown to contain differentially methylated regions (DMR). These CpG-rich regions can be methylated on the active or silenced allele and can contain various types of regulatory elements like enhancers, repressors and chromatin boundaries. Differential methylation results in allele-specific gene expression by modifying the protein-DNA interactions. Indeed, the addition of a methyl group to the cytosine modifies the aspect of the major groove of the DNA (which contains most of the DNA-protein recognition sites) (Constância et al., 1998; Fazzari & Greally, 2004) and can consequently:

- prevent binding of transcription factors,
- allow methyl CpG binding proteins (e.g. MeCP1, MeCP2) to bind. These proteins bind specifically to methylated DNA and mediate silencing through histone deacetylation and subsequent chromatin condensation (Jones, 1999; Jaenish & Bird, 2003).

**Antisense transcripts**

Fifteen percent of known imprinted genes have an antisense transcript. Amazingly, this antisense gene is also imprinted and (almost) always maternally silenced, whether the sense transcript is maternally or paternally expressed (Reik & Walter, 2001a). The hypothesis that the antisense transcript is important for the regulation of the sense gene was recently proven for Air (antisense Insulin-like growth factor 2 receptor) (Sleutels, Zwart & Barlow, 2002) and Kcnqlotl (antisense Kcnq1) (Thakur et al., 2004).

It is amazing to notice that if (almost) all imprinted antisense genes are paternally transcribed, the majority of the DMRs are maternally methylated. Reik & Walter (2001b) have linked these observations to the fact that the genome-wide demethylation occurring after fertilization is active on the paternally inherited chromosomes but passive on the maternally inherited ones. Hence, they suggested that this active demethylation could be an attempt from the mother’s side to remove paternal imprints. Paternal imprints would then have evolved towards another type of silencing mechanism *i.e.* antisense transcripts.

**Asynchronous replication**

Kitsberg and co-workers (1993) studied the timing of replication of imprinted genes by *in situ* hybridization to interphase nuclei. They showed that replication of imprinted genes is asynchronous and even allele-specific as the paternal allele always replicates before the maternal allele. Hence, Kitsberg and co-workers suggested that the different replication time is a necessary imprint to establish
allele-specific gene expression (e.g. by changing the accessibility of the DNA for methyltransferases).

Meiotic recombination
The observation that regions actively transcribed during gametogenesis are more prone to recombination led Thomas and Rothstein (1991) to the hypothesis that the sex-specific recombination frequencies observed at certain places of the genome might be caused by sex-specific gene expression e.g. imprinting. Paldi, Gyapay & Jami (1995) suggested that chromatin is organized in higher-order structures that are responsible for:
- Asynchronous replication.
- Different frequency of meiotic recombination between sexes: actively transcribed regions have a more “open” chromatin structure which allows more recombinations (possibly because enzymes initiating crossing-overs have a better access to the DNA).
- Imprinting: genes display allele-specific expression if:
  - They have specific signals in their sequence.
  - Modifying enzymes can access those signals in one sex but not in the other because of the different chromatin environment.

Regulation of expression at the IGF2 locus
IGF2 is part of a cluster of imprinted genes located on HSA11p15.5 (corresponding to SSC2p1.7). Two ICRs (imprinted control regions) regulate the imprinting of these genes. The first one, Kv ICR, controls the imprinting of the centromeric subcluster which contains KCNQ1, KCNQ1OT1 and CDKN1C. The second one, H19 ICR, is located 2-4 kb upstream of the H19 promoter and controls imprinting at the telomeric subcluster which contains H19 and IGF2 (Du et al., 2003). These two genes are reciprocally imprinted so that in most tissues H19 is maternally expressed (Bartolomé, Zemel & Tilghman, 1991) and IGF2 is paternally expressed (DeChiara, Robertson & Efstratiadis, 1991; Nezer et al., 1999). Furthermore, Li et al. (1998) suggested that the human H19 gene is an antagonist of IGF2 expressivity in trans. The mechanisms controlling imprinting at the IGF2-H19 domain are complex and appear to be tissue-specific. In the endoderm, expression of IGF2 and H19 depends on activation of their promoters by a set of shared enhancers located 3’ of H19 (Leighton et al., 1995). On the maternal chromosome, the unmethylated H19 ICR is bound by CCCTC-binding factors (CTCF). This creates a chromatin boundary which isolates the IGF2 promoters from the endodermal enhancers and results in silencing of IGF2 and expression of H19. On the paternal chromosome, methylation of the ICR prevents CTCF from binding which results in activation of the IGF2 promoters by the endodermal-specific enhancers and IGF2 transcription (Bell & Felsenfeld, 2000; Hark et al., 2000; Kanduri et al., 2000a, b). In addition, the methylated ICR directs methylation and subsequent silencing of the H19 promoter (Srivastava et al., 2000) (Figure 1). The situation appears to be more complex in mesodermal tissues. In addition to mesodermal-specific enhancers located 3’ of H19 (Ishihara et al., 2000), a series of other control elements have also been found as e.g. a silencer located in DMR1 (IGF2 intron 3) (Eden et al., 2001) and a muscle-
specific silencer situated in the IGF2-H19 intergenic region (Ainscough et al., 2000).

At imprinted loci, the level of transcription of the transcribed allele is still controlled by transcription factors like at non-imprinted loci. Most information available on the regulation of IGF2’s four promoters (P1-4) (Sussenbach 1989; van Dijk et al., 1991) comes from studies made in different human liver cell lines. Hence, promoter P1 has been shown to be activated by the ubiquitous transcription factor Sp1 (Rodenburg, Holthuizen & Sussenbach, 1997) and by the CCAAT/enhancer binding protein (C/EBP) (van Dijk et al., 1992). Promoters P3 and P4 have been shown to be regulated by the zinc finger transcription factors Egr-1 and WT1. These two proteins bind to the same DNA element but binding of Egr-1 results in transcriptional activation whereas binding of WT1 results in transcriptional repression (Bae et al., 1999; Lee et al., 1998). In addition, Rietveld et al. (1999) have demonstrated that transcription from promoter P3 responds to AP-2 binding so that overexpression of AP-2 results in activation of P3 in cells with low endogenous level of AP-2 and repression of P3 in cells with high endogenous level of AP-2. P3 has also been shown to be activated by the zinc finger oncogene PLAG1 (Zatkova et al., 2004). Finally, Sp1 was shown to bind to promoter P4 and to cooperate with Egr-1 to mediate maximal activity of this promoter (Lee, Park & Lee, 2001). The regulation of IGF2 transcription in skeletal muscle cells has been less studied and is poorly understood. However, Erbay et al. (2003) have demonstrated that the Ser/Thr kinase mTOR initiates myoblast differentiation by regulating the expression of IGF2. In addition, Zhang et al. (1998) have suggested that AP-2 may contribute to IGF-II overexpression in an embryonal skeletal muscle tumor (rhabdomyosarcoma).
Aims of the thesis

The objectives of this thesis were:
- To identify the causative mutation(s) for a major QTL in the pig influencing muscle growth, fat deposition and heart size located on SSC2p.
- To characterize the molecular mechanism(s) through which the mutation exerts its effects.
Methods

I. Transient transfection

This method can be used to determine if a DNA element is involved in transcriptional regulation of gene expression e.g. if it acts as promoter or enhancer. Firstly, the element of interest is cloned in a plasmid containing a reporter gene which expression can be easily assayed (e.g. luciferase, green fluorescent protein, chloramphenicol acetyl transferase). If the element is a putative promoter, it will be inserted in a plasmid containing a strong enhancer (e.g. SV40 enhancer). On the other hand, if the experimenter wants to test a supposed enhancer or silencer, it will be cloned in a plasmid containing a promoter. In this case, it is generally recommended to use the homologous promoter. Secondly, the plasmid is transfected into a suitable cell line. This can be done by a biochemical (e.g. cation lipid, calcium phosphate), physical (e.g. electroporation) or virus-mediated method. It is important to simultaneously transform the cells with a control vector (expressing a different reporter) to be able to monitor differential cell growth and transfection efficiency. Finally, the activity of the reporter is assayed after one to four days incubation and the observed reporter signal is normalized to the control.

II. Electrophoretic Mobility Shift Assay (EMSA)

EMSA is a standard biochemical *in vitro* method to detect protein-DNA interactions (Fried & Crothers, 1981). This assay is based on the fact that migration of DNA through a native polyacrylamide gel is retarded upon protein binding. First, a short double stranded DNA fragment (called the probe) is radioactively labelled. Second, the probe is incubated with proteins to allow DNA-protein complexes to form. Proteins from diverse origins can be used in EMSA e.g. nuclear or whole-cell extracts from cells or tissues (Dignam, Lebovitz & Roeder, 1983) and purified recombinant proteins. Third, the protein-DNA binding reactions are electrophoresed on a native polyacrylamide gel to separate free and protein-bound DNA. After autoradiography, the band corresponding to the DNA-protein complex appears higher on the gel compared to the free probe; the probe has also been “shifted” (Fig. 2). The migration of the DNA-protein complex depends mainly on the charge, shape and multimeric state of the protein. The specificity of the obtained complexes has to be tested by the addition of an excess of cold probe to the binding reaction. If the protein binds specifically it has the same affinity for the cold and for the radioactively labelled probe. Hence, both probes will compete for its binding and as the cold probe is in large molar excess there will be no protein left to bind to the labelled probe, which results in the disappearance of the complex.
The major advantages of EMSA are its simplicity and its high sensitivity. In addition, it allows the detection of complexes of different compositions, each complex appearing as a band with a specific migration. Furthermore, EMSA gives the possibility to check the identity of proteins included in a complex by using antibodies. Hence, a specific antibody can be added to the protein-DNA binding reaction and its binding to the protein will result in an antibody-protein-DNA complex which gel migration will be even more retarded (this is called a supershift). Alternatively, the antibody can cover the DNA binding site of the protein and thereby prevent the formation of the complex, resulting in the disappearance of the shifted band (Carey & Smale, 2000).

**III. DNase I Footprinting**

DNase I footprinting allows the detection of protein-DNA interactions *in vitro* (Galas and Schmitz, 1978). This method is based on the principle that DNA regions bound by proteins are protected from digestion by DNase I. Basically, a double-stranded DNA probe corresponding to the region of interest is radioactively labelled on one end and used to set up two parallel reactions: one with proteins, the other without. After incubation, a specific amount of DNase I is added to both reactions so that each DNA molecule is cut only once. DNase I cuts randomly, and in the absence of proteins the probe will be digested in a series of labelled fragments ranging from one bp to full length probe. On the other hand, if DNA-protein complexes form, the DNA bound by the proteins will not be accessible to
the enzyme and this will result in the absence of DNA fragments of specific sizes. Both reactions are then run on a denaturing polyacrylamide gel to separate the DNA fragments according to their length. After autoradiography of the gel, the probe incubated without proteins will appear as a continuous series of bands. The probe incubated with proteins will also appear as a series of bands, but if protein-DNA complexes have formed the areas corresponding to the complexes will be devoid of bands. Those regions are called “footprints” (Carey & Smale, 2000). Usually, a Maxam-Gilbert sequencing reaction of the probe is run together with the DNase I digestions to enable the localization of the footprints (Fig. 3). The main advantage of this method is that it gives the approximate binding site of each protein binding to the probe. Furthermore, it is possible to analyse a quite long DNA region in a single experiment (the probes are generally at least 300 bp long).

Figure 3 Hypothetical DNase I footprinting experiment allowing the detection of two protected regions. Footprint 1 appears as a region devoid of bands whereas in footprint 2 the bands are only weakened compared to the ones from the probe incubated without proteins. A ladder (which is usually a Maxam- Gilbert A+G sequencing reaction) is electrophoresed together with the digested probes to allow localization of the footprints.
Results and discussion

I. Comparative sequence analysis of the INS-IGF2-H19 gene cluster in pigs (Paper 1)

The aim of this work was to further characterize the region containing the mutation(s) causing the QTL. Here we report the sequence analysis of two pig contigs. The first one is 32 kb long and contains the five last exons of TH (Tyrosine hydroxylase) as well as the entire INS (insulin) and IGF2 genes. The second one contains H19 and covers 56 kb.

We started by characterizing the order and the structure of INS, IGF2 and H19 in pig and showed that they were identical to the ones in human. Hence, the gene order is as follows: TH - 1.9 kb - INS - 0.7 kb - IGF2 - 88.1 kb - H19. IGF2 is composed of ten exons (1-9 and 4b) that display high sequence identity between human and pig (Fig. 4).

![Genomic structure of porcine IGF2](image)

Figure 4. Genomic structure of porcine IGF2. The ten exons of IGF2 are represented as boxes, black boxes correspond to the translated exons. The four promoters of IGF2 (P1-4) are represented by arrows.

Nezer et al. (1999) showed that the coding region of IGF2 was identical between pigs with different QTL genotypes. We therefore suspected the causative mutation(s) to lie in (a) regulatory element(s). Such elements tend to be well conserved between species. This results from natural selection as individuals carrying mutations in a regulatory element might display erratic gene expression and lower fitness. Consequently, we compared our pig sequence with available human and mouse sequences to find these conserved regions. We report 59 evolutionary conserved elements (outside exons, promoters and simple repeats) in the INS-IGF2 region and 38 in the H19 region. Most of them have an unknown function but some have been assigned an important role in regulating the expression of IGF2 in human and mouse e.g. DMR1, CTCF binding sites, endodermal enhancers (see introduction).

As expected from phylogenetic studies, the overall sequence similarity was higher between pig and human than between pig and mouse or human and mouse. The pig sequence displays an amazingly large number of CpG islands: nine in the INS-IGF2 region and sixteen in the H19 region. This can be put into relation with the imprinting of the region. Indeed, CpG islands are more often found associated with imprinted than with non-imprinted genes. The sequence is also characterized by its low abundance of interspersed repeats. Once again, this could be related to
the imprinting of IGF2 and H19 as the introduction of foreign sequence might perturb the complex regulatory mechanisms controlling their expression.

Another important part of this work was to characterize IGF2 transcripts and promoter usage in different fetal and adult tissues. We found that IGF2 transcription is tissue- and development-specific and that it can be initiated from four promoters (P1-4 located upstream of exons 1, 4, 5 and 6, respectively).

II. A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig (Paper 2)

Nezer and co-workers (2003) refined the position of the QTL to a 250 kb-long interval between the markers 370SNP6/15 and SWC9 (located in the 3’ untranslated region of IGF2). The only known paternally expressed genes mapping to this region were insulin and IGF2. Therefore, we decided to re-sequence 28.6 kb covering these two genes on 15 chromosomes which QTL status could be determined by progeny testing and marker-assisted segregation analysis. One of the chromosomes (H254) appeared to be recombinant and allowed us to localize the QTL downstream of the first exon of IGF2. Among the 258 polymorphisms differentiating the 15 chromosomes, we only found one SNP co-segregating perfectly with the QTL status of the chromosome. Therefore, this SNP, a G→A transition at position IGF2-intron 3-nt 3072, has to be the causative mutation. The wild type allele (G) is associated with lower muscle mass and was therefore named "q" while the mutant allele (A) causes higher muscle development and was called "Q". The quantitative trait nucleotide (QTN) is located in an evolutionary conserved CpG island of unknown function. Consequently, we set up EMSA and transient transfection experiments to uncover its mechanism of action. In addition, we studied the methylation status of the CpG island by bisulphite sequencing.

We carried out EMSA with nuclear extracts from three different cell types (C2C12 murine myoblasts, HepG2 human hepatocytes and HEK 293 human embryonic kidney cells) and three different 27 bp-long probes:

- q: wild-type probe
- Q: mutant probe
- q*: wild-type probe with a methylated CpG at the QTN. As methylation is important for expression of imprinted genes we designed this probe in order to test the influence of methylation on in vitro binding to the QTN.

We demonstrated the existence of a specific complex forming only with the wild-type probe but not with the mutant probe nor with the methylated probe.

We transfected C2C12 myoblast cells with reporter plasmids expressing firefly luciferase under the control of the thymidine kinase minimal promoter (TK) and a 578 bp-long fragment corresponding to the q or Q genotype at the QTN. After normalization, we found that the q insert doubles the basal TK transcription whereas the Q insert increases it seven times.
Figure 5 Results of transient transfection experiments carried out with reporter plasmids expressing firefly luciferase under the control of TK and an insert corresponding to the q (q+TK) or Q (Q+TK) allele at the QTN. Results were normalised to a Renilla luciferase control plasmid and are expressed as relative activities to a plasmid expressing firefly luciferase under the sole control of TK. The triple asterisk indicates that the differences observed between the three plasmids are highly significant (P<0.01).

These results were quite difficult to conciliate with the results of the EMSA. For this reason, we replaced the TK promoter with IGF2 promoter 3 (P3) and repeated the experiment. We choose P3 because it is the most actively transcribed promoter in muscle cells and because it is influenced by the QTN in vivo (see below). This time, we found that q reduces the basal P3 transcription with 70%, whereas Q only reduces it with 30%.

Figure 6 Results of transient transfections carried out like in Fig. 6 but plasmids express luciferase under the control of the P3 instead of TK.

Taken together with the EMSA these new results suggest that the QTN abrogates the binding of a repressor to a cis-element. In addition, our transient transfection experiments illustrate how important it is to use a homologous rather then a heterologous promoter in this kind of experiments.

We analyzed the methylation of the CpG island containing the QTN and found that it is independent from the genotype at the QTN and from the parental origin of the allele. However, we found that it is tissue-specific. Hence, on average, 26% of the CpG dimucleotides are methylated in liver but only 3.4% are methylated in skeletal muscle. Interestingly, the effect of the QTN is observed in muscle i.e. in a
non-methylated tissue where the putative repressor is able to bind, but no effect is seen in liver which is more methylated and where the putative repressor might consequently not be able to bind so efficiently (according to the EMSA results). It would be very exciting to analyze the situation in other tissues to check if this association between methylation status and QTN effect holds.

We quantified the expression of IGF2 in vivo and found a significantly higher expression in postnatal muscle samples from QQ and Q\textsuperscript{Qq}Q\textsuperscript{Qq} animals compared to q\textsuperscript{q}Q\textsuperscript{q} and qq animals. A weaker (but significant) difference could also be observed in postnatal heart samples but not in postnatal liver nor in any tested prenatal tissue sample. Furthermore, we showed that IGF2 transcription was increased from all three promoters located downstream of the QTN (i.e. P2-4).

Finally, we genotyped the progeny of 13 heterozygous sires (Qq) and of 50 homozygous sires (QQ or qq) and used this data in segregation analyses. We found evidence for segregation in all heterozygous families but we could not find any indication of segregation among progeny sired by homozygous males. Furthermore, Jungerius et al. (2005) showed that the QTN also controls the QTL for backfat thickness found in a Meishan x European Whites cross. In conclusion, we demonstrated that the SSC2p QTL is caused by a G→A transition at position IGF2-intron 3(nt 3072 and that this mutation influences IGF2 expression. Hence, we showed that, in addition to its well-known fetal role, IGF2 is involved in postnatal muscle development.

III. IGF2 antisense transcript expression in porcine postnatal muscle is affected by a quantitative trait nucleotide in intron 3 (Paper 3)

The aim of this study was to search for an IGF2 antisense (IGF2-AS) gene in pig, and upon its existence to:
- Characterize and quantify its transcript(s).
- Examine its imprinting status.
- Determine whether its expression was influenced by the IGF2 QTN.

We have shown by RT-PCR and RNase protection assay (RPA) that IGF2-AS indeed exists in pig. Furthermore, we have shown that it has two different transcription start sites; a major site located around IGF2 intron4-nt70 and a minor site located approximately at IGF2 intron3-nt2294 (RPA results) or intron3-nt2205 (5’ RACE results). The 3’ end of the transcripts was mapped to position IGF2 intron2–nt1236 by RACE. We found three different transcripts originating at the major start site. These transcripts contain from three to five exons and share all the same first and last exons (Fig. 7).
Figure 7. Genomic structure of porcine IGF2 and multiple IGF2-AS transcripts (a-d). Exons are represented as boxes, black boxes correspond to the translated exons. Promoters are represented by arrows.

Hence, we can conclude that the structure of IGF2-AS is not well conserved between pig, human and mouse as this gene only has three exons in human and four in mouse. However, some of the exons are quite well conserved between species (Table 1).

Table1. Sequence identities between IGF2-AS exons in pig and human or mouse.

<table>
<thead>
<tr>
<th>Pig exon</th>
<th>Human exon</th>
<th>Mouse exon</th>
<th>% Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>72</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>3</td>
<td>60</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>63</td>
</tr>
</tbody>
</table>

Northern blot analyses revealed the existence of three transcripts in fetal muscle (4.7 kb, 3.3 kb and 2.1 kb), two in fetal liver (3.5 kb and 2.1 kb) and one in fetal kidney (3.3 kb). Surprisingly, the shortest band observed on the northern blot is longer than the longest transcript predicted by the RT-PCR and 3’ RACE results. This could result from preferential amplification of short truncated transcripts by the nested RACE PCR or from a real heterogeneity of the transcripts.

Next, we examined the imprinting status of IGF2-AS in liver and muscle samples from fetal, 3-weeks- and 4-months-old piglets. This was done by sequencing an A to C transversion at IGF2-AS exon2-nt32 which allowed us to discriminate between Q and q alleles at the QTN. We found that IGF2-AS is imprinted and only expressed from the paternal allele. This reflects the status at other imprinted loci as most antisense transcripts found in imprinted genes are maternally silenced. However, we should note that in muscle of 4-months-old pigs
we could detect some transcription from maternal origin which indicates that imprinting is partially released.

Finally, we used real-time PCR analysis to quantify *IGF2-AS* transcripts in muscle and liver from fetal, 3-weeks- and 4-months-old pigs carrying the q or Q allele at the QTN. Firstly, we found that the expression of *IGF2-AS* decreases noticeably after birth. Secondly, we found that in 3-weeks- and 4-months-old pig muscle it depends on the genotype at the QTN. Indeed, at these stages, *IGF2-AS* expression was significantly higher in Q than in q muscle samples. Hence, the putative repressor binding at the QTN seems to influence both *IGF2* and *IGF2-AS* expression.

The function of *IGF2-AS* is still unknown, but it has been suggested that it could take part in the regulation of *IGF2* expression. Indeed, it is noteworthy that in pig as well as in human, the first and the last exon of *IGF2-AS* overlap part of *IGF2* exon 4 and the entire exon 3, respectively. Consequently, *IGF2-AS* could interfere with transcripts originating from P1 and P2.

**IV. Molecular Characterization of a Region in *IGF2* Intron 3 harbouring a Quantitative Trait Nucleotide affecting Muscle Growth in the Pig (Paper 4)**

The aim of this study was to:
- Determine the binding site of the transcription factor binding to the QTN.
- Characterize the CpG island containing the QTN.
- Search for polymorphisms in the CpG island in human.

We used EMSA to determine which nucleotides were important for the binding of the putative repressor described in paper 2. Firstly, we performed EMSA with a series of mutated probes to determine if these mutations could abolish formation of the specific complex obtained with the wild-type probe. Secondly, we used the mutated oligonucleotides in a competition assay to determine if they were still able to compete against the wild-type probe. Taking these results together, we found that the core binding site of the transcription factor is: 5'-GCTCG-3'. New database searches did not reveal any factor with similar binding capacities.

DNase I footprinting of the 333 bp surrounding the QTN revealed two protected regions, FP1 and FP2:
- FP1 covers ~50 bp (from nucleotide position ~3027 to ~3076) and includes the core binding site of the putative repressor and a perfect consensus AP-2 binding site.
- FP2 is a ~30 bp-long footprint located ~20 bp upstream of FP1 (from nucleotide position ~2981 to ~3008) and covers a putative Sp1 binding site.

AP-2 has previously been shown to regulate *IGF2* transcription from P3 and it would be particularly interesting to confirm it’s binding to FP2. Even more interesting would be to know if it binds DNA as a heteromer with the putative repressor. Hence, if they bind together, one could try to purify the repressor by co-immunoprecipitating it with AP-2.
Comparative sequence analyses of the CpG island containing the QTN have revealed a high sequence conservation between pig and human. This led us to the hypothesis that the human DNA sequence could contain cis-elements involved in muscle development, as is the case in pig. Therefore, we resequenced this region in individuals with low versus high muscle mass. We discovered three SNPs and one insertion/deletion. Interestingly, one of the mutations (C→T, at position IGF2-intron 3-3462) lies in the putative AP-2 site included in FP2. Consequently, we analysed the in vitro binding capacity of this site by running EMSA with a wild-type and a mutant probe of the region. We were able to detect binding of specific complexes with similar mobility using both wild-type and mutant sequences. It would be very interesting to complement this experiment by a supershift assay with AP-2-specific antibodies to confirm the involvement of this transcription factor.
Future prospects

This thesis summarizes the work that led to the identification of the causative mutation for a QTL influencing muscle development, fat deposition and heart size in pig. In addition, it presents how we have started to uncover the molecular mechanisms by which this QTN mediates its effects. We have shown that the mutation abrogates the binding of a putative repressor element, which results in increased expression of $IGF2$ and $IGF2-AS$. We have also demonstrated that other transcription factors bind DNA in the immediate vicinity of this repressor. The obvious next step in this project will be to clone and characterize this repressor. Different methods could be considered to achieve this goal e.g. one-hybrid screen, in vitro expression library screening, biochemical purification. In order to choose the most appropriate approach, it is important to collect as much information as possible on this transcription factor and on its possible interactions with the neighbouring DNA-binding factors. It is for example essential to know if it binds DNA by itself or as a heteromer (as could be suspected from the DNase I footprinting results). Indeed, some of the methods cited above are only capable to deal with proteins binding as monomer, homodimer or homopolymer. In addition, identifying a known factor binding DNA together with the repressor could be extremely useful; antibodies against this factor could be used to co-immunoprecipitate both factors as a first step towards purification of the repressor.

The identification of the gene coding for a transcription factor might be a difficult task, but it would be worth the effort as this repressor could have fascinating therapeutic uses in the future. Indeed, it could be inactivated to increase the expression of $IGF2$ in specific tissues. Our study has demonstrated that it is active in skeletal muscle and heart but not in liver. In addition, it may also be active in adipocytes as the $IGF2$-QTL also influences backfat thickness. Hence, one could imagine to inactivate it in order to increase the muscle mass of patients suffering from muscular degenerative diseases. Furthermore, it could be transiently inactivated in patients confined in bed to avoid the muscle loss accompanying long periods of inactivity. In addition, if its activity is confirmed in adipocytes, it could even be used to treat obese patients.

Finally, the identification of the causative mutation for the $IGF2$-QTL makes it very easy to select pigs with the favourable Q allele to breed for the next generations and hence to fix the mutation in populations were it is present. In addition, it will also facilitate the introgression of the mutant allele in pig breeds where it is absent. Indeed, when the introgression of a favourable gene variant is based on linked markers there is always a risk to loose it because of a recombination between the markers and the actual causative mutation. This problem has now been eliminated and makes the $IGF2$-QTL a very attractive candidate for introgression into certain commercial populations.
References


Rietveld, P.A., Barlow, D.P. & Hogan, B.L. 1994. Mouse embryonic germ (EG) cell lines: transmission through the germline and differences in the methylation imprint of insulin-like growth factor 2 receptor (Igf2r) gene compared with embryonic stem (ES) cell lines. Development 120, 3197-3204.


ACKNOWLEDGEMENTS

I would like to take thank the following persons:

My supervisors:
Leif Andersson. Thank you for giving me the chance to work in the fascinating field of molecular genetics; but most of all, thank your enthusiasms, your optimism and for the great independence you gave me.
Göran Andersson. Thank you for always taking the time to answer my questions and for making me see the world from your transcriptional point of view.

Valérie Amarger. C'est toi qui m'as mis le pied à l'étrier, merci!

The past and present members of the lab:
Gabriella Lindgren, Carl-Gustaf Thulin, Stefan Marklund, Siw Johansson, Elisabella Giuffra, Jin-Tae Jeon, Anna Törnsten, Göran Hjälm, Anna-Karin Fridolfsson, Maria Moller, Örjan Carlborg, Gabriella Nyman, Teresa Casals, Jessica Petersson
Amélie Johansson. I am looking forward to the next ISAG conference.
Ana Fenandez and Meiying Fang. You saved me each time I was looked out of my office.
Carolyn Fitzsimmons. I will miss these Ph.D. student parties.
Emmanuelle Bourneuf. You have a great sense of humor (et merci d’avoir sauvé mon gel!)
Erik Bongcam-Rudloff. You saved my computer from falling out through the window.
Frida Berg. Where are you taking this amazing energy from?
Lina Jacobsson and Susanne Kerje. I would never have made it to the Big Apple without you.
Martin Braunschweig. The IGF2 group was not the same anymore after you left
My brother, Ben. It was great having you here for four month (même si Voldo s’est pris quelque raclées).
My office-mate, Hee-Bok Park. I enjoyed sharing the office with you and the “little guy”, good luck in Malmö.
Nicolette Sammon Hillbertz. I will miss the hugs, the week-ends at the countryside, the dogs and all the rest.
Sonchita Bagchi. You are a real sunshine.
The first person I met when I arrived, Gerli Pielberg. You have been a real friend during these five years. Thank you for defending your thesis two month before me in order to figure out all the administrative mess!
The new generation of Ph.D. students: Ullrika Gunnarsson, Per Wahlberg and Ellen Markljung. Time is flying so fast that you will soon be writing your own little book; enjoy pipetting as long as it lasts!
Ulla Gustafson and Gudrun Wieselander. The lab would not survive without you.
All the Swedes. Tack för att ni hjälpte mig att lära mig språket!
All the people in my “adoptive” labs:
Cihan Cetinkaya, Natalie von der Lehr and Fuad Bahram. I will never forget that
the cells are “my children”.
Susanna Tronnersjö. You made me discover the fascinating world of the budding
yeasts.
Beat Bornhauser. Thank you for always having the time to show me how to use
the nice machines you have down there.

The AgriFunGen program at SLU,
I am grateful for my founding, but also for all the interesting courses and for the
great meetings!

Ma famille et mes amis.
Merci pour tous ces bons moments passés ensemble pendant mes vacances et pour
vos encouragements pendant le restant de l’année.