QTL Analysis in the Pig

From the Identification of Quantitative Trait Loci to the Understanding of Molecular Mechanisms

ELLEN MARKLJUNG
Dissertation presented at Uppsala University to be publicly examined in B41, BMC, Husargatan 3, Uppsala, Friday, March 13, 2009 at 13:15 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English.

Abstract

Domestic pigs have become very different form the wild ancestors they originate from. Selection for muscle growth and meat quality has made the pig a good model for genetic studies of muscle development.

The first part of this thesis presents a genome-wide scan for quantitative trait loci (QTL) in a cross between Landrace and Hampshire pigs. Traits such as body composition, fat deposition, body length, meat quality and weight measurements of individual muscles were investigated. In total we identified 15 different QTLs that reached genome-wide significance. The three most significant QTLs were for carcass length on chromosome 17 and two overlapping QTLs on chromosome 1 for body composition and weight of \textit{M. biceps femoris}, respectively. A strong candidate gene for the body composition QTL is melanocortin 4 receptor (\textit{MC4R}). We also identified several QTLs for sizes of different muscles, fat deposition and meat quality traits.

In a previous study using a cross between the domestic Large White and wild boar, the mutation underlying a major QTL for muscle growth and fat deposition was identified as a single nucleotide substitution (QTN) in intron 3 of the \textit{IGF2} gene. The QTN disrupts the binding of a repressor affecting \textit{IGF2} mRNA expression. In the second part of this thesis, the identification of the repressor is presented. The repressor, named ZBED6, is a previously unknown mammalian member of the BED-domain protein family. We could show that Zbed6 specifically binds the wild-type but not the mutated sequence surrounding the QTN. Further studies of silenced \textit{Zbed6} in the mouse myoblast cell line C2C12 showed that it represses transcription in a luciferase reporter assay and affects \textit{Igf2} mRNA transcription and proliferation. ZBED6 shows very high sequence conservation and has a broad tissue distribution of expression suggesting that ZBED6 also has important biological function outside the muscle cell.

Keywords: pig, quantitative genetics, QTL, QTN, muscle development, meat quality, IGF2, ZBED6, Landrace, Hampshire

\textit{Ellen Markljung, Department of Medical Biochemistry and Microbiology, Box 582, Uppsala University, SE-75123 Uppsala, Sweden}

\copyright\ Ellen Markljung 2009

ISSN 1651-6206
ISBN 978-91-554-7412-6
urn:nbn:se:uu:diva-9556 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-9556)
To Martin
List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.


Reprints were made with permission from the respective publishers.
**Contents**

Introduction................................................................................................... 11

*Sus scrofa* ........................................................................................................ 11
  Domestication of the pig ........................................................................ 11
  Pig as a model for muscle development ............................................ 11
Muscle development .................................................................................. 12
  Myogenesis ..................................................................................... 12
  Muscle fiber types ........................................................................ 13
  IGFs in muscle development .......................................................... 14
Transcriptional regulation ......................................................................... 16
  Epigenetics ................................................................................ 16
  Epigenetic regulation of the *IGF2/H19* locus .................................... 19
Monogenic and quantitative traits .............................................................. 21
  Linkage mapping ...................................................................... 21
  QTL analysis ........................................................................ 23

Aim of this thesis .......................................................................................... 27

Present investigations .................................................................................. 29
  QTL analysis of a cross between Hampshire and Landrace (paper I and II)
  Background ........................................................................ 29
  Results and Discussion ................................................................ 31
  Future prospects ..................................................................... 35
Identification of *ZBED6*, a novel mammalian repressor regulating *IGF2*
expression in muscle (paper III)................................................................. 36
  Background ........................................................................ 36
  Results and Discussion .............................................................. 38
  Future prospects ................................................................ 41

Acknowledgments ........................................................................................ 45

References..................................................................................................... 47
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3C</td>
<td>Chromosome conformation capture</td>
</tr>
<tr>
<td>4C</td>
<td>Circular chromosome conformation capture</td>
</tr>
<tr>
<td>ADP</td>
<td>Adenosine diphosphate</td>
</tr>
<tr>
<td>AIL</td>
<td>Advance intercross line</td>
</tr>
<tr>
<td>AMPK</td>
<td>Adenosine monophosphate activated protein kinase</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>BEAF</td>
<td>Boundary element–associated factor</td>
</tr>
<tr>
<td>CFTR</td>
<td>Cystic fibrosis conductance receptor</td>
</tr>
<tr>
<td>ChIP</td>
<td>Chromatin immunoprecipitation</td>
</tr>
<tr>
<td>cM</td>
<td>Centi Morgan</td>
</tr>
<tr>
<td>CpG</td>
<td>Cytosine-phosphate-guanine</td>
</tr>
<tr>
<td>CTCF</td>
<td>CCCTC-binding factor</td>
</tr>
<tr>
<td>DIP</td>
<td>DNA immunoprecipitation</td>
</tr>
<tr>
<td>DMR</td>
<td>Differentially methylated region</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DNMT</td>
<td>DNA methyltransferase</td>
</tr>
<tr>
<td>DREF</td>
<td>DNA replication-related element factor</td>
</tr>
<tr>
<td>EMSA</td>
<td>Electrophoretic mobility shift assay</td>
</tr>
<tr>
<td>ES cells</td>
<td>Embryonic stem cells</td>
</tr>
<tr>
<td>GH</td>
<td>Growth hormone</td>
</tr>
<tr>
<td>HLA</td>
<td>Human leukocyte antigen</td>
</tr>
<tr>
<td>ICR</td>
<td>Imprinting control region</td>
</tr>
<tr>
<td>IGF</td>
<td>Insulin like growth factor</td>
</tr>
<tr>
<td>IGF2-AS</td>
<td>IGF2 antisense</td>
</tr>
<tr>
<td>IGF2R</td>
<td>Insulin like growth factor 2 receptor</td>
</tr>
<tr>
<td>IGFBP</td>
<td>Insulin like growth factor binding protein</td>
</tr>
<tr>
<td>IP</td>
<td>Immunoprecipitation</td>
</tr>
<tr>
<td>LCMS</td>
<td>Liquid chromatography mass spectrometry</td>
</tr>
<tr>
<td>LOI</td>
<td>Loss of imprinting</td>
</tr>
<tr>
<td>MAPK</td>
<td>Mitogen-activated protein kinase</td>
</tr>
<tr>
<td>MAR3</td>
<td>Matrix attachment region 3</td>
</tr>
<tr>
<td>MAS</td>
<td>Marker assisted selection</td>
</tr>
<tr>
<td>Mb</td>
<td>Mega base pairs</td>
</tr>
<tr>
<td>MC4R</td>
<td>Melanocortin 4 receptor</td>
</tr>
<tr>
<td>MHC</td>
<td>Myosin heavy chain</td>
</tr>
<tr>
<td>MHC</td>
<td>Major histocompatibility complex</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>miRNA</td>
<td>Micro RNA</td>
</tr>
<tr>
<td>MLC</td>
<td>Myosin light chain</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger RNA</td>
</tr>
<tr>
<td>Myf5</td>
<td>Myogenic factor 5</td>
</tr>
<tr>
<td>MyoD</td>
<td>Myogenic differentiation antigen</td>
</tr>
<tr>
<td>ncRNA</td>
<td>Non-coding RNA</td>
</tr>
<tr>
<td>PAGE</td>
<td>Polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PI3K</td>
<td>Phosphatidylinositol-3-kinase</td>
</tr>
<tr>
<td>PRC2</td>
<td>Polycomb recessive complex 2</td>
</tr>
<tr>
<td>PRKAG3</td>
<td>Protein kinase, AMP-activated, gamma 3 non-catalytic subunit</td>
</tr>
<tr>
<td>QTL</td>
<td>Quantitative trait locus</td>
</tr>
<tr>
<td>QTN</td>
<td>Quantitative trait nucleotide</td>
</tr>
<tr>
<td>RFLP</td>
<td>Restriction fragment length polymorphism</td>
</tr>
<tr>
<td>RN</td>
<td>Rendement Napole</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RYR</td>
<td>Ryanodine receptor</td>
</tr>
<tr>
<td>SELEX</td>
<td>Systematic evolution of ligands by exponential enrichment</td>
</tr>
<tr>
<td>SILAC</td>
<td>Stable isotope labeling by amino acid in cell culture</td>
</tr>
<tr>
<td>siRNA</td>
<td>Small interfering/silencing RNA</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>SSC</td>
<td><em>Sus scrofa</em> chromosome</td>
</tr>
<tr>
<td>Suz12</td>
<td>Suppressor of zeste 12</td>
</tr>
<tr>
<td>ZBED</td>
<td>Zinc finger, BED-type containing</td>
</tr>
<tr>
<td>ZC3H11A</td>
<td>Zinc finger CCCH type containing 11A</td>
</tr>
</tbody>
</table>
Introduction

Sus scrofa

Domestication of the pig
The pig was first domesticated ~9000 years ago in the Near East. Studies of mitochondrial sequence data have shown that domestication occurred at multiple locations throughout Europe and Asia\(^1,2\). The European wild boar is the principal ancestor of the European domestic pigs\(^2,3\).

Once domesticated, human-mediated selection began. Phenotypic variants, already present in the natural population or new occurring variants, which normally would have been under negative or no selection, were observed and selected for by humans, thereby increasing their frequency in the domesticated population. One of the first traits selected for was coat color\(^4\). Besides coat color, the domestic pig has been selected for traits such as behavior, growth, fertility and meat production traits. In Britain, organized meat production started during the 18\(^{th}\) century which led way for the selected breeding of production traits\(^5\). Introgression of the Chinese domestic pig occurred in the European domestic pig breeds during the 19\(^{th}\) century, which introduced traits such as increased fatness and reduced age and size at maturity\(^5,6\). During this time and forward, until the middle of the 20\(^{th}\) century, the domestic pigs were selected for fatness\(^5\). But after the Second World War consumers changed their preferences and demanded lean meat while the production industry required increased growth rates\(^5\). Present day breeds, such as Yorkshire (or Large White), Landrace, Duroc and Hampshire have all been selected for lean meat growth in their recent history.

Pig as a model for muscle development
The pig is used for meat production throughout the world except for most Muslim countries. There are over 730 breeds in the world although five breeds dominate the meat production industry \textit{i.e.} Large White, Duroc, Landrace, Hampshire and Piétrain\(^7\). The genetic diversity across breeds is high as shown by the allelic diversity present at the major histocompatibility complex locus (MHC)\(^8\). The recent selection for lean meat in the pig breeds has increased allele variants for muscle growth and reduced those variants
promoting fat deposition. Therefore, there are probably many loci affecting muscle development in pig that has been under strong selection recently. By studying muscle growth in pig we can obtain knowledge about muscle development. We can identify new genes and/or learn more about how the genes involved in muscle development are regulated. Mutations affecting muscle growth in the pig have so far been identified in three different genes, demonstrating that pig is a well-suited model for studying muscle development. A single nucleotide substitution affecting postnatal muscle growth in pig was identified in intron 3 of the \textit{IGF2} gene\textsuperscript{9}. The mutation increased skeletal muscle mass by 3-4\% and enlarged the heart. IGF2 was known to affect muscle development, but the study identified an unknown regulatory region of the gene involved in regulating muscle-specific expression of \textit{IGF2}\textsuperscript{9}. Another study identified a mutation in the \textit{ryanodine receptor gene} (\textit{RYR1}) causing the recessive disorder malignant hyperthermia\textsuperscript{10}. The mutation also increased lean muscle growth and had therefore been under selection in many pig breeds\textsuperscript{10}. The third example is a missense mutation in the \textit{PRKAG3} gene. The gene encodes a previously unknown muscle-specific isoform of the regulatory \(\gamma\) subunit of the adenosine monophosphate activated protein kinase (AMPK)\textsuperscript{11}. The mutation affected the energy metabolism in muscle where mutants showed \(~70\%\) increase in muscle-stored glycogen compared with the wild-type\textsuperscript{11}. Introducing the same mutation into mice and also creating a knockout gave further insight into the regulation of energy metabolism in muscle. The mutant mouse showed protection against developing insulin resistance when exposed to a high fat diet, demonstrating that it could be a drug target for treatment of type 2 diabetes\textsuperscript{12}.

**Muscle development**

This thesis has a major focus on using pig as a model for muscle development. I will therefore give a brief introduction to muscle development with focus on muscle development in pig and the involvement of the insulin-like growth factors.

**Myogenesis**

Myogenesis begins during early embryogenesis from mesodermal cells. Cells committed to the myogenic lineage express the transcription factors myogenic differentiation antigen (MyoD or Myf-3) and myogenic factor 5 (Myf-5), both belonging to the basic helix-loop-helix group of proteins\textsuperscript{13}. Myoblasts proliferate and subsequently exit the cell cycle and fuse into multinucleated myotubes or myofibers\textsuperscript{13}. Two waves of muscle fiber formation take place during the embryonic development. First, the primary fibers form early during embryogenesis\textsuperscript{13,14}. Later, at about the same time as inner-
vation of the embryo occurs, the secondary fibers form using the primary fibers as a scaffold to grow around\textsuperscript{13,14}. In human and large animals such as the pig, there is also a third generation of muscle fibers. In human this occurs between 3.5 to 4 months of gestation and in pig in the early post-natal period\textsuperscript{15,16}. The total number of myofibers formed in a specific muscle during the prenatal period seems to be fixed some time late in gestation, consequently it is unclear if the third generation of muscle fibers affect the total number of myofibers\textsuperscript{13,17}. The number of myofibers affects the muscles capacity to grow postnatally. Thus, factors that affect myoblast proliferation and differentiation are important for the ability of the muscle to grow. Consequently, this process is of importance in meat producing animals\textsuperscript{13}.

In the adult muscle there are undifferentiated cells, known as satellite cells, which are activated in response to injury or training. In adult muscle, fiber growth occurs when protein synthesis is higher than the protein degradation. The muscles grow by increasing the size of myofibers, not by increasing their number\textsuperscript{18}. It has been shown that the ratio between DNA i.e. nuclei and protein remains constant, so in order to grow, the muscle fibers acquire more nuclei\textsuperscript{19}. Satellite cells fulfill this role by proliferating and fusing to existing myofibers, thus increasing the number of nuclei in the fiber\textsuperscript{18}.

Figure 1. Mouse myoblast C2C12 cells. Left: Undifferentiated cells. Right: Differentiated cells.

Muscle fiber types

Myosin is the main protein that builds up the muscle\textsuperscript{15}. Myosin and actin together with troponin and tropomyosin form the contractile unit\textsuperscript{20}. Myosin consists of four light chains (MLC) and two heavy chains (MHC)\textsuperscript{15}. Muscle fibers can be divided into four different types (I, IIa, IIb and IIx) defined by their myosin heavy chain. They show differences in color, energy metabolism and contractility properties. Type I fibers are red, slow-twitch fibers with an oxidative metabolism. The type IIb and IIx are fast-twitch white muscle fibers with a glycolytic metabolism. The type IIa fibers are intermediate between type I and IIb/IIx. They are red with an oxidative metabolism but belong to the fast-twitch fibers\textsuperscript{21}. However, the classification is not that simple, since
more than one MHC type can be expressed within the same muscle fiber, the major MHC type within one fiber determines its properties\textsuperscript{15}. Further, different muscles in the body have different composition of fiber types. Some muscles have more red type I and type Ia fibers such as \textit{Musculus (M.) trapezius} and \textit{M. triceps brachii}. Other muscles are more white and contain predominantly type IIb fibers, such muscles are \textit{M. longissimus dorsi} and \textit{M. biceps femoris}\textsuperscript{20}. The fiber type in a muscle is not fixed and should be considered dynamic. Factors such as age, hormones and training can change the fiber distribution within muscles\textsuperscript{20,22}. Type II fibers have a larger cross-sectional area that is correlated with growth of the muscle. Hence, selecting for growth would result in a shift from type I fibers to type II fibers. Comparing domestic pigs with wild boars show that wild boars have more type I and type Ia fibers and that domestic pigs have more type IIb fibers\textsuperscript{23,24}. The fiber types in the muscles also affect meat quality. Type I fibers have been shown to have positive effects on meat quality, measured as increased redness, increased protein solubility and a higher pH after slaughter, all of which are related to a better meat quality\textsuperscript{25}. Hence, only selecting for increased growth in the domestic breeds would result in lower meat quality\textsuperscript{20}.

**IGFs in muscle development**

The insulin like growth factors, IGF1 and IGF2, are important for muscle development and muscle growth. Growth hormone (GH) stimulates growth by inducing IGF1 and IGF2 production\textsuperscript{13}. The IGFs interact with three different receptors: the IGF1-receptor, the IGF2-receptor and the insulin receptor. The IGF1-receptor mediates most of the action of both IGF1 and IGF2, and IGF1 binds the receptor with 2-15 fold higher affinity than IGF2\textsuperscript{26}. IGF2 binds the IGF2-receptor with high affinity and IGF1 binds with 500-fold lower affinity. Upon binding the IGF2-receptor, the IGFs are internalized and degraded, so this receptor appears to mainly have a clearing role. The insulin receptor binds insulin, but also IGF2 at a 10 times lower affinity. The insulin receptor and IGF1-receptor can also form heterodimers\textsuperscript{26}. Besides the circulating IGFs that originate from the liver, both IGF1 and IGF2 are expressed locally in tissues. Thus, IGFs function both in an endocrine and paracrine/autocrine fashion\textsuperscript{13,14}. The importance of IGF1 and IGF2 are shown by \textit{in vivo} studies of knockout mice. Igf1(-/-) or Igf2(p-) knockouts result in a mouse with only 60% of normal body size at birth. Double knockouts Igf1(-/-)/Igf2(p-) produce a mouse only 30% of its normal size, which dies at birth with a decrease in cell number in many tissues, including muscle\textsuperscript{27}.

During myogenesis in pigs, the \textit{IGF2} mRNA expression peaks in growing muscle fibers during late primary fiber formation. \textit{IGF1} mRNA expression increases slightly at the same time, but increases more during late secondary fiber formation and reaches its highest level in the neonate pig\textsuperscript{28}. The IGFs are involved in both proliferation and differentiation of muscle cells, as
shown by various cell culture systems (figure 1)\textsuperscript{13}. IGF1 is important for both proliferation and differentiation of myoblasts\textsuperscript{29}. IGF2 is important for the transition from proliferation to differentiation of myoblasts\textsuperscript{30}.

Two signaling pathways are involved in mediating the effect of the IGFs via the IGF1-receptor in muscle cells, the phosphatidylinositol-3-kinase (PI3K) pathway and the mitogen-activated protein kinase (MAPK) pathway\textsuperscript{31}. Inhibition of the MAPK pathway in cells stimulated by IGF1 inhibited the proliferative effect of IGF1, whilst differentiation to myotubes was enhanced. When the PI3K pathway was inhibited, proliferation was slightly reduced whilst differentiation was completely repressed. This suggests that the MAPK pathway is involved in IGF1 induced proliferation while the PI3K pathway is mainly involved in differentiation in response to IGF1\textsuperscript{31}. IGF2 has also been shown to induce differentiation via the PI3K pathway\textsuperscript{32}.

The insulin-like growth factor binding protein (IGFBP) family has six members, IGFBP-1 to IGFBP-6. They are bound to circulating IGFs, prolonging the half-life of these growth factors and act by modulating the availability of IGFs and activity of IGF-signaling pathways. IGFBPs have also been shown to have biological effects without interaction with IGFs\textsuperscript{33}. IGFBP-3 has been shown to have pro-apoptotic activity without binding IGFs\textsuperscript{34} and IGFBP-5 can induce cell migration of smooth muscle cells independently of IGFs\textsuperscript{35}. Primary muscle cells from many species grown in culture, including pig, express IGFBPs. In porcine embryonic myoblast cells, IGFBP-3 is expressed during proliferation but is reduced to one third during the differentiation phase, increasing again in fully differentiated cells to levels three times higher than during proliferation. It is thought that IGFBPs have an autocrine inhibitory effect by binding the IGFs\textsuperscript{36}.

IGFs are also involved in postnatal muscle growth. Both circulating and locally expressed IGF1 has been shown to be involved in postnatal muscle growth in many studies\textsuperscript{13}. Circulating IGF1 levels in postnatal pig have been associated with growth and the infusion of IGF1 into the blood of neonatal pigs increased the protein synthesis in muscle\textsuperscript{37,38}. On the other hand, not many studies have reported on the effect of IGF2 on postnatal muscle growth, although there have been some indications of circulating IGF2 affecting fat deposition\textsuperscript{37}. In a previous study by our group, it was shown that an increase of local IGF2 expression in pig muscle, caused by a point mutation at a regulatory site in the IGF2 gene, increased the muscle mass in postnatal pigs by 3-4\%\textsuperscript{9}. Further characterization of the mutation in the IGF2 gene is presented in this thesis (paper III).
Transcriptional regulation

Gene expression is regulated at many levels. Complicated regulatory networks initiated by stimuli from the environment or within a cell give signals to increase or decrease the expression of a particular gene or group of genes. The structure of DNA, *i.e.*, chromatin, determines if a region of the genome is accessible for transcription or not mainly via DNA methylation and various histone modifications. Further, the DNA code contains non-coding regulatory elements such as promoters, enhancers and repressor elements. DNA-binding transcription factors bind the regulatory elements and other factors interact with them and RNA polymerase machinery to regulate RNA transcription. DNA should not be regarded as a linear molecule but as a complex three-dimensional structure, where sequences located within or close to the gene being transcribed or even in different parts of the genome come in contact and interact to regulate transcription. Using circular chromosome conformation capture (4C) the H19 imprinting control region was shown to interact with up to three other loci in the genome simultaneously\(^{39}\). The RNA is subjected to posttranscriptional modifications such as splicing and the RNA stability is regulated by proteins and regulatory RNA molecules (miRNA and siRNA) that determines if and how much protein the RNA translates.

Epigenetics

Epigenetics are the changes in the chromatin not caused by differences in DNA sequence that regulates gene expression. The changes are passed on during mitosis from mother to daughter cell. The epigenetic changes include DNA methylation and posttranslational modifications of proteins such as histones\(^{40}\).

All the cells in the body contain the same genome (except lymphocyes) but their appearance and function differ between cell types. These differences are due to their individual gene expression profiles. Once differentiated, normal cells do not revert and they keep their expression profile during cell divisions\(^{41}\). Recently, the global changes of epigenetic marks during differentiation have been studied using whole genome analysis of DNA methylation and histone modification\(^{42,43}\).

The epigenetic changes occurring in the cells during differentiation are widely accepted and it has become evident that perturbed epigenetic changes are involved in development of diseases such as cancer\(^{40}\). However more controversial is the epigenetic inheritance through generations. Kaati *et al.* showed that food availability, *i.e.* starving or surfeit of food, during the paternal grandfather’s slow growth period (just before puberty) influence the grandchildren’s risk of dying of cardiovascular disease or diabetes\(^{44}\). However, they have not showed any molecular data proving the inheritance of
epigenetic marks yet. It remains unclear whether the effect detected in their epidemiological study is due to epigenetic inheritance through meiosis.

**DNA methylation**

DNA methylation is an epigenetic mark associated with transcriptional inactivity. A methyl group is covalently added to the 5’ position of a cytosine (C) in a CpG dinucleotide by a DNA methyltransferase (DNMT). In mammals there are three DNMTs: DNMT1, DNMT3a and DNMT3b. DNMT1 is a maintenance enzyme that methylates the newly synthesized DNA strand during cell division. DNMT3a and DNMT3b are responsible for the *de novo* methylation that occurs in the embryo and during development. They have also been shown to be active in embryonic stem cells, where they aid DNMT1 in maintenance methylation of CpG dense regions of repetitive DNA where DNMT1 activity is insufficient. In the genome there are regions with few CpG dinucleotides, i.e. CpG poor regions and regions with more CpG dinucleotides. CpG poor regions are often located in intergenic or in intronic regions. A region ≥500bp, with a GC content ≥55% and with a ratio ≥0.65 between the observed number of CpG and the expected number is referred to as a CpG island. CpG islands are often found in the vicinity of the 5’end of genes and about 40% of genes have CpG islands in their promoter regions. The CpGs in CpG poor regions are generally hypermethylated while the CpG islands are generally hypomethylated, although there are exceptions (see below).

DNA methylation is important for the silencing of repetitive elements originating from transposons, retrotransposons and other parasitic DNA elements and the maintenance of X chromosome inactivation and imprinting. The inactivated X chromosome is hypermethylated, including the otherwise hypomethylated CpG islands. However, there are some genes that remain active and are expressed from both X chromosomes in females. The CpG islands associated with those genes remains unmethylated.

During cell differentiation some genes change their methylation status. When comparing methylation status of promoter regions between sperm and somatic cells about 4% change their methylation pattern, many of which were germline-specific genes. Another study obtained similar results when doing whole genome bisulphite sequencing comparing neural progenitor cells with astrocytes. However, both studies conclude that methylation is not the main mechanism of developmental regulation of gene expression.

**Histone modifications**

The nucleosome that make up the smallest basic unit of chromatin consists of eight histone proteins, two each of the four histones H2A, H2B, H3 and H4, around which 147bp of DNA is wrapped. The histones have a highly conserved globular structure with N-terminal tails. There are over 60 known sites where histones are modified, many of them targeting the N-terminal
Acetylation of lysine, methylation of lysines and arginines, phosphorylation of serine and threonine, ubiquitylation and sumoylation of lysines, ADP ribosylation of glutamate, deimination of arginine and proline isomerization are the different classes of modifications. Some histone modifications are associated with silent heterochromatin such as methylation of certain lysines on H3 (H3K9, H3K27 and H3K20), while others are associated with actively transcribed euchromatin such as trimethylation of other lysine residues on H3 (H3K4, H3K36 and H3K79). Mikkelsen et al. studied genome-wide chromatin state (H3K4 vs. H3K27 methylation) in embryonic stem cells comparing with the chromatin state in lineage-committed neural progenitor cells and embryonic fibroblasts. The chromatin marks at promoter regions changed when cells became committed to a lineage and the chromatin marks were associated with the expression levels of the genes. Thus the chromatin state at lineage-specific genes could reveal the differentiated state of a cell.

Imprinting

Imprinting occurs in a small subset of genes in the genome. According to Ideraabullah et al. (2008) there are about 90 genes known to be imprinted but there are probably more yet to be identified. An imprinted gene shows a parent-of-origin expression where it is either the paternal or the maternal allele that is always being expressed. Several of the imprinted genes are found in clusters with two or more imprinted genes. A cluster of imprinted genes is regulated by an imprinted control region (ICR), which is epigenetically modified in a parent-of-origin manner. There are different types of ICRs, some function as insulators and others as promoters for regulatory non-coding RNA (ncRNA). The H19/IGF2 locus is a well-studied example of a cluster of imprinted genes where the ICR acts as an insulator. The imprinted regulation of this locus is described below. Two clusters that are regulated by ncRNA are the IGF2R and the KCNQ1 loci. Generally, the protein coding genes in both clusters are maternally expressed and the ncRNA is paternally expressed. The promoters of the ncRNA are maternally hypermethylated leading to repression of the ncRNA expression and the protein-coding genes are expressed. On the paternal allele the ncRNA promoters are hypomethylated, the ncRNA expressed and the protein-coding genes repressed.

The evolution of imprinting in mammals is explained by the conflict theory. The mother supplies all the resources for the offspring’s growth during pregnancy through the placenta and during the postnatal period before weaning. The more resources the offspring acquire the more it will grow and thereby increases its chances of survival. The mother is equally related to all her offspring and thus equally interested in their survival. On the other hand, the father to one of her offspring is not necessarily the father to her other offspring. Therefore there is a conflict of interest in the resource allocation.
for a single offspring. The paternally inherited imprinted genes tend to enhance embryonic growth and the maternally inherited imprinted genes tend to reduce growth. *IGF2* and the *IGF2-receptor* are examples of such genes where *IGF2* is paternally expressed and increases prenatal growth while the *IGF2-receptor* is maternally expressed and reduces growth, as shown by their respective knockout in mice.

**Epigenetic regulation of the *IGF2/H19* locus**

The *IGF2* locus has been under close examination during many years since it was one of the first loci proved to have parent-of-origin expression. As mentioned above, *IGF2* is a growth factor important for pre- and postnatal growth. *H19* encodes an untranslated RNA whose exact function is still unknown. Removing the *H19* gene and replacing it with the gene encoding luciferase resulted in no clear phenotype. *H19* has been shown to have both tumor suppressor and tumorigenic properties. Recently *H19* was shown to encode a miRNA named miR-675, but its function is yet to be determined.

Most *Igf2/H19* studies have been performed in mice, hence the discussion here is based on those findings and it cannot be excluded that the regulation of this locus is slightly different in other mammalian species such as human and pig. However, in all mammals investigated (therians *i.e.* placental mammals and marsupials), *Igf2* is paternally and *H19* maternally expressed. The parent-of-origin expression is controlled by a differentially methylated domain (DMD) or ICR, situated between the two genes, acting as an insulator. If the ICR is removed the imprinting status of *Igf2* and *H19* is revoked and the expression becomes biallelic. The ICR is paternally methylated and maternally unmethylated. When the ICR is unmethylated, on the maternal chromosome, the CCCTC-binding factor (CTCF) binds the ICR and enhancers located downstream of *H19* are prevented to act on *Igf2* transcription. Thus, *Igf2* is not transcribed and the enhancers interact with the *H19* promoter activating transcription. On the paternal chromosome the ICR is methylated preventing CTCF from binding. The enhancers are then free to act on *Igf2* transcription and *H19* is silenced by promoter methylation. Several different models have been proposed to explain how the ICR insulator mechanism works. Using the chromosome conformation capture (3C) method interactions between the ICR, DMRs, promoters, enhancers and sequences in between have been studied (figure 2). The two DMRs located in the *Igf2* gene, are both methylated on the paternal chromosome. DMR1 is a silencer and DMR2 is an enhancer of *Igf2* transcription. Kurukuti et al. presented a chromatin loop model where the ICR on the maternal chromosome interacts with DMR1 and matrix attachment region (MAR3) simultaneously, thus positioning *Igf2* in an inactive chromatin loop. On the paternal chromosome the enhancers have access to the entire region and can act
on Igf2 expression\textsuperscript{65}. Yoon \textit{et al.} suggested that, like many other genes, active transcription is always dependent on interaction between promoters and enhancers\textsuperscript{66}. The ICR interacts with promoters and enhancers to prevent their association, thus acting as a decoy\textsuperscript{66}. Further, Qui \textit{et al.} presented a knot model on the maternal chromosome\textsuperscript{67}. The ICR interacts with DMR1, enhancers located between \textit{Igf2} and \textit{H19} and the enhancers downstream of \textit{H19} thereby forming multiple chromatin loops and positioning the \textit{Igf2} gene and \textit{H19} enhancers away from each other\textsuperscript{67}. An enhancer tracking method was proposed by Engel \textit{et al.}\textsuperscript{68}. In their model the enhancers downstream of \textit{H19} track along the chromosome until they encounter an available promoter. On the maternal chromosome CTCF binding to the ICR prevents the enhancer to reach the \textit{Igf2} promoters while on the paternal chromosome there is no blockage by CTCF at the ICR enabling the enhancers to reach the \textit{Igf2} promoters\textsuperscript{68}. Further studies are required to unravel the model of how this locus is regulated.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure2.png}
\caption{Overview of the Igf2 locus in mouse. The figure is adapted from references in the text\textsuperscript{63,65,67}. Not drawn according to scale. Black boxes are DMR1, DMR2 and the ICR. Grey-shaded boxes represent exons and black arrows indicate promoters. P0-P3 refers to the Igf2 promoters. Approximate location of MAR3 is given\textsuperscript{65}. White boxes represent the enhancers.}
\end{figure}

Other epigenetic modifications, besides methylation, are also involved in the regulation of the Igf2/\textit{H19} locus. Two recent studies showed that histone modifications are involved in the allele-specific expression of \textit{Igf2} and \textit{H19}\textsuperscript{69,70}. Han and colleagues showed that the actively expressed maternal \textit{H19} region carried active chromatin marks e.g. H3K9 acetylation and methylation of H3K4, while the paternal had inactive chromatin marks H3K9 and H3K27 methylation and histone 2 variant A1. At the \textit{Igf2} locus the chromatin marks were reversed and the paternal allele carried active chromatin marks and the maternal inactive modifications. As shown in mice carrying mutations of the CTCF binding sites in the ICR the chromatin modification pattern was dependent on CTCF binding the ICR\textsuperscript{69}. This was confirmed by Li \textit{et al.}\textsuperscript{70}. They also showed that CTCF bound to the ICR interacts with the \textit{Igf2} promoters on the maternal chromosome and recruits the polycomb recessive complex 2 (PRC2). CTCF interacts with one of its components Suz12, leading to methylation of H3K27 and repressed transcription. Silenc-
ing Suz12 resulted in biallelic expression of Igf2, which shows that the chromatin modifications are necessary for the imprinted repression at this locus\textsuperscript{70}.

Maintaining the imprinting status of the IGF2 locus is important as shown by its involvement in many cancers and other diseases\textsuperscript{71-73}. Loss of imprinting (LOI) leading to biallelic expression of IGF2 has been found in various human tumor types\textsuperscript{71,72}. In a mouse model of colorectal cancer it has been shown that LOI at the Igf2 locus not only increases Igf2 expression but also increases the sensitivity to Igf2 signaling\textsuperscript{74}. LOI of IGF2 due to mutations in two bindings sites of CTCF in the ICR has also been found in some patients with Beckwith-Wiedermann syndrome\textsuperscript{73}.

Monogenic and quantitative traits

Monogenic traits are those where a simple association exists between the allele at one locus and the phenotypic character. The Austrian monk Gregor Mendel (1822-1884) used monogenic traits in his studies that resulted in the first description of the fundamental laws of inheritance\textsuperscript{75}. For example he studied the pea shape with the two alleles round and wrinkled. In domestic animals monogenic coat color traits have been widely studied. The dominant white coat color present in many pig breeds (e.g. Landrace, figure 3) and the belt coat color in the Hampshire breed (figure 3) are caused by different alleles at the KIT locus\textsuperscript{76,77}. Monogenic diseases have also been well studied in human e.g. cystic fibrosis, where the disease is caused by mutations in the cystic fibrosis conductance regulator gene (CFTR) gene\textsuperscript{78}. Most traits are however not as simple and have a more complex mode of inheritance. Complex traits are influenced by alleles at multiple loci as well as environmental factors. Examples of such traits include height, weight, growth and body composition. Complex traits are often quantitative i.e. they do not form discrete classes as for example coat color traits. In contrary, the trait values are continuously distributed within a population. There are also complex diseases, e.g. type 2 diabetes, where alleles at many loci give a predisposing risk of developing disease. Environmental factors such as diet and other lifestyle factors also influence the risk of developing type 2 diabetes\textsuperscript{79}.

Linkage mapping

A linkage map displays the relative order of markers and the genetic distance between marker loci. The linkage map is used for mapping phenotypic traits in relation to the marker loci.

Early in the history of genetic research, phenotypes such as eye color and wing shape in Drosophila melanogaster, i.e. the common fruit fly, were used
as markers\textsuperscript{80}. Later, immunological phenotypes such as blood groups (AB0) and human leukocyte antigen (HLA) and other biochemical phenotypes were used\textsuperscript{80}. However, such markers are limited in number and to cover the entire genome of a species many more markers are needed. The first DNA-based genetic markers to be used for genetic mapping were restriction fragment length polymorphisms (RFLP)\textsuperscript{81}. This revolutionized genetic research because it enabled the construction of linkage maps including hundreds of markers. Subsequently, genetic markers such as microsatellites and single nucleotide polymorphisms (SNP) have replaced RFLP. Microsatellites are short tandem repeats of usually two to four repeated nucleotides present in multiple copies \textit{e.g.} (CA)\textsubscript{n}, whilst a SNP is a single nucleotide variation (gactgac vs. gacegac). Microsatellites can have multiple alleles due to a relatively high mutation rate compared to SNPs that are generally biallelic. Until recently, the high information content of microsatellites made these the genetic marker of choice in genome mapping experiments. Genome sequencing projects, in combination with resequencing efforts in different populations such as the HapMap project\textsuperscript{82}, have led to the discovery of millions of SNPs. In conjunction with new molecular techniques to cost-effectively analyze thousands of SNPs simultaneously, the marker of choice has lately shifted from microsatellites to SNPs.

To be able to map a marker with linkage analysis it has to exist in at least two different variants denoted as alleles. The alleles must also segregate in the pedigree used to map the marker. In linkage mapping recombination events are exploited. During meiosis, when the haploid gametes are formed, homologous chromosome pairs line up and recombine. An individual’s genetic material, half inherited from each parent, will be mixed during meiosis, forming new combinations of allelic variants that will be passed on to the next generation. If two markers are located on different chromosomes they will show independent segregation. If two marker loci are located very close together on the same chromosome, the marker alleles at these loci will always be inherited together as a haplotype and show complete linkage. However, if two marker loci are located on the same chromosome, but at some distance apart, recombination events can sometimes separate the marker alleles and the markers will show partial linkage. The further away two markers are, the higher the probability of a recombination event occurring between them. Loci located on the same chromosome but very far apart will be separated by recombination and the markers will appear unlinked. By scoring recombination events between markers in a pedigree and calculating the recombination frequencies, the markers can be ordered into a linkage map. The construction of the map is complicated by the fact that there could be more than one recombination event occurring between two markers. If two recombination events occur between two adjacent markers on the same chromosome no recombination event will be observed between the markers. This leads to an underestimation of the recombination frequency. Mapping
functions have been developed to overcome this problem. The first was proposed by Haldane in 1919 and later extended by Kosambi in 1944 (Reviewed by Dronamraju 1987)\textsuperscript{83}. The Kosambi function is the most widely used and, unlike the Haldane mapping function, it takes interference into account, meaning that a recombination event at one location inhibits a second recombination event occurring nearby. The Kosambi function was used to estimate the linkage map in paper I and II. The best way to establish an accurate map is to use a very dense set of markers that can detect most (if not all) recombination events. Today with the availability of large sets of SNP markers and high throughput genotyping techniques this is possible.

Genetic map distance is measured in centi Morgans (cM), where 1cM corresponds to a 1% recombination frequency between two markers. The genetic distance is correlated to the physical distance between markers on the chromosomes, but the correlation is not perfect. The discrepancies are caused by variation in recombination rates (cM/Mb) across the genome. In some genomic regions recombination events are more frequent, generally referred to as hotspots of recombination, and in other regions less frequent (i.e. cold spots) compared to the genomic average\textsuperscript{84}. In many species, including pig, there is also a sex difference in recombination rate which leads to different map lengths for males and females\textsuperscript{85}.

**QTL analysis**

The objective of quantitative trait locus (QTL) analysis is to identify genomic regions that influence the phenotypic variation of a trait. The ultimate goal is to find the gene and the causative mutation(s) that explain the variation in phenotype. A QTL is a genomic region containing one or several sequence polymorphisms that influence the trait. No prior knowledge of the genes influencing the trait is needed to perform a genome-wide QTL analysis.

There are several different methods to identify QTLs. In the first QTL mapping studies performed, single point QTL analysis was used. In such analysis, each marker is tested independently for association with the phenotype and the linkage map (if available) is used to assign identified QTLs to a genomic location. A simple statistical test such as t-test, ANOVA or simple regression analysis could be used\textsuperscript{86}. The limitation of single point QTL mapping is that there is no separate estimate of QTL effect and QTL location. The QTL effect is estimated at the marker position and the estimate will decrease from the true value as a function of the distance between the marker and the actual QTL position. If the markers are widely spaced the power to detect QTLs located far from a marker will be low. Moreover, all individuals with missing genotype data for one marker, will be excluded from the analysis of that marker\textsuperscript{87}. 
Lander and Botstein introduced the interval mapping method for inbred lines using maximum likelihood in 1989. In interval mapping, the genetic linkage map is used to estimate the QTL genotype at every position (e.g. at every 1 or 2 cM across the genome) for each individual using the information from the flanking markers. The QTL analysis is then performed at every position in the genome and more accurate QTL positions and QTL effects can be estimated. Haley and co-workers further developed the interval mapping method using least squares and this method was also modified to be used for outbred line crosses. In paper I and II we used the method developed by Haley et al. for outbred line crosses.

Using outbred line crosses or existing family data from outbred populations, presents two problems for QTL mapping that are not present when inbred lines are investigated. First, the markers will not be fully informative in a cross between two outbred populations, such as two pig breeds, as the markers may segregate in both populations. In inbred lines the QTL genotypes are deduced by only using the flanking markers. For outbred line crosses this is insufficient. Instead, an algorithm is used that estimates the QTL genotypes by including all markers in an entire linkage group, usually all markers on a chromosome. Second, in a standard QTL mapping method, the lines or breed are assumed to be fixed for alternative QTL alleles. This will not always be true for outbred populations, as a QTL allele could segregate in only a few families within the pedigree, or the QTL allele could segregate in both outbred populations. If the assumption of fixation of alternative QTL alleles is false, the power to detect QTLs will decrease.

In the interval mapping method by Haley and colleagues, coefficients of a putative QTL genotype (QQ, Qq, qq) at every preset position in the genome (e.g. every 1 cM), are calculated for every F2 individual using the probabilities of inheriting the allele from any of its grandparents. The phenotypic values are then regressed onto the coefficients and an F-statistic value calculated. The additive and dominance effects are estimated using ordinary least squares. The additive effect is defined as the estimated phenotypic difference between individuals homozygous for one of the QTL alleles (QQ or qq) compared with the mean of the two homozygotes ((QQ+qq)/2). The dominance effect is defined as the phenotypic difference of the heterozygotes (Qq) compared with the mean of the two homozygotes ((QQ+qq)/2). Environmental or other factors known to affect the trait can be included in the statistical model. The inclusion of environmental factors in the statistical model reduces the residual error variance, which in turn both increases the power to detect QTLs and the accuracy of the estimate of QTL effects. A clear advantage of using an outbred line cross is that it also enables the estimation of a parent-of-origin effect (imprinting), provided that it is possible to distinguish if the QTL allele is transmitted from the paternal or maternal F1 parent (e.g. Q^{nat}/q^{mat} or q^{pat}/Q^{mat}). If a parent-of-origin effect is detected there is a difference in effect on the phenotype if the favorable QTL allele was inherited through the F1 sire or dam.
A problem with genome-wide screens is that of multiple testing. As one test is performed for each trait at every position in the genome, the possibility of obtaining a statistical significance by chance *i.e.* false positives (type I error) increases as the number of tests increase. To overcome this, a threshold level is calculated, usually allowing for 1% or 5% false positives in a whole genome scan. A common way to calculate the threshold level is by permutation and this method was used in paper I and II. Permutation involves a randomization of the data, disconnecting phenotype and genotype, and the QTL analysis is repeated\(^9\). The randomization is done multiple times, often \(10^3\)-\(10^4\). The threshold is set so that 99 or 95% of the observed F values, using the randomized data, are below the threshold.
Aim of this thesis

I Perform a genome scan to identify QTLs for length, body composition, muscle weight, fat deposition and meat quality traits in a cross between Landrace and Hampshire. (Papers I and II)

II Identify the repressor binding to the CpG island in intron 3 of the \textit{IGF2} gene, controlling \textit{IGF2} expression in pig muscle. Show that the factor binds specifically to the wild-type (q) sequence, but not the mutated (Q) sequence. Show that it represses transcription from promoter 3 in a luciferase reporter assay and further characterize the repressor. (Paper III)
Present investigations

QTL analysis of a cross between Hampshire and Landrace (paper I and II)

Background

In meat production the characteristics of the meat as well as the production efficiency are important. Today consumers tend to favor lean and tender meat. The Hampshire and Landrace pig breeds have both been selected for muscle growth and are both used in pig meat production in Sweden and worldwide. The breeds have different histories and it is likely that to a certain degree different QTLs have responded to selection, although they have been selected for similar traits during breeding. In this study we took advantage of this and tried to identify QTLs affecting fat and muscle characteristics in the pig, such as lean meat content, fat deposition, size of different muscles, length of the pig, and different meat quality measurements.

In Swedish meat production, Hampshire, Landrace and Yorkshire are used in a three way cross that gives the Pigham meat. Hampshire is used as the sire line and the sow is a crossbred of Landrace and Yorkshire. This meat is thought to have excellent meat characteristics in terms of tenderness, lean-ness and moisture.

The meat quality in this three breed cross is strongly influenced by a mutation in the PRKAG3-gene (RN-locus) common in the Hampshire breed. This gene encodes the regulatory γ3 subunit of adenosine monophosphate-activated protein kinase (AMPK), a muscle-specific isoform. The mutation is located in a highly conserved region of the gene and causes a non-conservative amino acid substitution changing an arginine (R) to a glutamine (Q), at position 225\(^{11,12}\). Carriers of the R225Q mutation (RN\(^{-}/RN^{-}\) or RN\(^{-}/rn+\)) show a large increase of glycogen storage in skeletal muscle leading to a lower ultimate pH after slaughter, increased water content, decreased water-holding capacity and increased cooking loss\(^{92}\).

Later, a second mutation was discovered only three base pairs away from the first mutation. The mutation changes a valine (V) to an isoleucine (I), V224I or \(rn*\). Three alleles at the RN-locus have so far been discovered in pig, 224V-225R (\(rn^+\) or wild-type), 224V-225Q (\(RN^-\)) and 224I-225R (\(rn*\)). The V224I mutation is widely spread and is found in many breeds including
Hampshire, Large White, Landrace, Berkshire, Duroc and wild boar. Its effect on meat characteristics is opposite that of R225Q. The 224I allele is associated with decreased storage of glycogen, higher pH after slaughter and better color scores of the meat.

Figure 3 The Landrace x Hampshire cross. a. Left: Landrace. Right: Hampshire. b. Overview of the cross. Landrace (L) boars were crossed with Hampshire (H) sows producing the F1 generation (LH). The F1 intercross produced 136 F2 pigs (K). Reciprocal backcross, where F1 boars were crossed with Hampshire sows giving 112 (KH) offspring and Hampshire boars were crossed with F1 sows giving 72 (HK) offspring.

The high allele frequency (~60% in 1997) of the RN-mutation (R225Q) in the Swedish Hampshire population was undesirable. To reduce the fre-
quency, the Swedish breeding company Quality Genetics crossed Hampshire with Landrace. The pedigree used in paper I and II are the first three generations of that cross. It is designed as a combined intercross/backcross pedigree with a total of 527 animals (figure 3). Eight Landrace boars were mated to Hampshire sows producing the F1-generation. F1 animals were then either intercrossed or backcrossed to Hampshire. The intercross (F1 x F1) produced 136 F2 offspring. The backcrossing to Hampshire was done reciprocally. F1 boars were mated with Hampshire sows producing 112 offspring and Hampshire boars were mated with F1 sows producing an additional 72 offspring. For an overview of the cross see figure 3. The cross has also been used to study the effects of the two RN-mutations ($RN^-$ and $rn^*$)\textsuperscript{95-97}.

Paper I and II present the results from a genome scan using this cross.

Results and Discussion

The aim of this study was to identify QTLs affecting various meat production traits in the Landrace x Hampshire cross. Landrace has been used in QTL studies before\textsuperscript{98-101} but this was the first time the Hampshire breed was used in a QTL analysis. Therefore, this was a great opportunity to identify previously undetected QTLs in the pig.

A total of 82 phenotypic measurements were analyzed in this study. They included standard measurements such as growth, lean meat content, carcass weight and various meat quality measurements included in most QTL studies, but also included specific measurements of different muscles and body parts. Therefore we could identify QTLs affecting single muscles or body parts that would be missed in other studies due to lack of resolution in phenotyping.

In total, we identified 15 different QTLs reaching at least the 5% genome-wide significance level. This was clearly more than expected by chance. The carcass traits are presented in paper I and the meat quality traits are presented in paper II. All QTLs identified are presented in table 1.

The most significant QTL (F-value of 25.2, P<0.001) identified in this study was for carcass length and was found on \textit{Sus scrofa} chromosome 17 (SSC17). The QTL effect was traced to a specific haplotype ($SW2441$ allele 171, $S0292$ allele 135) inherited from one of the Landrace parental boars. F2 offspring inheriting this haplotype were 1.2 cm longer than F2 offspring originating from this sire that did not inherit the haplotype. Another study using an Iberian x Landrace cross has also found a QTL for length on chromosome 17\textsuperscript{102}. That QTL is only suggestive but the intervals are overlapping, so it may be the same QTL.
**Table 1 Genotype-wide significant QTLs in the Landrace x Hampshire cross**

<table>
<thead>
<tr>
<th>Chr</th>
<th>Trait</th>
<th>Position (cM)</th>
<th>F-value</th>
<th>Mean±SE</th>
<th>Additive effect±SE</th>
<th>Dominance effect±SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Weight GM</td>
<td>9</td>
<td>9.2*</td>
<td>1.0±0.1</td>
<td>-0.24±0.06</td>
<td>-0.25±0.06</td>
</tr>
<tr>
<td></td>
<td>Lean meat content (%)</td>
<td>99</td>
<td>12.0**</td>
<td>66.5±2.2</td>
<td>-1.21±0.30</td>
<td>-0.05±0.39</td>
</tr>
<tr>
<td></td>
<td>Weight of fat on shoulder</td>
<td>92</td>
<td>9.0*</td>
<td>0.7±0.1</td>
<td>0.04±0.01</td>
<td>-0.00±0.02</td>
</tr>
<tr>
<td></td>
<td>Meat and bone in back (%)</td>
<td>99</td>
<td>14.6**</td>
<td>80.5±3.2</td>
<td>-1.97±0.43</td>
<td>-0.14±0.57</td>
</tr>
<tr>
<td></td>
<td>Weight of fat in back (kg)</td>
<td>99</td>
<td>15.0**</td>
<td>1.3±0.3</td>
<td>0.14±0.03</td>
<td>0.01±0.04</td>
</tr>
<tr>
<td></td>
<td>Fat in back (%)</td>
<td>99</td>
<td>14.5**</td>
<td>19.5±3.2</td>
<td>1.97±0.43</td>
<td>0.15±0.57</td>
</tr>
<tr>
<td></td>
<td>Backfat thickness (mm)</td>
<td>83</td>
<td>13.4**</td>
<td>10.7±1.6</td>
<td>1.07±0.22</td>
<td>0.40±0.30</td>
</tr>
<tr>
<td></td>
<td>Meat and bone in ham (%)</td>
<td>99</td>
<td>8.7*</td>
<td>81.8±2.6</td>
<td>-1.17±0.33</td>
<td>-0.11±0.43</td>
</tr>
<tr>
<td></td>
<td>Weight of fat in ham (kg)</td>
<td>81</td>
<td>8.4*</td>
<td>2.2±0.4</td>
<td>0.16±0.04</td>
<td>0.08±0.06</td>
</tr>
<tr>
<td></td>
<td>Fat in ham (%)</td>
<td>97</td>
<td>8.6*</td>
<td>18.2±2.6</td>
<td>1.13±0.32</td>
<td>0.12±0.14</td>
</tr>
<tr>
<td></td>
<td>Weight MBF/weight ham</td>
<td>105</td>
<td>12.2**</td>
<td>13.1±0.7</td>
<td>1.09±0.46</td>
<td>1.82±0.48</td>
</tr>
<tr>
<td></td>
<td>Weight MBF/weight of meat and bone in ham</td>
<td>109</td>
<td>9.5*</td>
<td>15.9±0.7</td>
<td>1.21±0.43</td>
<td>1.72±0.45</td>
</tr>
<tr>
<td></td>
<td>Weight MBF/weight of dissected muscles in ham</td>
<td>116</td>
<td>8.9*</td>
<td>26.1±0.9</td>
<td>1.91±0.57</td>
<td>2.38±0.60</td>
</tr>
<tr>
<td></td>
<td>Weight glycolytic muscles/weight of meat and bone ham</td>
<td>101</td>
<td>9.0*</td>
<td>36.3±1.6</td>
<td>0.81±1.02</td>
<td>2.38±1.07</td>
</tr>
<tr>
<td>2</td>
<td>Weight meat and bone back/weight meat and bone ham</td>
<td>76</td>
<td>9.0*</td>
<td>66.5±1.4</td>
<td>0.52±0.19</td>
<td>0.42±0.28</td>
</tr>
<tr>
<td>3</td>
<td>Protein content in LD (%)</td>
<td>156</td>
<td>9.1*</td>
<td>21.1±1.6</td>
<td>0.41±0.1</td>
<td>-0.3±0.2</td>
</tr>
<tr>
<td>4</td>
<td>Weight ST/weight of dissected muscles in ham</td>
<td>70</td>
<td>9.0*</td>
<td>7.6±0.7</td>
<td>-0.03±0.47</td>
<td>-0.69±0.48</td>
</tr>
<tr>
<td>6</td>
<td>Water content in LD (%)</td>
<td>51</td>
<td>9.5*</td>
<td>76.3±0.8</td>
<td>0.4±0.1</td>
<td>-0.3±0.2</td>
</tr>
<tr>
<td></td>
<td>Drip loss in LD day 3-7 p.m. (%)</td>
<td>69</td>
<td>9.3*</td>
<td>6.7±1.4</td>
<td>0.61±0.1</td>
<td>0.3±0.2</td>
</tr>
<tr>
<td></td>
<td>ΔpH/h in LD 45 min- 3h p.m.</td>
<td>61</td>
<td>9.0*</td>
<td>0.2±0.1</td>
<td>0.05±0.01</td>
<td>0.01±0.02</td>
</tr>
<tr>
<td></td>
<td>Chewing resistance</td>
<td>119</td>
<td>12.3*</td>
<td>40.5±11.7</td>
<td>9.9±3.5</td>
<td>-25.1±5.2</td>
</tr>
<tr>
<td></td>
<td>Tenderness</td>
<td>119</td>
<td>11.0*</td>
<td>53.8±14.7</td>
<td>-11.3±4.6</td>
<td>31.3±6.8</td>
</tr>
<tr>
<td></td>
<td>Chewing time</td>
<td>119</td>
<td>11.0*</td>
<td>56.3±9.9</td>
<td>8.8±3.1</td>
<td>-20.5±4.6</td>
</tr>
<tr>
<td>9</td>
<td>Weight of fat in ham (kg)</td>
<td>77</td>
<td>8.4*</td>
<td>2.2±0.4</td>
<td>0.05±0.04</td>
<td>-0.18±0.05</td>
</tr>
<tr>
<td>10</td>
<td>Weight of fat in ham (kg)</td>
<td>93</td>
<td>9.8*</td>
<td>2.2±0.4</td>
<td>-0.20±0.05</td>
<td>-0.26±0.08</td>
</tr>
<tr>
<td></td>
<td>Weight SMA (kg)</td>
<td>105</td>
<td>8.7*</td>
<td>1.7±0.18</td>
<td>0.45±0.13</td>
<td>0.54±0.13</td>
</tr>
<tr>
<td></td>
<td>Weight ST (kg)</td>
<td>101</td>
<td>8.6*</td>
<td>0.4±0.1</td>
<td>0.15±0.04</td>
<td>0.18±0.04</td>
</tr>
<tr>
<td>12</td>
<td>Carcass length</td>
<td>38</td>
<td>8.7*</td>
<td>95.2±9.3</td>
<td>-1.52±0.38</td>
<td>-1.52±0.49</td>
</tr>
<tr>
<td>13</td>
<td>Weight MBF/weight of dissected muscles in ham</td>
<td>65</td>
<td>9.9*</td>
<td>26.1±0.9</td>
<td>2.17±0.52</td>
<td>2.43±0.55</td>
</tr>
<tr>
<td>16</td>
<td>Freezing and cooking loss in LD (%)</td>
<td>41</td>
<td>9.6*</td>
<td>30.2±2.8</td>
<td>-0.0±0.2</td>
<td>-1.2±0.3</td>
</tr>
<tr>
<td></td>
<td>Weight of fat in ham/weight half carcass</td>
<td>83</td>
<td>8.9*</td>
<td>6.0±0.2</td>
<td>0.33±0.15</td>
<td>0.90±0.21</td>
</tr>
<tr>
<td>17</td>
<td>Carcass length (cm)</td>
<td>30</td>
<td>25.2**</td>
<td>95.2±9.3</td>
<td>-1.59±0.26</td>
<td>0.17±0.31</td>
</tr>
</tbody>
</table>

**1% genome-wide significance
*5% genome-wide significance
p.m.-post mortem
GM – M. Gluteus
SMA – M. semimembranosus et adductor
ST – M. semitendinosus
MBF – M.biceps femoris
LD – M. longissimus dorsi
Another QTL for length was detected on SSC12. It was significant at the 5% genome-wide level and as before the Landrace allele was associated with the longer carcass length, consistent with breed differences. No previous studies have identified a QTL influencing body length on this chromosome. The QTL is located in the region harboring the *growth hormone (GH)* gene, a gene known to affect height in humans making this an obvious candidate gene for this QTL.

On SSC1, two large overlapping QTLs were found. Between 81-99 cM a QTL affecting fat and meat content on the carcass was identified. The Landrace allele was associated with a lean phenotype (more muscle, less fat). The QTL affects the ratio between fat and meat on the carcass rather than the size of the carcass. The QTL shows a purely additive effect. This is compared to the other QTL between 105-116 cM affecting the weight of *M. biceps femoris*. It has an overdominant effect where the heterozygous class was associated with increased weight of the muscle. Another study identified a QTL for rump back fat in this region using a Piétrain x Large White intercross. This may represent the same locus as the fat/meat ratio QTL at 81-99 cM. The *melanocortin 4 receptor (MC4R)* gene is located in this region and it has previously been reported to be associated with fat content and daily gain in pigs. In a subsequent report this gene was evaluated as a candidate gene for the QTL in our cross. A previously known non-conservative mutation changing an aspartate (D) to a asparagine (N), D298N, was typed and analyzed in our cross. When including this mutation as a fixed effect in the analysis the fat/meat ratio trait curves decreased below significance while the QTL effect on *M. biceps femoris* traits did not. This gives further evidence that there are two QTLs in this region and that *MC4R* is a strong candidate gene for the fat/meat ratio QTL.

The QTL between 105-116 cM on chromosome 1 affects a single muscle, *M. biceps femoris*. Interestingly, in this study, we identified several QTLs affecting single muscles or fat deposition in specific body parts showing that our refined measurements are advantageous. QTLs were found at 9 cM on SSC1 affecting weight of *M. gluteus*, on SSC4 at 70 cM affecting *M. semitendinosus* in relation to the other muscles in the ham, on SSC10 affecting weight of *M. semitendinosus* and *M. semimembranosus* and adductor between 101 and 105 cM, and finally a QTL on SSC13 at 65 cM affecting the weight of *M. biceps femoris*. QTLs affecting fat deposition were found on SSC9 peaking at 77 cM and on SSC10 at 93 cM both for weight of fat in ham and finally a QTL on SSC16 at 83 cM affecting the weight of fat in ham in relation to the carcass weight.

For meat quality traits, presented in paper II, four QTLs located on three chromosomes were identified at the 5% genome-wide significance level. Two previously undetected QTLs were identified on chromosome 3 and 16. On SSC3 at 156 cM, a QTL for protein content in *M. longissimus dorsi* was detected, where the Hampshire allele was associated with increased protein content. On SSC16 at 83 cM, a QTL for protein content in *M. longissimus dorsi* was detected, where the Hampshire allele was associated with increased protein content.
content. On SSC16 a QTL for freezing and cooking loss was found peaking at 41 cM. The QTL showed overdominance i.e. the heterozygotes at this position had reduced freezing and cooking losses.

On SSC6, two different QTLs were detected for six measured traits. In the interval 51-69 cM significant QTL effects were observed for the three traits water content, drip loss day 3-7 and pH decline 45 min to 3h post mortem. The Hampshire allele is associated with higher water content, drip loss and pH decline after slaughter. Thus, there seems to be another locus apart from the RN-locus, which is located on chromosome 15, affecting water content and water loss in the meat as well as pH decline after slaughter in this cross. However, no effect on glycogen content was detected for this QTL. In this region three previous studies have found QTLs for similar traits\textsuperscript{108-110}. A candidate gene in this area is the Ryanodine receptor gene (RYR1). A mutation, C1843T, in the gene is possibly the causative mutation in two of the studies where carriers of the mutation from the Piétrain breed were used\textsuperscript{108,110}. The presence of this mutation was tested in our pedigree but all pigs proved to be non-carriers so a different QTL was detected here. In the third study, non-carriers of the RYR1-mutation were also used and we have possibly detected the same QTL\textsuperscript{109}.

The second QTL identified on SSC6 affected three highly correlated traits, all peaking at the same position 119 cM, chewing resistance, tenderness and chewing time. The Hampshire allele was in this case associated with less tender meat and increased chewing resistance and chewing time. A trained panel of persons scored these traits but only 54 pigs were included in the analysis, decreasing the power to detect QTLs for these traits. The QTL explains about 35\% of the residual phenotypic variance. This huge effect could explain why we were able to detect the QTL in such a low number of animals, but this could also be an overestimation and the results should be interpreted with caution. However, another study using a Norwegian commercial slaughter pig cross including Duroc, Norwegian Landrace and Yorkshire, has also detected a QTL for tenderness overlapping the QTL detected in our study\textsuperscript{111}. In a more recent study, microarray data was used to identify differentially expressed genes associated with differences in Warner-Bratzler shear force measurement, a standard method to measure tenderness\textsuperscript{112}. The differentially expressed genes were then mapped to the genome and several are located in the QTL interval on SSC6. The genes are creatine kinase isoform M (CKM), selenoprotein W1 (SEPW1), ribosomal protein L13a (RPL13a), PIH1D1 (or NOP17), josephin domain containing 2 (JOSD2), PTEN induced putative kinase 1 (PINK1) and gap junction protein alpha 4 (connexin 37 or GJA4)\textsuperscript{112}.

Although, we measured 39 different meat quality traits we only detected four QTLs and none at the 1\% genome-wide significance level. Possibly there are no more QTLs with major effects to be found in the Landrace x Hampshire cross. Another explanation is that we did not have sufficient
power to detect additional QTLs. Many of the meat quality traits were not measured in as many animals as the carcass traits. There were also several chromosomes where no QTLs were detected for either carcass or meat quality traits, possibly due to low marker coverage and QTLs could have been missed in those regions. As discussed in the introduction, muscle fiber type affects the meat quality. Red muscle fibers type I and IIA are associated with better meat quality and the white muscle fibers type IIB with reduced meat quality25. In a study by Nii et al. using a cross between Japanese wild boar and Large White, several QTLs affecting specific muscle fiber types as well as meat quality traits were identified113. It would have been interesting to include muscle fiber determination in our study.

Future prospects

The ultimate goal of a QTL study is to identify the underlying genetic variations that explain differences in phenotype between different breeds. Further, we would like to understand how genetic variation causes the phenotype in a functional context. This is challenging since it requires great efforts and resources. As seen with the IGF2 project (see next section and paper III) this approach can lead to major discoveries. However, it is not necessary to identify the actual causative genetic variation to use genetic information in commercial breeding. A linked marker to the mutation is sufficient to select for the trait in marker-assisted selection (MAS).

When this project was started in the late nineties, microsatellites were still the marker of choice due to the availability of markers and the superior information content compared to single nucleotide polymorphisms (SNP). The cost for typing a microsatellite marker in an entire cross was about the same as typing a SNP marker. Since then, new technologies to screen large number of SNPs at low cost have become available, making SNP typing a more common and advantageous choice. Also, the number of SNP markers identified has increased for most species through whole genome sequencing projects combined with shotgun sequencing of multiple breeds. In pig however, with the limited genome sequence available there is still no high density SNP array available. This work is underway and will probably be available during 2009114. The SNP array could be used to fine map the QTL region of existing QTLs in our cross but possibly also identify new QTLs in regions where our marker information was limited. This could reduce our QTL intervals enough to be used for MAS in breeding.

With the new high throughput genome sequencing technologies available today, where whole genomes can be sequenced in a very short time at a much lower cost than with the conventional Sanger sequencing technology, the possibilities are unlimited. It would not be surprising if future genome-wide association studies were to be performed by whole genome sequencing of all individuals within an entire pedigree. In this study, these techniques
can be used to sequence entire QTL regions of a selected number of informative individuals to identify, if not the causative genetic variation, the shortest shared haplotype of individuals carrying the favorable allele of a trait.

Applying these new techniques to reduce the QTL intervals is possible, but probably insufficient. In a three generation cross where the recombination events occurring in the F1 gametes are exploited, the expected QTL intervals are still large. To reduce the QTL intervals more recombination events are needed. There are two ways of increasing the informative meiosis events in a cross, by generating a backcross or an advance intercross line (AIL). Quality Genetics has continued breeding on the Hampshire x Landrace cross and this hybrid population now resembles an AIL. These animals could be used to further decrease QTL intervals by measuring phenotypes again and scoring genotypes using a dense SNP array when it becomes available.

The QTL for carcass length on SSC17 was the most significant QTL we identified. It would be very interesting to further study this QTL and the other length QTL on SSC12, since after the introduction of Landrace into Hampshire, an increase in body length has been observed in the pigs. The length can easily be scored on living pigs and genetic material collected by blood sample. SNP markers could be scored to cover the QTL intervals on SSC17 and SSC12 or a candidate approach could be employed. Interestingly, three recent genome-wide association studies conducted on different human populations, using densely spaced SNP markers and a large number of individuals, showed association between human adult height and markers on multiple human chromosomes, including chromosome 20 and chromosome 17[15-17]. Our QTL on SSC17 is located in the region corresponding to human chromosome 20[18] and the pig chromosome 12 corresponds to human chromosome 17[19]. The regions identified in these studies could easily be covered by markers and interesting genes nearby evaluated as candidate genes for our QTLs.

Identification of ZBED6, a novel mammalian repressor regulating IGF2 expression in muscle (paper III)

Background

This project was started in the late 1980 and was the first major QTL study performed in an outbred species. Using a three-generation intercross between the European wild boar and the domestic Large White, QTL analyses for carcass, growth, fatness and meat quality traits were performed[20,21]. Among the QTLs identified there was one located on the distal end of SSC2 for muscle mass, hearth weight and back-fat thickness[21]. The QTL showed a
strong parent-of-origin effect where the domestic allele, when inherited from the sire, was associated with a 3-4% increase in muscle mass, increased heart size and a reduction of back-fat thickness\textsuperscript{122}. The same QTL was also identified by another group using a cross between Large White and Piétrain, although in their cross the Piétrain allele was associated with increased muscle mass and decreased back-fat thickness\textsuperscript{123}. By joint efforts the QTL region was reduced to a 250kb region containing only two paternally expressed imprinted genes, \textit{insulin} and \textit{insulin-like growth factor 2 (IGF2)}\textsuperscript{124}. The most likely candidate was \textit{IGF2} but no mutations were found in the \textit{IGF2} coding sequence\textsuperscript{123}. The QTL genotype for boars from several different breeds was determined by progeny testing and marker-assisted segregation analysis. By resequencing a 28.6kb region in these pigs, one single nucleotide was determined to be the causative mutation\textsuperscript{9}. This quantitative trait nucleotide (QTN), G3072A in intron 3 of the IGF2 gene, is found within an evolutionary conserved CpG island with unknown function\textsuperscript{9}.

Expression studies showed that the mutation affects \textit{IGF2} mRNA expression in skeletal muscle but not in liver. A three-fold difference in expression level was detected in postnatal skeletal muscle while the expression in fetal skeletal muscle was unaffected by the QTN genotype\textsuperscript{9}. \textit{IGF2} is known to be important in fetal growth and \textit{IGF2} mRNA expression in muscle decreases ten-fold after birth. These results showed that IGF2 is also important for regulating postnatal growth of muscle\textsuperscript{9}. The \textit{IGF2} mRNA expression level from promoters 2, 3 and 4 were all affected by the QTN and the highest expression was seen for promoter 3 transcripts, the main promoter active in skeletal muscle. A smaller increase in expression was also seen in heart, consistent with the phenotype\textsuperscript{9}.

An electrophoretic mobility shift assay (EMSA) with nuclear extracts from mouse C2C12 myoblast cells was used to detect a protein complex binding the QTN region. In EMSA, doublestranded oligonucleotides are mixed with nuclear extracts and then separated on a non-denaturing PAGE gel. Proteins binding the oligonucleotides are detected as a band or gel shift. The wild boar sequence surrounding the QTN, q or wild-type, showed the formation of a specific protein complex or shift, whilst the Large White sequence carrying the mutation, Q, did not detect this protein complex. This showed that the mutation prevents the binding of a nuclear factor. Also, a q oligonucleotide methylated at the QTN CpG did not form the complex, showing that the nuclear factor does not bind a methylated sequence\textsuperscript{9}. Further, in four-month old pigs, the CpG island was shown to be virtually unmethylated in skeletal muscle (3.4%) while it was more methylated in liver (26%) regardless of the QTN genotype and the parent of origin. The QTN did not influence the imprinting status of \textit{IGF2}\textsuperscript{9}. This indicates that neither the CpG island nor the nuclear factor is directly involved in imprinting of \textit{IGF2} and that the methylation status of the CpG island is a possible mode of
regulating the accessibility of the nuclear factor to act on IGF2 mRNA expression.

The effect of the QTN on transcription was shown by transient transfection of luciferase reporter constructs, containing the wild-type (q) or mutated (Q) sequence of the QTN region fused to the IGF2 promoter 3, into mouse C2C12 myoblast cells. The construct containing the q sequence showed a 75% reduction of luciferase activity compared to the construct containing only the promoter 3 sequence, whilst the Q sequence construct only showed a reduction of 30%. Consequently, the nuclear factor is a repressor.

The expression of the IGF2 antisense transcript (IGF2-AS) in pig was also shown to be imprinted, paternally expressed and affected by the QTN genotype in the same way as the sense transcripts. Thus, the repressor affects transcription from four promoters all located downstream of the CpG island harboring the QTN.

A 16bp sequence where the QTN position is nucleotide 15, is completely conserved in the eight mammals whose sequence was available in 2003. Further studies have shown that the putative binding site of the repressor seems to cover at least 14 of the bases ending with the QTN position. There was no convincing match between the QTN sequence and any known transcription factor binding site, which suggested that an unknown transcription factor was binding to the QTN.

In paper III the identification of the transcription factor that represses transcription of IGF2 in postnatal skeletal muscle is presented.

Results and Discussion

The identification of the repressor binding the QTN site in the CpG island of the IGF2 gene intron 3, proved to be a difficult task. Several different approaches, such as one-hybrid screening, Southwestern screening of an expression library and biochemical purification were initially applied but failed. Finally, the repressor was identified using a combination of oligonucleotide affinity capture of proteins and quantitative mass spectrometry.

We used the fact that the two oligonucleotides, q and Q, differs only at one position yet the repressor only binds one of them i.e. the q oligonucleotide. Mouse C2C12 cells were grown with either stable-isotope labeled amino acids (^{13}C_6-L-Arg and ^{13}C_6^{15}N_2-L-Lys) producing “heavy” proteins or with their natural versions producing “light” proteins. This technique is called stable isotope labeling by amino acids in culture (SILAC). “Heavy” and “light” nuclear proteins were extracted separately. Biotinylated double stranded q- and Q-oligonucleotides were used to capture proteins in the “heavy” and “light” nuclear extracts respectively. All proteins binding the q-oligonucleotide would have the stable-isotope labeled amino acids incorporated while the Q-oligonucleotide binding proteins would have the naturally occurring versions of the two amino acids. After removing all proteins not
binding the oligonucleotides, the captured proteins were mixed and size separated, before digestion with trypsin and analysis by liquid chromatography mass spectrometry (LCMS). The “heavy” and “light” peptides were separable in the spectra by the different isotopes and the comparative enrichment of proteins by the two different oligonucleotides could be quantified. Six peptides of a protein where there was a nine-fold enrichment with the q-oligonucleotide compared to the Q-oligonucleotide were identified and proved to be the long sought repressor.

The identified protein was a predicted protein “similar to Zinc finger BED domain containing protein 4 (Zbed4)”, thus this was a previously unknown protein. We named it ZBED6 since it is the sixth mammalian protein carrying the BED domain. The BED domain was first discovered in two proteins in *Drosophila melanogaster*, **BEAF** and **DREF**, which also named the domain. ZBED6 contains two BED domains and a hATC dimerization domain. Although the BED domain was first identified in *Drosophila* and other BED proteins are found throughout the evolutionary tree, ZBED6 seems thus far, with the available genomic sequences, to be exclusive for placental mammals.

The transcript corresponding to ZBED6 was visualized in UCSC genome browser as an alternative splice variant of ZC3H11A, another zinc finger protein. ZC3H11A is poorly characterized and belongs to the large CCCH zinc finger family of proteins with 58 members in mouse and 55 in human. We found that ZBED6 is a retrogene inserted into intron 1 of ZC3H11A. Northern blot analysis showed that Zbed6 is co-expressed with Zc3h11a as a ~13kb long alternative transcript and this was confirmed by sequencing of RT-PCR products. In this way, Zbed6 “hitch-hikes” on the Zc3h11a transcription, a phenomenon not uncommon for functional integrated retrogenes. The DNA binding BED domains among the ZBED proteins are well conserved while the sequence conservation outside this domain is poor. The two BED domains in ZBED6 are more similar to each other compared to the BED domains in other BED proteins. Therefore, a possible scenario of evolution is that ZBED proteins have shared the BED domain by domain shuffling. ZBED6 has then duplicated its BED domain prior to the integration of the reverse transcribed mRNA in intron 1 of ZC3H11A. This reasoning indicates that there should be a host gene of ZBED6 whose spliced mRNA sequence was integrated in the genome. We searched the available genomes but were unable to find the original ZBED6 gene, thus it seems to have been lost during evolution.

The Zbed6 sequence contains two possible translation start sites. We developed a polyclonal antibody against the BED-domains in Zbed6 and could show that both alternative translation start sites are functional. The two isoforms of Zbed6 are 122 and 116 kDa and are denoted Zbed6a and Zbed6b, respectively.
Northern blot analysis and real-time PCR analysis of several mouse tissues revealed that Zbed6 mRNA is expressed in many tissues, including skeletal muscle and heart.

To show that ZBED6 is the repressor that binds the QTN region in the IGF2 intron 3 CpG island, we expressed the DNA-binding BED domains of Zbed6 in bacteria and tested it in EMSA. We could show that Zbed6 specifically bound the wild-type q sequence but not the mutant Q sequence. Further, EMSA using nuclear extracts from C2C12 cells produced the previously detected protein complex and by adding our newly developed anti-Zbed6 antibody the complex was supershifted. This shows that the complex previously detected in EMSA, that specifically binds the q but not Q oligonucleotide, contains the Zbed6 protein.

No clear effect was observed when Zbed6 was over-expressed in C2C12 cells, possibly because the endogenous expression was sufficient to achieve the biological effect. Instead we analyzed the effect of silencing Zbed6 in various assays. The silencing effect on mRNA expression was >75% and protein levels after silencing were undetectable by immunostaining of C2C12 cells and in western blotting. The previously used luciferase reporter assay constructs were used in transient transfection of C2C12 cells treated with scrambled siRNA as control or Zbed6 siRNA. As expected the control cells reproduced the previous results where the wild-type q sequence reduced luciferase expression to a larger degree compared with the mutant Q sequence in combination with the IGF2 promoter 3 (figure 4, left). In Zbed6 silenced cells the reduction of luciferase expression by the wild-type q sequence had completely disappeared while the luciferase expression from the other two constructs remained unaffected (figure 4, right). This shows that Zbed6 is able to repress expression from IGF2 promoter 3 after binding the CpG island.

The interaction between Zbed6 and the QTN site in IGF2 intron 3 in C2C12 cells was also confirmed by chromatin immunoprecipitation (ChIP). There was a clear difference in enrichment of the QTN site DNA in cells treated with scrambled siRNA as control or Zbed6 siRNA. As expected the control cells reproduced the previous results where the wild-type q sequence reduced luciferase expression to a larger degree compared with the mutant Q sequence in combination with the IGF2 promoter 3 (figure 4, left). In Zbed6 silenced cells the reduction of luciferase expression by the wild-type q sequence had completely disappeared while the luciferase expression from the other two constructs remained unaffected (figure 4, right). This shows that Zbed6 is able to repress expression from IGF2 promoter 3 after binding the CpG island.

The effect of the IGF2 mutation on muscle growth in pigs was only detected postnatally\(^9\), in mature muscle cells, and therefore we studied the effect of Zbed6-silencing during differentiation of the mouse myoblast cell line, C2C12. No effect on Igf2 expression was detected the first days after initiating differentiation, however at day 6 when the cells have differentiated and formed myotubes there was a clear increase in Igf2 mRNA expression in Zbed6-silenced cells compared to cells treated with scrambled siRNA. This shows that Zbed6 decreases Igf2 expression in differentiated muscle cells but not in undifferentiated muscle cells. Zbed6-silenced cells showed increased proliferation compared to controls. Zbed6-silenced cells were also studied using a wound healing assay where cells grown in a monolayer are scratched
to locally remove cells from the plate and the repopulation of cells in that area is observed. There was a clear difference as Zbed6-silenced cells showed a faster wound healing process compared with controls.

![Figure 4 Luciferase assay. C2C12 cells transiently transfected with reporter constructs containing pig IGF2 promoter alone or in combination with the wild-type q or the mutant Q sequence from the CpG island in intron 3 of the IGF2 gene. Left, C2C12 cells treated with scrambled siRNA. Right, C2C12 cells treated with ZBED6 siRNA. Firefly luciferase levels are normalized to Renilla luciferase levels expressed from a co-transfected control plasmid. Error bars s.e.m. (** P<0.01).]

In summary, we have identified ZBED6 as the repressor acting on IGF2 expression through the interaction with the QTN region in IGF2 intron 3. The functional data obtained is in agreement with the phenotype observed in pig where the mutation prevents the binding of a repressor leading to increased IGF2 mRNA expression and increased muscle growth. The high sequence conservation between placental mammals of both ZBED6 and the QTN region imply that ZBED6 is involved in IGF2 regulation and muscle growth in the whole eutherian taxonomic group. Further, exercise is known to induce muscle growth. We therefore analyzed ZBED6 expression in a human dataset where muscle biopsies from 12 individuals were collected before and after three weeks of endurance training. For nine subjects there was no difference in ZBED6 expression before and after training but for three subjects showing a higher ZBED6 expression prior to training the expression was reduced after training. The dataset is very small but the result is interesting since differential regulation of ZBED6 could explain differences in muscle growth in response to training.

Future prospects
We have identified the repressor regulating IGF2 mRNA expression in postnatal muscle. The repressor, ZBED6, turned out to be a highly conserved, previously unknown and ubiquitously expressed mammalian gene.
The effects in pig, where the binding of ZBED6 to one site in the genome increases growth of skeletal muscle and heart, shows that this factor has at least one important function. The human data shows that ZBED6 probably has the same function in human. In pigs a reduction in back fat thickness was also observed, although this could be an indirect effect of the increase in muscle mass. Training and increase in muscle mass have positive effects on insulin resistance and type 2 diabetes. This suggests ZBED6 could be a target for obesity and type 2 diabetes treatment. The tissue distribution further suggests that ZBED6 has many more target sites and other functions in the body and it will be very exciting to unravel the secrets of this transcription factor.

The production of Zbed6 knockout mice is the first step to confirm the effects seen on muscle, heart and fat tissue in pigs but also to reveal other functions of Zbed6. Due to the high sequence conservation of Zbed6 it is possible that a homozygous knockout is lethal. Therefore, a conditional knockout, where Zbed6 could be removed from muscle tissues or any other tissue, is the preferred alternative. A Zbed6 knockout could answer many questions such as: Is Zbed6 necessary for survival? Has it other major effects and on other tissues besides muscle? Is it a tumor suppressor gene? Is it involved in imprinting of Igf2? Does it regulate other imprinted genes? Does it affect establishment of any chromatin or DNA modifications? Does it always function as a repressor or could it act as an activator too?

Before the availability of the knockout mice, there are many studies that can be pursued to further understand the mechanism of how ZBED6 regulates IGF2 mRNA expression and also to identify and understand other functions of ZBED6. To detect other target sites of Zbed6 in the genome we could do chromatin immunoprecipitation using the antibody against Zbed6 followed by high-throughput sequencing (ChIP-seq). This experiment in C2C12 cells is already underway and will provide information of possible target genes that are interesting for further studies. There is a tissue-specific aspect that has to be taken into consideration when doing ChIP-seq. Transcription factors often have many target sites in different genes but all target sites are not functionally active in all tissues. ChIP-seq should either be done on multiple types of cultured cells or on multiple tissues. Confirmation of Zbed6 binding sites could either be done by creating oligonucleotides of the identified regions and testing in EMSA, or preferably by performing ChIP of Zbed6-silenced and control cells, followed by quantitative PCR using primers amplifying the “new” target sites.

As a complement to the ChIP-seq experiment, a microarray analysis of Zbed6-silenced and control cells could be done to study the genes whose transcription is affected by Zbed6. This could also indicate if Zbed6 always acts as a repressor.

The ChIP-seq experiment will also provide in vivo data on binding site specificity. The critical positions for binding of Zbed6 besides the QTN have
been defined by EMSA using mutated oligonucleotides\textsuperscript{126}. Further \textit{in vitro} experiments such as DIP-chip/seq\textsuperscript{134} or SELEX\textsuperscript{135} could provide us with a consensus binding site of Zbed6. The Zbed6 full-length protein or the BED domains could be used to retrieve all possible binding sites of Zbed6 from a pool of randomized double-stranded sequences or sheared whole genome DNA. The sequences that Zbed6 binds could be hybridized to a genomic DNA microarray or cloned and sequenced or subjected to direct sequencing, either way providing a consensus binding-site.

Identification of proteins interacting with Zbed6 is also important. From the mass spectrometry assay we have indications of two non-DNA binding proteins that interact with Zbed6, since these proteins were also enriched with the q oligonucleotide. Traditional immunoprecipitation (IP) or full-length Zbed6 coupled to beads could be used to isolate the proteins followed by verification of candidate proteins by Western blot or identification of new proteins by mass spectrometry analysis. Yeast two-hybrid screening experiments could also be done to identify new Zbed6-interacting proteins. Zbed6 would be used as “bait” to screen an expression library from e.g. C2C12 cells or muscle tissue.

Another task is to define the mechanism for how ZBED6 represses transcription of \textit{IGF2}. Does it interact with chromatin modifying enzymes to change the status of chromatin? We did not see a difference in ChIP of methylation of lysine 27 in histone 3 (H3K27) when Zbed6 was silenced compared with controls. H3K27 is a chromatin modification associated with repression of gene expression. Other chromatin modifications known to repress transcription such as methylation of H3 lysine 9 (H3K9) or H4 lysine 20 (H4K20) could be analyzed to see if Zbed6 possibly affects them\textsuperscript{50}.

The QTN mutation had an effect on expression from \textit{IGF2} promoter 2, 3 and 4 as well as the antisense transcript promoter. It is possible that ZBED6 affects transcription from these promoters by binding the QTN wild-type sequence and then represses transcription by interacting with proteins binding the promoter sequences. The interaction should be detectable by repeating the ChIP using the Zbed6 antibody in Zbed6-silenced and control cells followed by quantification of the promoter sequences by real-time PCR.

Since all four \textit{IGF2} promoters are located within a 4kb region, another possible way to affect transcription from all four promoters at the same time is to interact with a loci located downstream of the promoters, forming an inactivation loop around the promoters. If the QTN region interacts with other loci, regardless of how that interaction is done, the loci could be identified using the circular chromosome conformation capture (4C) method\textsuperscript{39}. In 4C, proteins are cross-linked to the DNA preserving the interactions between different regions of DNA. The DNA is then digested with restriction enzymes and diluted followed by circular ligation of interacting DNA sequences. The dilution ensures that ligation is limited to regions that are interacting. After ligation the cross-linking is reversed and the regions amplified by PCR using
primers located in the “bait” in our case the QTN region. The PCR fragments containing the regions interacting with the QTN region are then cloned and subsequently sequenced\textsuperscript{39}. To show that detected interaction are Zbed6 dependent, libraries of cloned fragments from both Zbed6-silenced and control cells could be produced. Sequences detected only or at a much higher frequency in control cells compared to Zbed6-silenced cells, are probably interactions dependent on Zbed6.

ZBED6 is a repressor of \textit{IGF2} transcription and overexpression of \textit{IGF2} due to loss of imprinting (LOI) has been shown to be involved in tumor development, making ZBED6 a possible tumor suppressor gene. As an initial assessment to evaluate ZBED6 as a potential tumor suppressor gene, tumor cells or tumor tissues could be screened by Northern blot, quantitative PCR or immunostaining of tissue arrays to determine whether aberrant expression patterns of ZBED6 is correlated with tumor type or severity.

Further experiments are required to address what post translational modifications, such as phosphorylation and acetylation, ZBED6 is subjected to and if such modifications are used to regulate, the activity, or stability of ZBED6. Also, how the co-expression of \textit{ZC3H11A} and ZBED6 is regulated and if the proteins cooperate in any way, needs to be addressed. Except for ZBED1, the other BED-family members have not been studied in any detail in mammalian species and it would be very interesting to further explore their function as well.
Acknowledgments

I would like to thank the following people:

**Leif Andersson**, main supervisor. Your passion for science and ability to see new opportunities and possibilities in the field of research has been very inspiring. Your calmness and patience, when nothing was working in the lab, kept me from panicking. I guess you always knew we would get there in the end.

**Göran Andersson**, co-supervisor. You have been a great support on the IGF2 project. Always there when I needed to discuss some experimental procedure or to hear some encouraging words.

**Göran Hjälm**, co-supervisor. You have been my link to the world of proteins. Without you the IGF2 project would have been seriously impaired.

**Stefan Marklund**, co-supervisor. Mainly there during my first years in the lab. Thank you for introducing me to Taqman.

**Frida, Lina, Susanne** and **Ulrika**. Your friendship has meant a lot to me during these years. I hope we will stay in touch and continue our “girl-dinners”.

**Ulla**, always a dependent person in the lab. You have been a great support in the lab during these years.

**Åsa**, thank you for your help with the expression library screenings. Unfortunately one of the failed attempts to identify ZBED6 but with the answer at hand the experiment was doomed to fail due to the nature of ZBED6.
**Per**, my roommate. Thank you for putting up with my oddities during these years and for leading the way into the thesis-writing process. Good luck in Paris!

I would like to thank all the collaborators on the Hampshire x Landract project. Especially **Martin Braunschweig**, working with me on the HxL project even after you left the lab, checking seemingly endless inheritance errors during the evenings and nights.

I would like to thank my predecessor on the IGF2 project **Anne-Sophie**, thank you for introducing me to EMSA, my successor **Lin**, I hope you will reveal all the secrets of ZBED6! **Anders**, thanks for bringing the epigenetic aspect into the project. I would also like to express my sincere thanks to the **Broad group** for finding the long sought repressor and to all the other collaborators on the ZBED6 paper.

**Jennifer**, thank you for proofreading and improving the English language in this thesis.

All the present and former **members of the lab**. None mentioned and none forgotten. You have made the lab a nice place to work in.

All friends from my studies at “**biomedicinar programmet**”, the **UGSBR** year and the **SNiB**-committee. It has been great fun!

**Puh**, my friends from home. Always good to see you at midsummer, weddings and other occasions. I hope we stay friends forever.

**My family**. Thank you for letting me find my own way and always supporting me. You are always there if I need you and that means a lot to me.

**Martins family**. Thank you for welcoming me into your home. I always enjoy a glass of wine and a game of cards in your company.

**Martin**, my love, my life. This would not have been possible without you. Your hugs when I am coming home from a really bad day in the lab are lifesavers. Tack för alla Lyckomolekyler! Thank you for your support during these years and for taking care of me and our home during the writing of this thesis. Plupp, plupp!
References


47


93. Ciobanu, D. et al. Evidence for new alleles in the protein kinase adenosine monophosphate-activated gamma(3)-subunit gene


Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 420

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine”.)