Functional characterization of the biological significance of the ZBED6/ZC3H11A locus in placental mammals

SHADY YOUNIS
Dissertation presented at Uppsala University to be publicly examined in B/B42, Biomedicinskt centrum (BMC), Uppsala, Monday, 30 October 2017 at 13:15 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English. Faculty examiner: Docent Ola Hansson (Department of Clinical Sciences, Malmö University Hospital, Lund University).

Abstract


The recent advances in molecular and computational biology have made possible the study of complicated transcriptional regulatory networks that control a wide range of biological processes and phenotypic traits. In this thesis, several approaches were combined including next generation sequencing, gene expression profiling, chromatin and RNA immunoprecipitation, bioinformatics and genome editing methods in order to characterize the biological significance of the ZBED6 and ZC3H11A genes.

A mutation in the binding site of ZBED6, located in an intron of IGF2, disrupts the binding and leads to 3-fold upregulation of IGF2 mRNA in pig muscle tissues. The first part of the thesis presents a detailed functional characterization of ZBED6. Transient silencing of ZBED6 expression in mouse myoblasts led to increased Igf2 expression (~2-fold). ChIP-seq analysis of ZBED6 and histone modifications showed that ZBED6 preferentially binds active promoters and modulates their transcriptional activities (paper I). In the follow-up studies using CRISPR/Cas9 we showed that either the deletion of ZBED6 or its binding site in Igf2 (Igf2^{ΔGGCT}) led to more than 30-fold up-regulation of Igf2 expression in myoblasts. Differentiation of these genetically engineered cells resulted in hypertrophic myotubes. Transcriptome analysis revealed ~30% overlap between the differentially expressed genes in Zbed6^-/- and Igf2^{ΔGGCT} myotubes, with significant enrichment of muscle-specific genes. ZBED6-overexpression in myoblasts led to cell cycle arrest, reduced cell viability, reduced mitochondrial activities and impaired the differentiation of myoblasts (paper II). Further studies on cancer cells showed that ZBED6 influences the growth of colorectal cancer cells with dramatic changes in the transcription of hundreds of cancer-related genes (paper III). The phenotypic characterization of Zbed6^-/- and Igf2^{ΔGGCT} mouse models showed that the ZBED6-Igf2 axis has a major effect on regulating muscle growth and the growth of internal organs. Transcriptome analysis demonstrated a massive up-regulation of Igf2 expression (~30-fold) in adult tissues, but not in fetal tissues, of transgenic mice (paper IV).

In the second part of the thesis we investigated the cellular function of Zc3h11a, the gene harboring ZBED6 in one of its first introns. The function of the ZC3H11A protein is so far poorly characterized. We show that ZC3H11A is a novel stress-induced protein that is required for efficient mRNA export from the nucleus. The inactivation of ZC3H11A inhibited the growth of multiple viruses including HIV, influenza, HSV and adenoviruses (paper V).

Keywords: ZBED6, IGF2, ZC3H11A, Muscle development, Transcriptome analysis, CRISPR/Cas9, mRNA export

Shady Younis, Department of Medical Biochemistry and Microbiology, Box 582, Uppsala University, SE-75123 Uppsala, Sweden. Science for Life Laboratory, SciLifeLab, Box 256, Uppsala University, SE-75105 Uppsala, Sweden.

© Shady Younis 2017

ISSN 1651-6206
urn:nbn:se:uu:diva-329190 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-329190)
In memory of my father,
List of Papers

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


*equal contributions

Reprints were made with permission from the respective publishers.
Additional papers not included in the thesis

Abbreviations

AMPK  5' adenosine monophosphate-activated protein kinase
bHLH  Basic-helix–loop–helix
Cas9  Crispr associated protein 9
ChIP-seq Chromatin immunoprecipitation sequencing
CLIP  Cross-linking immunoprecipitation
CPM  Counts per million
CRISPR Clustered regularly interspaced short palindromic repeats
DE    Differential expression
DMR   Differentially methylated region
DNA   Deoxyribonucleic acid
EMSA  Electrophoretic mobility shift assay
FDR   False discovery rate
FISH  Fluorescent in situ hybridization
FPKM Fragments per kilobase of transcript per million mapped reads
GH    Growth hormone
GO    Gene ontology
HAdV-5 Human adenovirus
HITS-CLIP High-throughput sequencing of RNA isolated by CLIP
HIV   Human immunodeficiency virus
HSV-1 Herpes simplex virus type 1
IF    Immunofluorescence
IGF1  Insulin-like growth factor-I
IGF1R IGF1-receptor
IGF2  Insulin-like growth factor-II
IGF2R IGF2-receptor
IGFBP IGF-binding protein
IR    Insulin receptor
kb    Kilobase
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOI</td>
<td>Loss of imprinting</td>
</tr>
<tr>
<td>MAPK</td>
<td>Mitogen-activated protein kinases</td>
</tr>
<tr>
<td>MRFs</td>
<td>Myogenic regulatory factors</td>
</tr>
<tr>
<td>NLS</td>
<td>Nuclear localization domain</td>
</tr>
<tr>
<td>NoLS</td>
<td>Nucleolus localization domain</td>
</tr>
<tr>
<td>ORF</td>
<td>Open reading frame</td>
</tr>
<tr>
<td>PI3K</td>
<td>Phosphatidylinositol-3-kinase</td>
</tr>
<tr>
<td>qPCR</td>
<td>Quantitative polymerase chain reaction</td>
</tr>
<tr>
<td>QTL</td>
<td>Quantitative trait loci</td>
</tr>
<tr>
<td>rAAV</td>
<td>Recombinant adeno-associated virus</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RNA-seq</td>
<td>RNA sequencing</td>
</tr>
<tr>
<td>SILAC</td>
<td>Stable isotope labeling of amino acids in cell culture</td>
</tr>
<tr>
<td>siRNA</td>
<td>Small interfering RNA</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>TF</td>
<td>Transcription factor</td>
</tr>
<tr>
<td>TREX</td>
<td>Transcription-export complex</td>
</tr>
<tr>
<td>TSS</td>
<td>Transcription start site</td>
</tr>
<tr>
<td>UTR</td>
<td>Un-translated region</td>
</tr>
<tr>
<td>WT</td>
<td>Wild-type</td>
</tr>
<tr>
<td>ZBED6</td>
<td>Zinc finger BED-type containing 6</td>
</tr>
<tr>
<td>ZC3H11A</td>
<td>Zinc finger CCCH containing 11a</td>
</tr>
</tbody>
</table>
Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>11</td>
</tr>
<tr>
<td>Transcriptional regulation</td>
<td>11</td>
</tr>
<tr>
<td>The transcriptome</td>
<td>12</td>
</tr>
<tr>
<td>Transcription factors</td>
<td>13</td>
</tr>
<tr>
<td>Epigenetic regulation</td>
<td>14</td>
</tr>
<tr>
<td>Histone modifications</td>
<td>14</td>
</tr>
<tr>
<td>DNA methylation</td>
<td>15</td>
</tr>
<tr>
<td>DNA demethylation</td>
<td>16</td>
</tr>
<tr>
<td>Epigenetic regulation of $H19/IGF2$ locus</td>
<td>17</td>
</tr>
<tr>
<td>Muscle development</td>
<td>20</td>
</tr>
<tr>
<td>Genetic factors regulate skeletal muscle development</td>
<td>21</td>
</tr>
<tr>
<td>Insulin-like growth factors (IGFs)</td>
<td>22</td>
</tr>
<tr>
<td>IGF2 in muscle development</td>
<td>22</td>
</tr>
<tr>
<td>IGF2 overexpression and cancer development</td>
<td>24</td>
</tr>
<tr>
<td>Background to the thesis projects</td>
<td>25</td>
</tr>
<tr>
<td>The discovery of ZBED6</td>
<td>25</td>
</tr>
<tr>
<td>$ZC3H11A$, the host gene of ZBED6</td>
<td>27</td>
</tr>
<tr>
<td>Aim of the thesis</td>
<td>29</td>
</tr>
<tr>
<td>Present investigations</td>
<td>30</td>
</tr>
<tr>
<td>Paper I</td>
<td>30</td>
</tr>
<tr>
<td>Paper II</td>
<td>33</td>
</tr>
<tr>
<td>Paper III</td>
<td>37</td>
</tr>
<tr>
<td>Paper IV</td>
<td>41</td>
</tr>
<tr>
<td>Paper V</td>
<td>44</td>
</tr>
<tr>
<td>Future perspectives</td>
<td>47</td>
</tr>
<tr>
<td>Summary in Swedish</td>
<td>50</td>
</tr>
<tr>
<td>Summary in Arabic</td>
<td>51</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>52</td>
</tr>
<tr>
<td>References</td>
<td>54</td>
</tr>
</tbody>
</table>
Introduction

Transcriptional regulation

A fundamental question in biology is how cells differentiate and maintain their identity from the same genome? It has been estimated that the human body contains more than 200 cell types, all differing in their function and properties, even though they contain the same genetic material as DNA sequences. The human genome contains approximately 20,000 protein-coding genes but not all of them are actively expressed in all cell types. The expression pattern of these genes behaves in a manner similar to an orchestra. In the same cell type some of the genes are expressed, at different levels, while the rest are silent. As a response to stimuli, growth or developmental conditions; the cells alter the expression pattern and some of the silent genes become actively transcribed. The transcriptional regulation allows the cells to reprogram themselves and quickly respond to changes in the environment by altering the pattern of gene expression. Dysregulation of the gene transcription leads to abnormal phenotypes and causes diseases, e.g. cancers and developmental disorders. Several components are involved in transcription regulation, including the RNA polymerase machinery, regulatory elements, transcription factors and epigenetic marks.

In 1958, Francis Crick proposed the central dogma model for the flow of the genetic information in the cells. He hypothesized that genetic information goes from DNA to protein synthesis through RNA molecules in a unidirectional way. Basically, the RNA polymerase machinery transcribes the protein coding sequences into messenger RNA (mRNA) that is translated into the corresponding proteins in the cytoplasm by ribosomal RNA (rRNA). This is the simplified explanation of the central dogma. In fact, the expression of a gene, starting from DNA until the translated protein, involves a series of complicated processes such as post-transcriptional modifications, an mRNA export machinery and post-translational modifications. These processes regulate the stability of the mRNA, determine the amount of mRNA to be exported from the nucleus to the cytoplasm, affect the translation rate and alter the function of the translated proteins by introducing certain modifications.
The transcriptome

The term transcriptome refers to the set of all RNA molecules that are produced in the cells, and exhibit distinct variations in their function, length and localization. There are three types of RNA polymerases (Pol-I, II and III) and each one is responsible for the transcription of certain types of RNA molecules. RNA Pol-I is responsible for the production of rRNA subunits 28S, 18S and 5.8S, which are the most abundant RNAs in the cells. RNA Pol-II transcribes the protein coding sequences into precursor mRNA (pre-mRNA), which undergo a series of processes to become a mature mRNA. The mRNA transcripts compose around 1.5% of the total RNA in the cells. RNA Pol-III produces the transfer RNA (tRNA) and the 5S subunit of the ribosome. RNAs can be classified into two main groups: coding and non-coding RNA. The mRNA transcripts are the only RNAs that contain the protein coding sequences and are translated into proteins by ribosomes. The non-coding RNAs (ncRNA) contain several species of RNA molecules that are not translated into proteins, but they are involved in various biological functions. For instance, rRNAs and tRNAs are responsible for protein synthesis, long non-coding RNAs (lncRNA) in gene regulation, micro RNAs (miRNA) play important roles in post-transcriptional regulation and gene silencing, small nuclear RNAs (snRNA) are one of the main components of the splicing machinery. The recent advances in sequencing technologies revealed the complexity of the transcriptome, and identified new types of RNA molecules beside the well-known ones (Table 1). Based on the encyclopedia of DNA elements (ENCODE), around 80% of the genome transcribes ncRNAs\textsuperscript{3,4}.

Table 1. Different species of RNA molecules\textsuperscript{4,5}.

<table>
<thead>
<tr>
<th>RNA type</th>
<th>Abbreviation</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Messenger RNA</td>
<td>mRNA</td>
<td>Protein coding</td>
</tr>
<tr>
<td>Ribosomal RNA</td>
<td>rRNA</td>
<td>Protein synthesis</td>
</tr>
<tr>
<td>Transfer RNA</td>
<td>tRNA</td>
<td>Carry amino acids to ribosomes</td>
</tr>
<tr>
<td>Long non-coding RNA</td>
<td>lncRNA</td>
<td>Gene regulation</td>
</tr>
<tr>
<td>Small nuclear RNA</td>
<td>snRNA</td>
<td>RNA splicing</td>
</tr>
<tr>
<td>Small nucleolar RNA</td>
<td>snoRNA</td>
<td>Processing of pre-rRNA</td>
</tr>
<tr>
<td>micro RNA</td>
<td>miRNA</td>
<td>Gene silencing</td>
</tr>
<tr>
<td>PIWI-interacting RNA</td>
<td>piRNA</td>
<td>Silence transposons in germ cells</td>
</tr>
<tr>
<td>Signal recognition particle RNA</td>
<td>srpRNA</td>
<td>Post-translational transport</td>
</tr>
</tbody>
</table>
Transcription factors

Transcription factors (TFs) are a diverse group of proteins that are crucial for transcriptional regulation. There are around 2,000 genes encoding TFs and they often produce large number of splice variants. The distinct feature of TFs is that they have DNA-binding domains (DBD) that give them the ability to bind DNA at specific sequences, consisting of a few base pairs, called binding sites. Moreover, most TFs can establish protein-protein interactions with other activators or repressors, and with co-activators and co-repressors that do not by themselves bind DNA. These features give the TFs the ability to interact with the transcription machinery and function either as activators or repressors of the transcription based on their effect.

The key step of the gene transcription is the assembly of a transcription pre-initiation complex (PIC) at a promoter. This process requires the participation of so called basal or general TFs (GTFs). The minimal PIC contains six GTFs (TFIIA, TFII B, TFIID, TFII E, TFII F and TFII H) that can directly bind the core promoter and help in assembly of a transcription complex with RNA Pol-II (Figure 1). The core promoter is defined as ~40 bp spanning the transcriptional start site (TSS), and is required for the initiation of transcription by the RNA Pol-II machinery. The most common DNA elements found in core promoters are: the TATA-box, initiator sequences (Inr), the TFII B recognition element (BRE) and the downstream promoter element (DPE). Genome-wide studies showed that the majority of the mammalian genes containing TATA-less promoters with multiple TSS. Moreover, promoters that lack the TATA box can rely on the DPE and Inr to recruit the PIC and initiate transcription. In case of TATA-containing promoters, the TATA binding protein (TBP) subunit of TFIID binds the TATA-box in the core promoter and initiates the assembly of the other GTFs and Pol II. In TATA-less promoters, TFIID interacts with the DPE and Inr through different TBP-associated factors (TAFs).

Furthermore, many TFs bind the proximal promoter and regulatory elements such as enhancers or silencers and thereby modulate the transcriptional activity. The proximal promoter extends ~200 bp upstream of the TSS and typically contains multiple TF binding sites. The enhancer or silencer sequences can be found far from the TSS, and contain binding sites for several TFs and function regardless of direction and orientation (Figure 1).

TFs are usually classified into groups based on their DBDs, for example: the basic-helix–loop–helix (bHLH) and zinc finger proteins. The bHLH motif is characterized by two α-helices containing basic amino acids residues linked by a loop structure. The bHLH proteins usually bind the DNA at a consensus hexanucleotide sequence known as the E box. Members of this family are involved in the regulation of several biological processes. For example, MyoD and Myogenin that are involved muscle development.
Zinc finger proteins, almost half of all human transcription factors are zinc fingers. These proteins bind a zinc ion to cysteine and histidine residues and form a finger-like domain\textsuperscript{12}, which facilitates the interaction with nucleic acids. Thus, zinc finger proteins can have a variety of functions such as DNA-binding, RNA-processing and protein-protein interaction.

Figure 1. Illustration of the transcription pre-initiation complex at the core promoter in combination with GTFs and RNA polymerase II. Regulatory elements are not shown according to scale. Adapted from Levine et al. 2003\textsuperscript{9}.

Epigenetic regulation

Every cell in our body contains the same genetic information, but how can the body then end up with different types of cells with different biological functions? This can be explained by the epigenetic regulation of the genome. The epigenome includes the methylation status of the DNA sequences and the post-translational modifications of proteins such as histones. These modifications affect the accessibility of the genes, and determine which genes need to be turned on and off in different cell types.

Histone modifications

In eukaryotic cells, the DNA can be packaged and form a complex of macromolecules called chromatin. The basic unit of chromatin is called a nucleosome and is composed of DNA associated with histone proteins. The
structure of each nucleosome consists of two sets of four histones (H3, H4, H2A and H2B), around which a 147 bp segment of DNA is wrapped. The nucleosomes are separated from each other by an around 20 bp of linker DNA. The histone proteins have a core globular structure with an unstructured N-terminal tail containing several amino acid residues available for post-translational modifications. There are more than 60 known residues where histones are modified, mainly on histone tails. These modifications include: acetylation, methylation, phosphorylation, ubiquitylation, SUMOylation, ADP-ribosylation, deimination and proline isomerization.

Histone modifications play important roles in transcriptional regulation, DNA replication and chromosome condensation. Histone modifications can act either by disrupting the contacts between nucleosomes in order to unravel chromatin and make it more accessible, or by recruiting non-histone proteins to bind. In regard to the transcriptional status, chromatin can be divided into two distinct types: the actively transcribed euchromatin, where DNA is more accessible for transcription, or the transcriptionally inactive heterochromatin, where DNA is compact and inaccessible. Each chromatin type is associated with certain histone modifications. For instance, acetylation and trimethylation of lysine residues (H3K4, H3K36 and H3K79) are associated with actively transcribed euchromatin. On the other hand, high level of methylation on lysine residues (H3K9, H3K27 and H4K20) and low acetylation are found at silent heterochromatin. Moreover, the promoters of actively transcribed genes are enriched with H3K4me3, H3K4me2, H3K27ac, H2BK5ac and H4K20me1. Furthermore, methylation at H3K36 and H3K9 negatively affect the transcription when they are found on promoters, whereas they positively affect the transcription when they found in a coding region. All these together indicate a central role of histone modifications in regulating gene transcription.

DNA methylation

Forty-seven years ago, the cytosine DNA methylation in eukaryotes was proposed as an inherited modification to the DNA that affects transcriptional regulation and cellular development. Since then, DNA methylation was considered as a central epigenetic mark and several investigations have been conducted in order to expand our understanding of DNA methylation in different aspects, for instance, the required pathway to establish and maintain the DNA methylation, the role of DNA methylation in transcriptional inhibition, the genomic imprinting, X-inactivation, embryonic development and tumorigenesis.

The DNA methylation process is carried out by DNA methyltransferase (DNMT) enzymes that add a methyl group to the 5’-position of a cytosine. In mammalian cells, there are four types of DNMT enzymes: DNA Methyltransferase 1 (DNMT1), DNA Methyltransferase 3 Alpha
(DNMT3A), DNA Methyltransferase 3 Beta (DNMT3B) and DNA Methyltransferase 3 Like (DNMT3L). The DNMT1 is known as the maintenance methyltransferase that transfer the DNA methylation patterns to the newly synthesized DNA strand during cell division\(^{22}\). DNMT3A and DNMT3B are \textit{de novo} methyltransferases and establish \textit{de novo} methylation during the development\(^{23}\). In contrast, DNMT3L is catalytically inactive but stimulates the \textit{de novo} methylation by DNMT3A\(^{24}\). DNA methylation was previously considered to only occur at CpG sites. This concept was re-examined by the genome-wide sequencing at single-nucleotide resolution, which provided evidences for DNA methylation at non-CpG sites\(^{25,26}\). A recent study showed that the large majority of DNA methylations (~ 98\%) occur on cytosines adjacent to guanine nucleotides in a CpG context in somatic cells, while ~ 25\% of the methylation in embryonic stem cells (ESCs) was identified in a non-CpG context, with enrichment in gene bodies. The methylated cytosine nucleotides in a non-CpG context disappeared after differentiation of the ESCs, which suggest that the ESCs may use different methylation mechanisms to regulate the expression of genes involved in pluripotency and differentiation\(^{26}\).

**DNA demethylation**

DNA demethylation is a multistep process that leads to a complete removal of the methyl group from the methylated cytosine (mC). This occurs through a consecutive oxidation of mC by ten-eleven translocation (TET) methylcytosine dioxygenase enzymes and produce three methylcytosine derivatives (Figure 2). First, the TET enzymes convert 5-methylcytosine (5-mC) to 5-hydroxymethylcytosine (5-hmC), then oxidize 5hmC to 5-formylcytosine (5-fC) and to 5-carboxycytosine (5-caC). Finally, the thymine-DNA glycosylase (TDG) in combination with the base excision repair (BER) pathway can convert the oxidized derivatives 5fC and 5caC into unmethylated cytosine\(^{27,28}\). Since the discovery of DNA demethylation in 2009, many studies were conducted to investigate if these 5-mC derivatives (5-hmC, 5-f0C, and 5-caC) have a functional role. Recent studies have revealed the genomic distribution of these derivatives in different tissues and cell lines. Interestingly, these derivatives are not randomly distributed and show a preference for certain genomic regions distinct from that of 5-mC\(^{29,30}\). For instance, 5-hmC is enriched at enhancer elements, while the 5-fC and 5-caC are enriched at major satellite repeats in mouse embryonic stem cells (ESCs)\(^{31,32}\). In addition, more recent studies have observed a significant reduction in 5-hmC in human cancers, which indicate a functional role of 5-hmC in tumor development\(^{33,34}\).
Figure 2. *The process of DNA methylation and demethylation.*

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

The *H19/IGF2* locus is one of the first discovered loci that showed a differential expression based on the parental origin of the inherited allele. This locus is located on human chromosome 11 (11p15.5) and mouse distal chromosome 7. It harbors both the *IGF2* and *H19* genes in close proximity, ~130 kb and ~75 kb apart in human and mouse, respectively. Insulin-like growth factor-II (IGF2) is a growth-promoting factor that plays central roles in embryonic development and cell growth. *H19* encodes a long non-coding RNA that is highly expressed in fetal tissues, and shows a reduced expression in adult tissues. The biological function of *H19* remains unclear, but some studies considered it as a tumor suppressor gene. On the other hand, other studies proposed the opposite and suggested an oncogenic role of *H19*. These contradictions are due to the differential expression of *H19* in many tumors. For instance, *H19* is highly expressed in breast, colon and gastric cancers\(^35,36\), while no expression was detected in Wilm’s tumor, rhabdomyosarcoma or Beckwith–Wiedemann syndrome\(^37–39\).

In 1991, DeChiara and colleagues showed that *Igf2* is exclusively expressed from the paternal allele, while the maternal allele is transcriptionally inactive\(^40\). This was concluded based on their studies using transgenic mice carrying an *Igf2* knock-out mutation. Transmission of the mutant allele from the males to the offspring resulted in a severe growth deficiency of the heterozygous offspring, while no effect was observed in the...
growth of the heterozygous offspring when they received the mutant allele from their mother. In contrast to Igf2, H19 is paternally imprinted and only expressed from the maternal allele\textsuperscript{41}. This monoallelic expression is controlled by an imprinted control region (ICR) located between the H19 and Igf2 genes, acting as an insulator. The deletion of the ICR disturbs the imprinted status of the H19/IGF2 locus and their expression become biallelic\textsuperscript{42}. The precise mechanism of how the ICR regulates the imprinted expression pattern of H19 and Igf2 has been debated, and several distinct models have been suggested. The main proposed mechanisms include the enhancer-blocker model and the chromatin loop model\textsuperscript{43–46}.

The enhancer-blocker model is based on the interaction between the ICR, CCCTC-binding factor (CTCF), enhancers located downstream of H19, and the H19 and Igf2 promoters. The ICR is a hyper-methylated region on the paternal allele and unmethylated on the maternal allele. CTCF binds the unmethylated ICR on the maternal chromosome, prevents it from de novo methylation, and acts as an insulator or enhancer blocker between Igf2 and the enhancers. As a consequence, the enhancers interact with the H19 promoter on the maternal allele and activate transcription of H19, but not Igf2\textsuperscript{43}. The methylation of the ICR on the paternal chromosome prevents CTCF from binding. Therefore, the enhancers can access the Igf2 promoters and activate the transcription, while the H19 promoter remains silenced (Figure 3A). Further studies have reported that the binding of CTCF to the ICR is essential for imprinting maintenance, but not establishment at the H19/IGF2 locus\textsuperscript{47,48}.

In 2004, Murrell \textit{et al.} proposed chromatin looping as a mechanism for the imprinted expression pattern of the H19 and Igf2 genes\textsuperscript{44}. As mentioned above, both Igf2 and H19 share the same enhancer elements and ICR site. Furthermore, the mouse Igf2 contains three differentially methylated regions (DMRs). The DRM0 exhibits biallelic hyper-methylation in fetal tissues and is maternally methylated in the placenta\textsuperscript{49}. The other two DMRs are methylated on the paternal chromosome and located upstream of promoter 1 (DMR1) and exon 6 (DMR2)\textsuperscript{50,51}. DMR1 functions as a silencer and DMR2 acts as an activator of Igf2 transcription\textsuperscript{51,52}. The proposed chromatin looping model is based on the physical interaction between ICR and Igf2 DMRs. On the paternal allele, the ICR interacts with Igf2 DMR2 and forms a chromatin loop that positions Igf2 promoters in close proximity to the enhancers. This leads to transcription activation at Igf2 promoters but not the H19 promoter, which is silenced by DNA methylation. On the maternal allele, the unmethylated ICR binds CTCF, interacts with Igf2 DMR1 and forms two chromatin domains. This places H19 and the enhancers in an active domain and keeps Igf2 away from the enhancers in an inactive chromatin loop\textsuperscript{44} (Figure 3B). In agreement with the model proposed by Murrell \textit{et al.}\textsuperscript{43}, Kurukuti and colleagues presented a chromatin loop model that involves an exclusive interaction between the ICR, Igf2 DMR1 and the matrix attachment region (MAR3) on the maternal chromosome\textsuperscript{45}. This
interaction forms a tight loop around the *Igf2* locus on the maternal allele, thereby silencing *Igf2* expression."}

**Figure 3.** Schematic representation of the epigenetic regulation at the *IGF2/H19* locus in the mouse genome. Black boxes represent the exons, blue boxes are enhancers, grey boxes are the *Igf2* DMRs and H19 imprinted control region (ICR), black arrows indicate the promoters. Red boxes represent the coding sequences of *Igf2* (A). The proposed enhancer-blocker model. CTCF binds the unmethylated ICR on the maternal allele and acts as an insulator between the enhancers (blue boxes) and *Igf2* promoters (P1-P3). On the paternal allele, CTCF does not bind the methylated ICR (filled orange circles with M), thereby the enhancers can access the *Igf2* promoters and activate the transcription. (B) The chromatin looping model. On the paternal allele, the methylated ICR interacts with *Igf2* DMR2 and forms a chromatin loop that positions *Igf2* promoters in close proximity to the enhancers and leads to transcription activation at *Igf2* promoters. On the maternal allele, CTCF binds the unmethylated ICR and interacts with *Igf2* DMR1. This places H19 and the enhancers in an active domain and keeps *Igf2* away from the enhancers in an inactive chromatin loop. The figure is adapted from references in the text, not drawn according to scale.
Additional models have been proposed in order to explain the imprinted regulation at this locus. For instance, Yoon et al. suggested that the ICR acts as a decoy and competes with promoters for enhancer interactions. It is consistent across the conducted studies that the ICR is essential for the parent-of-origin expression at this locus, but the precise mechanism requires further studies.

DNA methylation is not the only epigenetic modification that regulates the imprinted expression pattern at the H19/IGF2 locus. Recent studies have presented that certain types of histone modifications are involved as well. For instance, histone methylation of H3K4 and acetylation of H3K9 were found at the actively transcribed maternal H19 allele, while the paternal allele carried repression histone marks such as methylation on H3K9 and H3K27 lysine residues. The reversed histone modifications were found at the Igf2 region and the maternal allele carried repression modifications such as trimethylation of H3K27, while the actively expressed paternal allele carried active histones marks. This histone modification pattern at the H19/IGF2 locus is suggested to be dependent of the interaction between CTCF and the ICR.

It is clear that the H19/IGF2 locus undergo complicated and tight regulations in order to maintain the imprinting status. Loss of imprinting (LOI) of Igf2 leads to biallelic expression, which is involved in many diseases including cancer development.

Muscle development

The formation of skeletal muscle involves a series of complex developmental processes that occur throughout prenatal and postnatal life, and are highly controlled by genetic factors. Skeletal muscle basically originates from the mesoderm layer, one of the three germ layers that give rise to the specialized tissues in the adult. During embryonic development, the mesoderm produces myogenic progenitor cells and other cell types. Thereafter, the progenitors develop to form a population of proliferative mononucleated myoblasts. These embryonic myoblasts exit the cell cycle and differentiate to form multinucleated cells called primary myotubes. This phase called primary myogenesis occurs between weeks 8–10 in humans and embryonic day E12.5 in mice. The primary myotubes form the basic muscle pattern and serve as a scaffold for the second phase of the myogenesis, which occurs between weeks 10–18 in human and embryonic days E14.5–E17.5 in mice. In the secondary myogenesis phase, the fetal myoblasts stop dividing and differentiate to form secondary myotubes on the surface of primary myotubes and then elongate to form individual fibers. Once the two phases of myogenesis are completed, no further proliferation
occurs and the total number of fibers in muscle tissues is fixed. Therefore, the postnatal growth occurs hypertrophic rather than hyperplastic.

Small proportions of undifferentiated myoblasts remain proliferative in adult muscle tissues and are located extracellular to muscle fibers. These cells are called satellite cells and were first discovered in the 1960s. Satellite cells have the capacity to regenerate and repair damaged myofibers or form new ones as a response to injury, disease or training.

Muscle fibers are the major components of skeletal muscle tissues. There are three types of muscle fibers that differ from each other in energy metabolism and contractile properties. Type I fibers are slow twitch fibers, rich in mitochondria and generate ATP mainly from oxidative metabolism. The type II fibers are fast twitch fibers divided into two subtypes, IIa and IIb. The type IIa fibers are fast twitch that generate ATP from both oxidative and glycolytic metabolism. The type IIb fibers are fast twitch and generate ATP mainly from anaerobic metabolism. The fiber types differ in their myoglobin content; the oxidative fibers are rich myoglobin while the glycolytic fibers have a low amount of myoglobin. As a consequence, the oxidative fibers tend to be red and are therefore called the red muscles, while the glycolytic fibers are called the white muscles. The proportions of the three fiber types are changeable and responsive to exercise, hormones, age and some diseases. For instance, there is a positive correlation between obesity and a high proportion of type IIb fibers.

Genetic factors regulate skeletal muscle development

The regulatory network of muscle formation during prenatal and postnatal development is controlled by a specific class of transcription factors called myogenic regulatory factors (MRFs), which act together with other regulatory factors to coordinate muscle development. MRFs include four muscle-specific transcription factors: myoblast determination protein (MYOD), myogenic factor 5 (MYF5), muscle-specific regulatory factor 4 (MRF4) and myogenin. The MRFs belong to the basic-helix–loop–helix (bHLH) class of DNA-binding proteins, and activate many downstream genes. MRFs play crucial roles in muscle determination and development throughout the lifespan. For instance, MYOD and MYF5 determine skeletal muscle lineage and induce the fibroblasts to differentiate into myoblasts. Moreover, MYOD can convert other cell lines into muscle-like cells when it is activated in these cells. MYOD and MYF5 carry out an overlapping function, thus muscle develops normally if one of them is absent, while the absence or inactivation of both the MYOD and MYF5 genes in the embryo results in complete lack of muscle formation. Myogenin expression induces the differentiation of myoblasts into myotubes. The myogenin knock-out mice exhibited severe muscle deficiency. MRF4 expression in skeletal
muscle of postnatal animals maintains the skeletal muscle fibers. MRFs interact with other bHLH transcription factors such as E2A proteins and form heterodimers. These heterodimers activate the transcription of muscle-specific genes by binding to the E-boxes of target genes. The E-box (5'-CANNTG-3') is a DNA sequence located in the promoter regions of genes controlled by bHLH TFs. The dimerization between MRFs and E2A proteins is thought to be the mechanism how MRFs regulate muscle differentiation.

Insulin-like growth factors (IGFs)

In the middle of the 20th century, two single-chain polypeptides with high similarities to insulin were identified and showed cell growth promoting activities. These factors were non-suppressible by insulin antibodies and thus called non-suppressible insulin-like activity (NSILA I and II). In 1976, Rinderknecht & Humbel proposed the designation of NSILAs and insulin, and named them insulin-like growth factor-I (IGF1) and insulin-like growth factor-II (IGF2). Over the past decades, the IGFs have received much attention by researchers, and several studies were conducted in order to understand the biological roles of IGFs in prenatal and postnatal growth, and their potential role in cancer development.

IGF2 in muscle development

IGFs are part of a complex system consisting of IGF1, IGF2, insulin, six types of IGF binding proteins (IGFBP-1 to -6) and three cell surface receptors: the insulin receptor (IR), the IGF1-receptor (IGF1R) and the IGF2-receptor (IGF2R). The interaction between IGFs and these components mediates the biological functions of IGFs, which subsequently influence various cellular processes including cell proliferation, differentiation and survival. The production of circulating IGFs in serum is stimulated via growth hormone (GH) and IGFs are mainly secreted from liver. Moreover, IGFs are produced locally in skeletal muscle tissue and non-muscle tissues. Thus, IGFs can carry out their function through both an endocrine and an autocrine/paracrine mode of action.

Several in vivo and in vitro studies demonstrated the important roles of IGFs in muscle development throughout the lifespan. For instance, Igf1-null or Igf2-null mice resulted in growth-deficient newborn mice (60% of normal birth weight). Moreover, the Igf1r-null mice were 45% of the normal size and died at birth. The double knock-out of both IGF1 and IGF2 produced mice with only 30% of the normal size. These results clearly demonstrated the fundamental role of IGF1 and IGF2 during embryogenesis, and also suggest that the majority of IGFs actions are mediated through IGF1R. This
was confirmed by the knock-out of IGF1R specifically in the mouse skeletal muscles, which resulted in defective skeletal muscle development with reduction in the number of muscle fibers\textsuperscript{83}.

IGFs interact with their receptors at different affinities (Table 2). The insulin receptor exists in two isoforms (IR-A and IR-B) due to alternative splicing of exon 11. IR-A is highly expressed in fetal and cancer cells, and binds both IGF2 and insulin at high affinities, but not IGF1\textsuperscript{84}. Moreover, IGF2 can bind IR-B (the isoform containing exon 11) but with lower affinity than the IR-A. The interaction between IGF2 and IR-A is important for normal embryonic development. Using transgenic mouse models, Louvi et al. showed that the growth promoting activities of IGF2 during embryogenesis are mediated through IGF1R and IR, while IGF1 functions exclusively through IGF1R\textsuperscript{85}. The IR-A and IR-B can be found as homodimers or heterodimers (IRA/IRB) when they are co-expressed in the same cell. This IRA/IRB hybrid receptor behaves similarly to IR-A, and insulin and IGF2, but not IGF1, can bind it at high affinities\textsuperscript{86}. Moreover, the IRA-B form functional hybrid receptors with IGF1R, and this can be either IGF1R/IR-A or IGF1R/IR-B. It has been suggested that these insulin-IGF1 hybrid receptors behave similarly to IGF1R\textsuperscript{87}. However, the factors regulating the formation of hybrid receptors and their functional roles in cellular development remain unclear\textsuperscript{88}. In contrast to IR and IGF1R, IGF2R does not have intrinsic signaling transduction capacity and binds IGF2 at high affinities, but not IGF1 or insulin. Functionally, IGF2R binds IGF2 and sequesters it from activating IGF1R, and targets it for degradation in lysosomes\textsuperscript{89,90}. Therefore, IGF2R is known as a clearance receptor for IGF2. Transgenic mice lacking the IGF2R exhibited overgrowth and an increased level of circulating IGF2 in serum and in tissues as well\textsuperscript{91}.

Beside the cell surface receptors, IGFBPs are involved in the modulation of IGFs function. The circulating IGFBPs can positively or negatively modulate the action of IGFs in complex ways. IGFBPs bind IGFs in serum and regulate the amount of IGF that is available for signaling, either by preventing the binding to IGF1R or by releasing more IGFs to bind the IGF1R. Moreover, the interaction between IGFs and the IGFBPs can prolong and increase the half-life of IGFs. For instance, the majority of serum IGFs exist in heterotrimeric complexes of ~150-kDa consisting of IGF1 or IGF2, IGFBP3 and acid-labile subunit (ALS). This large complex cannot cross the vascular endothelium and thereby restricts the IGFs to the circulation and increases their serum half-lives\textsuperscript{92–96}. Hence, muscle cells produce IGFBPs during the proliferation and differentiation at different amounts as a response to the expression level of IGFs in order to modulate the available IGFs for cell signaling\textsuperscript{97}.
Table 2. *The IGFs and the binding affinities to their receptors.*

<table>
<thead>
<tr>
<th>Hormone</th>
<th>Receptor</th>
<th>Affinity</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGF1</td>
<td>IGF1R</td>
<td>Highest</td>
</tr>
<tr>
<td></td>
<td>IGF2R</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>IR-A</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>IR-B</td>
<td>Low</td>
</tr>
<tr>
<td>IGF2</td>
<td>IGF1R</td>
<td>Intermediate</td>
</tr>
<tr>
<td></td>
<td>IGF2R</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>IR-A</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>IR-B</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Insulin</td>
<td>IGF1R</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>IGF2R</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>IR-A</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>IR-B</td>
<td>High</td>
</tr>
</tbody>
</table>

*IGF1 receptor (IGF1R), IGF2 receptor (IGF2R), insulin receptor isoform A (IR-A), insulin receptor isoform B (IR-B).*

IGF action is carried out through two different intracellular signaling pathways, phosphatidylinositol-3-kinase (PI3K) and mitogen-activated protein kinases (MAPK). Signaling by these two pathways regulate fundamental cellular processes, such as cell proliferation and differentiation. Inhibition of the MAPK pathway in myoblasts resulted in the inhibition of IGF-stimulated proliferation and improved the differentiation. In contrast, inhibition of the PI3K pathway caused a minor inhibition of myoblasts proliferation and a complete abrogation of differentiation. Further studies identified serine-threonine kinase Akt as a downstream component of the PI3K pathway in myoblast differentiation. Activation of Akt reversed the negative effects of PI3K inhibition on myotubes formation. These data suggest an essential role for the PI3K/Akt pathway in IGF-stimulated differentiation of myoblasts.

**IGF2 overexpression and cancer development**

The role of IGF2 in carcinogenesis has been extensively investigated. Several studies have demonstrated a positive correlation between increased serum level of IGF2 and the risk of developing various cancer types such as colorectal, breast, prostate and lung cancer. In general, these studies revealed strong evidence confirming that IGF2 overexpression promotes cancer development and progression. It has been shown that dysregulation of the DMR on the maternal allele of *Igf2* causes loss of imprinting (LOI) and leads to upregulation of *Igf2* expression. This overexpression of IGF2 promotes cell growth and anti-apoptosis processes via activation of the downstream signaling pathways such as AKT and MAPK/ERK1/2.
Overexpression of IGF2 due to LOI is found to be common in various cancers including colon\textsuperscript{56,105}, breast\textsuperscript{106}, oesophageal\textsuperscript{107} and ovarian cancer\textsuperscript{108} and acute myeloid leukaemia\textsuperscript{109}. Biallelic expression of Igf2 was also detected in Wilms tumor, a childhood kidney cancer\textsuperscript{102}. Moreover, a study on breast cancer has found a biallelic expression of IGF2 in 67\% of the patients\textsuperscript{110}. Experiments on animal models confirmed the important role of IGF2 in cancer development. For instance, transgenic animals overexpressing Igf2 developed earlier tumors and aggressive cancers\textsuperscript{111,112}. Furthermore, transgenic mice overexpressing Igf2 developed similar features as present in Beckwith–Wiedemann syndrome (BWS)\textsuperscript{113}.

LOI is not the only mechanism for overexpressing IGF2 in cancer development. Additional defects might be involved as well, including changes in promoter activities, receptor dysregulation or loss of repressor function\textsuperscript{114–116}. For instance, loss of function of IGF2R is an early evidence of tumorigenesis\textsuperscript{114}. In normal conditions, IGF2R acts as a clearance receptor for IGF2 that targets IGF2 to the lysosomes for degradation and regulates its availability. The mouse Igf2r gene is imprinted and is exclusively expressed from the maternal allele while the imprinting of human IGF2R is polymorphic and only occurs in 50 \% of the individuals\textsuperscript{117}. Genetic polymorphisms of IGF2R are linked to an increased risk of developing cancers, such as oral\textsuperscript{118} and colon cancer\textsuperscript{119}. Moreover, IGF2 signaling via IGF1R is implicated in tumor progression as well. An increased expression of IGF1R has been suggested to be associated with IGF2 overexpression in cancer cells that tend to metastasize\textsuperscript{120}.

Background to ZBED6 and ZC3H11A projects

The discovery of ZBED6

During last 60 years the consumer demand for lean pork meat with more protein and less fat led to intense selection for enhanced muscle growth and less fat deposition in pigs\textsuperscript{121}. As a consequence, the modern domestic pigs have a greater proportion of skeletal muscle than the European wild boar. In the beginning of the 1990s, the intercross between the European wild boar and Large White domestic pigs was generated\textsuperscript{122}. The pedigree was used to map quantitative trait loci (QTL) in order to identify the loci that have responded to this selection pressure. The project revealed that a paternally expressed QTL with a major effect on muscle mass mapped to the IGF2 locus\textsuperscript{123}. By 2003, a single nucleotide polymorphism (SNP) was discovered in intron 3 of the IGF2 gene in an evolutionarily conserved CpG island, transferring the QTL into a QTN\textsuperscript{124}. This mutation is associated with higher muscle growth rates and lower fat deposition. Pigs that carry this SNP had a
3-fold upregulation in IGF2 mRNA in postnatal skeletal muscle and heart but not in liver, which resulted in an increased meat production by 3-4%\textsuperscript{124}. Electrophoretic mobility shift assay (EMSA) revealed that an unknown nuclear factor binds the wildtype sequence but not the mutant sequence. Oligonucleotides corresponding to the wild-type and mutant QTN sequences and the stable isotope labeling of amino acids in cell culture (SILAC) technique were used to identify this unknown nuclear factor. Mass spectrometric analysis of the captured peptides revealed a previously unknown protein containing two zinc-finger BED domains and a hATC dimerization domain, thus this new protein was named ZBED6 (Figure 4).

ZBED6 is encoded by an open reading frame (ORF) located within intron 1 of the Zc3h11a gene. This ORF was previously identified as a splice variant of Zc3h11a but bioinformatics analysis showed no sequence similarity between the ZC3H11A and ZBED6 proteins. Further analysis revealed that ZBED6 is a domesticated DNA transposon that integrated in the mammalian genome more than 200-million years ago, and lost the ability to transpose. ZBED6 is specific to placental mammals\textsuperscript{125}. Immunocytochemistry (ICC) in mouse C2C12 myoblasts using an anti-ZBED6 antibody revealed a nuclear localization of ZBED6 with enrichment in the nucleolus. Chromatin immunoprecipitation sequencing (ChIP-seq) identified a strong ZBED6 peak over an intron of Igf2 (Figure 5). Furthermore, ChIP-seq analysis identified about 2,500 putative downstream targets of ZBED6 throughout the genome. Gene ontology (GO) analysis of these targets revealed significant enrichment
for genes associated with growth regulation, development, transcriptional regulation and cancer development\textsuperscript{125}.

**Figure 5.** ChIP-sequencing peak of ZBED6 around the intron of Igf2, overlapping with Igf2 opposite strand (Igf2os). The ZBED6 binding motif in blue. Adapted from Markljung et al. 2009\textsuperscript{125}.

**ZC3H11A**, the host gene of **ZBED6**

Zinc finger CCCH containing 11A (**ZC3H11A**) is a protein-coding gene located on chromosome 1 in the human and mouse genome. It contains 17 exons and encodes a poorly characterized zinc finger protein. Zinc finger proteins were first discovered in the middle of the 1980s in the *Xenopus* genome as DNA-binding proteins required to transcribe the 5S ribosomal RNA gene\textsuperscript{126,127}. Since then, several types of zinc finger proteins have been discovered. These proteins form zinc finger domains in their structure through binding a zinc ion to cysteine and histidine residues\textsuperscript{12}. These domains facilitate protein binding to nucleic acids. Therefore, zinc finger proteins are highly involved in DNA-protein, RNA-protein and protein-protein interactions\textsuperscript{128}. In eukaryotes, there are fourteen different families of zinc finger proteins, and considered as largest multigene family of TFs\textsuperscript{12}. Zinc finger proteins are classified into families based on the coordination of the zinc ion and the finger motifs. ZC3H11A belongs to the CCCH-type zinc finger protein family, which forms a zinc finger domain through binding of three cysteines (CCC) and one histidine (H) to a zinc ion, and thereby forms a zinc finger domain\textsuperscript{129}. This family is divided into six sub-families based on the number of copies of the CCCH-type zinc finger domain\textsuperscript{129}. The majority of these domains are of the C-X\textsubscript{7-8}-C-X\textsubscript{5}-C-X\textsubscript{3}-H type, where X is any amino
acid residue\textsuperscript{129}. ZC3H11A contains three tandem CCCH zinc finger domains, and belongs to a subfamily that consists of nine members, CPSF4, MKRN3, ZC3H4, ZC3H6, ZC3H7A, ZC3H7B, ZC3H8, ZC3H10 and ZC3H11A. Most of the members of this family are largely uncharacterized\textsuperscript{129}. However, zinc finger proteins are involved in essential molecular processes such as gene transcription, translation, mRNA processing, protein folding and chromatin remodeling.

The expression of both ZC3H11A and ZBED6 are initiated from an upstream common promoter, and the regulation of ZBED6 and ZC3H11A expression depends on intron retention (Figure 6)\textsuperscript{125}. When intron 1 remains un-spliced, the end product is only the ZBED6 protein due to the stop codon directly downstream of the ZBED6 coding sequence. The mechanism regulating alternative splicing of ZC3H11A and ZBED6 remains unknown. In this thesis, we aimed to also characterize the molecular function of ZC3H11A (paper V).

Figure 6. ChIP sequencing data for the mouse ZC3H11A/ZBED6 upstream region generated by the ENCODE project using the B-cell lymphoma (CH12) and erythroleukemia (MEL) cell lines, adapted from the UCSC Genome Browser (http://genome-test.cse.ucsc.edu/). The location of the RNA polymerase II binding site (Pol2) is indicated.
Aims of this thesis

I. Characterize the functional roles of ZBED6 in muscle growth using a myoblast cell line, and to explore to which extent ZBED6 relies on IGF2 signaling (Paper I & II).

II. Characterize the phenotypic and transcriptional changes associated with ZBED6 deletion in colorectal cancer cells (Paper III).

III. Establish and characterize ZBED6 and IGF2 transgenic mice and to investigate the phenotypic significance of the ZBED6-IGF2 locus (Paper IV).

IV. Characterize the molecular function of ZC3H11A that was found as the host-gene for ZBED6 (Paper V).
Present investigations

ZBED6 modulates the transcription of myogenic genes in mouse myoblast cells (paper I)

Background

ZBED6 was discovered as a nuclear protein that binds a site in intron 3 of the IGF2 gene, and represses IGF2 expression. A mutation in the binding site of ZBED6 in IGF2 disrupts the binding and leads to a 3-fold upregulation of IGF2 mRNA in pig muscle tissues. This finding suggests an important role of ZBED6 in muscle development. It is well established that IGF2 plays an important role in myogenesis and regulates myogenic factors. These observations encouraged us to investigate how ZBED6 influences the growth of muscle cells, and to identify downstream targets of ZBED6 in muscle cells. Moreover, does ZBED6 act only as a transcriptional repressor? Does ZBED6 require certain histone marks to affect the expression of target genes? In order to address these questions, we used mouse myoblasts and altered the expression of ZBED6, followed by transcriptome analysis.

Results and discussion

The mouse myoblast C2C12 cell line was used as a model for muscle cells. In this study, we altered the expression of ZBED6 in C2C12 cells and then evaluated the changes that occurred in the transcriptome. The Zbed6 transcript was targeted using three small interfering RNA (siRNA) oligonucleotides. These siRNA oligos were transiently transfected into cells using Lipofectamine 2000 transfection reagent. Thereafter, the silencing of ZBED6 was validated at the mRNA and protein level. Immunoblot analysis showed efficient silencing of the ZBED6 protein, and qPCR quantification detected more than 60% decrease in Zbed6 mRNA. The overexpression of Zbed6 was conducted by cloning the coding sequences of Zbed6 into the pTRE-tight expression vector. Transient transfection was performed and the overexpression was validated by western blot (WB) and qPCR analysis, which showed more than 5-fold overexpression of ZBED6.

ZBED6 is identified as a DNA binding protein that represses IGF2 expression. We aimed to explore the global changes in gene expression when Zbed6 is knocked-down in order to understand the regulatory role of ZBED6 in muscle cells. Therefore, we performed whole transcriptome analysis in Zbed6-silenced and control cells at two time points, two and four days post transfection. The SOLiD sequencing platform was used to
sequence RNA-seq libraries, and we obtained approximately 25 million reads uniquely mapped to the mouse genome (mm9). The expression analysis of Zbed6 and Igf2 in Zbed6-silenced cells revealed a 70% down-regulation of Zbed6 mRNA and a 2-fold upregulation of Igf2 mRNA. Moreover, we identified around 1,000 differentially expressed (DE) genes at day two and 4,000 DE genes at day four (FDR <0.001). This increase in the number of DE genes at day four indicates possible secondary effects of Zbed6 silencing. In order to scrutinize the direct effects of ZBED6 on gene expression, we focused on DE genes that showed consistent changes at both time points. Based on that, we identified 780 significantly DE genes when Zbed6 was knocked-down. Gene ontology (GO) analysis was performed using Database for Annotation, Visualization and Integrated Discovery (DAVID) to define which functional categories these DE genes belong to. GO analysis revealed significant enrichment for muscle function categories (FDR <0.05). The most significant categories were genes encoding muscle proteins, contractile fibers, myofibrils, or involved in heart development or muscle contraction. Furthermore, among the DE genes there were 20 small nuclear RNA (snoRNA) genes that showed up-regulation after Zbed6 silencing. Interestingly, five out of twenty DE snoRNAs were identified as ZBED6 targets and contained ZBED6 binding sites within 10 kb of the TSS. This suggests that ZBED6 is directly or indirectly involved in the repression of snoRNA expression.

ZBED6 has more than 2,500 putative target sites in C2C12 cells. In this study, we performed new ChIP-seq analysis and combined it with RNA-seq data to distinguish between the genes that are directly or indirectly regulated by ZBED6. We found that 28% of the DE genes had ZBED6 binding sites within 5kb of the transcription start site (TSS), and most of them were up-regulated when Zbed6 was knocked-down. This indicates that ZBED6 primarily functions as a transcriptional repressor rather than an activator, even though we cannot exclude the possibility that it may acts as an activator at some loci. Therefore, we investigated this possibility on the Twist2 gene that showed significant down-regulation in Zbed6-silenced cells. Twist2 encodes a basic-helix–loop–helix transcriptional factor that binds to the E-box consensus sequence and inhibits transcriptional activation of muscle-specific genes such as Myod1, Mef2a and Mef2c. ChIP-seq revealed a ZBED6 peak 1 kb upstream of Twist2. This site has the ZBED6 consensus motif, GGCTCG, that is highly conserved among all placental mammals. This interaction between ZBED6 and the binding site on Twist2 was validated by a luciferase reporter gene assay and an electrophoretic mobility shift assay.

Myoblasts differentiate into myotubes under the control of myogenic factors. One of these factors is Myogenin (Myog), which induce the differentiation of myoblasts when it is expressed. The proliferative myoblasts express both Igf2 and Myog at low levels, and both of them are highly induced during differentiation. We explored the effects of ZBED6 on
Igf2 and Myog expression during differentiation of C2C12 cells. WB analysis and qPCR quantification revealed a minor increase in ZBED6 expression during differentiation. Overexpression of ZBED6 led to 40-50% reduction in the expression of Igf2 and Myog mRNA in undifferentiated and differentiated cells compared with their expression in control cells. In contrast, transient silencing of Zbed6 resulted in more than 2-fold upregulation in Igf2 and Myog mRNA in undifferentiated and differentiated cells. This upregulation of Igf2 and Myog in undifferentiated cells remain small compared with the huge increase in their expression that is induced by differentiation (Figure 5, paper I). This interesting finding indicates that ZBED6 regulates the level of expressed Igf2 and Myog, and does not determine whether the genes are on or off, which means that ZBED6 functions as a transcriptional modulator of targeted genes that are already actively transcribed in myogenic cells.

In order to obtain further insight about how ZBED6 represses or activates the genes, we overlapped ZBED6 binding sites with ChIP-seq signal of histone marks in myoblasts. Interestingly, we identified a significant enrichment between ZBED6 binding sites and histone modifications H3K4me3, H3K4me2 and H3K27ac, which are known to be associated with active promoters. On the other hand, there was no overlap between ZBED6 binding sites and repressive histone mark H3K27me3. These results suggest that ZBED6 binds the active promoters and modulates the transcription level of the targeted genes.
Changes in myotube development and mitochondrial activity in mouse C2C12 cells after ZBED6 silencing is largely due to increased IGF2 expression (Paper II)

Background

In paper I, we used siRNA to achieve a transient down-regulation of Zbed6 in myoblasts. This resulted in about 2-fold up-regulation of Igf2 expression, along with hundreds of genes that were differentially expressed before and after silencing. IGF2 is a well-known growth-promoting factor that is involved in the proliferation and differentiation of muscle cells. Therefore, we aimed to explore to which extent phenotypic changes associated with altered expression of ZBED6 is due to its interaction with the Igf2 locus. To address this question, we employed the recently discovered genome editing tool CRISPR/Cas9 to introduce specific genetic modifications into mouse myoblast C2C12 cells. In paper II, we generated two models of engineered C2C12 cells, a Zbed6 knock-out and a deletion of the ZBED6 binding site in the first intron of the Igf2 gene. These two models allowed us to study the effect of complete deletion of ZBED6 during myoblast differentiation.

Results and discussion

Almost the entire coding sequence of Zbed6 (2.5 kb out of 2.9 kb) was deleted in C2C12 cells using the CRISPR/Cas9 system guided with two specific guide RNAs (gRNA) as indicated in figure 7A. The genotyping of single clones revealed an efficient targeting of Zbed6 (Figure 7A). The ZBED6 binding motif 5′-GGCTCG-3′ in the first intron of Igf2 was targeted as well using a gRNA that cuts between C and T nucleotides in the binding site (Figure 7B). After genotyping several single clones, we obtained a clone with a four-nucleotide deletion (GGCT) of the binding motif in both alleles. Thus, we named it Igf2_{GGCT} and used it in the downstream analysis together with the Zbed6\(^{-/-}\) clones. The validation of both clones showed a complete deletion of ZBED6 in the Zbed6\(^{-/-}\) clones and normal expression in the Igf2_{GGCT} clone. Furthermore, both the Zbed6\(^{-/-}\) and Igf2_{GGCT} clones exhibited more than 30-fold up-regulation of Igf2 mRNA in comparison to the wild-type (WT) C2C12 cells (Figure 7D). In order to validate that the elevated expression of Igf2 mRNA in transgenic cells was due to the deletion of ZBED6 or its binding site, we generated a GFP-ZBED6 expression construct and introduced it into the transgenic cells. The transient overexpression of GFP-ZBED6 in Zbed6\(^{-/-}\) clones significantly down-regulated the elevated expression of Igf2, while no changes were observed when GFP-ZBED6 was over-expressed in the Igf2_{GGCT} clone. These results
indicate that ZBED6 represses Igf2 expression exclusively through interaction with its binding site in Igf2 intron 1. The real-time measurement of the growth rate revealed that both the Zbed6−/− and Igf2ΔGGCT clones grow faster than the WT cells.

Figure 7. (A) Schematic description of Zbed6 targeting using CRISPR/Cas9. Red scissors indicate the targeted sites of Zbed6 using two guide RNAs. Blue arrows indicate the location of the PCR primers that were used for genotyping of the KO clones (below). (B) Schematic description of the targeted ZBED6 binding sequences in Igf2 (bold), the scissor indicates the cleavage site using specific gRNA sequences (yellow) adjacent to the PAM sequences (blue). Black arrows indicate Igf2 promoters, red boxes are the coding sequences of Igf2. (C) Immunoblot validation of Zbed6−/−, Igf2ΔGGCT and WT cells. (D) Expression analysis of Igf2 mRNA by qPCR in the indicated cells.
Myoblasts can be induced to differentiate and form myotubes. Therefore, we evaluated the differentiation profile of Zbed6−/− and Igf2^ΔGGCT^ myoblasts. Both Zbed6−/− and Igf2^ΔGGCT^ myoblasts formed hypertrophic myotubes, with significant increase in the differentiation index in comparison to WT cells. The immunofluorescence (IF) staining of the differentiated myotubes showed increased expression of myosin heavy chain, myogenin and α-sarcomeric actin in Zbed6−/− and Igf2^ΔGGCT^ myotubes. In contrast, transient overexpression of ZBED6 (ZBED6-OE) in WT myoblasts impaired differentiation and resulted in poor myotube formation. This was associated with a remarkable reduction in cells expressing myogenin and myosin.

In order to explore the potential transcriptional targets of ZBED6 during myogenesis, we performed transcriptome analyses of Zbed6−/−, Igf2^ΔGGCT^ and WT cells before and after differentiation into myotubes. We focused our analysis on genes that showed differential expression in both Zbed6−/− and Igf2^ΔGGCT^ myotubes in order to explore to which extent the observed changes in gene expression in Zbed6−/− clones are secondary effects due to increased IGF2 expression. Counting of uniquely mapped reads identified ~12,500 expressed genes with at least one read count per million (CPM). Expression analysis of differentiated Zbed6−/− and of Igf2^ΔGGCT^ myotubes identified ~2,600 DE genes in each of them (log fold change >0.5; P<0.05), in comparison to WT myotubes. The dissection of the DE genes based on the direction of the change revealed a highly significant ~30% overlap between DE genes in Zbed6−/− and Igf2^ΔGGCT^ myotubes. The GO analysis of the up-regulated genes in both Zbed6−/− and Igf2^ΔGGCT^ myotubes revealed a significant enrichment of muscle-specific genes. This included genes for myosin heavy and light chains, troponins, myomesins, alpha actinin, leiomodin, titin and myoglobin. In contrast, genes involved in cell division and cell cycle regulation, such as cyclins, were enriched among down-regulated genes in both Zbed6−/− and Igf2^ΔGGCT^ myotubes.

Furthermore, transcriptome analysis of ZBED6-OE and control cells after three days of differentiation revealed ~1,500 down-regulated and ~1,100 up-regulated DE genes in ZBED6-OE versus the control cells. The GO analysis of down-regulated genes presented a significant enrichment of muscle-specific genes, while the up-regulated genes were primarily related to cell cycle regulation and cell division. Obviously, this is the opposite trend compared with Zbed6−/− cells. Moreover, we identified 463 genes that were significantly down-regulated in ZBED6-OE cells and significantly up-regulated in Zbed6−/− myotubes, which represent about 40% of the up-regulated genes in Zbed6−/− myotubes. The GO analysis of those 463 genes revealed a striking enrichment in the muscle-specific categories. Furthermore, the enriched KEGG pathways included cardiac muscle contraction, hypertrophic cardiomyopathy (HCM), and calcium, insulin and AMPK signaling. Interestingly, the key components of these pathways were found to be up-regulated in Igf2^ΔGGCT^ myotubes as well. For instance, the expression of phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic...
subunit beta (Pik3cb) and AMP-activated protein kinase alpha2 (Prkaa2), beta2 (Prkab2) and gamma3 (Prkag3) subunits were found to be up-regulated in Zbed6−/− and Igf2AGGCT myotubes, and down-regulated in ZBED6-OE differentiated cells. All these data together strongly indicate that the ZBED6-IGF2 interaction has an essential role in muscle development and influences muscle metabolism.

In contrast to Zbed6−/−, transient ZBED6-OE leads to significant down-regulation of Igf2 mRNA and reduced cell viability. Furthermore, we attempted to establish a stable ZBED6-OE cell line but we did not succeed. The cell cycle analysis of ZBED6-OE versus the control cells displayed a significant difference in the proportion of the cells in different cell cycle phases. Around 82% of the ZBED6-OE cells were found in the G0/G1 phase and 10% in S-phase, whereas the corresponding proportions in control cells were 58% and 35%, respectively. The transcriptome analysis of undifferentiated ZBED6-OE myoblasts identified thousands of DE genes. The GO of down-regulated DE genes revealed a significant enrichment in genes involved in cell cycle and cell division processes. Out of 98 down-regulated cell cycle–related genes, 21 genes were previously identified as putative direct targets of ZBED6125,132.

The altered expression of ZBED6 leads to significant changes in cell proliferation and differentiation, accompanied with differential expression of IGF2. On the other hand, the correlation between mitochondrial activities and cell proliferation is well established133–135. This encouraged us to look closer at mitochondrial activities in response to ZBED6-overexpression or ablation.

Using the MitoTracker staining of active mitochondria in living cells, we detected a significant reduction in mitochondrial mass in ZBED6-OE cells and an increase in mitochondria in Zbed6−/− cells, while no change was observed in Igf2AGGCT cells. Moreover, the mitochondrial oxidation rate (OCR) assay revealed an increased respiration rate in Zbed6−/− and Igf2AGGCT myoblasts. This indicates a possible role of IGF2 in the increased respiration since both cell models overexpress IGF2. This hypothesis was investigated by measuring the IGF2 dose-response on oxygen consumption rate of WT cells. Treatment with high concentrations of recombinant IGF2 resulted in a 2-fold increase in the respiration rate of myoblasts. This indicates a positive correlation between the amount of IGF2 protein in myoblasts and the oxygen consumption rate.

In summary, this study conclusively demonstrates that the interaction between ZBED6 and its binding site in Igf2 plays an essential role in regulating the development of myogenic cells. Furthermore, the disruption of the ZBED6 binding site in Igf2 resulted in very similar phenotypic effects as observed in the Zbed6 deletion. This suggests that the phenotypic effects caused by Zbed6 inactivation in myoblast cells are largely mediated through the regulation of Igf2 expression.
The transcriptional modulator ZBED6 affects cell cycle and growth of human colorectal cancer cells (paper III)

Background

Overexpression of IGF2 is found to be associated with cancer progression and increased growth of tumor cells. IGF2 action is carried out through the activation of the PI3K/Akt signaling pathway, which results in enhanced cell growth and mitogenesis. For instance, somatic mutations in PI3K pathway members such as PIK3CA and PTEN are found in late colorectal cancer progression and leads to increased tumor cell growth and invasivity. As described before, ZBED6 represses the expression of IGF2 through its binding site in an intron of IGF2. ChIP-seq analysis of mouse cells revealed thousands of ZBED6 binding sites, with enrichment in genes involved in cell cycle, proliferation, development and tumorigenesis. The regulatory role of ZBED6 on IGF2 expression, plus the importance of IGF2 and PI3K signaling pathways in cancer development encouraged us to investigate the biological function of ZBED6 in tumorigenesis. For that purpose, we knocked-out ZBED6 in two colorectal cancer cell lines, RKO and HCT116, and then investigated the phenotypic and transcriptional changes associated with ZBED6 deletion.

Results and discussion

A recombinant adeno-associated virus (rAAV) gene targeting construct was used to generate ZBED6-/- knock-out cell lines. The coding sequences of ZBED6 was targeted by (i) replacing Q173 with a stop codon (CAA>TAA), (ii) deleting a single nucleotide (GCA) of A178 and (iii) inserting a scar of LoxP sequences containing several stop codons. We targeted both ZBED6 alleles in HCT116 and RKO cell lines, and validated the loss of ZBED6 by immunoblot analysis. The independent ZBED6-/- knock-out clones were selected based on complete deletion of ZBED6, and intact expression of ZC3H11A. Based on these criteria we selected three ZBED6-/- knock-out clones of the HCT116 and RKO cells and used them in this study. The knock-out of ZBED6 resulted in up-regulation of IGF2 expression in both RKO and HCT116 cell lines.

We evaluated the phenotypic consequences of ZBED6 knock-out in colorectal cancer cells by measuring the growth rate and cell cycle regulation. The parental RKO and HCT116 cells together with three ZBED6-/- knock-out clones from each cell line were assessed by real-time measurement of growth rate. The ablation of ZBED6 in HCT116 cells had a growth-inhibitory effect. In contrast, the knock-out of ZBED6 in RKO
resulted in a consistent increase in the growth rate (Figure 8A). This increased proliferation in ZBED6−/− RKO clones were suppressed by re-expressing ZBED6 in the knock-out clones using an AcGFP-Zbed6 construct. ZBED6−/− RKO cells stably expressing AcGFP-Zbed6 grew slower than both the parental and ZBED6−/− RKO cells. Furthermore, the cell cycle analysis of RKO cells revealed higher fraction of cells in the S-phase in ZBED6−/− RKO clones than in their parental cells.

Whole transcriptome analysis was performed using three independent ZBED6−/− knock-out clones of both RKO and HCT116 cells. Alignment with TopHat (2.0.10) identified ~11,000 expressed genes in RKO and ~11,900 genes in HCT116. Differential expression analysis using Cufflinks revealed ~2,700 and ~2,200 DE genes reaching the significance level (P<0.05) of multiple testing analyses in RKO and HCT116, respectively. Shared DE genes between the cell lines were ~600 genes. The dissection of DE genes based on the direction of the change showed ~1,300 up-regulated and ~1,400 down-regulated DE genes in RKO versus ~1,400 up-regulated and ~850 down-regulated DE genes in HCT116. GO analysis displayed a significant enrichment of genes associated with cell cycle regulation among the up-regulated genes in ZBED6−/− RKO clones. On the other hand, genes involved in negative regulation of cell proliferation and cell death tended to be up-regulated in ZBED6−/− HCT116 clones (Figure 8B). These patterns of altered transcription are in agreement with the striking differences in phenotypic change after ZBED6 deletion, whereas RKO and HCT116 cells exhibit increased and decreased growth rate, respectively.

In order to identify the putative direct targets of ZBED6 in colorectal cancer cells, we performed ChIP-seq analysis in parental HCT116 cells. ChIP-seq analysis identified more than 7,000 peaks overall in the genome, 70% of the peaks were found within 1 kb of a TSS. Moreover, ZBED6 binding sites were significantly enriched at the up-regulated DE genes versus down-regulated genes, which imply a direct effect of ZBED6 on several up-regulated genes. This observation is in agreement with the previous identification of ZBED6 as a transcriptional repressor. In this study we aimed to identify the genes that are directly regulated by ZBED6 in colorectal cancer. Therefore, we combined the RNA-seq with ChIP-seq data and selected candidate genes that had ZBED6 binding sites close to their TSS, and that were up-regulated ≥ 1.5-fold in both RKO and HCT116 knock-out clones (Table 3). For experimental validation of putative direct ZBED6 targets, we picked out ten genes (ARL4C, FOSL2, MYBL1, PMEPA1, ROCK2, SGK1, SPTBN1, TCF7, WWC1, WWTR1) that all had strong ZBED6 peaks at their respective TSS, and we validated their expression change by qPCR in RKO and HCT116 cells. As a rescue experiment, we knocked-down ZBED6 in parental HCT116 cells using siRNA against the ZBED6 transcript. We obtained about 50% down-regulation in ZBED6 mRNA and expression changes similar to those found in the RNA-seq of RKO and HCT116 ZBED6−/− knock-out clones (Figure 9).
In this study, we showed that the inactivation of ZBED6 influenced the growth of cancer cells and led to dramatic changes in the transcriptome, with high proportion of secondary effects caused by IGF2 upregulation. Furthermore, ZBED6 appears to directly regulate a subset of genes involved in important pathways for cancer growth. For instance, ARL4C$^{139}$, WWTR1/TAZ and WWCl/KIBRA$^{140}$ are involved in the Hippo pathway; ARL4C and TCF7 are involved in the Wnt pathway; SPTBN1$^{141}$, PMEPA1$^{142}$ and ROCK2$^{143}$ are involved in the TGFβ pathway and MYBL1$^{144}$ is involved in the cell cycle regulation pathway$^{145}$. These findings together suggest an important role of ZBED6 in tumor development, acting as a transcriptional modulator. Thus, its effect depends on the transcriptional state of the cells.
Table 3. Candidate genes as direct ZBED6 targets in RKO and HCT116 cells, based on RNA-seq and ChIP-seq data.

<table>
<thead>
<tr>
<th>Genes</th>
<th>Fold-change</th>
<th>ChIP-seq</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RKO</td>
<td>HCT116</td>
</tr>
<tr>
<td><strong>Up-regulated genes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPTBN1</td>
<td>1.5</td>
<td>2.2</td>
</tr>
<tr>
<td>COL13A1</td>
<td>4.3</td>
<td>2.1</td>
</tr>
<tr>
<td>WWC1</td>
<td>1.6</td>
<td>2.3</td>
</tr>
<tr>
<td>WWP1</td>
<td>1.6</td>
<td>1.5</td>
</tr>
<tr>
<td>TCF7</td>
<td>2.0</td>
<td>2.3</td>
</tr>
<tr>
<td>PMEPA1</td>
<td>2.3</td>
<td>12.0</td>
</tr>
<tr>
<td>ARL4C</td>
<td>3.0</td>
<td>2.7</td>
</tr>
<tr>
<td>MYBL1</td>
<td>2.0</td>
<td>2.5</td>
</tr>
<tr>
<td>FOSL2</td>
<td>1.8</td>
<td>1.9</td>
</tr>
<tr>
<td>ROCK2</td>
<td>1.5</td>
<td>1.7</td>
</tr>
<tr>
<td>KCTD1</td>
<td>1.5</td>
<td>1.7</td>
</tr>
<tr>
<td>MAP3K14</td>
<td>1.6</td>
<td>1.7</td>
</tr>
<tr>
<td>SGK1</td>
<td>1.7</td>
<td>3.3</td>
</tr>
<tr>
<td>TUBB3</td>
<td>2.8</td>
<td>2.0</td>
</tr>
<tr>
<td>MB2I12</td>
<td>2.0</td>
<td>2.3</td>
</tr>
<tr>
<td>CRIM1</td>
<td>3.2</td>
<td>2.4</td>
</tr>
<tr>
<td><strong>Down-regulated genes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPAT</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>UPP1</td>
<td>0.0</td>
<td>0.5</td>
</tr>
<tr>
<td>DDIT4</td>
<td>0.4</td>
<td>0.2</td>
</tr>
<tr>
<td>ANKRD50</td>
<td>0.3</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Figure 9. Expression of ZBED6 and putative direct targets in HCT116 cells after siRNA knock-down of ZBED6 measured by qPCR. Error bars indicate SD. *P < 0.05; **P < 0.01; ***P < 0.001.
The ZBED6-IGF2 axis has a major effect on growth of skeletal muscle and internal organs in placental mammals (Paper IV)

Background

In the first three papers, we used mammalian cell lines as a model to characterize the molecular function of ZBED6. The repression effect of ZBED6 on IGF2 expression was consistent in mouse and human cell lines. Transient silencing or complete inactivation of ZBED6 resulted in significant upregulation of IGF2 expression. In this paper, we have established transgenic mouse models to characterize the significance of the ZBED6-IGF2 interaction in the growth of placental mammals.

Results and discussion

We generated two transgenic mouse models, ZBED6 knock-out (Zbed6-/-) and IGF2 knock-in mice (Igf2pAmG) and crosses between them. The Igf2pAmG mice carry exactly the same mutation as identified originally as a QTN in IGF2 mutant pigs with a single nucleotide substitution (G > A) in the ZBED6 binding site124 (Figure 10). Using this experimental design, we are able to evaluate which phenotypic effects present in the Zbed6-/- mice are caused by altered IGF2 regulation. Both Zbed6-/- and Igf2pAmG mice were viable and did not exhibit any obvious malformations. The phenotypic characterization of the Zbed6-/- and Igf2pAmG mice showed a higher growth rate, higher muscle growth and larger heart, kidney and liver. These phenotypic effects were more pronounced in females than in males, probably because testosterone is promoting growth in males. These phenotypic changes were accompanied with a very striking up-regulation of Igf2 expression (~30-fold) in all tissues that were examined. Interestingly, nearly identical phenotypic changes were observed both as regards to growth and weight of internal organs and increased Igf2 expression in the Zbed6-/- and Igf2pAmG mice, and combining the two genetic changes did not result in more extreme phenotypes.

Transcriptome analysis of skeletal muscles of Zbed6-/- and Igf2pAmG mice revealed a significant upregulation of Igf2 and Igf2os transcripts, while no significant changes were detected in the expression of H19. The expression of ~100 genes were found to be significantly changed in each of the Zbed6-/- and Igf2pAmG mice. Genes involved in cell signaling and metabolic processes were found among the differentially expressed genes. MicroRNA-seq of Zbed6-/- muscles showed a remarkable upregulation of miR-483. This
microRNA is derived from an intronic region in IGF2 and acts as a positive regulator of IGF2 expression\textsuperscript{146}.

The expression analysis of Igf2 in embryonic muscles indicated a minor role of the ZBED6-IGF2 mechanism in regulating Igf2 prenatally. RNA-seq analysis of muscle tissues from Zbed6\textsuperscript{-/-} embryos only showed modest changes in Igf2 expression. This is in contrast to what we have observed in adult muscles of Zbed6\textsuperscript{-/-} mice.

![Diagram](image)

**Figure 10.** Strategy for generating Zbed6\textsuperscript{-/-} and Igf2 mutant mouse models using homologous recombination in ES cells. (A) Two homology arms were used to insert loxP sites flanking the Zbed6 coding sequence (red). The floxed mice were crossed with Cre-expressing mice to delete the sequences between the loxP sites. (B) The Igf2 mutant mouse model (KI) carried a single nucleotide substitution G to A in the ZBED6 binding site GGCTCG in the CpG island (green box) located in the first intron of Igf2. Blue arrows indicate Igf2 promoters (P1-3). Two loxP sites and a single FRT remained in the transgene used in the present study. ChIP-seq data from paper I to pinpoint the binding site of ZBED6 over Igf2 in the mouse genome\textsuperscript{132}.

Furthermore, the expression of Enho and Tmem8 were significantly upregulated in both fetal and adult Zbed6\textsuperscript{-/-} muscles. Both Enho and Tmem8 are putative direct targets of ZBED6 based on ChIP-seq analysis\textsuperscript{132}, which indicate that ZBED6 acts as a repressor in fetal tissues as well. These results imply that the minor effect of ZBED6 on regulating Igf2 expression in fetal tissues is due to the action of its microRNA target, miR-483.
tissues is locus-specific, because ZBED6 may not be able to access its binding site in \( \text{Igf2} \) to repress \( \text{Igf2} \) expression in embryonic tissues. The isoform quantification analysis revealed that the expression of \( \text{Igf2} \) in fetal tissues was arising mainly from the P2 promoter. Surprisingly, the disruption of the ZBED6-IGF2 interaction led to increased transcriptional activities from both the P1 and P2 promoters in adult muscles (Figure 11).

The phenotypic changes caused by \( \text{Zbed6} \) inactivation and the \( \text{Igf2} \) regulatory mutation in transgenic mice showed both high similarities and differences compared with domestic pigs carrying the \( \text{IGF2} \) mutation\textsuperscript{124}. The increased muscle growth is consistent between the mutant pigs and transgenic mice. Moreover, there was an excellent agreement between transgenic mice and mutant pigs as regards the minor role for ZBED6 in regulating \( \text{Igf2} \) expression in fetal tissues in contrast to the major role in postnatal tissue. Taken together, this study demonstrates the ZBED6-IGF2 axis as a novel mechanism that is contributing to the regulation of the \( \text{IGF2} \) expression, which in turn results in a major effect on growth of skeletal muscle and internal organs in placental mammals.

**Figure 11.** Differential expression of \( \text{Igf2} \) isoforms generated from different \( \text{Igf2} \) promoters (P1-3) in fetal tissues (left) and adult muscle (right) of WT and mutant mice.
Multiple viruses rely on the stress-induced protein ZC3H11A for efficient replication (Paper V)

Background

The ZBED6 gene is located in one of the first introns of ZC3H11A, and the two genes are transcribed from the same promoter. The expression of the two alternative transcripts is determined by intron retention that allows ZBED6 expression. The zinc-finger protein ZC3H11A is well-conserved among vertebrates but its function has remained largely unknown. In this study we knocked-out ZC3H11A from human cell lines using the CRISPR/Cas9 system and explored the potential molecular function for this gene.

Results and discussion

We used the CRISPR/Cas9 system to knock out the ZC3H11A gene in HeLa cells to study its function. A specific guide RNA targeting a region in exon 6 of ZC3H11A was used to introduce a frame shift mutation in the coding sequences. The ZC3H11A−/− clones were validated by immunoblot analysis (Figure 12). Deletion of ZC3H11A did not show a distinct phenotype under normal growth conditions, as it was measured by real-time growth rate. It has been reported that some of CCCH zinc finger proteins are accumulated in the stress granules as a response of cellular stress147. Therefore, we challenged the ZC3H11A−/− cells in various ways, including virus infection and heat-shock stress. We infected the ZC3H11A−/− cells with human adenovirus (HAdV-5), a well characterized virus that infects a wide range of human cells and put the host cells under huge stress. Interestingly, the production of some adenovirus capsid proteins was drastically reduced in ZC3H11A−/− cells compared to WT cells and resulted in a three-fold reduction in virus titre in ZC3H11A−/− cells. The endogenous expression of ZC3H11A mRNA did not show any increase, while ZBED6 mRNA levels were significantly down-regulated in HAdV-5 infected cells. Furthermore, the steady-state amount of the ZC3H11A protein increased in WT cells during virus infection. This finding was unexpected since adenovirus elicits an efficient shut down of most host cell proteins during infection. Based on this initial discovery we initiated a detailed characterization of ZC3H11A and its role during virus infection. In order to explore if this host-virus interaction is unique for adenovirus, we infected the ZC3H11A−/− cells with other viruses including the influenza H1N1 strain A/WSN/33, two human immunodeficiency virus strains (HIV-1 IIIB, HIV UG29A) and human
herpes simplex virus type 1 (HSV-1). All tested viruses showed a significant reduction in their growth in \( ZC3H11A^{-/-} \) cells.

The host-virus interaction was explored in some detail with regards to an HAdV-5 infection. We compared the localization of ZC3H11A in both uninfected and HAdV-5-infected HeLa cells. ZC3H11A primarily localized in the nucleus and accumulated in so-called nuclear speckles, which are associated with RNA splicing proteins. During HAdV-5 infection, ZC3H11A changed its localization and surrounded the nuclear regions where viral replication takes place. This remarkable change in the localization was verified using a construct encoding a GFP-tagged ZC3H11A (GFP-ZC3H11A) to exclude any artifact from the primary anti-ZC3H11A antibody (Figure 13). The co-localization of ZC3H11A with SRSF2 (also named human splice factor SC35) in nuclear speckles indicates that ZC3H11A is an RNA-binding protein. Two recent large-scale studies have indicated that ZC3H11A is associated with the TRanscription-EXport (TREX) complex that plays a crucial role for mRNA processing and nuclear export of a subset of mRNAs\(^{148,149}\). TREX complex proteins selectively bind RNA transcripts and exports them to the cytoplasm. Thus, we performed HITS-CLIP (High-Throughput Sequencing of RNA isolated by CLIP) analysis in order to identify the RNA molecules associated with ZC3H11A. The HITS-CLIP results showed an exclusive interaction between ZC3H11A and protein-coding transcripts, with enrichment in the coding sequences and 3’UTR of mRNAs. Moreover, the data showed that ZC3H11A binds short purine-rich sequences. HITS-CLIP analysis of HAdV-5 infected cells indicated that ZC3H11A directly bind the 5’UTR and protein-coding sequences of the major late mRNA transcripts of HAdV-5. Transcriptome analysis of HAdV-5 mRNA revealed a significant differential expression of certain viral transcripts in the cytoplasm of \( ZC3H11A^{-/-} \) cells. For instance, 52,55K and...
Fiber were down-regulated while hexon did not show differential expression in ZC3H11A−/− cells. Fluorescent in situ hybridization (FISH) analysis of poly(A)+ RNA indicated that the ZC3H11A silencing resulted in a significant decrease in poly(A)+ RNA accumulation in the cytoplasm.

Figure 13. Cellular localization of ZC3H11A. Immunofluorescence staining of wild-type HeLa cells using antibodies against ZC3H11A and SRSF2 before and after infection with HAdV-5 (left). Immunofluorescence staining with GFP-ZC3H11A and an anti-72k adenovirus protein antibody as a marker for HAdV-5 replication centers (right).

In addition to being up-regulated during HAdV-5 infection, we also found that ZC3H11A expression is induced under other stress conditions such as heat shock. Heat shock at 42°C for 1 hour showed a two-fold increase in endogenous ZC3H11A expression in WT cells without an increase in mRNA levels, suggesting that the effect was caused by protein stabilization. This interpretation was supported by the observation that a similar accumulation of the ZC3H11A protein is observed after treatment with a proteasomal inhibitor (MG132). A previous large-scale proteomic study supported our finding that ZC3H11A is induced by heat stress and further suggested that ZC3H11A becomes one of the most highly SUMOylated proteins in human cells after heat shock. We have confirmed that ZC3H11A is SUMOylated after heat stress and during virus replication, which suggests SUMOylation may function to protect ZC3H11A from degradation, explaining why we observe a higher amount of the ZC3H11A protein during virus infection and after heat stress.

In summary, these results suggest that ZC3H11A is critical for maintaining nuclear export of mRNA transcripts during stressful conditions, and therefore nuclear replicating viruses take advantage of this mechanism to promote their own replication. This study establishes ZC3H11A as a validated drug target for the development of novel antiviral therapies.
Future perspectives

The overall goal with this thesis was to explore the molecular function of the recently discovered transcriptional factor ZBED6 and the previously uncharacterized ZC3H11A protein. In the first part of the thesis, we used mammalian cell lines and transgenic animal models in order to gain insight into the molecular functions of ZBED6 (papers I-IV). The second part was focused on the characterization of the ZBED6 “host” gene ZC3H11A (paper V).

Our experiments on mouse and human cell lines revealed remarkable phenotypic changes caused by the alteration of ZBED6 expression (papers I-III). These changes were associated with differential expression of hundreds of genes associated with essential cellular pathways involved in regulating cell growth and differentiation\textsuperscript{132,151}. An important question here is how much of these changes in the transcriptome are transmitted into the proteome of the cells? Therefore, we plan to run large scale proteome analyses and measure the downstream effects of ZBED6 on protein expression and signaling pathways. The ZBED6 knock-out cell lines can be used for stable isotope labelling (SILAC) followed by high-resolution mass spectrometry (LC-MS/MS) quantification. Furthermore, identification of ZBED6 interacting proteins is of considerable interest to understand how ZBED6 acts as a transcriptional repressor or activator. This can be explored by immunoprecipitation of GFP-tagged ZBED6 followed by MS identification.

The precise mechanism of how ZBED6 mediates transcriptional silencing remains unknown. The enrichment of ZBED6 in nucleoli suggests that it may mediate transcriptional silencing by moving the targets into the nucleolus. We examined this hypothesis by cloning the wildtype and mutant IGF2 QTN (the 145 bp around the ZBED6 binding site in IGF2) and P3 promoter into an episomal vector system and generating a stable cell line. Thereafter, we used a fluorescence DNA probe targeting the episomes to explore the localization patterns and investigate if ZBED6 moves its targets into the nucleolus. Counting the fluorescent signals from episomes carrying the wild-type or mutant IGF2 QTN allele revealed a significant enrichment of wild-type episomes in the nucleoli. These observations imply an important role of the nucleolar localization of ZBED6. These findings encouraged us to get a deeper insight into the mechanism of the ZBED6 action. The ZBED6 protein structure consists of six domains: the nucleolus localization domain (NoLS), the nuclear localization domain (NLS), two DNA-binding domains (BED1 and BED2), the hAT dimerization domain and Ribonuclease H-like domain\textsuperscript{152}. We have generated a series of deletion constructs targeting these domains. Thereby, we can investigate which
domain is required for ZBED6 action by comparing the functionality of the truncated constructs in repressing the targets of ZBED6.

In paper IV, we have established two transgenic mouse models, ZBED6 knock-out (Zbed6<sup>−/−</sup>) and IGF2 knock-in mice (Igf2<sup>αA/mG</sup>) with a single nucleotide substitution in the ZBED6 binding site in Igf2. The phenotypic characterization showed an increase body weight, muscle mass and internal organ size of the transgenic mice. The disruption of the ZBED6-IGF2 interaction resulted in a massive up-regulation of Igf2 expression. Surprisingly, ZBED6 ablation in the fetal tissues did not affect the expression of Igf2 to the same extent as in adult tissues. These observations proposed that either ZBED6 is functionally inactive in fetal tissues, or that it is not able to access the binding site in the IGF2 locus. The first possibility was explored by measuring the expression of other putative direct ZBED6 targets such as Enho and Tmem8 in fetal tissues. Interestingly, both of them were up-regulated after ZBED6 deletion both in fetal and adult tissues. Therefore, the second possibility could be that the epigenetic changes at the IGF2 locus interfere with ZBED6, and prevents it from the binding in fetal tissues. This raises an important question; does DNA methylation determine whether ZBED6 can get access to the binding site or not? This hypothesis can be examined by running bisulfite sequencing of the ZBED6 binding sites in fetal and adult tissues to investigate if changes in DNA methylation are an important mechanism regulating how ZBED6 can interact with its targets.

Post-transcriptional regulators of gene expression such as microRNAs control many developmental and cellular processes in eukaryotes. MicroRNAs are non-coding small RNAs that contain ~21 nucleotides in length and modulate the proteome by annealing to the 3'-UTR of target mRNAs and inhibiting protein translation or promoting mRNA instability. These microRNAs are often encoded from the intronic regions of protein-coding genes. The microRNA sequencing of skeletal muscle tissues of Zbed6<sup>−/−</sup> mice revealed an upregulation of miR-483 microRNA, which is derived from an intron of Igf2. It has been reported that the human miR-483 acts as a positive regulator of IGF2 expression and enhances tumorigenesis. Considering the huge upregulation of Igf2 expression after ZBED6 inactivation in both cell lines and animal tissues, follow-up studies are required to explore the feedback between ZBED6 and miR-483 in regulating Igf2 expression. Therefore, we plan to knock-out the corresponding sequences of miR-483 in the presence and absence of ZBED6 binding, and characterize the changes on Igf2 expression. This model will allow us to obtain further insight into the transcriptional regulation of Igf2. The phenotypic characterization of Zbed6<sup>−/−</sup> and Igf2<sup>αA/mG</sup> mice showed an increased growth of internal organs. On the other hand, overexpression of IGF2 is found to be associated with several types of tumors. Thus, it is important to explore the risk of developing tumors in Zbed6<sup>−/−</sup> and Igf2<sup>αA/mG</sup> mice. This can be accomplished by establishing an intercross between Zbed6<sup>−/−</sup> or Igf2<sup>αA/mG</sup> mice and mouse cancer models, for example a p53-
deficient mouse model. Characterizing these mouse models will allow us to investigate the potential role of the ZBED6-IGF2 axis in cancer development.

These findings suggest that the interaction between ZBED6-IGF2 may be an important therapeutic target for human diseases where anabolism is impaired. Hence, we will screen for small molecules that alter the expression level of ZBED6 or interfere between ZBED6-IGF2.

In the second part of the thesis we aimed to characterize the molecular function of ZC3H11A. In paper V, we showed that ZC3H11A is critical for maintaining nuclear export of mRNA and translation during stressful conditions, and that several nuclear replicating viruses take advantage of this mechanism to facilitate their replication. Further studies are required to reveal the cellular function of ZC3H11A. For instance, we plan to use mass spectrometry for in depth studies of post-translation modifications of ZC3H11A, as well as protein-protein interactions involving ZC3H11A in the absence and presence of different forms of stress, e.g. heat or lipopolysaccharides (LPS). We can also genetically engineer the ZC3H11A coding sequence to reveal structure-function relationships such as which putative SUMOylation sites that are most likely critical for function. The significance of ZC3H11A in the transcription/export (TREX) complex and other complex formations can be characterized using cell fractionations, blue-native PAGE immunoblots, and purification using overexpressed tagged versions of ZC3H11A in HeLa WT and ZC3H11A knock-out cells. These assays can be combined with immunofluorescence imaging using fixed cells, or GFP-tagged live imaging to coordinate any complex formation with the dynamic nuclear relocalization of ZC3H11A that occurs upon stress. These analyses may reveal other potential functions for ZC3H11A. We consider ZC3H11A a valid target for antiviral therapies since the growth of HAdV-5, HSV-1, influenza and HIV-1 is drastically reduced in ZC3H11A\(^{-}\) cells (Paper 5). Therefore, we will screen for small molecules that interfere with the specific ZC3H11A-protein or nucleic acid interactions in order to develop new anti-viral and anti-cancer therapeutic strategies of considerable medical significance.
Summary in Swedish

Senare års framsteg inom molekylär- och beräkningsbiologi har gjort det möjligt att studera de komplicerade transkriptionella nätverken som reglerar ett brett spektrum av viktiga biologiska processer och fenotypiska egenskaper. I denna avhandling har vi kombinerat flera tillvägagångssätt, bland annat Next Generation Sequencing, profilering av genuttryck, immunoprecipitering av kromatin och RNA, bioinformatik samt genomediteringsmetoder för att karaktärisera den biologiska betydelsen av generna för ZBED6 och ZC3H11A.

ZBED6 upptäcktes nyligen som en transkriptionell repressor som reglerar uttrycket av IGF2. En mutation av bindningsstället för ZBED6, lokalisat i ett intron i IGF2-genen, förhindrar bindningen av ZBED6, och leder till tre gångers uppreglering av IGF2-mRNA i muskelvävnad från gris. Den första delen av denna avhandling behandlar en detaljerad funktionell karaktärisering av ZBED6. Inhibiering av ZBED6-uttryck medelst siRNA i myoblaster från mus ledde till ungefär två gångers ökat IGF2-uttryck. ChIP-seq-analys av ZBED6 och histonmodifieringar visade att ZBED6 företrädesvis binder till aktiva promotorer, och modulerar deras transkriptionella aktivitet (publikation I). I uppföljningsstudierna, där vi använt CRISPR/Cas9, visade vi att en deletion av ZBED6 (ZBED6\(^{-/-}\)), eller dess bindningsställe i Igf2 (Igf2\(^{AGGCT}\)), ledde till mer än 30 gångers uppreglering av Igf2-uttryck i myoblaster. Differentiering av dessa genetiskt modifierade celler resulterade i hypertrofämyotuber. Transkriptomanalysen visade på ungefär 30 % överlapp mellan de differentiellt uttryckta generna i myotuber från ZBED6\(^{-/-}\)-och Igf2\(^{AGGCT}\)-celler, med signifikant överrepresentering av muskelspecifika gener. Överuttryck av ZBED6 i myoblaster ledde till cellcykelarrest, minskad cellöverlevnad, minskad mitokondriell aktivitet, och försämrad differentiering av myoblasters (publikation II). Ytterligare studier på cancerceller visade att ZBED6 påverkar tillväxten av kolorektala cancerceller, med dramatiska förändringar i transkriptionen av hundratals cancerrelaterade gener (publikation III). Den fenotypiska karaktäriseringen av musmodeller av ZBED6\(^{-/-}\) och Igf2\(^{pA/mG}\) (muterat bindningsställe för ZBED6 i Igf2-genen) visade att ZBED6 och Igf2 har en stor effekt på regleringen av muskeltillväxt och tillväxten av inre organ. Transkriptomanalys visade en massiv uppreglering av Igf2-uttryck (~30 gånger) i vävnader från vuxna, men inte i fostervävnader från transgena möss (publikation IV). I den andra delen av avhandlingen undersökte vi cellfunktionen hos Zc3h11a, genen i vilken ZBED6-genen är lokaliserad i dess första intron. Funktionen hos ZC3H11A-proteinet är hittills däligt karaktäriserat. Vi visar att ZC3H11A proteinet är ett tidigare okänt stressinducerat protein som krävs för effektiv mRNA-export från kärnan. Inaktivering av ZC3H11A inhiberade tillväxten av flera virus, inklusive HIV, influensa, HSV och adenovirus (publikation V).
التطورات الحديثة في تقنيات البيولوجيا الجزيئية والبرمجة الحيوية جعلت من الممكن دراسة التنظيمات المعقدة للتعبير الجيني. و التي تحكم في مجموعة واسعة من العمليات البيولوجية الأساسية والصفات المظهرية للكائن الحي. في هذه الرسالة نقوم باستخدام طرق Bioinformatics و التكنولوجيا كفل الهندسة الوراثية Genome Editing Next Generation Sequencing (NGS) لتسليط الضوء الوراثي. 
تم إنتاج ZC3H11A و ZBED6 في عام 2003، تم اكتشاف طفرة وراثية في أحد أفراد العائلة المسؤول عن إنتاج عامل النمو ZBED6، حيث تمت دراسة التعبير البشري للجين IGF2. ونتج عن ذلك زيادة كنلفة العضلات الهيكلية وزيادة حجم القلب. في هذه الرسالة تم دراسة الدور البيولوجي للعامل الوراثي ZBED6 وعلاقته بنمو الخلايا العضلية. باستخدام تقنية CRISPR/Cas9، تم إزالة التفريغ الوراثي المسؤول عن انتاج طفرة في الخلايا المنتجة للعضلات myoblasts. ونتج عن ذلك ارتفاع كمية IGF2. ونتج عن ذلك ارتفاع كمية JIGF2 - IGF2 في ZBED6. ونتج عن ذلك ارتفاع كمية IGF2. ونتج عن ذلك ارتفاع كمية IGF2 من الفوائد المطلوبة في الخلايا العضلية. علاوة على ذلك، هذه الخلايا المعدلة وراثيا أظهرت قدرة كبيرة على التماس وتشكيل أنابيب عضلية بشكل أسرع من الخلايا الطبيعية. تحليل التعبير الجيني الكلي لهذه الخلايا باستخدام RNA-sequencing وجدنا زيادة مغنية في التعبير الجيني للجينات المتعلقة بنمو العضلات. 
لفهم المزيد عن كيفية عمل ZBED6 وعلاقته بـ IGF2، قمنا بإنتاج خيول مهددة وراثيا لا تحتوي على زئبق الوراثية لـ ZBED6 وتحتوي على طفرة وراثية في الضرور الأول لـ ZBED6 (IGF2). التوصيف المائي لـ ZBED6 لـ ZBED6 (IGF2) وتحوي على زئبق ارتفاعاً مغنية في كمية ارتفاع RNA-sequencing الجيني الكلي لهذه الخيول باستخدام ارتفاع RNA-sequencing يظهر زيادة في النمو. بالإضافة إلى زيادة التعبير الجيني لبعض الجينات الهامة في التماثل الغذائي. صفا معدل النمو من الصفات والمكالمات وارتفاع النواتج الكلاسيكي، فهي الصفات التي تتحكم فيها العديد من الخصائص الوراثية من خلال هذا الربط أن نقص الصفات المكاثفة لمعدل النمو يمكن أن تتأثر ببطوره وراثي في تطويره وثبتنا هذا الربط في دراسة الوراثة الجزيئية. هناك ارتفاع إنتاج التكثيف في الوراثة الجزيئية، ساهم بشكل كبير في دراسة الوراثة الوراثية الممكنة عن الصفات بشكل أكثر تعمقاً ووضوحًا.

دراسة الدور الوراثي لـ ZC3H11A تزداد بشكل كبير في الخلايا عند تعرضها لعامل بيني مثل الإجهاد الحركي أو الإصابة. RNA بعدين الوراثي. هذه الزيادة تحدث على مستوى البروتين ولا يصاحبه أي تغيير في كمية الـ ZC3H11A. بدراسة متعمقة لهذا النقطة وجدنا أن البروتين ZC3H11A أكثر ثباتاً وقليل من معدل التكثير. هذا يعبر مثالاً آخر يوضح كيف أن العامل البيولوجي يمكن أن يحدث تغييرات مباشرة في تنظيم التعبير الجيني.
Acknowledgments

Leif Andersson, my main supervisor. There are no enough words to explain how grateful I am to get an opportunity to work with you and conduct the research under your supervision. Your passion for science and ability to see new opportunities and possibilities in the field of research has been very inspiring. Your aspiring guidance and inputs have always been helpful. It was a unique experience working as part of your team. I am thankful for your support and encouragement during these years.

Göran Andersson, my co-supervisor. You have been a great support for the ZBED6 project. Your endless knowledge in genetics and the extensive list of examples you had in mind, inspired me constantly. I still remember our first meeting when you introduced me to ZBED6 project. It has been the start of a long road full of excitement.

I would like to express my deep gratitude to Göran Akusjarvi for his precious contribution in “STINP” project. Thank you for the valuable suggestions and insights in molecular virology.

I would also like to thank Nils and Xuan for the fantastic collaboration over many years in ZBED6 project. I would also like to thank Juleen Zierath and her lab members at Karolinska Institutet, in particular Milena, Julie, Marie and Anna Krook for the fruitful collaboration in IGF2 project.

Wael, my Egyptian friend in IMBIM, we both came to Sweden as master students and started our PhD in about the same time, good time flies quickly. Mårten, thank you for the nice long discussions about science and life in general and your proofreading of my thesis. Ulla, always a dependent person in the lab. You have been a great support in the lab during these years.

I would to take this opportunity to thank Kerstin for being an inspiring personality and a source of motivation.

I would like to thank the present and former members of the ZBED6 project. Mårten, Muhammed, Rakan, Xiaofang, Daniel, Rajesh, Ola and Elisabeth for all your help and working together. Ola, thank you for introducing me to ChIP-seq and chromatin world. Elisabeth, thank you for your continuous positive attitude about the ZBED6-KO project, it was a fruitful collaboration and interesting experience as well 😊.

I would especially like to thank Calle, Mats, and Nima for invaluable discussions on science in general and bioinformatics in particular. I would
like to thank Matt, I learned a lot from your Perl course, it helped me in the bioinformatics analysis.

Nima and Sangeet, we have started together “8-years ago” as master students in Netherlands and obtained our PhD from the same lab and the same supervisor. I have been blessed with our friendship.

Completing this work would have been all the more difficult were it not for the support and friendship provided by the other members of the lab (D9 & D11:3). I am indebted to them for their help. Special thanks go to Jennifer, Susanne, “senior” Sergey, Ginger, Gerli, Eva, Jessica, Andreas, Anna, and Åsa. I have had some very good time with Iris, Doreen, Jessika, Sharadha, Junfeng, Chao, Fan, Nika, Chungang, Matteo, Sergey, Christina, Leonie, Anni, Mathilde and Gosia. 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.

Alvaro & Anna, the lovely couple in our lab, I wish you all the best and good luck Anna with your horses 🐴

I am also grateful to all the colleagues at IMBIM administration for taking care of all the practical aspects of my PhD program.

Special thanks also to my Egyptian friends in Uppsala, Wael, Farag, Mahmoud, El-kenawy, Taha and Assem. Heartfelt thanks for all your support throughout, which has played an important role in during my stay in Sweden. I would like to express my very great appreciation to my teachers and colleagues at Ain Sham university for their moral support, Hussein Mansour, Hany Gado, Manal El-sayed, Ahmed Galal, Nazmy Abdelhamed, Mohamed Fergani and the deceased Salah Galal.

My dear wife Shaymaa, thank you for the constant and overwhelming love, patience and endless support. My daughter Areej, your smile and laugh were more than enough to forget all the frustration from the lab and failed experiments, and helped me to move forward. My Mom, brothers and sister, thanks for your continued support and encouragement.
References


57. Martin-Trujillo, A. *et al.* Loss of imprinting of IGF2 characterises high IGF2


110. McCann, A. H. *et al.* Biallelic expression of the IGF2 gene in human breast


128. Hall, T. M. Multiple modes of RNA recognition by zinc finger proteins.


145. Di, X. et al. A high concentration of genistein down-regulates activin A,


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”.)