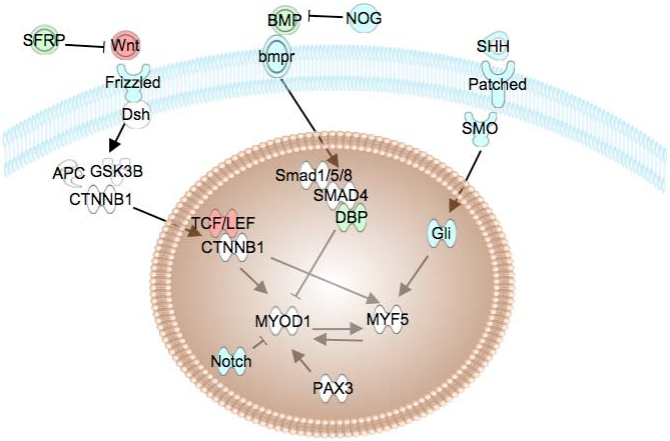
Swedish University of Agricultural Sciences
Faculty of Veterinary Medicine and Animal Science

Bioinformatics analysis of ZBED6: Interactome of Novel Transcription Factor ZBED6 in C2C12 Myoblasts

Awaisa Ghazal



Department of Animal Breeding and Genetics
Examensarbete 330
Uppsala 2010

Master's Thesis, 30 HEC

One-Year Master's Programme in Biology – Bioinformatics

Swedish University of Agricultural Sciences
Faculty of Veterinary Medicine and Animal Science
Department of Animal Breeding and Genetics

Bioinformatics analysis of ZBED6: Interactome of Novel Transcription Factor ZBED6 in C2C12 Myoblasts

Awaisa Ghazal

Supervisors:
Göran Andersson SLU, Department of Animal Breeding and Genetics
Erik Bongcam-Rudloff, SLU, Department of Animal Breeding and Genetics
Examiner:
Stefan Marklund, SLU, Department of Animal Breeding and Genetics

Credits: 30 HEC
Course title: Degree project in Animal Science
Course code: BI1021
Programme: One-Year Master's Programme in Biology
- Bioinformatics

Level: Advanced, A2E

Place of publication: Uppsala
Year of publication: 2010
Name of series: Examensarbete 330
Department of Animal Breeding and Genetics, SLU
On-line publication: http://epsilon.slu.se

Key words: Ingenuity Pathway Analysis (IPA), C2C12 myoblasts, Chromatin Immunoprecipitation (ChIP), Microarray, Myogenesis, ZBED6

ABSTRACT

ZBED6 is a novel mammalian transcription factor that was recently identified and shown to act as a repressor of *IGF2* transcription in skeletal muscle. Chromatin Immunoprecipitation (ChIP) sequence data in murine C2C12 myoblasts indicated that ZBED6 holds 2499 targeting sites. Whereas microarray data portrayed that ZBED6 differentially regulates almost 400 genes in C2C12 myoblasts. This data suggested that ZBED6 is targeting and regulating a vast array of genes, so there was a need to investigate system level knowledge of ZBED6. To elucidate the complete interactome of ZBED6 and particularly to build and visualize muscle-specific networks by using ChIP sequence and microarray data, Ingenuity Pathway Analysis (IPA) was employed. Networks of ZBED6-targeted genes suggested that ZBED6 mainly induces tissue development and is involved in development of cancer. These effects most likely involve the Wnt, human embryonic stem cell pluripotency and TGFβ canonical pathways. Many of ZBED-targeted genes like *IGF2, IGF1R, SRF, SMAD7, CDH2, CTNND2, PITX2, TRIO, WNT3a, WNT1, MSX1, PAX7, VEGFA, ACTN1, HEY1, SKI, E2F1, EP300, FGF9, MTOR, FGFR1, BMP7,* and *TGFβ* have established roles in skeletal muscle myogenesis. Whereas networks of ZBED6-regulated genes revealed that ZBED6 is mainly involved in organismal development and cell-to-cell interaction and signaling; and also engaged into hepatic fibrosis, clathrin-mediated endocytosis and tight junction signaling cascades. ZBED6-regulated genes including *BMP4, DBP, CDH2, AGT, IGF1, IGF2, THBS1, PDGFRA, MPP2, AKT1, HGF, MET, FGF4, TGFβ3, ACTN1, F2R,* and *VCAM1* have established roles in muscle proliferation, myogenesis and contraction. Our findings suggest that ZBED6 holds promise as a target to control and influence many cellular functions and canonical pathways and also controls many factors and cascades that are crucial for skeletal muscle myogenesis.

INTRODUCTION

Muscle Development

Muscles, the contractile tissues of animals, are classified into skeletal, cardiac and smooth muscles. During early embryogenesis, mesodermal cells give rise to muscles. Paraxial mesoderm is divided into somites which give rise to myotome. Myotome further produces myoblasts. In response to activation by certain growth factors like fibroblast growth factors, myoblasts start to proliferate [1]; when these factors are exhausted myoblasts secrete fibronectin that leads to cell cycle arrest. The action of fibronectin leads to muscle cell differentiation [1]. Differentiation is mediated by myoblasts alignment and fusion. Cell membrane proteins like cadherines, fibronectins, integrins mediate myoblast alignment [1]. After alignment, myoblasts are fused. Calcium ions and certain metalloproteinases are crucial for myoblast fusion [1] which further give rise to muscle fiber. Different stages of muscle formation are shown in Figure 1.1. Healthy adult skeletal muscles contain satellite cells, which are undifferentiated and mitotically quiescent cells. These cells are activated in response to injury and stimulate proliferation and differentiation to repair damaged fibers (Clow & Jasmin 2010).

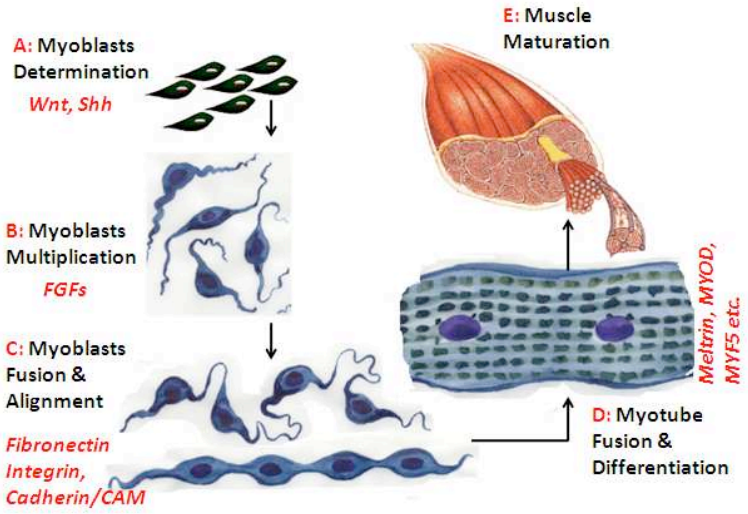


Figure 1.1: Differentiation of myoblasts into muscle cells. *Various stages of muscle development and factors mediating muscle development are illustrated.* ***A:*** *Myotome determination induced by Wnt, Shh* ***B:*** *Myoblast multiplication implied by growth factors primarily FGFs.* ***C-D:*** *Alignment and fusion, after growth factor depletion, myoblasts start to align and fuse to form myotubes that is induce by fibronectin, cadherin integrin.* ***E:*** *mature muscle.*

Myogenesis is controlled by a set of transcription factors denoted myogenic master regulators. In Figure 1.2 a schematic representation of the key transcription factors is shown (Figure 1.2). For skeletal muscle formation, myoblast differentiation is indispensable and is directed by regulation of such muscle-specific transcription factors including MyoD, myogenin and transcriptional co-regulators (Jeong et al., 2010). Myogenic regulatory factors (MRFS) including MyF5, MyoD, MRF4 and myogenin activate transcription of muscle-specific genes (Harada et al., 2010). MyoD and Myf5 are imperative for myogenic determination while myogenin and MRF4 are pivotal for terminal differentiation and lineage maintenance. *MyoD* and *MyF5* knockout mice embryo showed entirely ablated skeletal muscle myoblasts and myofibers (Kablar et al., 1999). MyoD belongs to the basic helix-loop-helix (bHLH) protein family of transcription factors. The 68 amino acid bHLH domain plays a crucial role in myogenesis (Weintraub et al., 1991). Only skeletal muscle and its precursors express MyoD while in non-muscle cells and tissues *MyoD* transcription remains silent due to methylation at CpG sites of distal enhancer (Brunk et al., 1996). The skeletal muscle-specific *MyoD* transcription is known to be regulated by a set of transcription factors including SRF (L'honore et al., 2003), MSTN (Langley et al., 2002), PAX3, PAX7 (Horst et al., 2006), IGF1 (Strle et al., 2004), WWRT1 (Jeong et al., 2010).

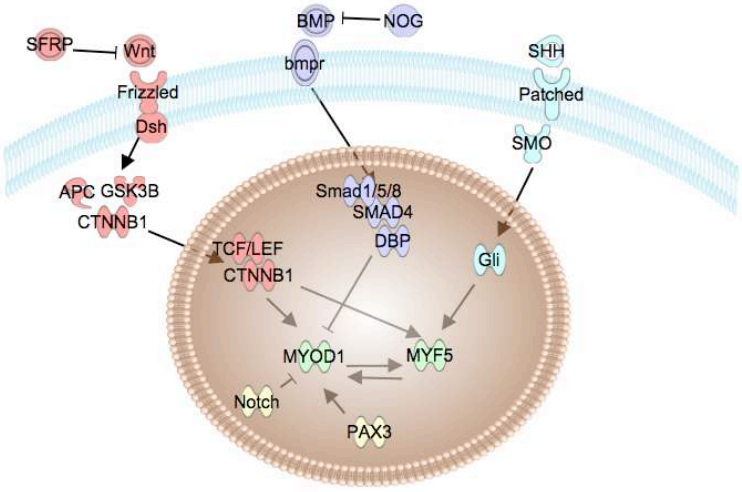


Figure 1.2: MyoD and MyF5 Regulatory pathway. *Wnt signaling (Ridgeway et al., 2000, Borello et al., 2006) positively regulate expression of MyoD and MyF5. Bone morphogenic proteins (BMP) signaling (Reshef et al., 1998) and Notch signaling (Hirsinger et al., 2001) act as antagonist for expression of MyoD. Sonic hedgehog (SHH) signaling (Gustafsson et al., 2002) and Pax3 (Sato et al., 2010) positively regulate expression of MyF5.*

IGFs and Muscle Development

Insulin-like growth factors (IGFs), are evolutionary conserved mitogenic proteins, encompasses IGF1 and IGF2. During embryogenesis, IGFs stimulate two biological events of myogenesis *i.e* myoblast proliferation and differentiation. All vertebrates studied to date elucidate critical role of IGFs in growth and development (Wood et al., 2005). The major organ for IGF expression is the liver but IGFs are ubiquitously expressed in most tissues. *IGF1* or *IGF2* knockout mice has birth weight 60% of wild type littermates while null mutations in both *IGF1* or *IGF2* leads to body weight 30% of their wild type litter mates and mice died shortly afterwards (Baker et al., 1993; Liu et al., 1993). IGF1 over expression in mice leads to 1.3 fold increase in body weight (Mathews et al., 1988). IGF-I peptide administrated to rats causes increase in protein synthesis and body growth (Tomas et al., 1992). IGFs stimulate various cellular processes including proliferation, differentiation, migration and survival. These biological actions are accomplished through binding of IGFs to IGF receptors. There are two IGF receptors, IGF1 receptor (IGF1R) and IGF2 receptor (IGF2R). IGF1R shows sequence and structure similarity with insulin receptor (IR). Once ligand is bound to IGF1R it results in autophosphorylation and activation of multiple signal transduction cascades including phosphatidylinositol 3-kinase (PI3K)-Akt cascade and the Raf-Mek-Erk1/2 cascade. Phosphatidylinositol 3-kinase (PI3K)-Akt signaling cascade is regulating glycogen synthesis, glucose transport, protein synthesis, cellular proliferation and apoptosis while Raf-Mek-Erk1/2 cascade is playing significant role in cellular proliferation and apoptosis (Dupont and LeRoith, 2001). *IGF1R* knockout mice have body weight 45% of wild littermates and die shortly after birth (Baker et al., 1993; Liu et al., 1993). IGF2R is different from IGF1R structurally and functionally. Mice inheriting loss of imprinted IGF2R/ M6P has body weight 25-30% greater than

wild type siblings and die around birth (Lau et al., 1994). Recent studies have elucidated that IGF2 binding with IGF2R stimulates Erk signaling by inducing sphingosine kinase (SK)-dependent transactivation of sphingosine 1 phosphate (S1P) receptors (El-Shewy et al., 2007).

In addition to IGFs and IGF receptors, other important components are IGF binding proteins (IGFBPs). To date six members IGFBP1- IGFBP6 have been identified. IGFBPs are secreted proteins and perform functions that are important to coordinate and regulate IGFs activities. IGFBPs are responsible for transport, prolong half-life, localization and interactions of IGFs (Jones & Clemmons, 1995).

IGFs play a vital role in skeletal muscle growth and differentiation and also in muscle regeneration. IGF-II levels increase dramatically during myogenesis. Endogenous IGF expression induces differentiation in murine C2C12 cells (Yoshiko et al., 2002). Inactivation of *IGF1* gene in mice causes reduced body size which was attributed to decrease in muscle, bone mass and multiple organs (Powellbraxton et al., 1993). Mice deficient in IGF1R developed hypoplasia including muscle and delay in ossification, abnormalities in central nervous system and epidermis while double knockout *IGF2/IGF1R* and *IGF1/IGF2* mice showed intense dwarfism (Liu et al., 1993). Virus-mediated over-expression of IGF1 induced increase in muscle mass and strength in adult mice (Barton-Davis et al., 1998). Over-expression of IGF1 in mice suffering from a muscle disease similar to duchenne muscular dystrophy led to 40% increase in muscle mass as IGF1 induces muscle regeneration and protein synthesis pathways (Barton et al., 2002). IGF-II antisense oligodeoxynucleotide complementary to first five codons of IGF2 abolished differentiation in cultured muscle cells in absence but not in presence of exogenous IGF2. IGFBP5 is the most abundant IGFBP secreted by skeletal muscle cells and it stimulates muscle differentiation and growth. Consequently, IGF signaling pathway encompassing IGFs, IGFRs and IGFBPs is critical for muscle differentiation and growth.

Regulatory Mutation in IGF2 Effects Muscle Growth

Meat is a substantial dietary constituent for humans. Animal domestication allowed production of meat and breeding of animals to improve meat production. Different domestic animals like chicken, sheep, goat, cattle and pig are principal source of meat. The pig (*Sus scrofa*) was domesticated around 9000 years ago (Chen et al., 2007). It is a meat production source throughout the world except in most Muslim countries. Lean meat is demanded by consumers and has consequently been a major selection criteria resulting in increased muscle growth caused by favorable alleles for muscle growth in most pig breeds used for meat production. Muscle growth studies unveil knowledge about muscle development. In pigs, three mutations have been identified that affect muscle growth. Firstly, a single point mutation identified in the porcine ryanodine receptor gene (*RYR1*) leads to a recessive disorder called malignant hyperthermia (Fujii et al., 1991). Secondly, a nonconservative substitution in the *PRKAG3* gene mediates high glycogen content in pig skeletal muscle (Milan et al., 2000). *PRKAG3* encodes a muscle-specific isoform of the regulatory γ subunit of adenosine monophosphate-activated protein kinase (AMPK). Thirdly, a single nucleotide substitution, a G to A transition in an evolutionary conserved region in intron 3 of

the paternally expressed porcine *IGF2* gene was identified, which results in a three-fold increased *IGF2* transcription skeletal muscle that results in increased postnatal muscle growth (Van Laere et al., 2003).

In the late 1980ies, a Quantitative Trait Locus (QTL) mapping study was performed in pigs. The experimental design for the QTL study was based on intercross between a European wild boar and domestic large white. The F2 individuals generated from this intercross were used for QTL mapping of genes influencing a number of growth-related traits including carcass weight, growth, fatness and meat quality and significant QTLs were identified on chromosome 2, 3, 4 and 8 (Andersson et al., 1994, Andersson-Eklund et al., 1994). A similar QTL study was performed using a Pietrain and large white intercross and significant score was obtained on chromosome 2 for muscle mass and fat deposition (Nezer et al., 1999). While in another study, siginificant QTLs for muscle, heart weight and backfat thickness were obtained at chromosome 2 in wild boar and large white intercross (Jeon et al., 1999). Consequently, the QTL on chromosome 2 was reduced to 250kb region (Nezer et al., 2003). This region embraces the insulin (*INS*) and *IGF2* genes. The *IGF2* gene is an imprinted gene with paternal expression. *IGF2* was therefore a very strong candidate gene for the observed QTL phenotype. Initially, all the coding exons were sequenced but no mutations could be identified in coding sequence of *IGF2* (Van Laere et al., 2003). Then a 28.6 kb region was resequenced and a single causative mutation was identified at G3072A in intron 3 of *IGF2* (Van Laere et al., 2003). This quantitative trait nucleotide (QTN) was located in an evolutionary conserved CpG island with unknown function. This mutation in *IGF2* did not affect the imprinting status of *IGF2*. The methylation status of this QTN was defined by using bisulphite sequencing in four month old pigs. The result of this sequencing revealed that the CpG island was methylated in liver but remained unmethylated in skeletal muscle (Van Laere et al., 2003). To reveal whether the QTN region functions as a transcriptional regulatory element biochemical and functional experiments were performed. First, DNA-protein interaction studies using electrophoretic mobility shift assays (EMSA) were performed. In EMSA double-stranded 20 bp oligonucleotides corresponding to a sequence spanning the QTN position (3072) for both wild type q (G3072) and mutant type Q (A3072) sequence were used. Q and q radioactively labelled double-stranded oligonucleotides were incubated with nuclear extracts prepared from murine C2C12 myoblast. A complex was obtained with the q probe but not with Q and methylated q (Van Laere et al., 2003). Slightly weak complex migration was also observed in HEPG2 hepatocytes and HEK293 embryonic kidney cells.

In C2C12 myoblasts, the functional effect of the *IGF2* mutation on transcription was analyzed by transient transfection assay. Q and q construct were made containing a 578 bp fragment of the region containing the QTN. This fragment was cloned upstream of the P3 promoter of IGF2 and inserted in front of a Luciferase reporter gene. Fragment q showed 25% reduced activity while Q showed 70% reduced activity compared with the activity of a reporter construct containing the P3 promoter alone driving luciferase (Van Laere et al., 2003). Expression studies for *IGF2* mRNA were carried out through real time PCR and northern blot analysis, these assays showed 3% difference in expression level of postnatal skeletal muscle *IGF2* mRNA. *IGF2* mRNA expression

initiated from the P2, P3, and P4 promoters of *IGF2*, were analyzed by northern blotting and real-time PCR. Transcription from all these promoters was affected by QTN but promoter 3 showed the highest expression because it is main promoter for *IGF2* in skeletal muscle (Van Laere et al., 2003). In heart muscle, there was a small increase in *IGF2* mRNA expression. Like the sense transcript, the *IGF2* antisense transcript (*IGF2-AS*) expression was imprinted paternally and was affected by QTN in pigs (Braunschweig et al., 2004). Combination of the obtained results suggested that a transcriptional repressor regulates expression of all four *IGF2* promoters.

ZBED6 Repressor of IGF2 During Myogenesis

QTN containing 16 bp sequence shares identity with eight mammalian species including pigs, bovine, horse, dog, rabbit, human, mouse, and rat (Van Laere et al., 2003). This QTN sequence did not have binding sites for any known transcription factor. Thus, a hitherto unknown transcription factor binding to QTN region function as a repressor for *IGF2* transcription in skeletal muscle. Therefore, the identification of this unknown repressor for *IGF2* transcription became a high priority. Oligonucleotide affinity capture of proteins and quantitative mass spectrometry were used to identify *IGF2* repressor (Markljung et al., 2009). For this purpose, stable isotope labeling of amino acid in culture (SILAC) technique was used to prepare two different nuclear extracts. In this technique heavy extract proteins (stable isotope labeled arginine and lysine amino acids) and light extract proteins (natural version of amino acids) were prepared. Biotinylated double-stranded oligonucleotides q and Q were used to capture protein in heavy and light nuclear extracts, respectively. Captured proteins of both extracts were mixed and analyzed by liquid chromatography mass spectrometry (LCMS). Proteins indicating highest enrichment by q was an alternative splice form of ZC3H11A, a member of zinc finger family of transcription factors (Markljung et al., 2009). Captured protein was encoded by an intronless gene located in intron 1 of *ZC3H11A*. Northern blot analysis showed that captured protein is expressed with ZC3H11A. Two BED and one hATC dimerization domain was observed in encoded protein. Bioinformatic analysis was used for identifying BED domains (Markljung et al., 2009). Drosophila melanogaster BEAF and DREF were the first proteins that indicated the presence of BED domains. ZBED1 to ZBED5 are already discovered so the captured protein was given the name ZBED6 (Markljung et al., 2009). Two DNA binding domains of ZBED6 are more similar with each other compared to other ZBED BED domains that indicates ZBED6 had duplicated its BED domains before integration into *ZC3H11A*. The BED domain in ZBED6 shares 100% similarity among 26 mammals; therefore ZBED6 is a highly conserved protein in placental mammals. ZBED6 contains two translation start sites and translates into two isoforms ZBED6a and ZBED6b of 122 and 116 KDa respectively (Markljung et al., 2009). *ZBED6* mRNA is expressed in many tissues including skeletal muscle and heart muscle, it was confirmed by Northern blot and real time PCR (Markljung et al., 2009).
EMSA showed that ZBED6 bound with q sequence but not with Q, these results proposed that ZBED6 is the *bona fide* repressor of IGF2 transcription that interacts with QTN region of *IGF2* (Markljung et al., 2009). It was also observed that ZBED6 is localized inside nucleus in the nucleolus (Markljung et al., 2009). To get knowledge about functional significance of ZBED6, *zbed6* mRNA was silenced in C2C12 cells by using siRNA. Quantitative PCR showed that *ZBED6* mRNA was decreased to 75% in ZBED6 silenced C2C12 cells (Markljung et al., 2009). Transient

transfection assays were performed on control and ZBED6-silenced C2C12 cells. Luciferase activity was reduced to a large extent in wild type q construct when compared with construct containing P3 alone. These results suggested that ZBED6 represses expression of *IGF2* promoter 3 after binding with the QTN in the CpG island (Markljung et al., 2009).

Interaction of ZBED6 with QTN site in IGF2 was also confirmed by chromatin immunoprecipitation (ChIP). Real time PCR showed that controlled C2C12 cells had clear difference in enrichment of QTN as compared with ZBED6 silenced C2C12 cells (Markljung et al., 2009). To get knowledge about ZBED6 activity during myogenesis, *IGF2* mRNA expression was measured in control and ZBED6-silenced C2C12. There was no effect in *IGF2* mRNA at first days but there was an increase in IGF2 expression at day 6 when myotubes are formed compared with controlled cells. ZBED6-silenced C2C12 cells showed increased proliferation and faster wound healing compared with control cells (Markljung et al., 2009). These results implicated that ZBED6 acts as a repressor for IGF2 and plays an important role in muscle myogenesis. Consequently IGF2 is a crucial factor for muscle development and ZBED6 acts as its repressor. We have ChIP sequencing and microarray data of ZBED6 in C2C12 myoblast indicating that ZBED6 regulate transcription of a large number of different genes besides IGF2. Thus, there was a need to elucidate system level knowledge of ZBED6 and in particular during IGF2-mediated muscle development in myoblasts.

Systems Biology and Network Analysis

Systems biology refers to a "study of an organism, viewed as an integrated and interacting network of genes, proteins and biochemical reactions which give rise to life [2]". Instead of individual system analysis, systems biology focuses on all components and their interaction and considers them as a part of one system. According to Kitano H. systems biology comprise of following four modules. 1: System Structures including gene interaction networks and biochemical pathways, 2: System Dynamics focusing on system behavior under various conditions over time, 3: Control Methods including mechanisms controlling cell state systematically, 4: Design Methods including strategies to construct and modify desired properties into biological system (Kitano 2002). The term systems biology was first time coined in 1993 by Zieglgansberger W & Tolle TR. Explosive progress in genome sequencing and massive data generated by DNA microarray, proteomics, transcriptomics, interactomics allows systems biology to integrate all this data. Systems biology is thus combination of omic approaches, data integration and modeling, and requires fusion of various disciplines including biology, computer sciences, mathematics etc. A critical component of system biology is software infrastructure. Analysis and modeling of complex biological processes at system level requires integrated databases that can provide properties of genes, proteins and also complex network. Molecular interaction networks reveal knowledge about complex roles of genes, gene products (RNA and proteins) and cellular environment during biological processess. In networks there are two important objects: nodes and edges; nodes depict genes and gene product while edges illustrate specific interaction between nodes (Baitaluk et al., 2006). An edge may be an illustration of transcription factor binding to a promoter region in protein-DNA network while it may represent evidence of co-immunoprecipitation or two-hybrid interaction of proteins in protein-

protein interaction (Baitaluk et al., 2006). Currently, a number of tools is available for network analysis, modeling and visualization including Osprey [3], Cytoscape [4], BiologicalNetworks [5], and Ingenuity Pathway analysis [6].

The aim of this study is to investigate the complete interactome of ZBED6 and specifically building and visualizing muscle specific networks participating into muscle development by using Chip sequencing and microarray data. Another objective of this study was to compare IPA versus other available gene ontology and network analysis tools.

MATERIALS AND METHODS

ChIP Seq Data

Chromatin Immunoprecipitation (ChIP) sequencing experiment for ZBED6 was performed in mouse C2C12 cells by using anti ZBED6 antibody. ZBED6-silenced and control mouse C2C12 cells were used for ChIP sequencing. To sequence ChIP DNA fragments AB SOLid technology was used. 2499 ZBED6 target sites were revealed.

Microarray Data

Microarray data was generated from ZBED6 microarray experiment. ZBED6-silenced and control mouse C2C12 cells were used for microarray experiment and expression was measured at two time points: day two and four. Illumina mouse ref-8 v2.0 array platform was used for generating microarray data. There were near about 400 genes on both day points that showed a P value less than 0.054 and were differentially expressed.

Ingenuity Pathway Analysis

Ingenuity Pathway Analysis (IPA) was used to construct and visualize networks and pathways, and to get insight into functional analysis of ChIP sequencing and microarray data. IPA is a commercially available service that is provided by Ingenuity systems, a company situated in Redwood City, CA. Networks and pathways are constructed by implementing unique analysis algorithms and ingenuity knowledge base. Novel algorithm approaches are developed and deployed by ingenuity team but no information about them is available publically while Ingenuity knowledge base, is a repository of functional annotations and biological interactions derived from relationship among genes, proteins, complexes, tissues, cells, diseases and drugs. These relationships are updated frequently and reviewed manually for accuracy. IPA is a tool that facilitates to understand, analyze, explore, visualize and model the complex biological and chemical systems. IPA quickly analyzes the experimental data by identifying interactions, functions, mechanisms, and relevant pathways. IPA is equipped with enhanced data mining, query language, filtering tools and graph manipulation;

IPA provides two main services:

- Search and explore
- Analyze and interpret data

Search and Explore

IPA contains search services: *genes and chemicals*, *functions and diseases*, *pathways and toxicity lists* which provides updated information about genes, chemicals, drugs, protein families, normal cellular functions and disease processes, signaling and metabolic pathways; and tox lists. IPA search services also helps in extracting information from scientific literature about genes, drugs, chemicals, biomarkers, cellular functions and diseases, signaling and metabolic pathways. Besides this, IPA search and explore services help to enrich and filter datasets, identify common and unique molecules among lists, pathways and to build and design pathways from datasets or gene list provided by user.

Analyses and interpretation of Data

IPA can analyze and interpret data generated from smaller-scale experiments that generate gene lists to larger scale experiments including SNP microarrays, gene expression and proteomics experiments. IPA supports a vast variety of array platforms including Affymetrix, Applied Biosystems, Illumina, Agilent and CodeLink; and also support vast variety of identifier types like Kyoto Encyclopedia of Genes and Genomes (KEGG), RefSeq, UniGene, UniProt, Affymetrix, Applied Biosystems, Illumina, Agilent and CodeLink, HUGO symbol, dbSNP, GenBank, Ingenuity and miRBase. IPA integrates different interaction databases including protein-protein, microRNA-RNA and many other interaction databases. Protein-protein interaction databases integrated by IPA are Biomolecular Interaction Network Database (BIND), Biological General Repository for Interaction Datasets (BiOGriD), Cognia, Database of Interacting Proteins (DIP), Molecular Interaction database (IntAct), Interactome studies, Molecular Interaction Database (MINT), Mammalian Protein-Protein Database (MIPs). IPA integrated microRNA-RNA interaction databases are (Argonaute 2, TARBase) and additional interaction databases are Gene Ontology (GO), GVK Biosciences, KEGG, Obesity Gene Map Database. IPA analyses and data interpret services helps in core, tox, biomarker, and metabolomics analysis.

IPA Tox analysis

IPA tox analysis enables to assess toxicity and safety of compounds by using toxicity function and relates the experimental data to clinical pathology endpoints. It helps to understand pharmacological responses.

IPA Biomarker analysis

IPA Biomarker analysis identifies relevant biomarkers from datasets generated at each step of drug discovery.

IPA Metabolomics analysis

IPA metabolomics analyzes metabolite data and provides important findings about cell physiology and metabolisms.

IPA Core Analysis

IPA core analyses helps to assess the molecular networks, signaling and metabolic pathways and biological processes which are significantly associated to uploaded datasets. The IPA core was used to analyze ChIP seq and Microarray data. ChIP seq data was taken and uploaded for new core analysis. File format was selected, as flexible format and in identifier type all identifiers were selected provided by IPA. Other parameters were used according to default settings. Genes that were differentially expressed on both day points were taken to make a dataset and was uploaded in IPA and was run for new core analysis by using reference set illumina mouseref-8 v2.0 as array platform. File format was selected, as flexible format and in identifier type all identifiers were selected provided by IPA. Expression value was added as an observation and threshold value for expression was set 0. Other parameters were used according to default settings. A schematic flow chart about how datasets were analyzed by using IPA is shown in Figure 2.1.

IPA Statistics

IPA also provides statistical calculation to evaluate the significance of results. Significance value is denoted by p-value and is measured by right tailed Fisher Exact Test. Significant value for functions/ pathways/ tox lists/ tox functions is a measurement of likelihood that functions/ pathways/ tox lists/ tox functions is associated with dataset by random chance. P value less than 0.05 is statistically significant. IPA indicates p-value for high-level functions in the form of range. IPA also provide ratio for canonical pathways and tox list. Ratio is calculated by dividing number of genes in dataset that are present in canonical pathway/ tox list with total number of genes present in that canonical pathway/ tox list. Ratio helps to determine which pathway/ tox list overlap with most of genes in uploaded dataset. Ratio describes the strength of association. Both ratio and significance value are insufficient to illustrate how the genes are associated with pathway, so to evaluate the function of affected genes in canonical pathway one has to look whether pathway is upregulated or downregulated.

RESULTS

ChIP Seq data Analysis

ChIP sequence data analysis, performed by using IPA, is divided into three components:

1. Networks
2. Canonical pathways

These components are illustrated below.

1: Networks

ChIP sequence data networks analysis comprises of:

I. Top five networks table
II. Top five networks figures

Network table and figures are shown below.

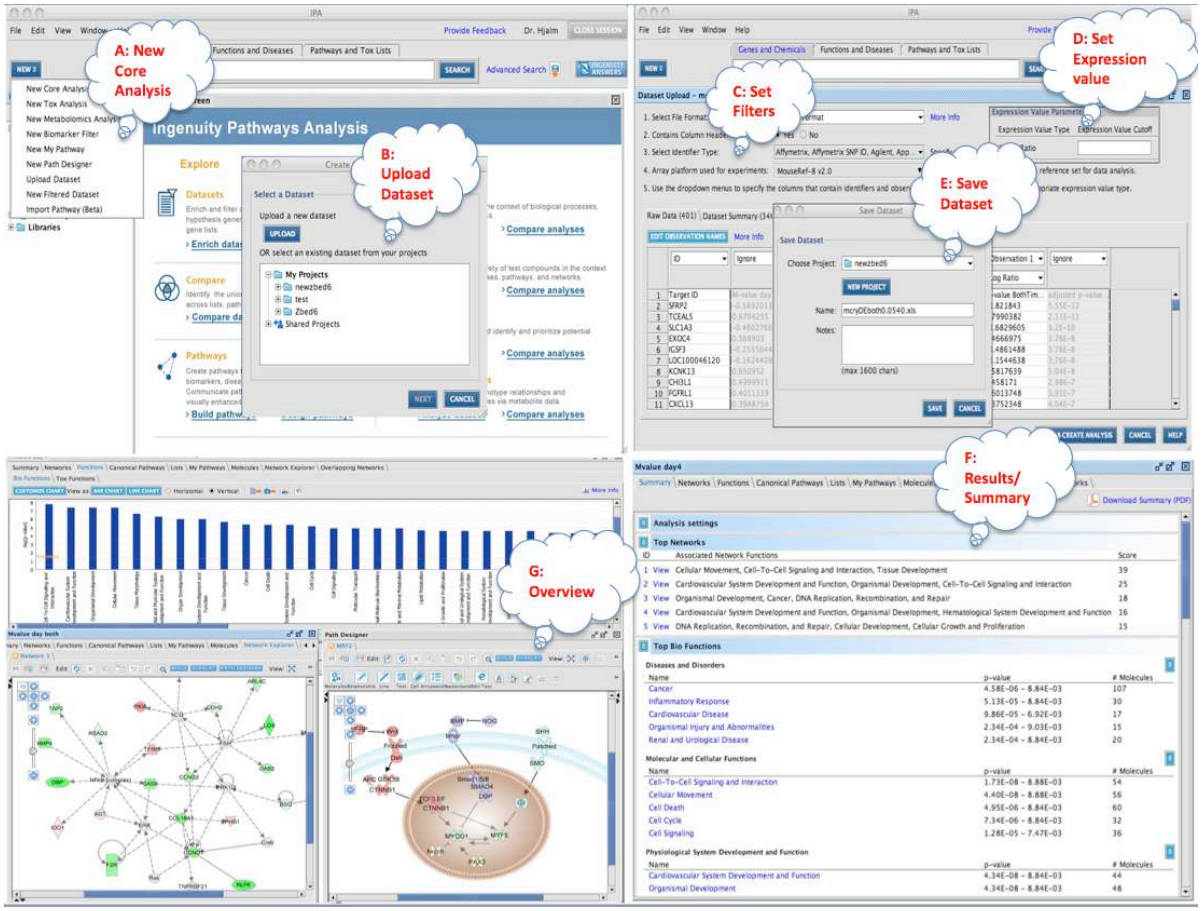


Figure 2.1: *This figure is depicting how dataset can be analyzed by using IPA core analysis service. A): New core analysis can be performed by NEW tab. B): Upload Dataset, excel or text file can be uploaded. C): Set filters, filters are file format, identifier type, and array platform. D): Set expression value if dataset is expression data. E): save dataset and run analysis. F): When IPA completes analysis, it creates results that contain summary, networks, canonical pathways and functions. G): By using IPA, networks, pathways and chart for function and pathways can be visualized.*

I: Top five networks table

Top five networks table from ChIP sequence data contains associated network functions and scores, which are shown in Table 1

Table 1: *ZBED6 ChIP sequence data networks table. Network associated functions and scores are shown below. Score indicates network significance; higher the score lower is the p-value.*

Network Associated Functions	***Score***
Tissue Development, Cancer, Embryonic Development	*32*
Cancer, Immunological Disease, DNA Replication, Recombination, and Repair	*32*
Cellular Growth and Proliferation, Tissue Development, Gene Expression	*30*
Cell Cycle, Gene Expression, Connective Tissue Development and Function	*28*
Cellular Development, Nervous System Development and Function, Cellular Movement	*25*

II: Top five network figures

Top five network figures constructed by using ChIP sequence data are illustrated and shown below.

1): Tissue Development, Cancer, Embryonic Development

This network encompasses genes for tissue development, cancer and embryonic development. Genes participating into above functions are illustrated below while network is shown in Figure 3.1.

Tissue Development

In this network there are 15 genes, which are targeted by ZBED6, which are participating into tissue development. *CDH2, CREBBP, FGF3, IKZF3, MSX1, PAX7, PITX2, RELN, RUNX1, SREBF1, T, TRIO, VEGFA, WNT1, WNT3A*.

Cancer

In this network there are 16 genes, which are participating into development of cancer: *CDH2, CREBBP, EHHADH, FGF3, JUP, MSX1, NR4A1, PAX7, PITX2, RBBP4, RELN, RUNX1, SCN5A, TRIO, VEGFA, WNT1*.

Embryonic Development

In this network there are 14 genes, which are playing role into embryonic development: *CDH2, CREBBP, CTNND2, FGF3, IKZF3, MED21, MSX1, PITX2, RELN, RUNX1, T, VEGFA, WNT1, WNT3A*.

2): Cancer, Immunological Disease, DNA Replication, Recombination, and Repair

This network encompasses genes for cancer and immunological diseases, and DNA replication, recombination and repair. Genes participating into above functions are illustrated below while the network is shown in Figure 3.2.

Cancer

In this network there are 10 genes, which are known to participate in cancer development: *BRD2, COL15A1, HBP1, MN1, PAWR, PDE4D, ROBO1, TRIB1, UPP1, VAV3*

Immunological Disease

In this network there are five genes, which have been implicated in Immunological disorders: *AKAP7, BRD2, PDE4D, ROBO1, VAV3*

3): Cellular Growth and Proliferation, Tissue Development, Gene Expression

This network encompasses genes for cellular growth and proliferation, tissue development and gene expression. Genes participating into above functions are illustrated below while network is shown in Figure 3.3.

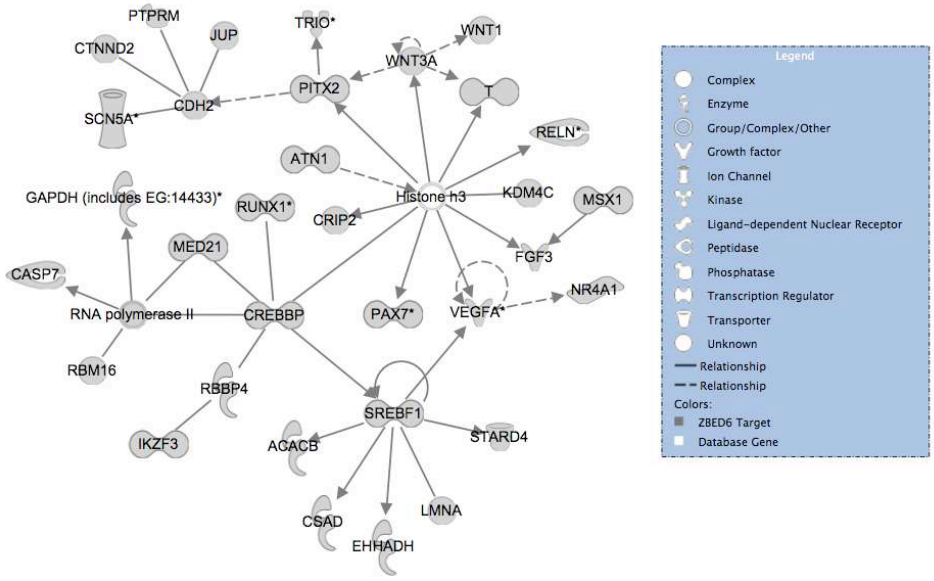


Figure 3.1: *A schematic diagram constructed by using IPA illustrating ZBED6 target gene networks. In this network there are 33 ZBED6 target genes, including 10 genes encoding for transcription factors. This network illustrates genes for tissue development, cancer and embryonic development, of which 15 have established roles in cancer development.*

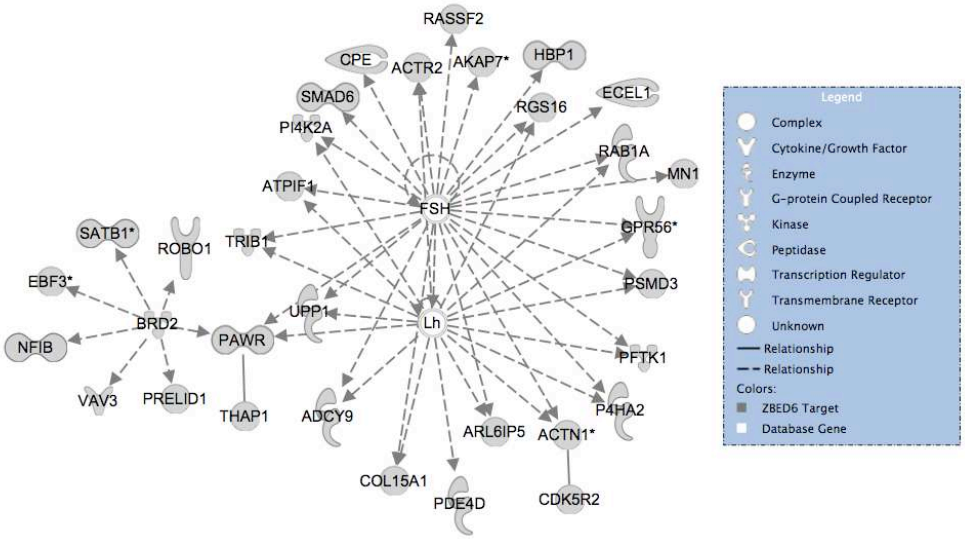


Figure 3.2: *A schematic diagram constructed by using IPA illustrating ZBED6 target gene networks. In this Network there are 32 ZBED6 target genes, including five genes encoding for transcription factors. This network illustrates genes for cancer and immunological diseases, and DNA replication, recombination and repair. Many of the genes (10) have established roles in cancer development.*

Cellular Growth and Proliferation

In this network there are 27 genes, which are participating into cellular growth and proliferation: *ADAM15, BCL2L11, BUB1, CADM1, CDKN2C, CITED2, CLIP1, ENPP1, ESPL1, FDFT1, FLI1, FOXO1, FOXO3, HEY1, HIF1A, IGF2, INHBB, MAP3K7, MAP3K7IP2, MKI67, PAFAH1B1, RUNX3, SKI, SMAD7, TGFB1, TGFBR3, ZEB2*

Tissue Development

In this network there are 16 genes, which are participating into tissue development: *ADAM15, BCL2L11, CITED2, ENPP1, FDFT1, FOXO1, HEY1, HIF1A, IGF2, INHBB, MAP3K7, SKI, SMAD7, TGFB1, TGFBR3, ZEB2*

Gene Expression

In this network there are 19 genes, which are participating in different aspects of gene expression: *CDKN2C, CITED2, FLI1, FOXO1, FOXO3, HEY1, HIF1A, IGF2, INHBB, KDM5B, KLF15, MAP3K7, MAP3K7IP2, RORA, RUNX3, SKI, SMAD7, TGFB1, TGFBR3*

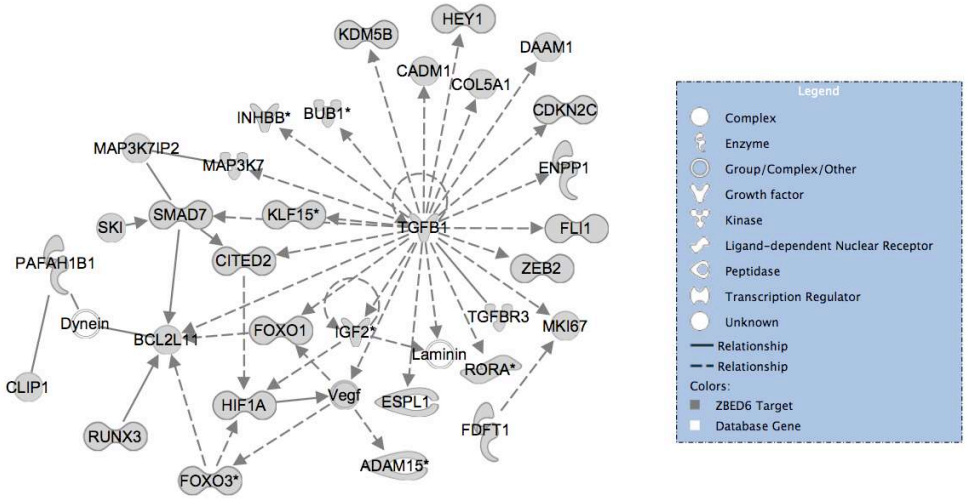


Figure 3.3: *A schematic diagram constructed by using IPA illustrating ZBED6 target gene networks. In this Network, 33 genes are ZBED6 target genes, including 12 genes that encode for transcription factors. This network illustrates genes for cellular growth and proliferation, tissue development and gene expression and the majority of these genes (27) have established functional roles in cellular growth and proliferation.*

4): Cell Cycle, Gene Expression, Connective Tissue Development and Function

This network encompasses genes for cell cycle, gene expression and connective tissue development and functions. Genes participating into above functions are illustrated below while network is shown in Figure 3.4.

Cell Cycle

In this network there are 12 genes, which are participating at different steps during the cell cycle: *CDKN1A, CEBPA, E2F1, E2F2, EP300, FGF9, FLT1, MTOR, NASP, RBL1, SMARCA2, TSG101*

Gene Expression

In this network there are 20 genes, which are participating in different aspects of gene expression: *CDKN1A, CEBPA, CITED1, CREG1, CTNNBIP1, DDX17, DUSP4, E2F1, E2F2, EGR2, EP300, HEXIM1, LHCGR, NPAS2, RBL1, SMARCA2, SMARCC2 (includes EG:6601), SOCS3, TSG101, ZMIZ1.*

Connective Tissue Development and Function

In this network there are four genes, which are participating into connective tissue development and function: *E2F1, EGR2, RBL1, SOCS3.*

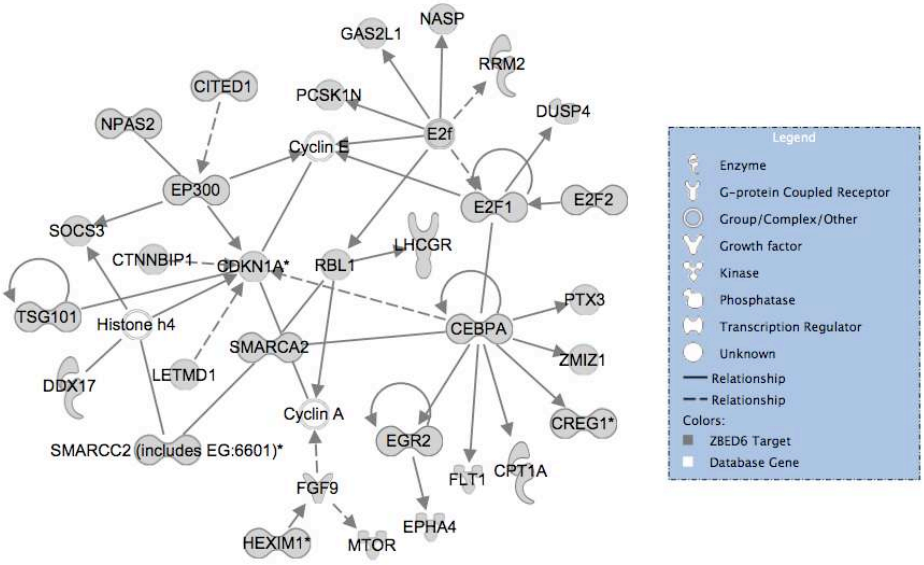


Figure 3.4: *A schematic diagram constructed by using IPA illustrating ZBED6 target gene networks. In this Network 31 genes are ZBED6 target genes, including 12 genes that encode transcription factors. This network illustrates genes for cell cycle, gene expression and connective tissue development and function most of these (20) genes have functional roles in gene expression.*

5): Cellular Development, Nervous System Development and Function, Cellular Movement

This network encompasses genes for cellular development, nervous system development and functions and cellular movement. Genes participating into above functions are illustrated below while network is shown in Figure 3.5.

Cellular Development

In this network there are 21 ZBED6 target genes, which are participating into cellular development: *ARRB2, BMP7, BMPR2, CCL2, CNR1, CXCL12, FGFR1, GDNF, IGF1R, IGFBP3, MMP14, NR2E1 (includes EG:7101), PLAU, PLCE1, POU4F1, POU4F2, RET, SIRT1, SRF, TBX21, TGFB2*

Nervous System Development and Function

In this network there are 13 genes, which are participating into nervous system development and function: *ARRB2, BMP7, CNR1, CXCL12, FGFR1, GDNF, MAPK10, PLAU, POU4F1, POU4F2, PPP1R1B, SIRT1, TGFB2*

Cellular Movement

In this network there are nine genes, which are participating into cellular movement: *BMP7, CCL2, CXCL12, GDNF, POU4F1, POU4F2, RET, SRF, TGFB2*

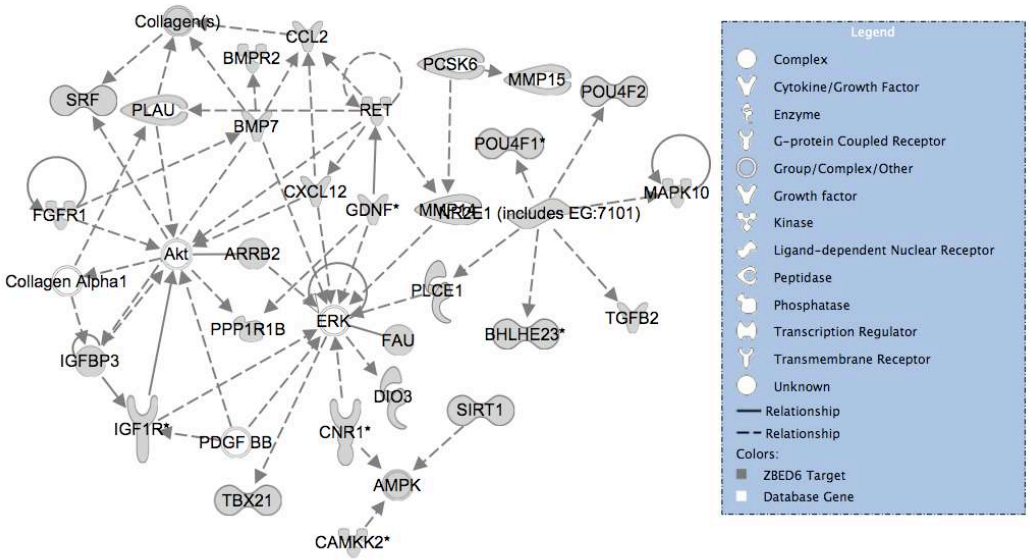


Figure 3.5: *A schematic diagram constructed by using IPA illustrating ZBED6 target gene networks. In this Network 29 genes are ZBED6 target genes, including six genes encoding for transcription factors. This network illustrates genes for cellular development, nervous system development and functions and cellular movement but many of genes (21) are related to cellular development.*

2: Canonical Pathways Analysis

ChIP sequence data canonical pathway analysis contains:

I. Top five canonical pathways table
II. Canonical pathways figures

Canonical pathway table and figures are illustrated and shown below.

I. Top five canonical pathways table

Top five canonical pathways table from ChIP sequence data contains name of canonical pathway, p-value and ratio, which are shown in table 2.

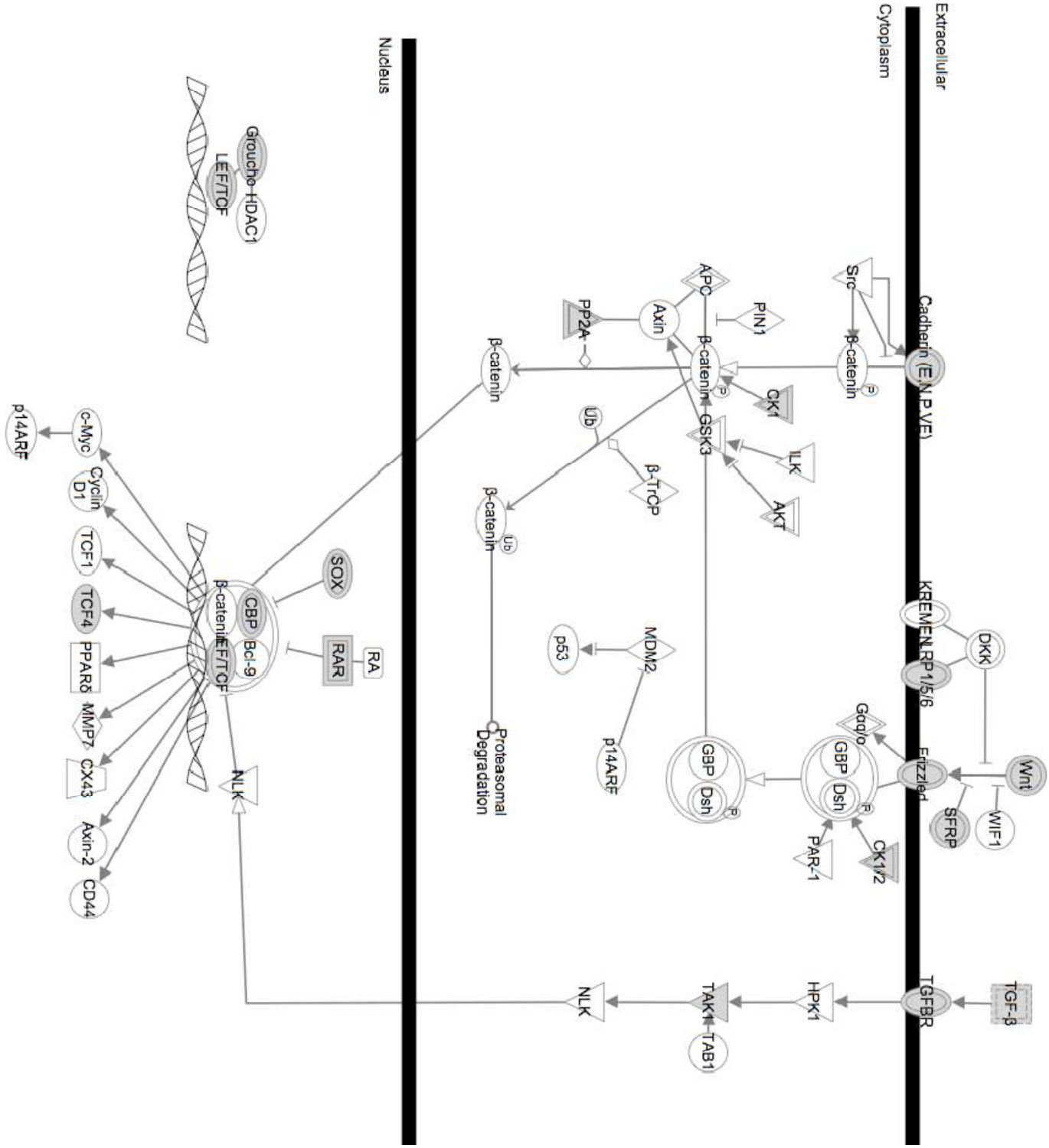
Table 2: *ZBED6 ChIP sequence data canonical pathway table. Canonical pathways names, p-value and ratio are shown below. Lower p-value and higher ratio indicates the significance of pathway.*

Name	***p-value***	***Ratio***
Wnt/β-catenin Signaling	*2.14E-09*	*37/167 (0.222)*
Human Embryonic Stem Cell Pluripotency	*4.93E-07*	*27/142 (0.19)*
Axonal Guidance Signaling	*6.92E-07*	*56/396 (0.141)*
TGF-β Signaling	*1.37E-06*	*20/83 (0.241)*
Role of Osteoblasts, Osteoclasts and Chondrocytes in Rheumatoid Arthritis	*2.61E-05*	*34/218 (0.156)*

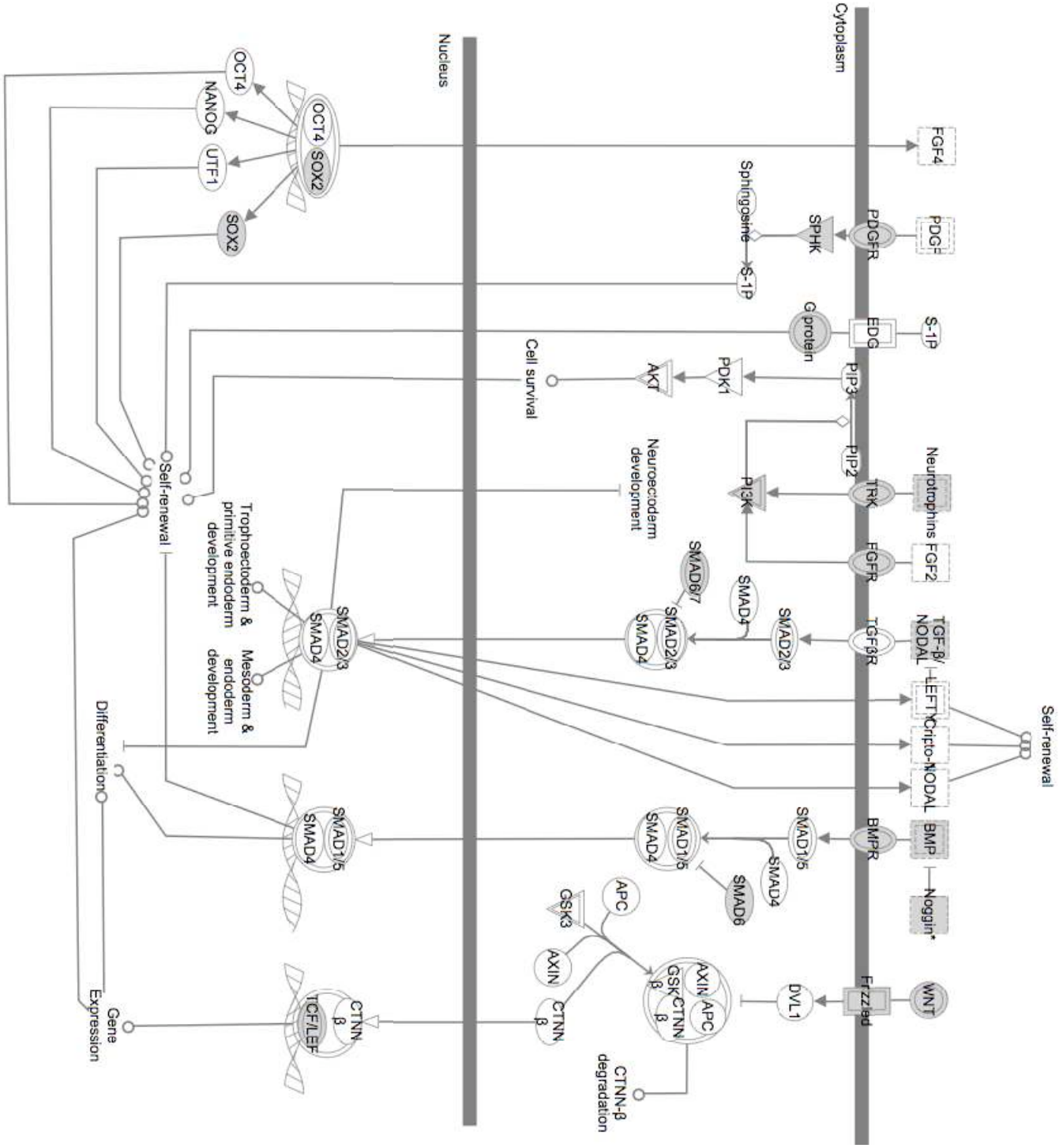
II. Canonical pathways figures

Top 3 canonical pathway figures are shown in Figure 3.6

Wnt/β-catenin Signaling



Human Embryonic Stem Cell Pluripotency



TGF-β Signaling

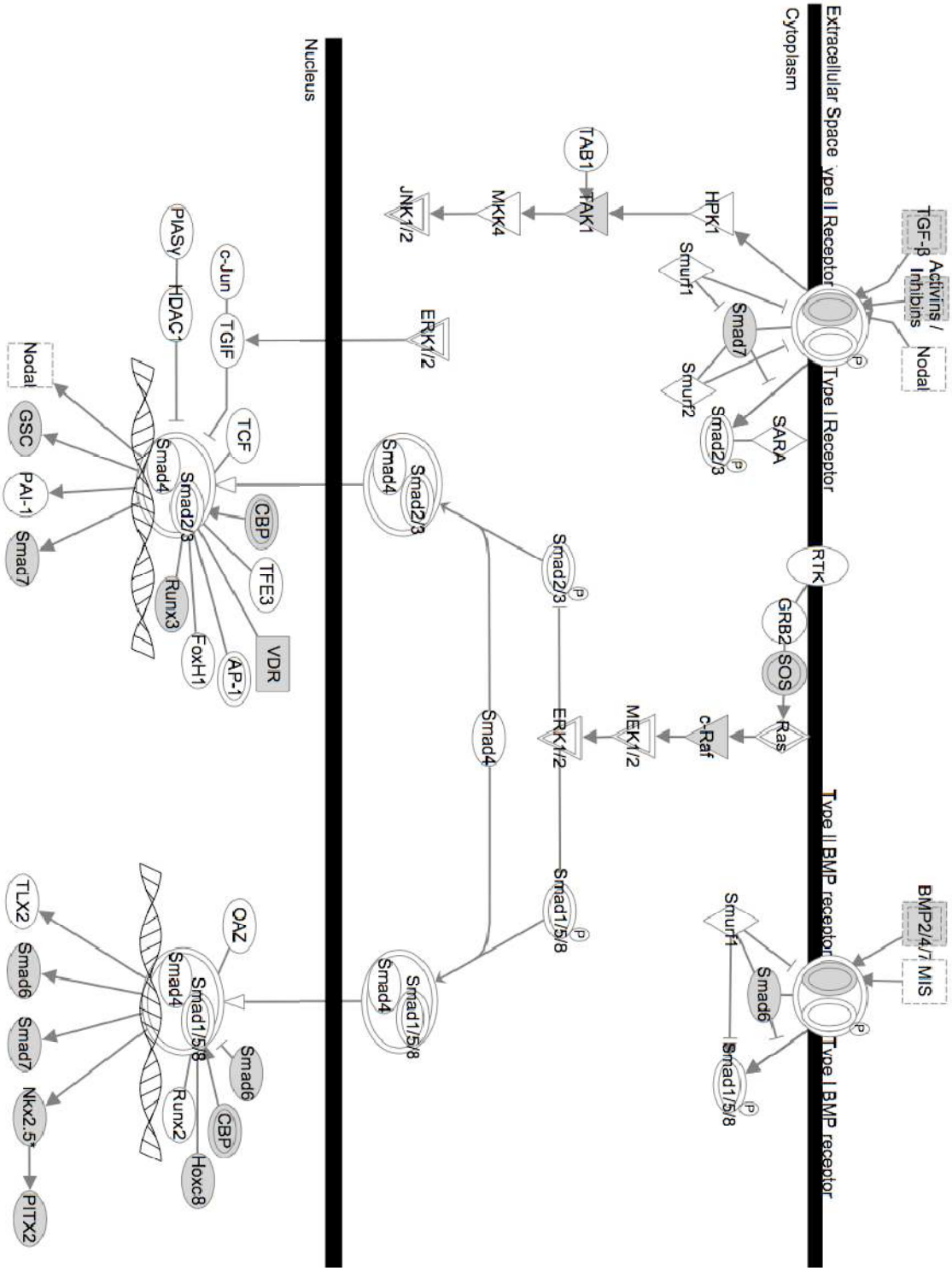


Figure 3.6: *In all canonical pathway figures ZBED6 target genes are shown with grey colored symbols. Wnt/β-catenin signaling cascade includes 37 ZBED6 target genes while human embryonic stem cell pluripotency contains 27 ZBED6 target genes and TGF-β Signaling covers 20 ZBED6 target genes.*

To elucidate how many ZBED6 targeted genes are shared between these cascades, IPA was used to get common genes between these three signaling cascades (Figure 3.7).

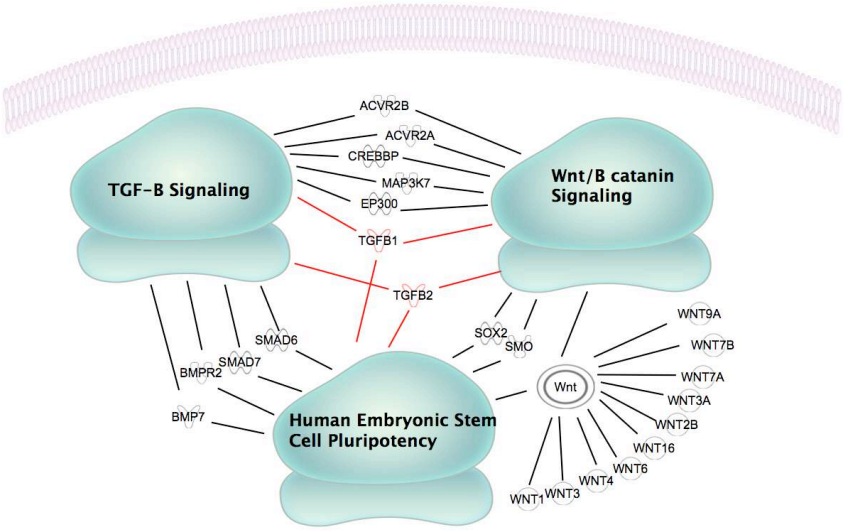


Figure 3.7: *Genes from ChIP sequence data that are common in Wnt, TGFβ and Human Embryonic Stem Cell Pluripotency signaling cascades. Genes depicted in red are playing role in all three pathways.*

ChIP sequence Gene Network for Skeletal Muscles

Data from ChIP sequencing of C2C12 cells were used to indicate genes participating into skeletal muscle development and used to construct a network (Figure 3.8).

Microarray data Analysis

Microarray data analysis, performed by using IPA, is divided into three parts:

1. Networks
2. Canonical pathways

These components are discussed below.

1: Networks

Microarray data networks analysis includes:

I. Top five networks table
II. Top five network figures

Network table and figures are illustrated below.

I: Top five networks table

Top five networks table from microarray data includes associated network functions and scores, which are shown in Table 4.

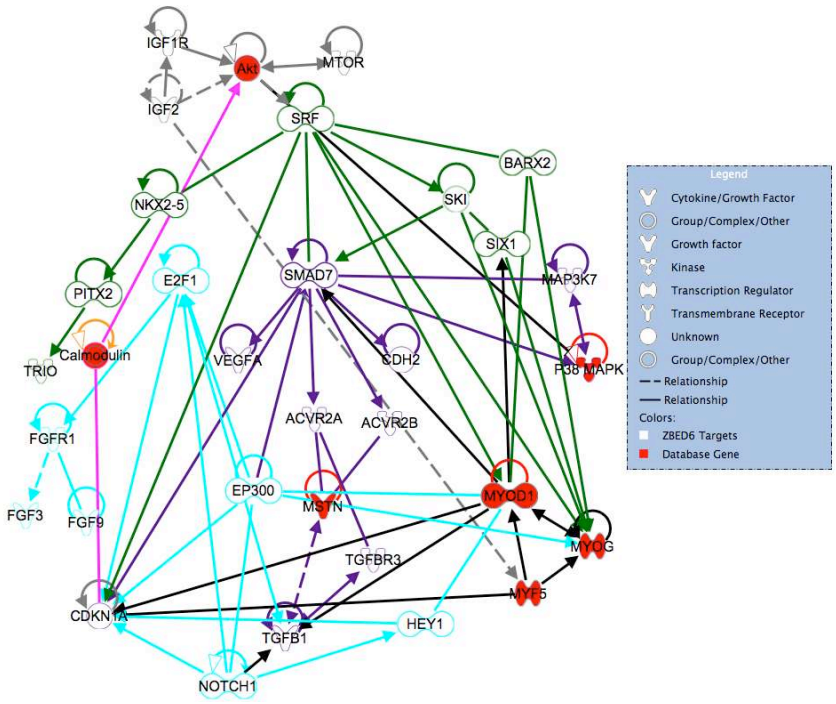


Figure 3.8: *A schematic diagram constructed by using IPA illustrating ZBED6 target genes participating into skeletal muscle development. Different colors are depicting different pathways. Green color molecule outlines and interactions are indicating SRF, NKX2-5, BARX2 and SKI network. Molecule outlines and interactions colored cyan are depicting EP300, NOTCH1, E2F1 network. Network including purple color molecule outlines and interactions are portraying SMAD7, ACVR2A, ACVR2B and TGFβR3. IGF2, MTOR, IGF1R molecule outline and interaction are shown with gray color. Molecules that are not ZBED6 direct target are filled with red color and their interaction are shown with black color.*

Table 4: *ZBED6 microarray data networks table. Network associated functions and scores are described below. Higher score is based on lower p-value and indicates the significance of network.*

Network Associated Functions	***Score***
Cellular Movement, Cell-To-Cell Signaling and Interaction, Tissue Development	*39*
Cardiovascular System Development and Function, Organismal Development, Cell-To-Cell Signaling and Interaction	*25*
Organismal Development, Cancer, DNA Replication, Recombination, and Repair	*18*
Cardiovascular System Development and Function, Organismal Development, Cell Death	*16*
Organismal Development, Antigen Presentation, Cellular Development	*16*

II: Top five network figures

Top five network figures constructed by using microarray data are illustrated and shown below.

1): Cellular Movement, Cell-To-Cell Signaling and Interaction, Tissue Development:

This network includes genes for cellular movement, cell-to-cell signaling and interaction, tissue development. Genes playing role into above functions are illustrated below while network is shown in Figure 3.9.

Cellular Movement:

In this network there are 13 genes, which are participating into cellular movement. *AGT, BMP4, BSG, CCND1, CDH2, CHI3L1, COL18A1, EPHB1, F2R, GAS6, LOX, TNFRSF21.*

Cell-to-Cell Signaling and Interaction:

In this network there are 13 genes, which has a functional role in cell-to-cell signaling and interaction. *AGT, BMP4, BSG, CCND1, CDH2, CHI3L1, COL18A1, EPHB1, F2R, GAS6, LOX, TFRC.*

Tissue Development:

In this network there are 16 genes, which are important for tissue development. *AGT, BMP4, BSG, CCND1, CCND2, COL16A1, COL18A1, DAB2, EPHB1, F2R, GAS6, LOX, PKRG2, RSAP2, TNFRSF2.*

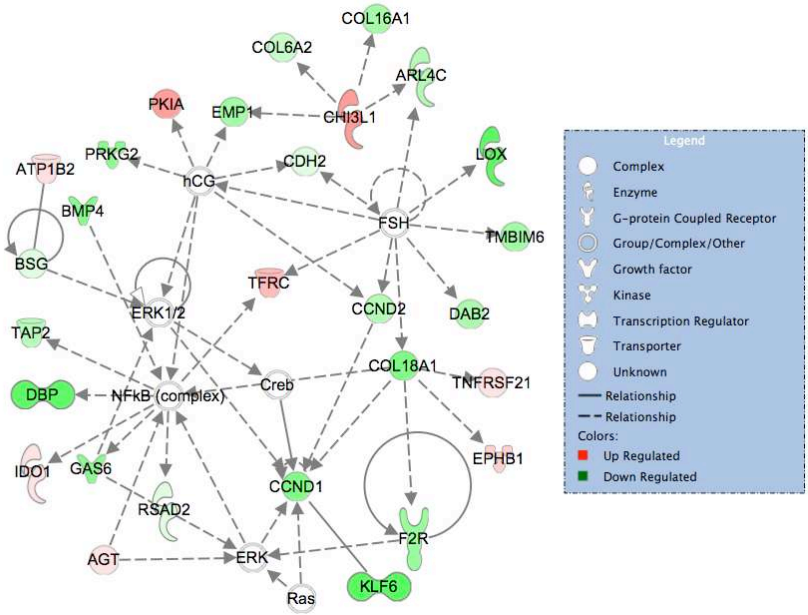


Figure 3.9: *A schematic diagram constructed by using IPA depicting network of ZBED6 upregulated and downregulated genes. In this network 28 genes are regulated by ZBED6, of which 20 are downregulated and 8 are upregulated. This network illustrates genes for cellular movement, cell-to-cell signaling and interaction, tissue development, of which 16 have established roles in tissue development.*

2): Cardiovascular System Development and Function, Organismal Development, Cell-To-Cell Signaling and Interaction

This network portrays genes for cardiovascular system development and function, organismal development, cell-to-cell signaling and interaction. Genes playing role into above functions are described below while network is shown in Figure 3.10.

Cardiovascular System Development and Function:

In this network there are 14 genes, which are known to participate into cardiovascular system development and function. *AKT1, ANPEP, HGF, IGF1, IGF2, IL18, IL1B, MET, MMP2, PDGFRA, SPARC, SPP1, THBS1, VCAM1.*

Organismal Development:

In this network there are 13 genes, which have been implicated into Organismal development. *AKT1, ANPEP, HGF, IGF1, IGF2, IL18, IL1B, MET, MMP2, PDGFRA, SPARC, THBS1, VCAM1*

Cell-To-Cell Signaling and Interaction:

In this network there are 16 genes, which are participating into cell to cell signaling and interaction. *AKT1, ANXA9, CCL5, CD28, HGF, IGF1, IGF2, IL7, IL18, IL1B, LBP, MMP2, SPARC, SPP1, THBS1, VCAM1*

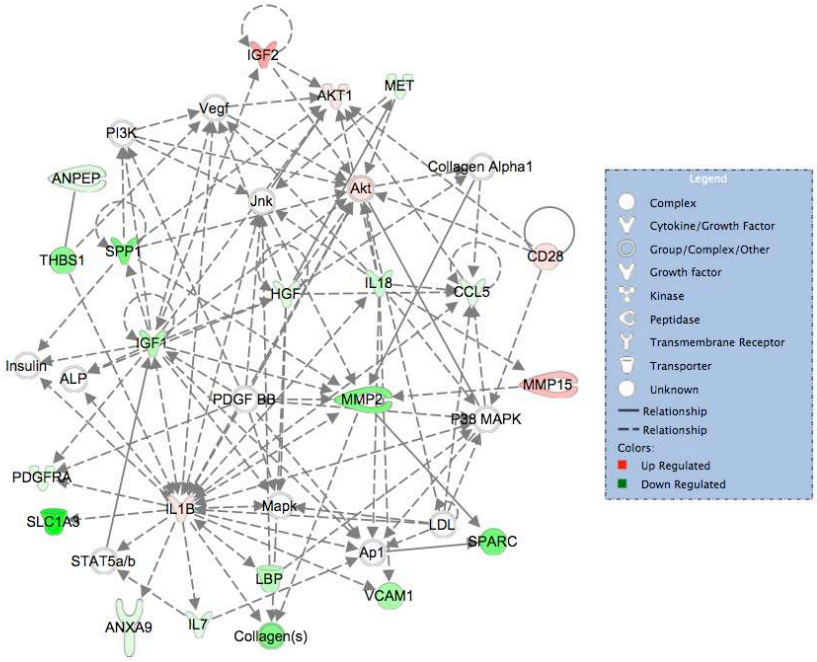


Figure 3.10: *A schematic diagram constructed by using IPA illustrating network of ZBED6 upregulated and downregulated genes. In this network, 21 genes expression is regulated by ZBED6, of which 15 are downregulated and 6 are upregulated. This network encompasses genes for cardiovascular system development and function, organismal development, cell-to-cell signaling and interaction. Many of genes (16) have established roles in cell-to-cell signaling and interaction.*

3): Organismal Development, Cancer, DNA Replication, Recombination, and Repair

This network encompasses genes for organismal development, cancer, DNA replication, recombination, and repair. Genes participating into above functions are illustrated below while network is shown in Figure 3.11.

Organismal Development:

In this network there are 9 genes, which have role into 0rganismal development. *BMP4, CCND1, CCND2, DBH, EPHB3, FGFR4, HGF, MMP2, STIM1.*

Cancer:

In this network there are 8 genes, which are known to participate into Cancer. *CCND1, CCND2, DPEP1, FGFR4, HGF, HTRA1, MMP2, STIM1.*

DNA Replication, Recombination, and Repair

In this network there are 4 genes, which are taking part into DNA Replication, Recombination, and Repair. *BMP4, HGF, CCND1, MMP2.*

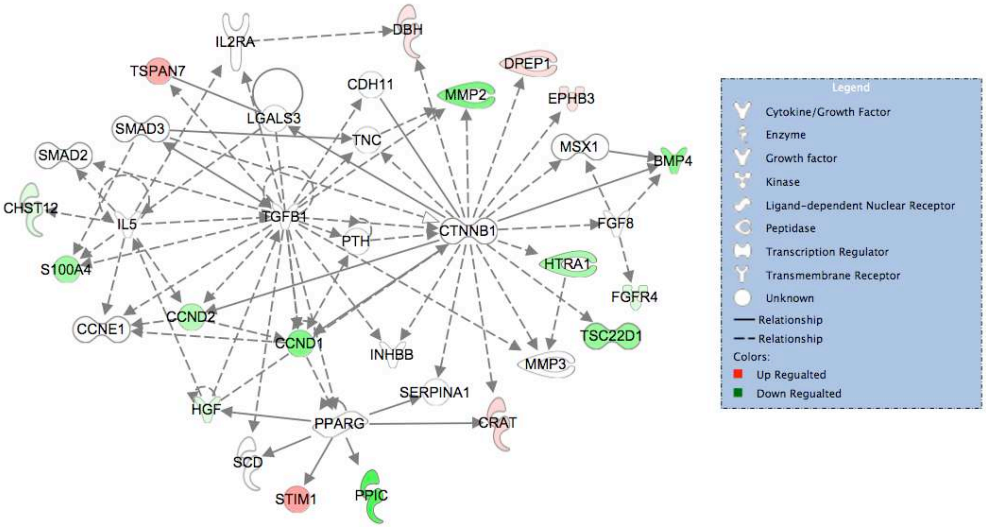


Figure 3.11: *A schematic diagram constructed by using IPA illustrating network of genes that are regulated by ZBED6. In this network, 17 genes are regulated by ZBED6, containing 11 downregulated and 6 upregulated genes. These genes are associated with cardiovascular system development and function, organismal development, cell-to-cell signaling and interaction. Majority of genes (9) have established functional role in organismal development.*

4): Cardiovascular System Development and Function, Organismal Development, Cell Death

This network includes genes for cardiovascular system development and function, organismal development, cell death. Genes playing role in above functions are illustrated below while network is shown in Figure 3.12.

Cardiovascular System Development and Function:

In this network there are 7 genes, which are known to be involved in cardiovascular system development and function. *ACTG2, ACTN1, CASQ2, COL18A1, MMP2, THBS1, THBS2.*

Organismal Development:

In this network there are 4 genes, which are participating into different stages of organismal development. *COL18A1, MMP2, THBS1, THBS2.*

Cell Death:

In this network there are 7 genes, which are participating into cell death. *COL18A1, EIF2AK2, MMP2, NEU2, TGFB3, THBS1, THBS2.*

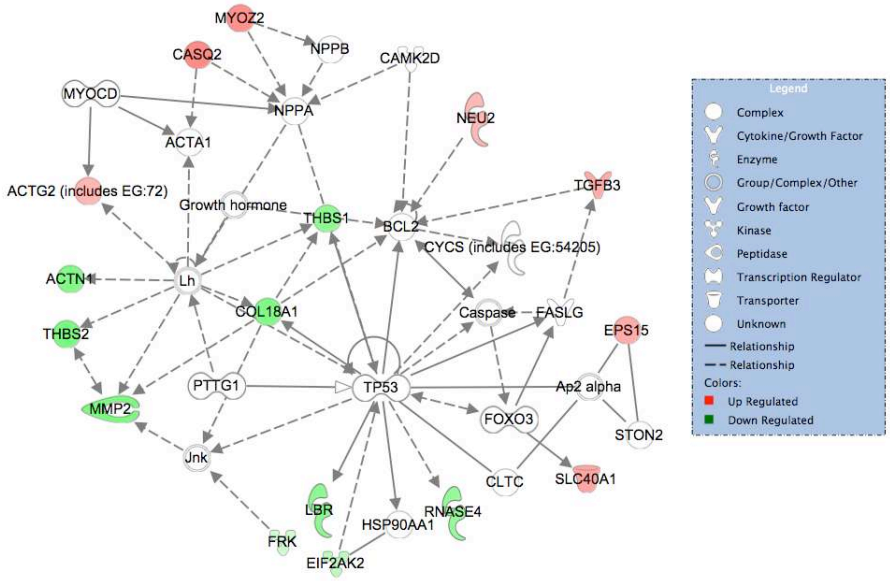


Figure 3.12: *A schematic diagram of genes regulated by ZBED6, IPA was used to construct this network. In this network, 16 genes are regulated by ZBED6, containing 9 downregulated and 10 upregulated genes. These genes are associated with cardiovascular system development and function, organismal development, cell death.*

5): Organismal Development, Antigen Presentation, Cellular Development

This network portrays genes for organismal development, antigen presentation, and cellular development. Genes playing role into above functions are described below while network is shown in Figure 3.13.

Organismal Development:

In this network there are six genes, which are known to be involved into organismal development. *ATP1B1, CCND2, GRB10, SERPINF1, SLC12A2, THBS2*

Cellular Development:

In this network there are three genes, which have role in cellular development. *CCND2, TCF7, THBS2.*

Antigen presentation:

This network contains seven genes, which are involved in antigen presentation. *AGT, CCND1, CCL5, HGF, IL18, JAM3, LBP.*

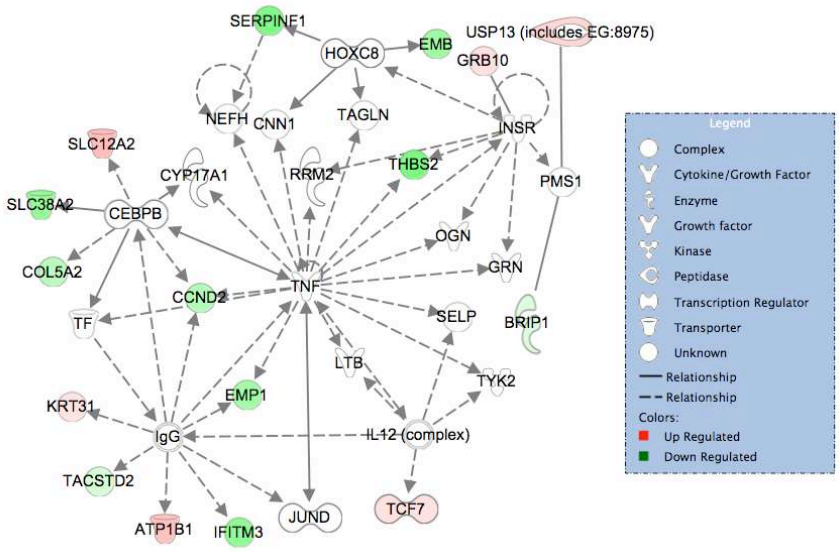


Figure 3.13: *A schematic diagram constructed by using IPA illustrating network of genes that are regulated by ZBED6. In this network, 16 genes are regulated by ZBED6, containing 10 downregulated and 6 upregulated genes. These genes are associated with organismal development, antigen presentation, and cellular development.*

2: Canonical Pathways Analysis

Microarray data canonical pathway analysis contains:

I. Top five canonical pathways table
II. Canonical pathways figures

Canonical pathway table and figures are illustrated and shown below.

I: Top five canonical pathways

Top five canonical pathways table from microarray data contains name of canonical pathway, p-value and ratio, which are shown in Table 5.

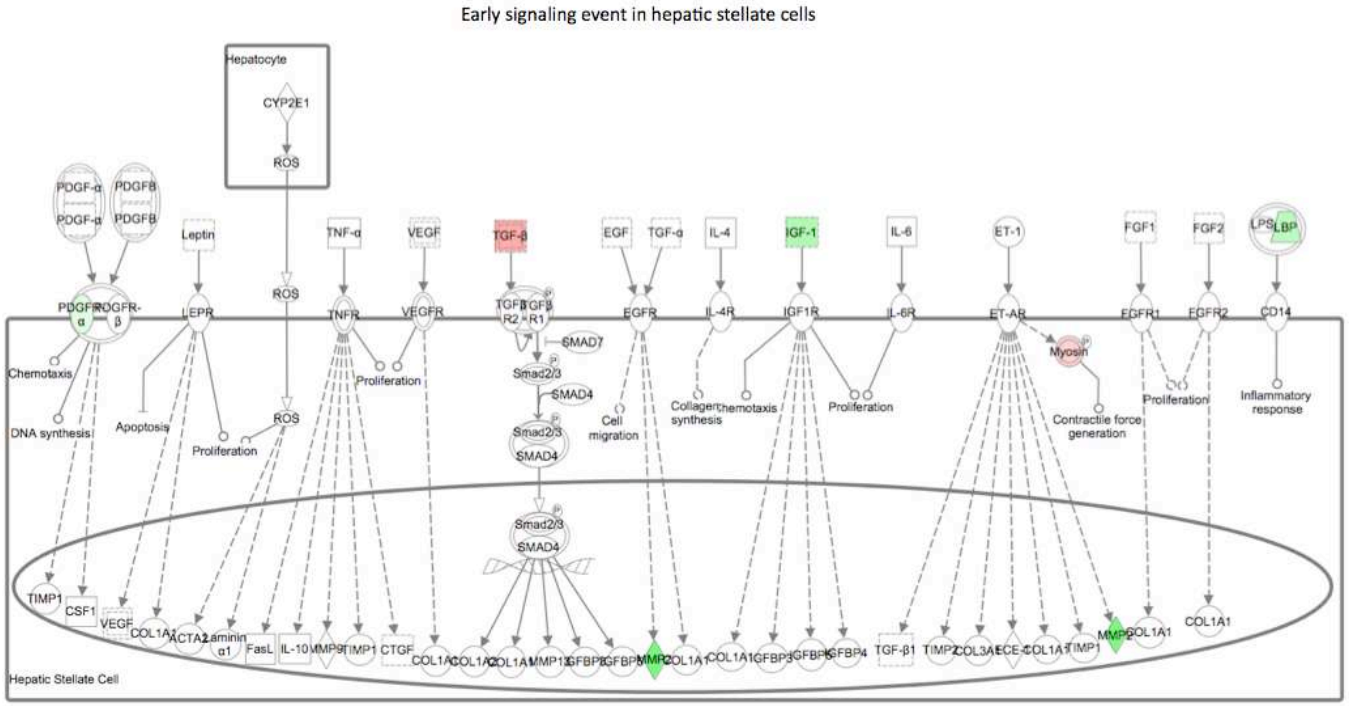
Table 5: *ZBED6 Microarray data canonical pathway table. Canonical pathway names, p-values and ratios are shown below. Lower p-value and higher ratio shows the significance of pathway.*

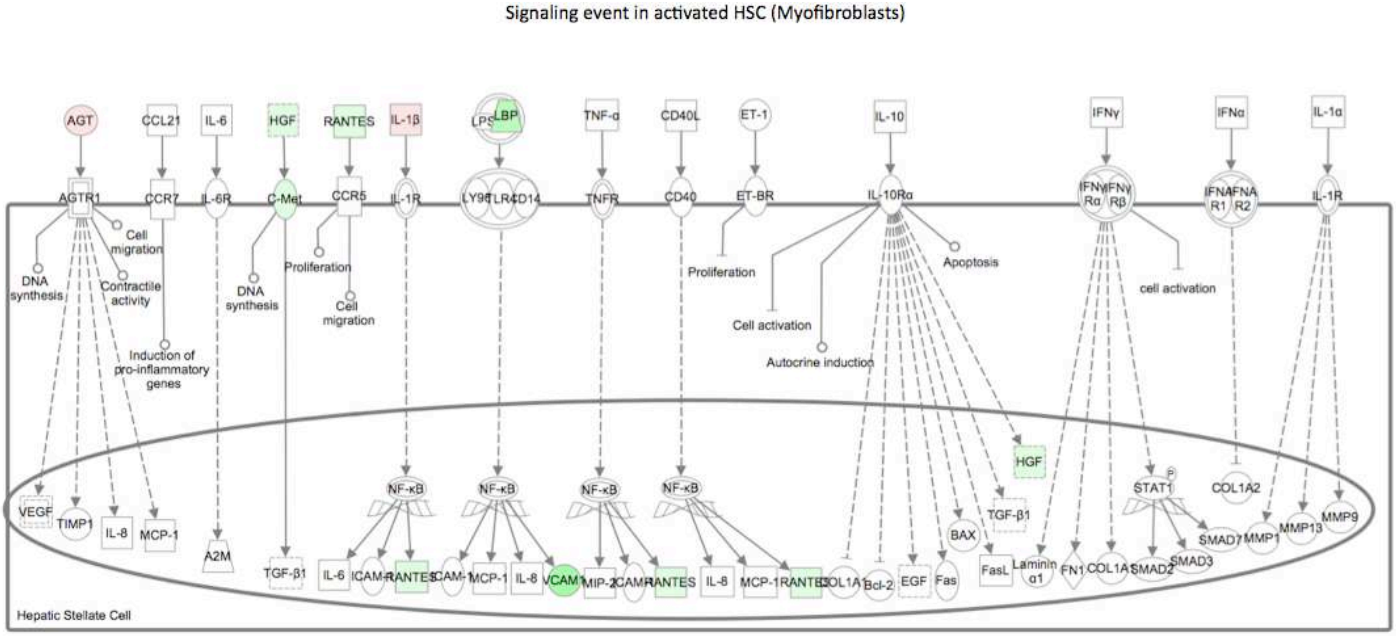
Name	***p-value***	***Ratio***
Hepatic Fibrosis / Hepatic Stellate Cell Activation	*2.86E-05*	*12/131 (0.092)*
Clathrin-mediated Endocytosis Signaling	*2.95E-03*	*10/164 (0.061)*
Tight Junction Signaling	*7.03E-03*	*9/164 (0.055)*
β-alanine Metabolism	*8.17E-03*	*5/53 (0.094)*
Human Embryonic Stem Cell Pluripotency	*8.44E-03*	*8/142 (0.056)*

II: Canonical Pathways Figures

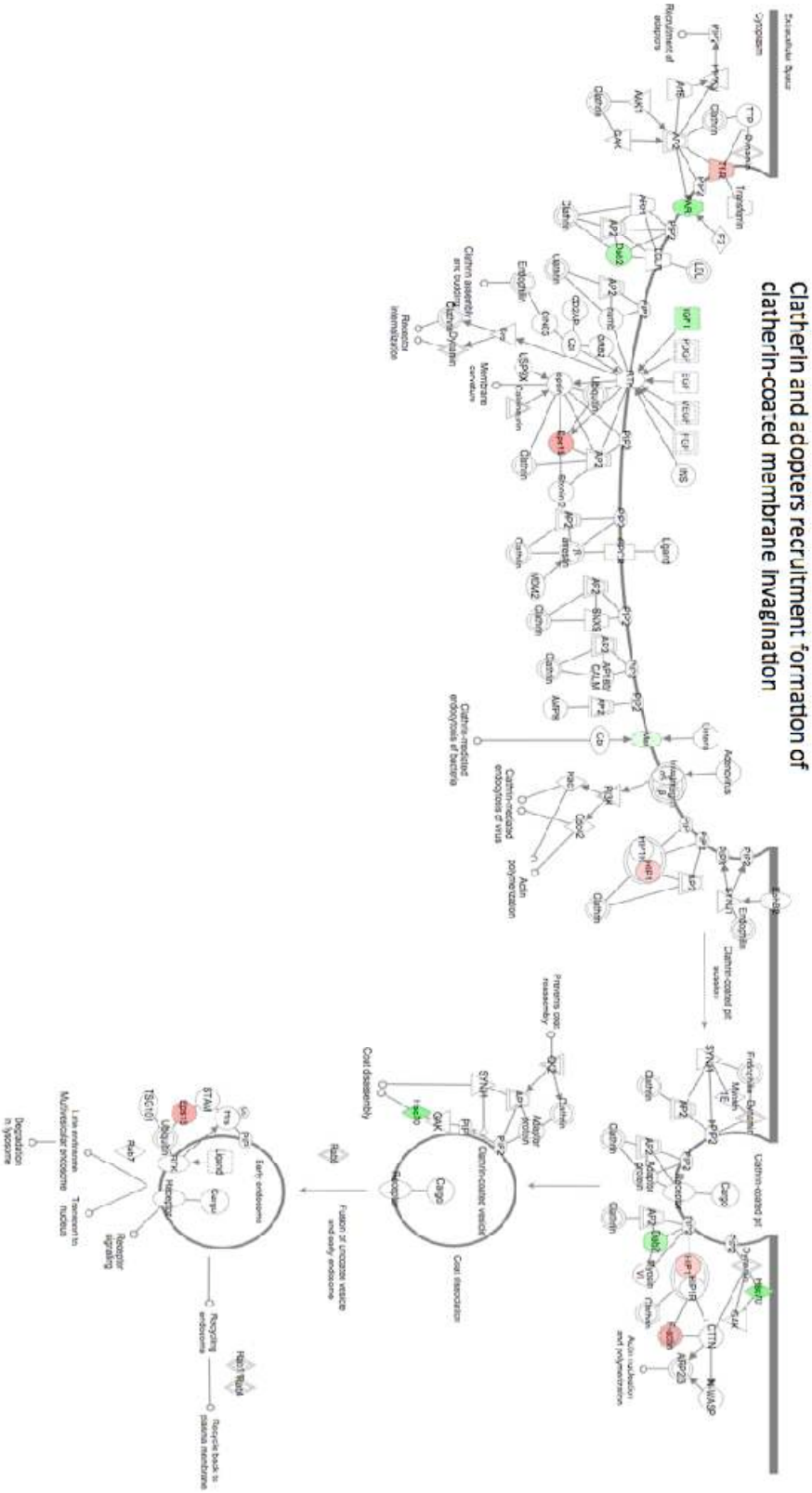
Canonical pathway figure are shown in Figure 3.15.

Hepatic Fibrosis / Hepatic Stellate Cell Activation





Clathrin-mediated Endocytosis Signaling



Tight Junction Signaling

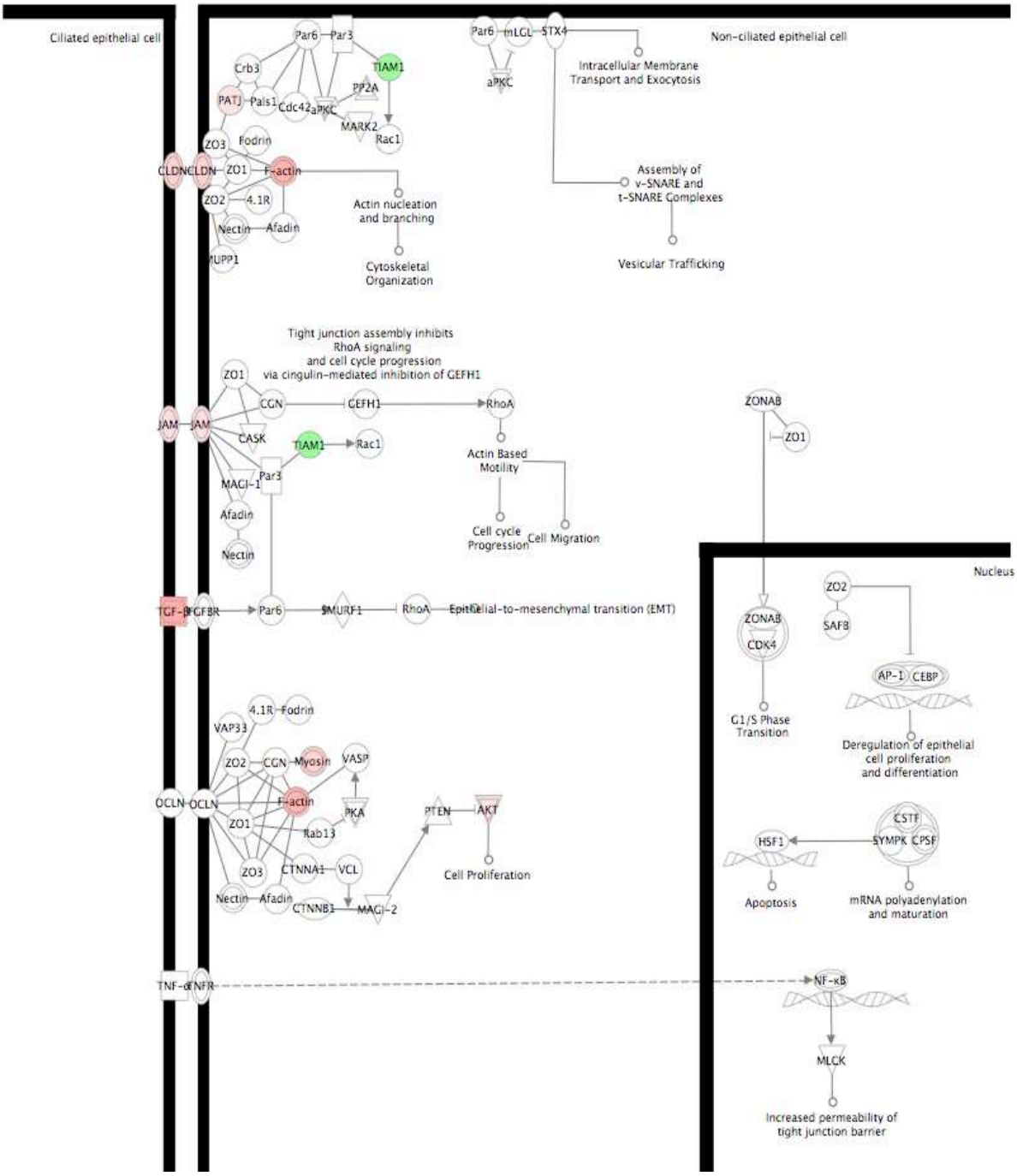


Figure 3.15: *In all canonical pathway figures ZBED6 regulated genes are shown with colored symbols, green for downregulated and red for upregulated. Hepatic Fibrosis / Hepatic Stellate Cell Activation contains 12 genes which are regulated by ZBED6; of which four are upregulated and six are downregulated. Clathrin-mediated Endocytosis Signaling includes 12 genes regulated by ZBED6, contains five upregulated and five downregulated while Tight Junction Signaling encompass nine ZBED6 regulated genes, including eight upregulated and one downregulated.*

To elucidate which genes are targeted or regulated by ZBED6 in figure 1.2, ZBED6 target and regulated genes are bring to light in figure 3.16.

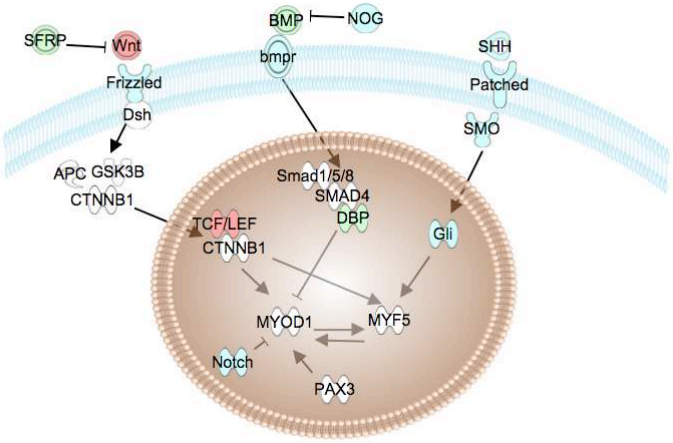


Figure 3.16: *This figure is a reflection of figure 1.2, but it highlights those factors, which are targeted or regulated by ZBED6. Colored genes are targeted by ZBED6 while green and red colors are indicating genes, which are negatively and positively regulated by ZBED6 respectively. White color genes are not targeted or regulated by ZBED6.*

Table 6: *Comparison of IPA with Other tools for Network Building*

Attributes	***IPA***	***Cytoscape***	***Osprey***
Interface	*It is a commercial service and provides user friendly interface, and consists of project manager, project workspace and quick start screen. Data can be shared through user account.*	*It is an open source software but is difficult to use due to different plug-in.*	*It is a public software and provides friendly interface.*
Data representation	*Data is represented in form of nodes, Relationships (connectors) and attributes. Nodes have different form for each type of molecules likewise connectors have different types for direct, indirect, catalysis, inhibitor etc. Different colors and gradients are available that can be applied on nodes and relationships.*	*Data is represented in form of nodes and connector but no different forms for nodes and connecters are available to differentiate nodes and connecters.*	*Data is represented in form of nodes and connector but no different forms for nodes and connecters are available to differentiate nodes and connecters.*
Input	*Excel files and tab delimited text files can be loaded.*	*SIF, SMBL, GML excel and tab delimited file can be loaded.*	*It accepts tab delimited text files.*
Output	*Local (Excel or tab delimited text) files, image export and printing*	*Local file and image printing*	*OSP file and image printing*

Search	*It provides search facility for genes and chemicals, function and diseases, pathway and tox list and also provide advance search options. It provides comprehensive gene ontology. Build and design pathway, add neighbours, and explores all published data about a gene interaction or pathway.*	*It provides search facility to find only node name on graph.*	*It provides search facility to find only node name on graph.*
Data integration	*It integrates huge number of databases KEGG, RefSeq, UniGene, BIND, BIOGRID, Cognia, DIP, INTACT, MINT, MIPs, Argonaute 2, TARBase, Gene Ontology, GVK Biosciences, Obesity Gene Map Database etc. and also provides various microarray platforms like illumina, applied biosystems, and affaymatrix.*	*It integrates GO, Pathway Commons, IntAct, BioMart, NCBI Entrez Gene, and PICR databases.*	*It integrate GO database.*
Filtering	*IPA is equipped with two type of filters stringent and relaxed which can be applied on data sources, species, tissue and cell lines, relationship types, molecule types, diseases, biofluids etc.*	*Flexible filtering is provided*	*Network and connection filters are available.*
Network operations	*It provides different network operation like: overlap networks, Network comparisons, build and design, grow, connect, trim, keep, add molecule/relationship to network, auto layout and sub cellular layout.*	*It provides different network layouts; various plug-ins are available for network operation.*	*It provides different network layouts like circumlunar, dual ring, global etc.*
Microarray Data	*It can perform various operations on microarray data: Import and Export microarray data, Building networks from microarray data, visually display of gene expression in Network, identifying canonical pathways harboring microarray data, Identifying bio functions (disease and disorders, molecular and cellular functions, physiological system development and function), tox function related to microarray data, also show statistical significance for each result.*	*Several plug-in are available*	*Not available*
Significance	*IPA finds maximum possible interaction and references. IPA is cited by almost 2651 scientific studies to this date.*	*Limited interactions and references.*	*Only finds direct interaction and limited references*

DISCUSSION

To elucidate the interactome of ZBED6, IPA was used which is a sophisticated and state-of-the-art tool for constructing and visualizing networks and pathways. At present it has been cited in 2651 scientific studies, and after comparison with other network building tools, it was found that IPA is

an outstanding network for constructing and analyzing tool due to its wide range of services and integrated databases, it is also clear from Table 6 where IPA is compared with Osprey and Cytoscape.

ChIP Sequence Data

ZBED6 targeted top five networks with significant score are shown in results. Network one encompasses ZBED6 targeted genes participating into tissue development, cancer and embryonic development, in mentioned functions number of ZBED6 targeted genes is 15, 16 and 14 respectively. *CDH2, CREEBP, FGF3, MSX1, PITX2, RELN, VEGFA, WNT1* are common in cancer, tissue development and embryonic development. This Network contains 10 ZBED6 targeted transcription factors. Network two indicates that 10 and 5 ZBED6 targeted genes are involved in cancer and immunological disorders, respectively while no ZBED6 targeted gene are involved in DNA repair, recombination and replication. In cancer and immunological disorders *BRD2, PDE4D, ROBO1, VAV3* are common. This network embraces five transcription factors. Network three indicates 27, 16 and 19 ZBED6 targeted genes take part into cellular growth and proliferation, tissue development and gene expression respectively. *CITED3, FOXO1, HEY1, IGF2, INHBB, MAP3K7, SKI, SMAD7, TGFβ1* are common in cellular growth and proliferation, tissue development and gene expression. This network contains 10 transcription factors. Network four shows 12, 20 and 4 ZBED6 targeted genes which have role in cell cycle, gene expression, and connected tissue development and function respectively. This network contains E2F1 and RBL1 common for cell cycle, gene expression and connected tissue development and function and also encompasses 12 transcription factors. Network 5 reflects 21, 13 and 9 ZBED6 targeted genes contribute for cellular development, nervous system development and function and cellular movement respectively. *BMP7, CXCL2, GDNE, POU4F2,* and *TGFβ2* are common in cellular development, nervous system development and function and cellular movement. This network contains 6 transcription factors.

From these networks it is revealed that ZBED6 targeted genes are involved in tissue development and cancer with tremendous number of genes: 31 and 26 respectively. In these networks *CDH2, CTNND2, PITX2, TRIO, WNT3a, WNT1, MSX1, PAX7, VEGFA, ACTN1, SMAD7, HEY1, IGF2, SKI, E2F1, EP300, FGF9, MTOR, IGF1R, FGFR1, SRF, BMP7, TGFβ2* has establish role in myogenesis (Figure 3.8) and it is discussed below.

Analysis to elucidate the role of ZBED6 target genes in muscle showed that ZBED6 is targeting various genes that are playing pivotal roles in muscle differentiation (Figure 3.8). Most mitogens are taking part into myoblasts proliferation but act as antagonist for myoblasts differentiation whereas IGFs are participating into both processes. IGF2 is playing an important role in muscle myogenesis. IGF2 is targeted by ZBED6 and regulates many other ZBED6 targeted genes downstream. IGF2 activates IGF1R and mediates activation of AKT signaling (Figure 3.8) subsequently induces myogenesis (Yoon & Chen 2008). AKT activates MTOR and SRF (Figure 3.8). MTOR participates into myogenesis by controlling IGF2 production (Yoon & Chen 2008). SRF plays a decisive role to initiate muscle differentiation by regulating transcriptional activation

of muscle specific genes (Kim et al., 2009). SRF induces expression of MRFs (MYOD1 and MYOG) and also interacts with many other muscle differentiation-mediating factors like BARX2, SKI, NKX2-5 and SMAD7 (Figure 3.8). BARX2 promotes skeletal myotube differentiation; its overexpression accelerates the onset of myotubes. In C2C12 during myotube formation expression of BARX2 is accelerated (Meech at al., 2003). SKI induces terminal muscle cell differentiation. SKI knockdown in C2C12 cells showed impaired differentiation. SKI induces transcription of MYOG by associating with SIX1 and EYA3 (Zhang & Stavnezer 2009). During C2C12 myoblasts differentiation and myotube formation NKX2-5 expression level is critical (Riazi et al., 2005) and it is activating expression of PITX2, which further upregulates TRIO. PITX2 contributes in morphogenesis and also ensures proper proliferation and differentiation of C2C12 (Gherzi et al., 2009). TRIO controls fetal skeletal muscle formation by mediating myoblasts localization or fusion (O'Brien et al., 2000). SMAD7 plays Pivotal role in muscle myogenesis by interacting with various important factors like VEGFA, CDH2, MAP3K7, ACVR2A, ACVR2B and EP300 (Figure 3.8). VEGFA increases angiogenesis and telomerase activity in skeletal muscles, endothelial and satellite cells (Zaccagnini et al., 2005). VEGFA induces increase myogenic differentiation in differentiating C2C12. Cadherins are important to exit cell cycle and to mediate differentiation of skeletal muscles. CDH2 named as N-cadherin makes a complex with p120 catenin and regulates activity of RHOA, which subsequently positively regulates MYOD expression and induces skeletal muscle myogenesis (Taulet et al., 2009). MAP3K7 is an upstream regulator of P38 signaling cascade, P38 MAP kinase regulates activation of MRFs thus controlling skeletal muscle differentiation (Lluis et al., 2005). Myostatin is a negative regulator of muscle myogenesis; it binds with ACVR2B and to lesser extent ACVR2A and mediates signals (Lee & McPherron 2001). Myostatin through SMADs regulates expression of MYOD. TGFβ1 induces myostatin expression, which conversely stimulates secretion of TGFβ1 in C2C12 myoblasts (Zhu et al., 2007). TGFβ inhibits expression of muscle-specific genes, thus myostatin and TGFβ acts as negative regulator of skeletal muscle myogenesis. EP300 named, as P300 is important for muscle myogenesis. Mouse embryos harboring mutation in p300 alluded to impaired MRFs expression and myogenesis (Roth et al., 2005). EP300 further interacts with NOTCH1 and E2F1 (Figure 3.8). NOTCH1 signaling leads to inhibition of myogenesis by suppressing P38 MAPK activity in C2C12 cells (Kondoh et al., 2007). NOTCH1 interacts with HEY1; HEY1 named CHF2 forms an inactive complex with MYOD and represses myogenesis transcriptionally (Sun et al., 2001). Thus, NOTCH1 either directly or through HEY1 can inhibit myogenesis. E2F1 also acts as negative mediator for inhibition of MYOD through retinoblastoma protein and cathepsin B complex (Li et al., 2000). These results unveil that ZBED6, by targeting the above mentioned factors, is an indispensable mediator for skeletal muscle myogenesis and is controlling both positive and negative regulator of muscle development simultaneously.

ZBED6 ChIP sequence data analysis by IPA revealed its role in various canonical pathways. ZBED6 targeted genes participate significantly in Wnt/β-catenin signaling, human embryonic stem (ES) cell pluripotency and TGFβ signaling. All of these pathways are very important and performing a vast array of functions. Wnt/β-catenin signaling is a an evolutionary conserved signaling cascade and plays a key role in stem cell maintenance, cell-cell adhesion, cell

differentiation and tissue development. Core components of this signaling cascade are the wingless-type mouse mammary tumor virus (MMTV) integration site family (WNT), disheveled (Dvl), axin, adenomatous polyposis coli (APC), glycogen synthase kinase 3beta (GSK3β), beta-catenin (CTNN-β) and T-Cell Factor (TCF) (Lee at al., 2003). WNT proteins bind with frizzled receptors and controls location and concentration of CTNN-β. CTNN-β and TCF complex is responsible for regulation of transcription of many genes. Absence of destruction complex leads to interaction of destruction complex (APC-axin- GSK3β) and CTNN-β. This interaction results in phosphorylation of CTNN-β which is further degraded by proteosomes (Mirams et al., 2010). Wnt is a complex, comprising many proteins of which ZBED6 targets are Wnt1, 2, 6, 16, 2B, 3A, 7A, 7B, 9A. In frizzled receptor complex ZBED6 target SMO whereas ZBED6 targets TCF4 in TCF complex. Secreted frizzled related proteins (SFRP) inhibit binding of Wnt with frizzled receptors. Among these SFRPs, SFRP2 and SFRP5 are targeted by ZBED6. In addition to these, some other Wnt signaling factors are targeted by ZBED6 like *CDH2, ACVR2A, 2B, CREBP, SOXs, TGFβ* and *LPRs* (Figure 3.6). These findings suggest that ZBED6 plays an important role in Wnt signaling by targeting various key factors.

Other important signaling cascade targeted by ZBED6 is human ES cell pluripotency. Studies of human ES cell pluripotency holds promise to provide knowledge about early development in humans and pluripotent ES cells could have high therapeutic potential. ES cells are derived from inner cell mass of blastocysts and can be differentiated into any cell type (Boyer et al., 2005). Human ES cell pluripotency signaling cascade includes different signaling pathways like WNT, Transforming growth factor beta (TGFβ), bone morphogenetic protein 2 (BMP), fibroblast growth factor 2 (FGF2) and platelet derived growth factor (PDGF) signaling which are responsible for gene expression, development, differentiation and self-renewal, respectively. ZBED6 targets many genes participating into human embryonic stem cell pluripotency like *Wnt (1, 2, 6, 16, 2B, 3A, 7A, 7B, 9A), SMO, BMP7, BMPR2, Noggin, TGFβ1, TGFβ2, SMAD6, SMAD7, PDGFRA,* and *SPHK1* (Figure 3.6). It proposes that ZBED6 is imperative for human ES cell pluripotency.

ZBED6 also takes part into TGFβ signaling. TGFβ signaling is involved in multiple different cellular activities including cell proliferation, differentiation, recognition, apoptosis, and specification of developmental fate both in embryonic and adult tissue (Shi & Massague, 2003). TGFβ signaling is mediated by interaction of TGFβ or BMP with type I and type II receptors. Receptor II phosphorylates receptor I which in turn phosphorylate sma and mad related proteins (SMADs) which are transcriptional regulators. There are different types of SMADS like receptor-regulated SMAD (R-SMAD), the Co-mediator SMAD (Co- SMAD), and inhibitory SMAD (I-SMAD). R-SMADs are SMAD, 2, 3, 5, and 8; Co- SMAD is SMAD4 while I-SMADs are SMAD6 and 7. R-SMADs are activated and make complex with co-SMAD and regulates transcription of genes, whereas I-SMADs compete with R-SMADs so I-SMAD act as antagonist for TGFβ signaling. Thus, SMAD proteins function either as transcriptional activators or repressors and their function as transcriptional regulators depend on their interaction with associated factors (Shi & Massague, 2003). *BMP7, BMPR2, TGFβ1, TGFβ2, I-SMADs* are targeted by ZBED6 (Figure 3.6). ZBED6 also targets some other factors which are known to implicate with TGFβ signaling like

EP300, VDR, RUNX3, ACVR2A, ACVR2B, MAP3K7, CREBBP, NKX2-5 and PITX2 (Figure 3.6). These results reflect that ZBED6 plays a significant role in TGFβ signaling.

In these three signaling cascades there are many genes that are shared. Analysis for common genes revealed that TGFβ1 and TGFβ2 are common among these three cascades (Figure 3.7). This is an important finding as ZBED6 by targeting these two genes can regulate both these signaling cascades. There are many genes shared between Wnt signaling and TGFβ signaling, TGFβ signaling and human embryonic stem cell pluripotency, and Wnt signaling and human ES cell pluripotency (Figure 3.7) These findings suggest that by regulating genes of one signaling cascade ZBED6 can also influence the other cascade.

Microarray data

ZBED6 regulated top five networks are showed in results**.** Network one includes total 28 ZBED6-regulated genes. 13, 13 and 16 genes are contributing for cellular movement, cell-to-cell interaction, and tissue development respectively, among these genes *AGT, BMP4, BSG, CCND1, COL16A1, COL18A1, DAB2, EPHB1, F2R, GAS6,* and *LOX* are common. 20 genes are upregulated while eight are downregulated by ZBED6 after silencing. Network two encompasses total 21 ZBED6-regulated genes. 14, 13 and 16 ZBED6-regulated genes are participating into cardiovascular system development and function, organismal development and cell-to-cell signaling and interaction respectively, and AKT1, HGF, IGF1, IGF2, IL18, IL1B are common ZBED6-regulated genes for these functions. 15 downregulated and six upregulated genes are present in this network. Network three contains a total of 17 ZBED6 regulated genes. In this network nine, eight and four ZBED6-regulated genes are playing role in organismal development, cancer and DNA replication, recombination and repair respectively, while HGF, CCND1, MMP2 are common ZBED6-regulated genes for these functions. This network shows ZBED6 downregulates 11 genes and upregulates 11 genes. Network four reflects total 16 ZBED6-regulated genes. This network includes seven, four and seven ZBED6-regulated genes which are known to be involved in cardiovascular development and function, organismal development, and cell death respectively, while Col18A1, MMP2, THBS1, THSB2 are common ZBED6-regulated genes among them. Nine genes are downregulated while seven are upregulated by ZBED6 in this network. In network 5, 16 genes are regulated by ZBED6 of which nine are downregulated while six are upregulated and; six, three and seven ZBED6 regulated genes are participating into organismal development, cellular development and antigen presentation respectively.
These networks illustrates that ZBED6 regulated genes are involved in organismal development and cell to cell signaling and interactions with highest number of genes 32 and 29 respectively. In these networks, BMP4, DBP, CDH2, AGT, IGF1, IGF2, THBS1, PDGFRA, MPP2, AKT1, HGF, MET, FGF4, TGFβ3, ACTN1, F2R, VCAM1 are playing role in muscle proliferation, myogenesis and contraction.

ZBED6 microarray data analysis by IPA divulged it as a key regulator in various canonical pathways. ZBED6-regulated genes are contributing significantly in Hepatic Fibrosis / Hepatic Stellate Cell Activation, Clathrin-mediated Endocytosis Signaling and Tight Junction Signaling.

These signaling cascades are executing many important functions. Hepatic fibrosis is a chronic liver disease that is caused by hepatotoxins including excessive ethanol, glucose, bile acids, free fatty acids and viruses. Hepatotoxins initiate a pro-inflammatory events cascade that leads to activation of Hepatic stellate cells (HSCs) followed by cytokine secretions that disseminate their activated state. Liver injury leads to accumulation of activated HSCs and myofibroblasts, which stimulate synthesis of large amount of extra cellular matrix (ECM) proteins, mainly collagen which conducts tissue fibrosis and eventually liver fibrosis. Imbalance between synthesis and degradation of ECMs causes liver fibrosis (Zou et al., 2007). ZBED6 positively regulates TGFβ3 and MYL6B, which are part of TGFβ and Myosin complex respectively; and ZBED6 negatively regulates *PDGFRα, IGF1, LBP and MMP2* during early stage in hepatic stellate cells (Figure 3.15). During signaling events in myofibroblasts, ZBED6 positively regulates *AGT, IL-1β* while negatively regulates *HGF, C-Met, RANETS (CCL5), LBP, VCAM*. It implies that ZBED6 may be a crucial factor for hepatic fibrosis either as positively or negatively regulator.

ZBED6 also regulates genes in clathrin-mediated endocytosis signaling cascade. Clathrin-mediated endocytosis signaling is also called receptor-mediated endocytosis. It is responsible for internalization of cargo, which may be hormones, nutrients or other signaling molecules from plasma membrane to interacellular organelles. Clathrin and adopter proteins help in internalization of cargo (Henne et al., 20101). Clathrin lattice is assembled at the plasma membrane by Adaptor protein 2 (AP2), which is connected to predominant phosphoinositide at plasma membrane. Cargo binds to respective receptor and Ap2 interact with various proteins including EPS15, epsin, HIP1, DAB2 and leads to formation of clathrin coated pits (CCP), Dynamin along with various proteins is recruited, activated and directs to release of clathrin coated vesicle. Vesicles fuse with early endosome after internalization and coat disassembly. During clathrin and adopter recruitment, and formation of clathrin coated membrane invagination, ZBED6 positively regulates *TFRC (TFR)*, *EPS15,* and *HIP1*, and negatively regulates, *IGF1, DAB2, PAR1 (F2R), C-Met* (member of Met complex) (Figure 3.15). During clathrin coated pit formation and coat dissociation ZBED6 positively regulates *ACTC1* and *ACTG2* (member of F-Actin complex) while negatively regulates *DAB2* and *HSPA8* (member of HSC70 complex), and during early endosome ZBED6 positively regulates *Eps15* (Figure 3.15). These results indicate that ZBED6 regulates clathrin-mediated endocytosis by influencing sufficient number of genes contributing in this signaling cascade.

ZBED6 regulated genes are involved in tight junction signaling. Tight junction signaling is important for cell-to-cell signaling, cellular assembly and organization and cell morphology. Tight junctions are composed of transmembrane proteins (claudins (CLDN), occludins (OCLN) and JAMs (JAM)), adaptors proteins (ZO1,2,3, PAR3,6, PATJ and MUPP1), regulatory proteins (aPKC, RAB13,3b, PP2A and PTEN) and transcriptional and post-transcriptional regulators (ZONAB, SYMPK and HSF1) (Matter & Balda 2003).In this signaling, CLDNs, OCLNs and JAMs are major players. These transmembrane proteins interact with other proteins and forms tight junction complexes that are involved in actin remodeling, cytoskeleton organization and epithelial polarization and ultimately mediate cellular assembly, cell-to-cell signaling, and organization and cell morphology. ZBED6 positively regulates *CLDN (CLDN2), JAM (JAM3), TGFβ (TGFβ3)*, F-

actin *(ACTC1, ACTG2s)* and myosin *(MYL6B)* while negatively regulates *TIAM1*. It concludes that ZBED6 plays an important role in tight junction signaling by regulating important factors.
According to previous studies Wnt, Bmp and Shh signaling are involved in regulation of embryonic regulations similarly these cascades may consign adult stem cell to myogenesis. ZBED6 is targeting and regulating various molecules in these cascades (3.16). ZBED6 also targets and regulates cadherins, and FGFs, which have role in myoblasts multiplication and fusion. These findings suggest that ZBED6 has a major role in skeletal muscle myogenesis.

CONCLUSION

ZBED6 is a recently identified transcription factor that is targeting and regulating a plethora of molecular factors. The interactome of ZBED6 was constructed by using ChIP sequence and microarray data through IPA. The ZBED6-targeted genes interactome elucidates that ZBED6 targeted genes are overrepresented in tissue development and cancer networks. ZBED6-targeted genes are engaged significantly into Wnt, human embryonic stem cell pluripotency and TGFβ canonical pathways. ZBED6 binds to many regulatory regions of genes, which have established roles in skeletal muscle myogenesis including: *IGF2, IGF1R, SRF, SMAD7, CDH2, CTNND2, PITX2, TRIO, WNT3a, WNT1, MSX1, PAX7, VEGFA, ACTN1, HEY1, SKI, E2F1, EP300, FGF9, MTOR, FGFR1, BMP7,* and *TGFβ2*. It is revealed from the interactome of ZBED6 differentially regulated genes that it is mainly involved in organismal development and cell to cell interaction and signaling. ZBED6 differentially regulated genes are contributing significantly in hepatic fibrosis, clathrin-mediated endocytosis and tight junction signaling cascades. ZBED6-regulated genes including *BMP4, DBP, CDH2, AGT, IGF1, IGF2, THBS1, PDGFRA, MPP2, AKT1, HGF, MET, FGF4, TGFβ3, ACTN1, F2R*, and *VCAM1* have established roles in muscle proliferation, myogenesis and contraction. Wnt, Bmp and Shh signaling are controlling regulation of MYOD and MYF5 which are important myogenic transcription factors for skeletal muscle differentiation; ZBED6 targets and regulates many factors in these signaling cascades. These findings suggest that ZBED6 is a crucial factor for many cellular functions and canonical pathways either by directing or differentially regulating the molecular factors.

FUTURE PERSPECTIVES

- ZBED6 networks will be elucidated and constructed in other cell lines like HepG2, it will facilitate in understanding the interactome of ZBED6 in other cell lines besides C2C12 myoblasts.
- The ZBED6 interactome will be identified and built from human tumor samples to elucidate the role of ZBED6 in human cancer as our results suggested that ZBED6 targeted genes are involved profoundly in cancer networks.

- Some of the ZBED6 targeted or regulated key pathways will be validate by employing functional assays. These wet lab experiment will provide us a base to accept our computational results.

ACKNOWLEDGMENTS

I feel pleasure to place on record my deep sense of thankfulness and indebtedness to my supervisor **Goran Andersson** and co-supervisors **Erik Bongcam Rudloff** and **Leif Andersson** for their supervision, kind attitude and in preparation of this report. I also want to express my thanks to ZBED6 group for providing me data and assistance. I feel to obligatory to express my thanks to my best friends Asim Ali, Romaisa Ghazal, Muhammad Kashif, Amara Jabeen, Naghmana Kanwal and Dr. Ishtiaq Qadri. They are always there whenever I need a good friend. I would like to thanks my family, my colleagues in bioinformatics program and all other friends. For funding, I want to express my gratitude and indebtedness to **Higher Education Commission Pakistan** (HEC).

REFERENCES

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

Andersson-Eklund L., Marklund L, Lundstrom K., Haley C. S., Andersson K., Hansson I., Moller M. and Andersson L. (1998) Mapping quantitative trait loci for carcass and meat quality traits in a wild boar x Large White intercross. *Journal of Animal Science*. **76**: 694-700.

Baitaluk M., Sedova M., Ray A., and Gupta A. (2006) BiologicalNetworks: visualization and analysis tool for systems biology. *Nucleic Acids Research*. **34**: 466–471.

Baker J., Hardy M.P., Zhou J., Bondy C., Lupu F., Bellve A.R., Efstratiadis A. (1996) Effects of an IGF1 gene null mutation on mouse reproduction. *Mol. Endocrinol*. **10**: 903-918.

Barton-Davis E.R., Shoturma D.I., Musaro A., Rosenthal N., Sweeney H.L. (1998) Viral mediated expression of insulin-like growth factor I blocks the agingrelated loss of skeletal muscle function. *Proc. Natl. Acad. Sci*. **95**: 15603–15607.

Barton E.R., Morris L., Musaro A., Rosenthal N., Sweeney H.L. (2002). Muscle specific expression of insulin-like growth factor I counters muscle decline in mdx mice. *J. Cell Biol.* **157**: 137–147.

Braunschweig M.H., Van-Laere A.S., Buys N., Andersson L., Andersson G. (2004) IGF2 antisense transcript expression in porcine postnatal muscle is affected by a quantitative trait nucleotide in intron 3. *Genomics*. **84(6)**: 1021-9.

Borello U., Berarducci B., Murphy P., Bajard L., Buffa V., Piccolo S., Buckingham M., Cossu G. (2006) The Wnt/beta-catenin pathway regulates Gli-mediated Myf5 expression during somitogenesis. *Development.* **133(18):** 3723-32.

Boyer A. L., Lee I. T., Cole F. M., Johnstone E. S., Levine S. S., Zucker P. J., et al. (2005) Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. *Cell.* **122:** 947–956.

Brunk B.P., Goldhamer D.J., Emerson C.P. **(**1996) Regulated demethylation of the myoD distal enhancer during skeletal myogensis. *Dev Biol.* **177(2):** 490-503.

Chen Kefei, Tara B., Muir W. M., Groenen M. A., and Schook L. B. (2007) Genetic Resources, Genome Mapping and Evolutionary Genomics of the Pig *(Sus scrofa). Int J Biol Sci.* **3(3):** 153–165.

Clow C., Jasmin B.J. (2010) Skeletal Muscle-derived BDNF Regulates Satellite Cell Differentiation and Muscle Regeneration. *Mol Biol Cell* [Epub ahead of print]

Dupont J., LeRoith D., (2001) Insulin and insulin-like growth factor I receptors: similarities and differences in signal transduction. *Horm. Res.* **55**: 22–26.

El-Shewy H.M., Lee M.H., Obeid L.M., Jaffa A.A., Luttrell L.M. (2007) The insulin- like growth factor type 1 and insulin-like growth factor type 2/mannose-6- phosphate receptors independently regulate ERK1/2 activity in HEK293 cells. *J. Biol. Chem.* **282**: 26150–26157.

Florini J.R., Magri K.A., Ewton D.Z., James P.L., Grindstaff K., Rotwein P.S., (1991)

Spontaneous differentiation of skeletal myoblasts is dependent upon autocrine secretion of insulin-like growth factor-II. *J. Biol. Chem*. **266:** 15917–15923.

Fujii J., Otsu K., Zorzato F., de Leon S., Khanna V.K., Weiler J.E., O'Brien P.J., MacLennan D.H. (1991) Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science*. **253(5018):** 448-51.

Gherzi R., Trabucchi M., Ponassi M., Gallouzi I-E, Rosenfeld M. G., Briata P. (2009) Akt2-mediated phosphorylation of Pitx2 controls Ccnd1 mRNA decay during muscle cell differentiation. *Cell Death and Differentiation.* **17:** 975–983.

Gustafsson M. K., Pan H, Pinney D.F., Liu Y., Lewandowski A., Epstein D. J., Emerson C.P. (2002) Myf5 is a direct target of long-range Shh signaling and Gli regulation for muscle specification. *Genes Dev.***16**, 114–126.

Harada A., Okada S., Odawara J., Kumamaru H., Saiwai H., Aoki M., Nakamura M., Nishiyama Y., Ohkawa Y. (2010) Production of a rat monoclonal antibody specific for Myf5. *Hybridoma (Larchmt).* **29(1):** 59-62.

Henne M. W., Boucrot E., Meinecke M., Evergren E., Vallis Y., Mittal R., McMahon H. T. (2010) FCHo Proteins Are Nucleators of Clathrin-Mediated Endocytosis. *Science* **328:**

Hirsinger E., Malapert P., Dubrulle J., Delfini M.C., Duprez D., Henrique D., Ish-Horowicz D. and Pourquie O. (2001) Notch signalling acts in postmitotic avian myogenic cells to control MyoD activation. *Development* **128**, 107–116.

Horst D., Ustanina S., Sergi C., Mikuz G., Juergens H., Braun T., Vorobyov E. (2006) Comparative expression analysis of Pax3 and Pax7 during mouse myogensis. *Int. J. Dev. Biol.* **50:** 47-54.

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

Jeong H., Bae S., An S.Y., Byun M.R., Hwang J.H., Yaffe M.B., Hong J.H., Hwang E.S. (2010) TAZ as a novel enhancer of MyoD-mediated myogenic differentiation. *FASEB J.* [Epub ahead of print]

Jones J.I., Clemmons D.R., (1995) Insulin-like growth-factors and their binding- proteins biological actions. Endocr. Rev. 16: 3–34.

Kablar, B. *et al.* (1999) Myogenic determination occurs independently in somites and limb buds. *Dev. Biol.* **206**, 219–231.

Kim J., Kee H.J., Kim J.Y., Joung H., Nam K., Eom G.H., Choe N., Kim H, Kim J.C., Kook H., Seo S. B., Kook H. (2009) Enhancer of Polycomb1 Acts on Serum Response Factor to Regulate Skeletal Muscle Differentiation. *JBC.* **284**: 6308 –16316.

Kitano H. (2002) Systems Biology: A Brief Overview. *Science*. **295:** 1662 – 1664.

Kondoh K., Sunadome K., and Nishida E. (2007) Notch Signaling Suppresses p38 MAPK Activity via Induction of MKP-1 in Myogenesis. *THE JOURNAL OF BIOLOGICAL CHEMISTRY*. **282**: 3058 – 3065.

Langley B., Thomas M., Bishop A., Sharma M., Gilmour S., Kambadur R. (2002) Myostatin inhibits myoblast differentiation by down-regulating MyoD expression. *J Biol Chem.* **277(51):** 49831-40.

Lau M.M.H., Stewart C.E.H., Liu Z.Y., Bhatt H., Rotwein P., Stewart C.L. (1994) Loss of the imprinted Igf2/cation-independent mannose 6-phosphate receptor results in fetal overgrowth and perinatal lethality. *Genes Dev*. **8:** 2953–2963.

Lee E**.**, Salic A., Krüger R., Heinrich R., Kirschner M.W. (2003) The roles of APC and Axin derived from experimental and theoretical analysis of the Wnt pathway. *PLoS Biol*. **1:** 116- 132

Lee S. and McPherron A. C. (2001) Regulation of myostatin activity and muscle growth. *PNAS.* **98**: 9306 –9311.

L'honore A., Lamb N.J., Vandromme M., Turowski P., Carnac G., Fernandez A. (2003) MyoD distal regulatory region contains an SRF binding CArG element required for MyoD expression in skeletal myoblasts and during muscle regeneration. *Mol Biol Cell.* **14(5):** 2151-62.

Liu J.P., Baker J., Perkins A.S., Robertson E.J., Efstratiadis A. (1993) Mice carrying null mutations of the genes encoding insulin-like growth factor-I (IGF-1) and type-1 IGF receptor (IGF1R). *Cell.* **75**: 59–72.

Li F.Q., Coonrod A., Horwitz M. (2000) Selection of a dominant negative retinoblastoma protein (RB) inhibiting satellite myoblast differentiation implies an indirect interaction between MyoD and RB. *Mol Cell Biol*. **20(14):** 5129-39.

Lluis F., Perdiguero E., Nebreda A.R., Muñoz-Cánoves P. (2006) Regulation of skeletal muscle gene expression by p38 MAP kinases. *Trends Cell Biol.* **16(1):** 36-44.

Markljung E., Jiang L., Jaffe J.D., Mikkelsen T.S., Wallerman O. et al. (2009) ZBED6, a Novel Transcription Factor Derived from a Domesticated DNA Transposon Regulates IGF2 Expression and Muscle Growth. *PLoS Biol*. **7(12):** e1000256. doi:10.1371/journal.pbio.1000256.

Mathews L. S., hammer e. R., behringer r. R., joseph d'ercole a., bell i. G. Brinster l. R. And palmiter D. R. (1988) Growth Enhancement of Transgenic Mice Expressing Human Insulin-Like Growth Factor I. *Endocrinology*. **123**: 2827-2833.

Matter K. & Balda M. S. (2003) Signalling to and from tight junctions**.** *Nature Reviews Molecular Cell Biology*. **4**: 225-237.

Meech Robyn, Makarenkova H., David E.B., Jones F. S. (2003) The Homeodomain Protein Barx2 Promotes Myogenic Differentiation and Is Regulated by Myogenic Regulatory Factors. *JBC.* **278**: 8269 –8278.

Milan D., Jeon J., Looft C., Amarger V., Robic A. Thelander M., Rogel-Gaillard C., Paul S., Iannuccelli N., Rask L., et al. (2000) A Mutation in PRKAG3 Associated with Excess Glycogen Content in Pig Skeletal Muscle. *Science.* **288:** 248 – 1251.

Mirams R.G., Byrne M.H., King R. J. (2010) A multiple timescale analysis of a mathematical model of the Wnt/β -catenin signalling pathway. *J. Math. Biol.* **60:** 131–160.

Nezer C, Moreau L, Brouwers B, Coppieters W, Detilleux J, Hanset R, Karim L, Kvasz A, Leroy P, Georges M. An imprinted QTL with major effect on muscle mass and fat deposition maps to the IGF2 locus in pigs. Nat Genet. 1999 Feb; 21(2): 155-6.

Nezer C., Collette C., Moreau L., Brouwers B., Kim J., Giuffra E., Buys N., Andersson L., and Georges M. (2003) Haplotype Sharing Refines the Location of an Imprinted Quantitative Trait Locus With Major Effect on Muscle Mass to a 250-kb Chromosome Segment Containing the Porcine *IGF2* Gene. *Genetics*. **165**: 277-285.

O'Brien S. P., Seipel K., Medley Q. G., Bronson R., Segal R., and Streuli M. (2000) Skeletal muscle deformity and neuronal disorder in Trio exchange factor-deficient mouse embryos. *PNAS.* **97:** 12074 –12078.

Powellbraxton L., Hollingshead P., Warburton C., Dowd M., Pittsmeek S., Dalton D., Gillett N., Stewart T.A. (1993). IGF-I is required for normal embryonic growth in mice. *Genes Dev.* **7**: 2609–2617.

Reshef R., Maroto M. & Lassar A. B. (1998) Regulation of dorsal somitic cell fates: BMPs and Noggin control the timing and pattern of myogenic regulator expression. *Genes Dev.***12**, 290–303.

Riazi A. M., Lee H., Hsu C., Arsdell G.V. (2005) CSX/Nkx2.5 Modulates Differentiation of Skeletal Myoblasts and Promotes Differentiation into Neuronal Cells in Vitro. *JBC.* **280:** 10716 –10720

Ridgeway A.G., Petropoulos H., Wilton S., Skerjanc I. S. (2000) Wnt SignalingRegulates the Function of MyoD and Myogenin. *JBC.*Vol. **275**: 32398 –32405.

Roth J., Shikama N., Henzen C., Desbaillets I., Lutz w., Marino S., Wittwer J., Schorle H., Gassmann M., and Eckner R. (2003) Differential role of p300 and CBP acetyltransferase during myogenesis: p300 acts upstream of *MyoD* and *Myf5*. *The EMBO Journal.* **22**: 5186 – 5196.

Sato T., Rocancourt D., Marques L., Thorsteinsdo´ ttir S., Buckingham M. (2010) A Pax3/Dmrt2/Myf5 Regulatory Cascade Functions at the Onset of Myogenesis. PLoS *Genet.* **6(4).**

Shi Y. and Massague J. (2003) Mechanisms of TGF-Signaling from Cell Membrane to the Nucleus. *Cell.* **113**: 685–700.

Strle K., Broussard S.R., McCusker R.H., Shen W.H., Johnson R.W., Freund G.G., Dantzer R., Kelley K.W. (2004) Proinflammatory cytokine impairment of insulin-like growth factor I-induced protein synthesis in skeletal muscle myoblasts requires ceramide. *Endocrinology.* **145(10):** 4592-602.

Sun J., Kamei C. N., Layne M. D., Jain M. K., Liao J. K., Lee M., Chin T. M. Regulation of Myogenic Terminal Differentiation by the Hairy-related Transcription Factor CHF2. *THE JOURNAL OF BIOLOGICAL CHEMISTRY*. **276**: 18591–18596.

Taulet N., Comunale F., Favard C., Charrasse S., Bodin S., Gauthier-Rouviere C. (2009) N-cadherin/p120 Catenin Association at Cell-Cell Contacts Occurs in Cholesterol-rich Membrane Domains and Is Required for RhoA Activation and Myogenesis. *THE JOURNAL OF BIOLOGICAL CHEMISTRY*. **284**: 23137–23145

Tomas F.M., Knowles S.E., Owens P.C., Chandler C.S., Francis G.L., Read L.C., Ballard F.J. (1992) Insulin-like growth factor-I (IGF-I) and especially IGF-I variants are anabolic in dexamethasone-treated rats. *Biochem J.* **282 (Pt 1):** 91-7.

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

Weintraub H., Davis R., Tapscort S., Thayer M., Krause M., Benezra R., Blackwell T., Turner D., Rupp R., Hollenberg S., Zhuang Y., Lassar A. (1991) The myoD Gene Family: Nodal Point During Specification of the Muscle Cell Lineage. *Science*. **251**: 761-766.

Wood A.W., Duan C., Bern H.A. (2005) Insulin-like growth factor signaling in fish. *Int Rev Cytol.* **243:** 215-85.

Yoon M. and Chen J. (2008) PLD regulates myoblast differentiation through the mTOR-IGF2 pathway. *Journal of Cell Science.* **121:** 282-289

Yoshiko Y., Hirao K. and Maeda N. (2002) Differentiation in C2C12 myoblasts depends on the expression of endogenous IGFs and not serum depletion. *Am J Physiol Cell Physiol.* **283:** 1278-1286.

Zaccagnini G., Gaetano C., Pietra L. D., Nanni S., Grasselli A. , Mangoni A., Benvenuto R., Fabrizi M., Truffa S., Germani A., Moretti F., Pontecorvi A., Sacchi A., Bacchetti S., Capogrossi M. C., and Farsetti A . (2005) Telomerase Mediates Vascular Endothelial Growth Factor- dependent Responsiveness in a Rat Model of Hind Limb Ischemia. *JBC*. **280**: 14790 –14798.

Zhu J., Li Y., Shen W., Qiao C., Ambrosio F., Lavasani M., Nozaki M., Branca M.F., Huard J. (2007) Relationships between transforming growth factor-beta1, myostatin, and decorin: implications for skeletal muscle fibrosis. *J Biol Chem.* **282(35):** 25852-63.

Zhang H. and Stavnezer E. (2009) Ski Regulates Muscle Terminal Differentiation by Transcriptional Activation of Myog in a Complex with Six1 and Eya3. *JBC.* **284**: 2867-2879.

Web Sites

- *http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=dbio&part=A3475 (NCBI, Developmental biology)*
- *http://www.systemsbiology.org*
- *http://biodata.mshri.on.ca/osprey/servlet/Index*
- *http://www.cytoscape.org/*
- *http://biologicalnetworks.net/*
- *http://www.ingenuity.com/*