Long noncoding RNAs in cardiac and skeletal muscle differentiation during mouse embryogenesis

---

A Dissertation Presented to

the Faculty of the Department of Biology and Biochemistry

University of Houston

In Partial Fulfillment
of the Requirements for the Degree Doctor of Philosophy

By

Linet George

August 2017

Long noncoding RNAs in cardiac and skeletal muscle differentiation during mouse embryogenesis

_____

Linet George

APPROVED:

_____

Dr. Robert J. Schwartz, Chairman

_____

Dr. Chin-Yo Lin

_____

Dr. Li Chen

_____

Dr. Daniel Frigo

_____

Dr. Austin J Cooney

_____

Dr. Dan Wells, Dean, College of Natural Sciences and Mathematics

Acknowledgements

I would like to express my sincere gratitude to my advisor Dr. Robert Schwartz for providing me the opportunity to undertake my thesis study under his guidance. It was indeed a privilege to work under his mentorship.

I would like to thank my committee members Dr. Chin-Yo Lin, Dr. Li Chen, Dr. Daniel Frigo and Dr. Austin Cooney for their valuable time, constructive criticism and support.

I am immensely thankful to Dr. Ashley Benham for her constant support and guidance throughout the project. Without her willingness to share her scientific expertise; this work would not have been possible.

I also thank Dr. Yu Liu, Dr. David Stewart, Dr. Dinakar Iyer, Dr. Benjamin Soibam, Dr. Jong Kim, Dr. Mathew Roberston, Dr. Wei and Dr. Steven Bark for their valuable support and insights with experiments.

I thank all other lab members and my friends for providing me a friendly and cooperative working environment.

I am grateful to my family, my pillar of strength. My parents who have taught me that there is no replacement for hard work and have always reminded me that if one dare to chase their dreams, success will follow. Thank you for being there through my ups and downs. My parents-in law who were always ready to extend their helping hand and willing to be there

when I needed that extra help. Without the understanding and constant motivation of Anoop my husband, my ambition to complete my thesis study would have just remained a dream. I am so fortunate and blessed to have you in my life. I thank my two precious gifts, Lia and Ian, for being so understanding at such a young age and for allowing me to be away from you for long hours every day.

Long noncoding RNAs in cardiac and skeletal muscle differentiation during mouse embryogenesis

---

An Abstract of a Dissertation

Presented to
the Faculty of the Department of Biology and Biochemistry

University of Houston

In Partial Fulfillment
of the Requirements for the Degree Doctor of Philosophy

By

Linet George

August 2017

Abstract:

Development is a multistep process that involves a close co-ordination of gene regulatory networks. Lineage specification is a crucial step involved in embryogenesis and understanding the significant steps involved in gene regulatory mechanisms is critical. Long non-coding RNAs, longer than 200 nucleotides have been identified as a new regulator of many molecular mechanisms involved in the development and pathological conditions.

From the genome-wide transcriptome analysis of *Mesp1*-lineage reporter mouse ESC line (UH3), we identified mesoderm specific lncRNAs, from which we selected a subset of 12 lncRNAs for functional characterization.

Of this, we identified lincRNA *Platr14* to have a positive expression *in vivo* in cardiac plate and somites of E9.5 mouse embryos. *Platr14* was seen enriched *in vitro* in the undifferentiated AB2.2 ES cell line, and the inhibition of *Platr14* showed a decrease in the beating percentage of embryoid bodies which correlated with a deregulation of the mesoderm and cardiac-specific genes.

Since *Platr14* was enriched in the region of the myotome of E9.5 Embryos, we studied its role in myogenesis. *Platr14* showed enrichment in skeletal muscle and the tongue in E15.5 embryonic mouse tissues. Knockdown of *Platr14* in C2C12 mesenchymal cell line showed a deregulation of myogenic markers as well as a reduction in the fusion rate for the myotube

formation. Though overexpression of *Platr14* showed an up-regulation of the main myogenic markers, it did not show any significant change in the myotube formation rate. Strikingly, *Platr14* overexpression in the terminal stages showed a repression in the expression of terminal myogenic markers. The whole transcriptome analysis by RNA-Sequencing of differentiating C2C12 myoblasts at day 2, showed a down-regulation of genes associated with the mesenchymal formation, somitogenesis, metabolism, cell migration, calcium transport and cell-cell signaling all related to developmental changes. *In-silico* analysis identified conservation of 3' UTR of *Platr14* across species as well as unique sites complementary to about 29 DNA binding sites of the mouse genome. Gene Ontology studies identified these DNA binding sites to have roles related to developmental processes.

In this dissertation, we identified lincRNA *Platr14*, to have a novel role in mesoderm lineage driving the cardiac and skeletal myogenic differentiation.

Table of Contents

Page

Page

Page

Page

Table of Contents for Figures

Page

Page

Table of Contents for Tables

Page

Chapter 1

Introduction

1.1 Embryogenesis:

An embryo forms as a result of fertilization of the ova or the egg and the sperm. During embryogenesis, an embryo undergoes division and differentiation. It is a highly complex process that involves a significant gene network circuitry. In mouse, embryogenesis takes only 19 to 21 days. Under the control of a set of well-coordinated gene circuitry a mouse, zygote develops to a blastocyst at day four post fertilization. A bilinear germ layer, comprising of the epiblast and the hypoblast forms by the end of embryonic stage 6 (E6.0) or day 6-post fertilization. At this point, the bilateral symmetry for the developing embryo sets in by the formation of a structure known as the primitive streak. Following this, the gastrulation process begins results in three germ layers the ectoderm, the mesodermal and the endoderm. This stage is also known as the ‘early streak' or the ES stage. During this phase, the endodermal and the mesodermal cell layer gets committed to forming their respective lineages. The ectoderm gives rise to epidermis and epidermal derivatives, most of the nervous system. The mesoderm gives rise to cardiac muscle, skeletal muscle, kidney tubules, hematopoietic derivatives and smooth muscle. The endoderm derives the hepatic cells, pancreatic cells and the inner lining of the digestive and respiratory system and the lining of the follicles of the thyroid gland.

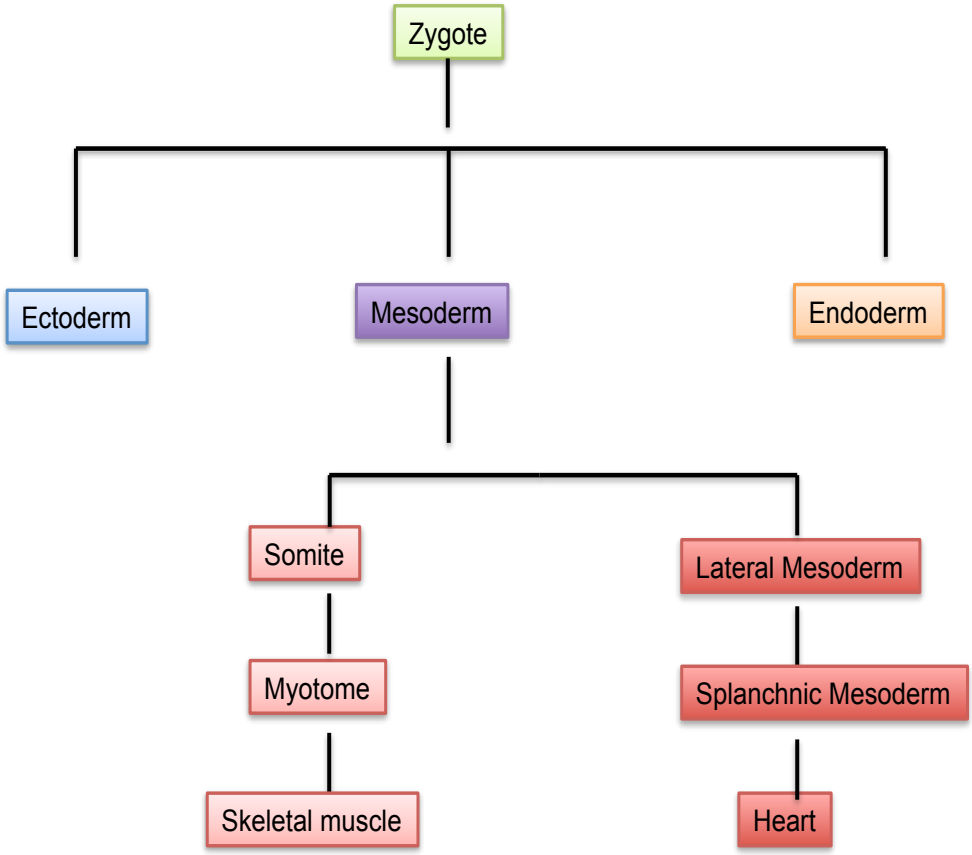


Figure 1.1: Schematic representation of the embryogenesis resulting in the three different germ layers.

1.2 Mesodermal lineage:

The mode of entry and the direction of migration of the epiblast cells into the primitive streak determine the formation of the muscle, heart, kidney and bone from the mesodermal cells. After the gastrulation process, the

undifferentiated mesodermal cells form the paraxial, intermediate and the lateral plate mesoderm.

Embryonic development requires many processes including cell division, cell migration, cell-cell signaling, cell adhesion and many other biological processes. A network of regulators controls these processes at the transcriptional level (Wagner, 2005).

During the early stages of embryogenesis, genes involved in cellular processes such as cell cycle regulation, apoptosis, DNA synthesis, transcriptional processes, protein folding gets expressed. While the factors involved in cell adhesion, energy metabolism, mitochondria, and structural processes involving organogenesis gets expressed towards the end of embryonic development. This shows a controlled expression of transcriptional regulation at different stages of embryonic development (Wagner, 2005).

During embryogenesis, animal establishes a segmentation pattern and the homeotic genes regulate these segments to differentiate into unique structures. The paraxial mesoderm is one such segment also known as the segmental plate. The paraxial mesoderm segments to form complete somites along the head to tail region (C.F Bentzinger e*t al*., 2016; Aulehla and Pourquine, 2006). This process of somitogenesis is required to differentiate further to form the myogenic precursors. The lateral

mesoderm further migrates and differentiates to form the cardiac plate, which later develops into the fully functioning heart.

1.3. Skeletal muscle development:

Skeletal muscle is a highly complex and heterogeneous tissue. Skeletal muscles constitute about one-third of our body mass and make up the largest tissue in our body. These are attached to the bones through tendons, and they are not only responsible for movement but also responsible for metabolism. Skeletal muscles can regenerate after injury with the help of the quiescent satellite cells. The satellite sales can enter the cell cycle to proliferate and myogenic differentiation (Chargé S. B. P, 2004).

Myogenesis is a process of muscle formation this involves several distinct phases (Tajbakhsh, 2009). The four stages of skeletal muscle growth or myogenesis are, the embryonic stage (E10.5–E12.5), fetal (E14.5-P0), postnatal (P)/neonatal (P0–P21), and adult (>P21) stage. (Murphy M, 2011). During embryogenesis, the undifferentiated mesenchymal cells give rise to early precursors of the mesodermal lineages. The individual somites of the embryo receive signals from the adjacent notochord, ectoderm, neural tube, and lateral plate mesoderm (LPM) triggering the differentiation process. The somites differentiate to give rise to different organ structures

between embryonic day E8.0 and E13.0 in mice (Eng D., 2013). The paraxial somitic mesoderm gives rise to the skeletal muscles of the trunk while the prechordal and non-somitic paraxial head mesoderm gives rise to the skeletal muscles of the head. The somites get categorized based on their localization such as the mesenchymal sclerotome, which forms the chondrocytes for the formation of cartilage and bone, and the dermomyotome that contains the epithelial cells. The ventral lip of the dermomyotome matures into the myotome, which contains committed muscle cells expressing high levels of MyoD, a basic helix-loop-helix transcriptional factor and Myf5 determining these cells as the terminal specification markers of muscle lineage. The appearance of these factors determines the terminal specification of the myotome cells into muscle lineage precursors (Powell *et al.,* 2002). The epaxial and the hypaxial part of the dermomyotome and the myotome form the muscles of the trunk and the limbs.

During embryogenesis mesoderm-derived structures give rise to the first muscle fibers of the body, the primary muscle fibers by E12.5 and the secondary waves of additional fibers are generated along these template fibers by E14.5 (Parker *et al*., 2003; Sambasivan and Tajbakhsh, 2007). Each myofiber assembles to form myofilaments expressing actin and myosin proteins.

1.3.1 Gene network in Myogenesis:

A broad range of signaling factors controls myogenesis during embryonic development and in postnatal life (Kuang *et al*., 2008; Bentzinger *et al.,* 2010). The highly conserved *MyoD, Myf5*, *myogenin* (MyoG), and *MRF4* genes are known as the myogenic regulatory factors (MRFs) since they are collectively expressed in the skeletal muscle lineage (Weintraub *et al.,* 1991; Rudnicki and Jaenisch, 1995). During embryonic development*, Myf5* is the first MRF to be expressed. It is up-regulated in the paraxial mesoderm and during the formation of myotome. It functions alongside the other MRFs during the formation of the myotome (Ott *et al*.,1991; Buckingham, 1992). Myogenin and MRF4 trigger the expression of myotube-specific genes, and these are directly involved in the differentiation process. *Myogenin* expression is required for terminal muscle differentiation (Gross M K, 2000) and this is followed by the expression of muscle contractile proteins.

The muscle cells or the myoblasts with remarkable migratory capacity are controlled specifically by a network of gene circuitry such as the members of the Wnt family (either through their canonical or non-canonical), Sonic hedgehog (Shh) as well as the Bone morphogenetic protein (BMP). While the former two has a positive regulation of myogenic genes, the latter has

an inhibitory role pathway. The sequence-specific transcriptional factors (SSTFs) co-ordinate the molecular mechanisms involved in the development of the myogenic precursors into fully developed functional muscle in a timely manner (Cheng, L, 2004). *Pax3* and *Pax7* cells are enriched in the somatic cells and form the precursors of limb development (Relaix, F and Duchossoy KL, 2005). Myogenesis begins in limbs from Pax3 and Lbx1 expressions to the expression of Myod or Myf5 expression and leads to Myog expression. Six1 and Six4 the sine oculis-related homeobox family members *Six4*, are required for the formation of the muscles of the limb and the back (R.Grifone,2007 and Niro, C, 2010).Mutation or absence of these transcriptional factors affects fetal myogenesis (R. Grifone, 2004 and Duchossoy KL, 2005). The four highly conserved basic helix-loop-helix transcription factors Myogenic Regulatory Factors (MRFs); the Myogenic factor 5 (Myf5), Myogenic differentiation 1 (Myod), Myogenic factor 6 (Mrf4), and Myogenin (MyoG) have the ability to transform non-muscle cells into myogenic cells.

Histone acetyl transferases (HATs) p300 and CBP co-operates with Myod and Myf5 to activate Myog and MRF4 (Chen J, 2013). Differentiation and proliferation in limb muscles are also controlled by wingless-related MMTV integration site proteins (Wnt), Fibroblast growth factors (FGFs), bone

morphogenetic proteins (BMPs), BMP inhibitors, sonic hedgehog (Shh), hepatocyte growth factor (SF/HGF). The skeletal muscle is composed of highly contractile sarcomeric units consisting organized cytoskeletal structures such as the actin and myosin proteins as well as the regulatory proteins like the troponin and tropomyosin.

Other specialized cytoskeletal filaments bundle up the myofibers to create the sarcoplasmic reticulum to store the Ca^{2+} ions, which are later, required for the transmission of nerve impulses for the muscle. Myogenesis is a stepwise process, which regulates the undifferentiated muscle cell precursors to enter the differentiation stage. This process is initiated by the exiting of the myogenic precursors from the cell cycle. There are several cyclins and CDKs, which are involved in the control of cell cycle progression. Once the differentiating myoblasts exit cell cycle they become myocytes, which they migrate and fuse to form myotubes. These myotubes later develop into the primary myofibers. Embryonic myogenesis occurs by a cascade of myogenic transcription factors. The significant ones are the paired box (PAX) family of transcription factors (Pax3/7), the basic helix-loop-helix (bHLH) myogenic regulatory factors (MRFs) such as Myf5, MyoD, myogenin, and MRF4, as well as the MEF2, SRF and SMADs (Lagha M., 2008; Bentzinger C.F, 2012). Through feed forward and

feedback mechanisms, these transcription factors can act either in synergy or antagonism. During embryonic myogenesis or muscle regeneration, activated myoblasts undergo active proliferation, controlled by several cyclins and CDKs. In proliferating myoblasts, the CDKs can prevent precocious differentiation. Cdkna1 (P21) and Cdknb1 (P27) are found to be elevated as the myoblasts enter differentiation (Zabludoff SD, 1998 and Zhang P, 1999). CyclinE/CDK2 and Cyclin *D/CDK4* inhibits differentiation and the transcriptional activity of MyoD (Ross SS, 1995). Cyclin *E/CDK2* blocks MyoD-induced gene expression (Skapex, 1995 and Peschiaroli A, 2002). Once sufficient numbers of myoblasts have formed, it exits the cell cycle to form the myocytes, which migrate and fuse with each other to undergo terminal differentiation. This is controlled mainly by the co-ordination of MyoD and MyoG. During the process of terminal muscle differentiation, the cell signaling activity of the various myogenic transcriptional factors regulates the outcome of the myogenic differentiation (Bentzinger C.F, 2012). During the final stages of myogenesis, the multinucleated myotubes will further mature to develop into myofibers with specialized structures known as the Neuromuscular Junction,(NJM). Each myofiber assembles to form myofilaments expressing actin and myosin proteins.

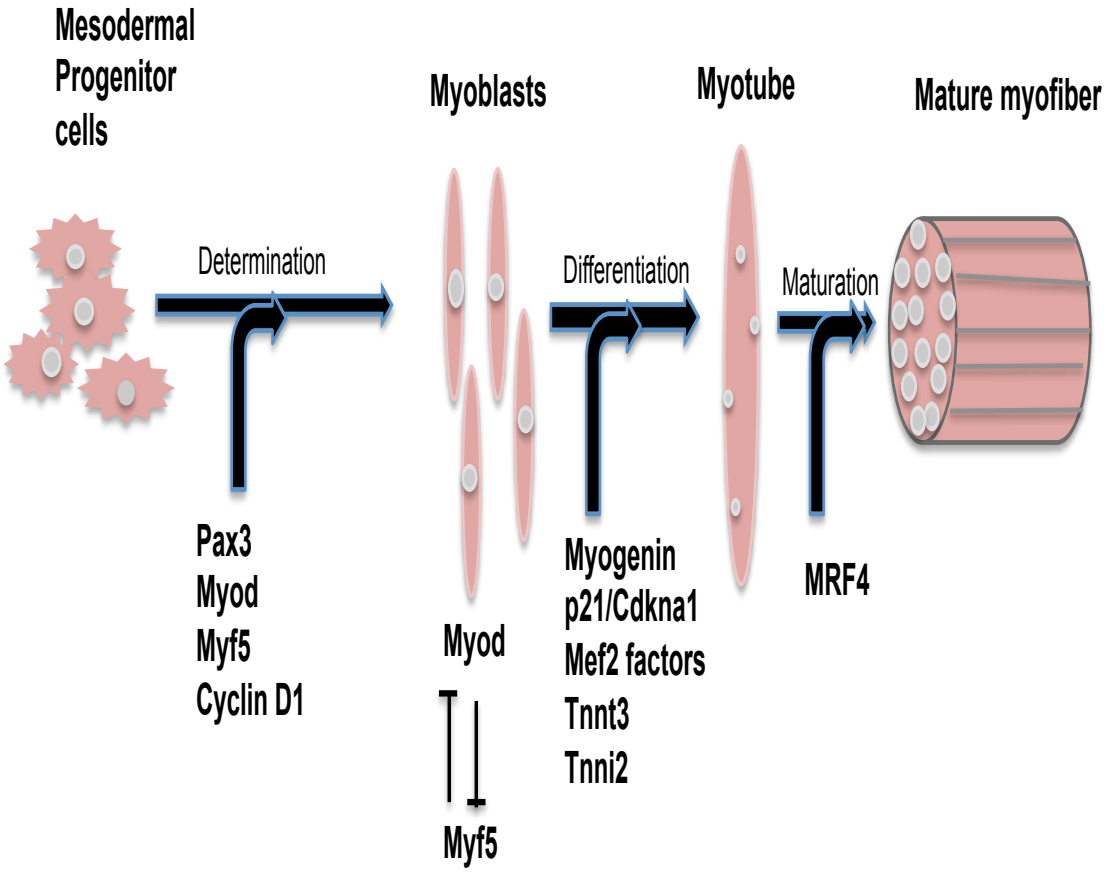


Figure 1.2: The levels of gene regulation in myogenesis.

1.4. Cardiac formation:

During the gastrulation process, the mesodermal cells that give rise to the heart becomes evident first. When the primitive streak forms, the heart precursors become located in the caudal epiblast and later lie adjacent to the cranial region of the primitive streak. This makes the cardiac progenitors the earliest cells to gastrulate (Evans *et al*., 2010). At

embryonic day E6.5 in mouse, the cardiac progenitor cells ingress while the primitive streak elongates cranially (Parameswaran M, 1995 and Lopez-Sanchez C, 2001). Late primitive streak stages in mouse localize the cardiac precursors as marked bilateral "fields" in the form of a cardiac crescent with differentiating cardiomyocytes (Parameswaran M,1995). The fundamental functional unit of the heart is the cardiomyocyte. Although cardiomyocyte cell formation is a developmental end point, a diversity of cardiomyocyte subtypes exists within the heart. There are a variety of cardiomyocyte subtypes that forms the outflow tract; atria right ventricle, left ventricle, caudal great veins, and specialized conduction system tissue, including the sinoatrial node, atrioventricular (AV) node, and His-Purkinje tracts.

1.4.1 Gene network in Cardiac formation:

Positive and negative signaling triggers the underlying endoderm to induce the progenitors to specify the cardiac lineage. Signaling cues from adjacent tissues induce cardiac differentiation in the cardiac precursors. Tissue-specific gene transcription is essential for the precise differentiation and patterning in cardiac formation. These cardiac transcription factors form transcriptional gradients regulating the different stages of heart

development such as terminal differentiation of cardiac cells, the formation of cardiac chambers and formation of patterning boundaries.

In early mesoderm formation, graded levels of the TGFB-family member, Nodal is essential for specifying different types of mesoderm (Brennan et al. 2001). Wnt required for the dorsal vessel formation, is one of the earliest markers of cardiac specification in mesodermal progenitors (Zaffran S, 2002).

As cells enter through the cranial primitive streak, they are exposed to signaling factors such as retinoic acid and FGFs. Bone Morphogenic protein (BMP) and Wnt signals control the early stages of cardiac differentiation. Though Wnt signaling initially promotes cardiogenesis, later it acts as an inhibitor as cardiac progenitors differentiate into the different cardiac derivatives (Naito *et al.,* 2006; Kwon *et al*., 2007). T-box transcription factor Eomes, (Eomesodermin) is the earliest marker expressed in committed cardiac precursors (Costello *et al*., 2011). Eomes activate Mesp1, another transcription factor whose expression was considered to be one of the earliest markers of cardiac differentiation and is expressed on the onset of gastrulation E6.5 to E7.5. But Mesp1 is also involved in various mesoderm-derived lineages. While Mesp1 is a downstream target of Fgfr1, Mesp1 is the upstream regulator of many cardiac specification transcription factors, such as Nkx2.5, Islet1, and

myocardin, which seems to be expressed at the crescent stage of cardiac formation.

After gastrulation, many molecular markers begin to express determining the location of the heart fields in the lateral plate mesoderm. During mouse embryogenesis, the expression of Nkx2.5 in the cardiac crescent is dependent on BMP2 signaling FGFs produced can induce expression of Nkx2.5 and ventricular myosin heavy chain (vMHC). But in the second heart field, Nkx2.5 down-regulates BMP2 signaling by direct interaction with Smad. This helps to control the number of cardiomyocytes that would differentiate to form the heart tube (Prall *et al.*, 2007). Also Shh hedge- hog receptor smoothened is also required for the early phase of Nkx2.5 expression in but not in later stages of heart formation. LIM homeodomain transcription factor Islet1 is an indicator of proliferating undifferentiated cardiogenic progenitors. This is required for proliferation, survival of second heart field progenitors and their migration into the heart. The canonical Wnt and Fgf8 signaling are also required for the proliferation of early cardiac progenitors and these two dictates Islet1 expression. Wnt/β-catenin-induced expansion of cardiac precursors requires down-regulation of Islet 1.

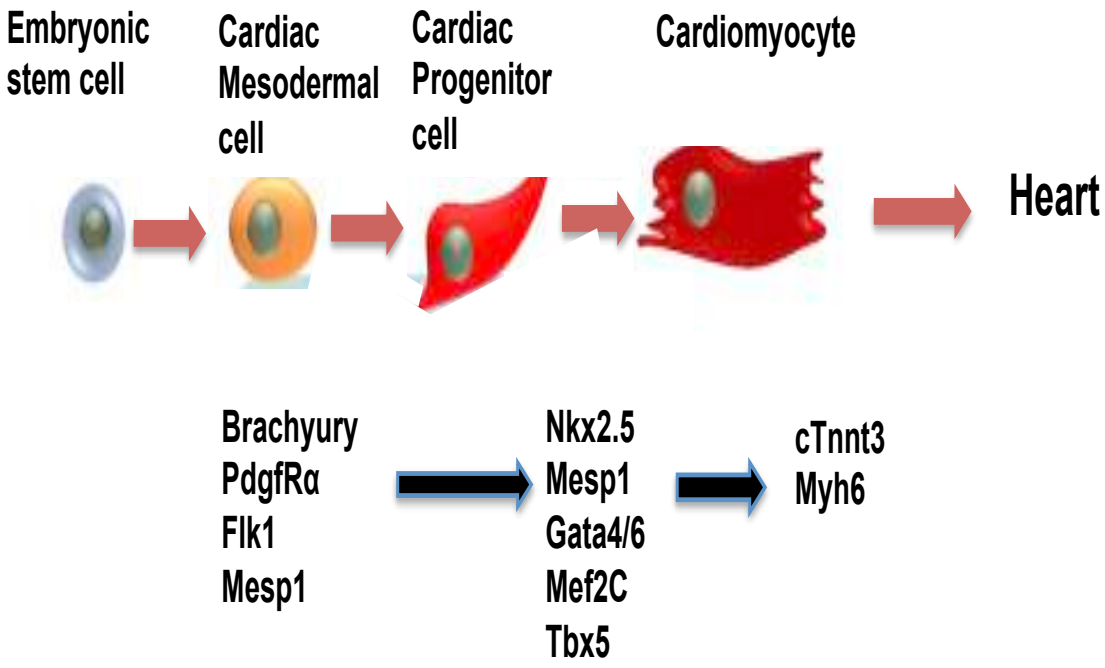


Figure 1.3: Schematic representation of the steps in cardiac morphogenesis.

Wnt/β-catenin-induced expansion of cardiac precursors requires down-regulation of Islet 1 (Isl1), another transcription factor that promotes cardiac differentiation (Kwon *et al*., 2009). BMPs help to regulate cardiac precursor development and differentiation (Prall *et al*., 2007). Transcription factor Mef2*c* is expressed at E7.5 in the pro cardiogenic mesodermal cells of the mouse embryo before the formation of the linear heart tube. Another factor is T-box transcription factor Tbx2, Tbx5, Tbx20 that is required for the growth and patterning of early cardiac development (Bruneau *et al*.,

2001). Mutations in the T-box transcription factors causes septal defects such as Arterial Septal Defects (ASD), Ventricular Septal Defects (VSDs) and conduction system diseases suggesting its significant role in early patterning in the heart. The onset of cardiac differentiation is regulated by Gata4 and Gata6 factors (Zhao *et al.,* 2008), where Gata4 is required for the cardiac fusion, and the serum response factors (SRF) regulate the expression of many cardiac genes (Niu *et al*., 2008). The basic helix loop helix, the Hand proteins control aspects of chamber differentiation (Firulli *et al*., 1998). These transcription factors interact to regulate downstream players of cardiogenesis. Chromatin localization studies have shown that transcriptional factors such as Nkx2-5, Gata4, Tbx5, and Mef2a bind together at genes that are co-regulated during heart development, thereby exhibiting a high degree of co-operation between cardiac transcription factors (He *et al*., 2011).

Along with these developmental genes, many noncoding RNAs also are expressed at different stages of differentiation. The noncoding RNA functions as regulators of the transcriptional factors involved in developmental processes. The most recent finding has shown the role of non-coding RNA, Long Non-Coding RNA, (lncRNA) in lineage specification in embryogenesis.

1.5 Long Non-Coding RNAs:

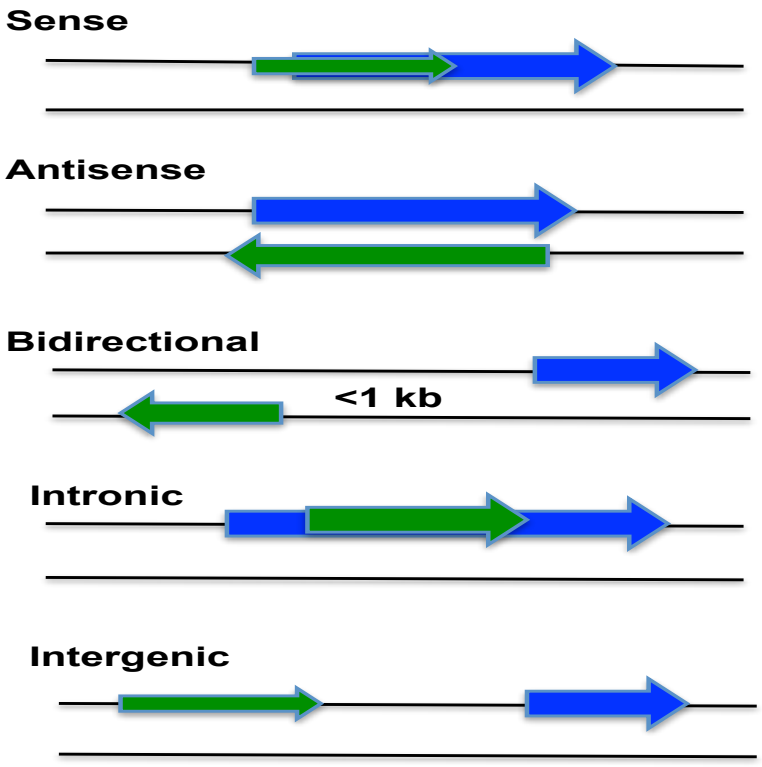
Even though transcription corresponds to more than ~80% of the genome, the human genome is composed of less than 2% of protein-coding genes (Lander E.S, 2001). Lee *et al*., in 1993 first identified small non-coding RNAs, and since then they have been studied extensively to understand their molecular mechanism of action and function. Currently, they represent the most prominent vertebral gene groups. In human more than ~100,000 genomic loci have been presumed to generate lncRNAs in the human genome (Zhao *et al.,* 2016) thereby making them the largest class of genes.

Many noncoding RNAs have been recently discovered. Non-coding RNAs greater that 200 nucleotides are classified as 'Long Non-Coding RNAs'. Like the mRNAs, the lncRNAs are RNA Polymerase II transcripts with a poly-A tail and 5' cap. Unlike the protein coding genes, these minimally expressed and long non-coding RNAs are not evolutionarily conserved. They also can assist in protein transport and trafficking and therefore are involved in various biological processes such as cellular transport and metabolism, cell differentiation, organ or tissue development (Kapoor BD, 2012).

1.5.1 Classification of Long non-coding RNAs:

1.5.1.1 Classification of Long non-coding RNAs based on genomic locations:

LncRNAs can be classified into several types based on their genomic locations such as the (1) intronic lncRNAs; (2) intergenic lncRNAs (lincRNAs); (3) sense-overlapping lncRNAs, and (4) antisense lncRNAs (Batista P.J, 2013). Intergenic lncRNAs are located between annotated protein-coding genes and are generally in proximity to neighboring protein-coding genes. Both the sense and antisense lncRNAs differ from each other by its transcriptional directions. They also partially overlap with exon sequences of annotated protein-coding genes, and only an intronic lncRNA transcribes in either the sense or antisense direction and overlaps the intronic region of a coding gene (Batista P.J, 2013).



Purple= Protein coding genes

Green = Long non coding RNA

Figure 1.4: Classification of long non-coding RNAs based on their genomic locations.

1.5.1.2 Classification of Long non-coding RNAs based on functional roles:

LncRNAs can also be categorized based on their functional roles. Nucleus enriched lncRNAs can function at the transcriptional level, either through

co-transcriptional interactions with transcriptional complexes or by the recruitment of chromatin modification enzymes to transcription sites in *cis* or *trans*. The *cis* nature of a lncRNA gives it the ability to act on a neighboring gene proximal to the site of transcription (Feng J *et al*., 2006; Martianov I. *et al.,* 2007; Maamar H *et al.,* 2013; Ng SY *et al.,* 2013).Many lncRNAs can function in *trans* to target gene loci distant from where the lncRNAs gets transcribed (Wang K C, 2011; Cabianca DS, 2012). These can activate or repress transcription through chromatin remodelers and modifiers, thus changing the chromatin architecture at the specific loci or the entire genome.

LncRNAs can also determine sub-nuclear architecture by regulating chromosome looping thus promoting or disrupting chromosomal interactions. Additionally, nuclear lncRNAs can act as scaffolds for the formation of nuclear compartments such as speckles, paraspeckles, and Polycomb bodies (Tripathi V, 2010; Mao YS, 2011; Yang L, 2011).

Meanwhile, cytoplasmic lncRNAs can regulate the translational rate. They can also activate or reduce mRNA decay. It has been suggested that LncRNAs may actively involve in the Staufen 1- and Staufen 2- (STAU1- and STAU2-) mediated mRNA decay (SMD). About one-third of lncRNAs are reported to contain at least one short interspersed element (SINE)

sequence (C Gong *et al*., 2011). SINE within the 3’ UTR of a protein-coding RNA can interact through intermolecular base pairing with a complementary SINE of one or more lncRNAs, resulting in double-stranded RNA (dsRNA). This can be recognized by Staufen 1 and Staufen 2 and degraded through SMD. Several mRNAs encode proteins, that play important roles during developmental processes have SINE sequences in their 3 UTRs.

Competing endogenous RNAs (ceRNAs) are another class of lncRNAs that can act as ‘molecular sponges' to regulate microRNAs and its mRNA targets (Mackowiak SD, 2013).

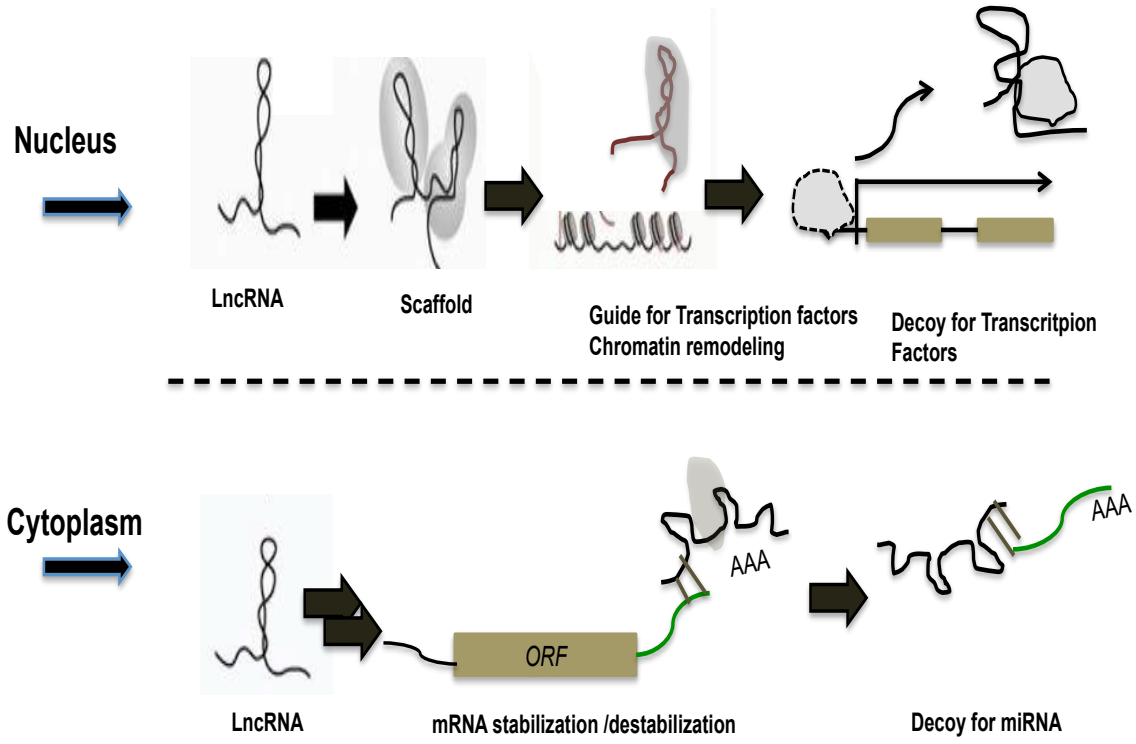


Figure 1.5: Mechanisms for long noncoding RNA (lncRNA) function.

1.5.2 Functional role of long non-coding RNAs:

The study of long noncoding RNAs originates from the interest in the non-protein-coding part of the genome and its function and evolutionary conservation dating back to the 1950s.The C-value paradox was identified by the scientists that brought a better understanding of having no correlation with the size of the organism or the developmental complexity and the DNA content of a cell. Ohno *et al.,* in 1972 identified that most of the organismal genome do not code for the entire genome. This noncoding

part of the genome comprised of the repeats, pseudogenes, transposons, and other elements with an unknown function constitutes about 50%-70% of the mammalian genome (de Koning, 2011). Following the discovery of heterogeneous RNAs with no protein-coding function originating from the repetitive and heterochromatic region (Yunis and Tsai, 1978), snRNAs and snoRNAs were identified. These are found to be involved in posttranscriptional regulation. As more advanced technologies emerged, such as the microarrays and next generation sequencing, more non-coding RNAs were identified. This enabled to better understand that about 70-90% of the genome is transcribed at specific time points during development, homeostasis or disease (Bertone *et al*., 2004; Mercer and Mattick, 2013). Some of the transcripts generated at his time maintain low copy numbers and lncRNAs can be classified under this class. Studies such as RNA-sequencing analysis and chromatin signatures studies such as histone modifications, TF binding, *DNase* I hypersensitivity assays have helped to show regulatory regions in these loci indicating that lncRNAs can be considered as novel transcripts with some predicted function (Guttman *et al*., 2009; Cabili *et al*., 2011; Iyer *et al*., 2015). Therefore noncoding RNAs can be considered a new layer of regulatory network in gene regulation.

An attempt to decipher the sequence-functional role of lncRNAs has led to the identification of active domains in these transcripts. LncRNAs have an abundance of transposable element (TE) sequences. Repetitive mobile sequence elements vary in origin and evolutionary age and comprise of about two-third of our entire genome (Lander *et al*., 2001; de Koning *et al*., 2011). Though these were earlier considered as parasitic elements these are now recognized for their significant roles in the cellular process and conferring genetic variability across genomes (Cordaux and Batzer, 2009). TEs that are inserted into the lncRNA exons can cause preformed structural and sequence features enabling lncRNA interaction with other molecules. TEs are hypothesized to be a driver in the generation of new lncRNA genes. Previously inactive genomic regions will be randomly inserted with novel TE promoter fragments, thereby driving the transcription of lncRNA transcripts, which will eventually acquire specific functions (Johnson, 2015). The presence of TEs can explain the potential for the evolution of lncRNA regulatory networks (Johnson, R *et al*., 2014).These functional domains can work in two ways: a) Formation of a specific secondary structure that enables to interact with a protein partner (Blackwell *et al*., 2012); and (b) sequence-based hybridization to another nucleic acid (Gong and Maquat, 2011). TEs have been identified to interact with wide range of self-coded and host-coded proteins (Goodier *et al*.,

2013). Recently identified protein classes seen to interact with RNPs, such as the Alu, LINES, SINES have seen to have relevance to the functional roles of few identified LncRNAs (Blackwell *et al*. 2012; Goodier *et al*. 2013).

TEs have the ability to interact specifically to the other multiple copies of the same repeat elements that exist, throughout the genome. Such interactions may occur in lncRNA inserted with TEs with DNA or RNA sequences through Watson-Crick base-pairing (Gong and Maquat, 2011). LncRNA embedded with TEs can be involved in sequence-specific RNA complementarity. lncRNAs have been identified that function in the similar manner where this targets mRNAs for Staufen-mediated decay through Alu-mediated complementary base-pairing (Gong and Maquat 2011).At the same time, through either Hoogsteen base pairing or the conventional Watson-Crick base-pairing it is predicted for LncRNAs to interact directly with genomic DNA sequence (Buske *et al*., 2012). Therefore, lncRNA regulatory function can be based on the number of these TEs embedded in its sequences. Such as in lncRNA, HOTAIR multiple protein binding sites are involved in the formation of protein complexes (Tsai *et al*., 2010).

This combination of RNA-binding domain with protein binding could assist in mRNA processing. Johnson *et al*., 2015 proposed a hypothetical RNA–

DNA adaptor configuration that would enable the recruitment of mRNAs or other ncRNAs) to specific genomic locations for mRNA processing. Thus lncRNA can be considered to have a role in transcriptional regulation by recruiting epigenetic complexes (Grote *et al.* 2013).

1.5.3 The evolution of complex long non-coding RNA regulatory networks:

lncRNA have played a significant role in the evolution of developmental gene regulatory networks (Pollard *et al.*, 2006; Mattick, 2009). Recent evidence supports a diverse role for lncRNA as regulatory key players of processes known to have evolutionally modified between mammals, such as the pluripotency of stem cells (Guttman *et al*., 2011), neurodevelopment (Ng *et al*. 2012), as well as immune function (Carpenter *et al*., 2013). The functional evolutions of the lncRNAs have not yet been well understood though recent evidence has addressed its evolution (Necsulea *et al*., 2014; Washietl *et al*., 2014). But it is believed that transposable elements would have likely played a significant step in both the processes.

1.6 Long non-coding RNA interactions:

Many lncRNAs likely interact with proteins, other RNAs, and DNA regions to perform their functions. The ncRNAs such as snoRNAs can form ribonucleoprotein complexes with proteins (S.R. Eddy, 2001). Therefore it is more than likely that lncRNAs might form similar complexes. Similarly, lncRNAs interacts with DNA regions along with proteins to confer epigenetic regulation. Recent advancements have led to different methodologies to unravel these lcnRNA interactions. Few of which used in this study are mentioned here.

1.6.1 *In-vitro* RNA pull-down assay:

Proteins that interact with a specific lncRNA of interest can be identified and characterized by an in-vitro RNA pull-down assay. LncRNA probes are synthesized and biotin labeled. The lncRNA probe is incubated with lysate prepared from an in-vitro sample or recombinant protein to form a specific lncRNA–protein complex. The protein complex is pulled down with streptavidin agarose or magnetic beads. Finally, the retrieved protein is identified by Western blot or mass spectrometry (MS) (Feng Y *et al*., 2016).

1.6.2.RNA antisense purification (RAP) RAP-MS:

RNA antisense purification (RAP) can be used to identify RNA interaction

with the genomic regions of chromatin (J.M. Engreitz, 2013). RAP uses overlapping long antisense RNA probes approximately 120 nucleotides long that are tiled across the entire lncRNA transcript. After the genomic DNA–RNA is isolated, to reduce the genomic DNA fragments to 300 bp or fewer; the complexes are partially digested with DNaseI. The products obtained are subjected to Mass Spectroscopy analysis to identify further proteins associated with lncRNAs *in-vivo*.

1.6.3.Chromatin Isolation by RNA Purification (ChIRP):

The ChIRP method uses biotinylated oligonucleotides complementary to the lncRNA of interest. The probes designed as ‘odd’ and ‘even ‘across the lncRNA sequences are used to pull down lncRNA-associated proteins and chromatin DNA. Once the cultured cells are cross-linked, the chromatin is extracted and sonicated. The tilling oligonucleotides that tile the entire lncRNA sequence are added and hybridized. The hybrids including target lncRNA, proteins, and chromatin DNA were eluted with magnetic streptavidin beads and subjected to qRT-PCR or deep sequencing for DNA analysis, or to Western blotting or MS for protein analysis (Chu *et al*., 2012). Also, the recently developed ChIRP-seq method allows global and high-throughput discovery of genomic DNA associated with lncRNA.

1.7 Long non-coding RNA in development:

Recent findings have shown that lncRNAs can be key players of cellular differentiation, cell lineage choice, organogenesis and tissue homeostasis. Many lncRNAs have been reported to be involved in the ectoderm, mesoderm and endoderm development and any deregulation in its expression results in associated developmental disorders (Schmitz, S.U, 2015).

1.7.1.Long non-coding RNA in Cardiac Differentiation:

Like microRNAs, lncRNA discovery in cardiac development is showing to be promising in therapeutics. Recent findings reveal very significant roles for lncRNAs in fine-tuning cardiac development along with other multiple modes of gene regulation (Grote *et al*., 2013; Ounzain *et al.,* 2014 and 2015). Long noncoding RNAs (lncRNA) are differentially expressed at the fetal and adult stages of cardiac development. LncRNAs are found to regulate gene expression through mRNA stabilization, mRNA decay, or microRNA (miRNA) sponge during the transition from fetal to adult heart. lncRNA expression can coordinate mRNA through epigenetic changes (H3K4me1 as enhancers and H3K4me3 as promoters). In the fetal heart, lncRNAs are associated with genes involved in development and programming, whereas in the adult heart lncRNAs are associated more

with genes involved in disease and dysfunction. Deregulated lncRNAs have been linked to cardiovascular diseases (Yap *et al*., 2010; Michalik *et al*., 2014; Ounzain *et al*., 2014; Wang *et al*., 2014; Yan *et al*., 2015), including coronary artery and other heart dysfunctions.

Table 1: Long non-coding RNAs identified to have cardiogenic role.

Name	Species	Genomic context	Function and mechanism	References
ALIEN	Human	Intergenic	Cardiovascular commitment	Kurian et al., 2015
ANRIL	Human	Antisense	Metabolism, coronary artery disease, myocardial infarction	Yap et al., 2010; Vausort et al., 2014
Bvht	Mouse	Intergenic	Cardiac mesoderm commitment Decoy	Klattenhoff et al., 2013
Carl	Mouse	Intergenic	Mitochondria, cardiomyocyte apoptosis; miRNA sponge	Wang et al., 2014b
CARMEN	Mouse/Human	Intergenic	Cardiac specification & homeostasis; enhancer-associated, cis- & trans-regulation	Ounzain et al., 2015a
CHAST	Mouse/Human	Antisense	Pro-hypertrophic; cis-regulation	Viereck et al., 2016
Chrf	Mouse	Intergenic	Cardiac hypertrophy; miRNA sponge	Wang et al., 2014a
Fendrr	Mouse/Human	Intergenic	Cardiac development; cis- & trans-regulation, Guide	Grote et al., 2013
H19	Mouse/Human	Antisense	Cardiac fibroblast proliferation, negative regulator of hypertrophy	Tao et al., 2016
HIF1A-AS2	Human	Antisense	Myocardial infarction	Vausort et al., 2014
KCNQ1OT1	Human	Antisense	Cardiovascular development, arrhythmia, myocardial infarction; Guide	Thakur et al., 2004; Korostowski et al., 2012; Vausort et al., 2014
LIPCAR	Human	Unknown	Heart failure	Kumarswamy et al., 2014
Malat1	Mouse/Human	Intergenic	Endothelial cell identity, myocardial infarction; Decoy	Michalik et al., 2014; Vausort et al., 2014
MEG3	Mouse/Human	Intergenic	Modulation of TGF-β pathway; Guide	Mondal et al., 2015
Mhrt	Mouse/Human	Intergenic	Cardiac hypertrophy; Decoy	Han et al., 2014
MIAT	Human	Intergenic	Myocardial infarction; miRNA sponge	Yan et al., 2015
Mm67/77/85/130/132	Mouse	Intergenic	Cardiac development & remodeling; Enhancer-associated lncRNA Cis-regulation	Ounzain et al., 2014
Novlncs	Mouse	Intergenic	Cardiac remodeling; Enhancer-associated	Ounzain et al., 2015b
NRF	Mouse	Intergenic	Regulated necrosis of cardiomyocytes; miRNA sponge	Wang et al., 2016b
PANCR	Human	Intergenic	Cardiac differentiation	Gore-Panter et al., 2016
PUNISHER	Human	Antisense	Endothelial cell identity	Kurian et al., 2015
PVT1	Mouse	Antisense	Cardiomyocyte cell size, possibly regulating hypertrophy	Yu et al., 2015
RNCR3	Mouse/Human	Intergenic	Atheroprotective; miRNA sponge	Shan et al., 2016
ROR	Human	Intergenic	Pro-Hypertrophic; miRNA sponge	Jiang et al., 2016
SENCR	Human	Antisense	Smooth muscle contractility	Bell et al., 2014
Smad7-lncRNA	Mouse	Antisense	Enhancer-associated lncRNA, Cis-regulation	Ounzain et al., 2014
SMILR	Human	Intergenic	Vascular smooth muscle cell proliferation	Ballantyne et al., 2016
TERMINATOR	Human	Intergenic	Pluripotency, cardiovascular development	Kurian et al., 2015
uc.167	Mouse	Antisense	Involved in cardiac differentiation	Song et al., 2016
UCA1	Human		Biomarker for acute myocardial infarction, anti-apoptotic (rat model)	Liu et al., 2015; Yan et al., 2016

1.7.2.Long non-coding RNA in myogenesis:

LncRNAs have been identified as important novel regulators of skeletal muscle biology and diseases. LncRNAs have been identified as enhancer-associated lncRNAs, acting in *cis* or trans, regulate specific transcription factors to modulate gene expression as well as act as molecular decoys during the proliferation and differentiation stages of myogenesis.

Table 2: Long non-coding RNAs identified to have myogenic role.

LncRNA	Site of action	Function	Effector molecule	Regulation during muscle differentiation	References
eRNAs (CE and DRRRNAs)	Nucleus	Transcriptional activation	MyoD	up	Mousavi K *et al* 2013
H19	Nucleus and cytoplasm	Epigenetic repression, miRNAs sponge	PRC2, let-7 miRNAs	up	Kallen AN *et al*,2013
Linc-MD1	Cytoplasm	miRNAs sponge	HuR	up	Cesana M,2011
Malat1	Nucleus	Epigenetic repression, pre-mRNA splicing	Cbx4 and SR family of splicing f	up	Watts R,2013
Neat1	Nucleus	Structural integrity of nuclear paraspeckles	Unknown	up	Sunwoo H *et al*,2009
Nctc1	Nucleus	Unknown	Various RNA-binding proteins	up	Eun B *et al* ,2013
SRA	Nucleus	Scaffold factor	MyoD	up	Caretti G *et al*,2006
SINE containing lncRNAs	Cytoplasm	mRNA decay	STAU1 and STAU2	up	Wang J *et al*,2013
Yams	Nucleus	Transcriptional activation	Unknown	up/down	Lu L *et al*,2013
LncMyoD	Nucleus	Transcriptional activation	Myod	up	Gong, C. G. *et al,2015*

Given that long non-coding RNAs have regulatory role in many aspects, we identified few lncRNAs that are uniquely expressed in mesoderm lineage. Understanding their functional role will help us to discover new regulatory mechanisms in lineage specification which will further help us to translate this knowledge to clinical applications.

Chapter 2

Materials and Methods

2.1 Cell Lines and Animal tissues:

2.1.1 ES cell lines: UH3 and AB2.2 cell lines (generous gift from Dr.Yu Liu, generated by Dr.Austin Cooney) were maintained in ES media constituting of the mLIF (ESGRO® Leukemia Inhibitory Factor, Sigma-Aldrich, St. Louis, MO) (1 µg/µL) in Gibco Knockout- Dulbecco's Modified enriched media (KO-DMEM™, Thermo Fisher Scientific, Waltham, MA) with 10% ES grade Fetal Bovine Serum (FBS; Atlanta Biologicals Inc. Flowery Branch, GA),1% antibiotic/antimitotic (Thermo Fisher Scientific, Waltham, MA) at 37 °C and 5% CO_2 in undifferentiated state. The media was replaced every 24 hours, and the cells were passaged every 3 days to maintain the undifferentiated state.

2.1.2 C2C12 cell line: C2C12 were cultured in undifferentiated state in regular growth media containing High Glucose Dulbecco's Modified enriched media (DMEM, Thermo Fisher Scientific, Waltham, MA) with 10% ES grade Fetal Bovine Serum (FBS; Atlanta Biologicals, Inc., Flowery Branch, GA) 1% antibiotic/antimitotic (Thermo Fisher Scientific, Waltham, MA) at 37 °C and 5% CO_2. The cells were maintained at 60-70% confluency to avoid differentiation. For the myogenic differentiation the cells were allowed to reach ≈ 90% confluency and allowed to differentiate in differentiation media (DMEM with 5% Horse Serum (Life Technologies,

Invitrogen), 1% antibiotic- antimycotic (Thermo Fisher Scientific, Waltham, MA) supplemented with Insulin (Insulin, Human Recombinant, Gibco, Grand Island, NY) at a concentration of 4 µg/µL. The differentiation media was replaced every 24 hours.

2.1.3 Animal tissues: Animal tissues were procured from wild-type mouse strains. The embryonic tissues were isolated from E15.5 and E17.5 embryos. Postnatal hearts were isolated from postnatal days 3,7, and 10 old mice cubs. Adult heart tissues were obtained from 6-week-old mice (n=3) in each category.

2.2 Embryoid Bodies: Embryoid bodies were generated with ES cells, UH3, and AB2.2 cell lines. 400 cells/20 µL of the ES media without mLIF were dropped on culture plate lids and allowed to differentiate as three-dimensional aggregates. The Embryoid bodies were maintained in hanging drop and on day 4 of differentiation they were collected and plated onto gelatinized culture plates and maintained for the required number of days and replenished with fresh media every 3 days.

2.3 RNA Isolation: The RNA Isolation was performed using the Qiagen RNA isolation kit (Qiagen, Netherlands). All the samples were treated with

DNase I and quantified by using the Nano drop spectrophotomer. For the total RNA isolation from E15.5, E17.5, postnatal day 3, 7, and 10, and 6week-old mouse heart tissues were homogenized by sonication (Bioruptor 300 Plus) in RNA extraction tubes (Diagenode), followed by RNA isolation by Trizol and isopropanol precipitation. This was followed by DNase I treatment before final RNA cleanup by ethanol precipitation.

2.4 cDNA Preparation: cDNA was prepared using the qscript cDNA Synthesis reagent kit (Quanta Biosciences, Gaithersburg, MD). The recipe was prepared on ice. qScript cDNA SuperMix (5X) was mixed with 1000 ng of RNA and made up to a final volume of 20 µL with RNase/DNase-free water. The reaction mix was centrifuged briefly and incubated with at 5 minutes at 25 °C, 30 minutes at 42 °C, 5 minutes at 85 °C incubation and held on at 4 °C. The total volume of cDNA prepared was diluted with 180 µL RNase/DNase-free water to obtain a final concentration of 5 ng/µL and stored at -20 °C.

2.5 Cloning of the shRNAs: The nucleotide sequences were obtained from the UCSC genome browser. Using the IDT website, the 19 nucleotide double-stranded oligonucleotides were designed. Five best shRNA hit were obtained. Three best shRNA predicted sites spanning across the whole lncRNA sequence were selected and further designed as double-

stranded oligonucleotides to be cloned into the pLL3.7 vector. The oligonucleotides were re-suspended in dd water at a concentration of 100uM.The oligonucleotides were annealed using the following reaction. For a total reaction volume of 18 µL; ddwater: 6.0 µL, Oligos-Sense: 1.0 µL, Oligos-Antisense: 1.0 µL, 2X Quick Ligation Buffer: 10.0 µL were added and heated to 94 °C. The reaction was immediately removed and incubated at room temperature for about 2 hours.

The inserts were phosphorylated as follows: To the above reaction volume of 20 µL, 1.0 µL of 10 mM ATP and 1.0 µL of T4 Kinase were added and incubated. The total reaction volume of 20 µL was incubated at 37°C for one hour and 65 °C for 10 minutes. Following phosphorylation, the reaction was centrifuged at 1000 rpm for 3 minutes. 1 µL of the total reaction was mixed with 19.0 µL of dd water. The reaction was incubated on ice.

2.5.1 pLL3.7 Vector preparation: The vector was prepared with dd water: 13.5 µL, pLL3.7 vector (85.0ng/µg): 3.5 µL, Antarctic phosphatase buffer 10X: 2.0 µL, Antarctic phosphatase: 1.0 µL. The reaction was mixed and incubated at 37 °C for 15 minutes followed by inactivation at 65 °C for 10 minutes.

2.5.2 Ligation reaction: Ligation reaction was performed fresh on the day of transformation. The reaction was as follows prepared with dd water: 6.0 µL, Diluted shRNA Oligos: 1.0 µL, 2x Quick Ligation buffer: 10.0 µL, pLL3.7 vector (from the above reaction): 2.0 µL and mixed thoroughly before adding ligase. Followed by 1.0 µL of Quick T4 Ligase. The reaction was incubated at room temperature for 15 minutes.

2.5.3 Transformation: Transformation was performed immediately into competent Stbl competent cells (One Shot™ Stbl3™ Chemically Competent E. coli, Thermo Fischer Scientific, Waltham, MA). The reaction was incubated on ice for a minimum of 30 minutes. The reaction volume added was always 10% of the cell number. The reaction was heat shocked at 42 °C for 45 seconds. The reaction was immediately removed and incubated on ice for 2 minutes. 250 µL of the S.O.C medium (2% tryptone, 0.5% yeast extract, 10 mM sodium chloride, 2.5 mM KCl, 10 mM MgCl, 10 mM magnesium sulfate, 20 mM glucose) (Invitrogen, Thermo Fischer Scientific, Waltham, MA) warmed to room temperature, was added to 50 µL of Stbl competent E.coli cells and incubated at 30°C for one hour in a shaker. After incubation, the entire volume of the culture was plated onto the pre-warmed LB plates with 100 µg/mL of Carbenicillin (Thermo Fischer Scientific, Waltham, MA) and incubated at 37 °C overnight. Well-rounded

colonies were selected and grown in 5 mL LB (Luria-Bertani) with 50 µg/mL Carbenicillin at 30 °C at about 250 rpm overnight.

2.5.4 DNA isolation: The DNA isolation was performed using the DNA isolation kit QiaPrep spin Miniprep kit (Qiagen, Netherlands) following the manufacturer's protocol and stored at -20 °C.

2.5.5 Restriction Digestion: For the validation of the inserts, the isolated plasmids were digested with XbaI (20 units/µL) (New England BioLabs Inc.UK) and Not I (40 units/µL) (Roche Life Science) at 37 °C for 2 hours. The digestion was run on 1.5% agarose gel using 1kb DNA as the ladder. The run was visualized under UV light to check for the inserts. The clones with the expected inserts that showed bands at 7200 bp and 550 bp were selected and stored at -20 °C. These clones were sequenced to confirm the orientation of the inserts. The results were analyzed using Serial Cloner 2.0, for the selection of clones.

2.6 Transfection:

2.6.1 Lentiviral transfection: To generate the lentiviral particles, pMDG1, pSAX2 and the expression vector pLL3.7 harboring the selected shRNA were introduced into 293T cells in OptiMem, using FuGENE® HD Transfection Reagent according to the manufacturer's instructions.

18 hours post transfection; regular media replaced the OptiMeM. 24 hours later the supernatant with viral particles were collected, filtered and the titer value was checked using Lenti-X™ GoStix™ (Clontech, Mountain View, CA), and the viral particles were stored at -80 °C.

2.6.2 Dsi transfection: Three predicted Dsi RNAs were designed and selected using the IDT website for targeting the entire nucleotide. The Dsi transfection was performed using Lipofectamine® RNAiMAX Transfection Reagent (Thermo Fisher Scientific, Waltham, MA), following manufacturer's protocol with 1nM of three Dsi RNA. The C2C12 cell lines were double transfected to make sure the required knockdown is maintained for the specific time points during the study.

2.7 Transduction: For both C2C12 and AB2.2 cell lines the protocol followed is as follows: To 1 mL of the cells suspension at a concentration of 3 x10^5 cells, 1 mL of filtered viral supernatant and 2 µL of 10 mg/mL of Polybrene®, (Sigma-Aldrich, Corp., USA) were mixed gently and plated onto a 6 well tissue culture plate. The cells were incubated overnight at 37 °C, 5% CO_2. The following day the cells were trypsinized and collected for RNA isolation, and RT-qPCR was performed to validate the knockdown efficiency of the shRNA. The cell line with the strongest knockdown was FACS Sorted based on the GFP tagging in the pLL3.7 Vector.

2.8 Overexpression of *Platr14*: The cDNA of *Platr14* was cloned into pLentiCMVtight eGFP Hygro pLenti (Addgene plasmid # 17446) with a Tet-On Advanced vector. Lentiviral particles were generated to transduce the Ab2.2 ES cell line and the C2C12 cell line. The transduced cells were selected under hygromycin and blasticidin. *Platr14* overexpression was induced by doxycycline (1µg/uL), and GFP expression after 48 hours was used as a maker for transfection efficiency. The expression levels of *Platr14* were validated to confirm overexpression of *Platr14*.

2.9 Primer design: Primer3 tool was used for the primer design. The FASTA format of the nucleotide sequences obtained from the UCSC genome browser was submitted following the instructions in Primer3. From the hits obtained the best hit was selected based on the GC content, the Tm value and the inclusion of the splice site regions in the primers. BLAST was performed with the forward and the reverse primers generated to confirm its accuracy with the nucleotide of interest.

2.10 Real Time-Quantitative Polymerase Chain Reaction (RT-qPCR): Real Time-Quantitative Polymerase Chain Reaction (RT-qPCR) was used to study the expression pattern of the genes of interest with ABI PRISM® Sequence Detection System (Applied Biosystems) and the SYBR Green PCR Kit (Qiagen, Netherlands). Each reaction mixture contained 2X

SYBR Green PCR master mix, 1 µL cDNA, 0.5 µM each of forward and reverse primers and 3 µL of RNase-free water (Qiagen, Netherlands) to a final volume of 10 µL. Each reaction was performed in triplicate in 384 and 96 well plates (Applied Biosystems) under the following thermocycling conditions: 15 min at 95 °C for initial activation followed by 40 cycles of 15 s at 94 °C, 30 s at 56 °C and 30 s at 72 °C. The fluorescent dye ROX was used as an internal reference for the normalization of SYBR Green fluorescence. A dissociation protocol ranging from 60 °C to 90 °C was performed when required to investigate the specificity of the primer and the presence of primer dimers after the final amplification cycle. Mouse β-Actin and GAPDH were used as endogenous controls where required. NTC (non-template control) without the cDNA were used as controls. Data analysis and relative expressions were determined by the comparative threshold cycle (ΔΔCt) method. Two-tailed Student's t-test was used for statistical analysis. The primers used are listed in the appendices.

2.11 Contractility Assay: Embryoid bodies were generated using the ES cell lines, control cell line AB2.2 with empty vector and the knockdown cell line AB2.2 cells with the desired shRNA. On day 4 post differentiation, the embryoid bodies were plated on pre-gelatinized 10cm culture dishes. 100 embryoid bodies were counted to study the beating percentage of the

embryoid bodies in both the cell lines at specific time points such as Day 5,8,10,12 post-differentiation. Three biological replicates were analyzed to calculate a beating rate for each cell line at each time point.

2.12 Whole mount *in situ* hybridization: Embryos were harvested at E5.5, E6.5, E8.5 and E9.5 from wild-type mice and fixed with paraformaldehyde/ glutaraldehyde. After fixation, embryos were rehydrated followed by proteinase K treatment. Target RNAs were hybridized at 65 °C with Digoxigenin (DIG)-labeled RNA probes that ranged in size from 300 to 500 nucleotides. After hybridization, the embryos were washed to remove the nonspecific binding and were then incubated overnight with alkaline phosphatase-DIG antibody for immunohistochemistry staining with the Purple substrate with a staining mean time of 18 hours. This was followed by subsequent washes with phosphate-buffered saline with Tween-20 (PBST twice. The embryos were re-fixed in 4% paraformaldehyde, washed again in PBST, and stored at 4 °C in PBST until imaging was performed.

2.13 Immunocytochemistry:

2.13.1 Tnnt2 Staining of AB2.2 Embryoid Bodies: Differentiated ES cells at day 10 post hanging drop were dissociated, and approximately 200,000 cells were plated onto 0.1% gelatin-coated, 18 mm coverslips. After 2 days of additional culture in mLIF-free media, cells were washed

once with PBS and fixed with 4% paraformaldehyde in PBS (Santa Cruz Biotechnology, Dallas, TX) for 10 minutes at room temperature. Cells were washed with PBS and permeabilized with PBS containing 0.25% TritonX-100 (Sigma-Aldrich Corp. St. Louis, MO). Next, cells were blocked in PBS with 5% goat serum (Sigma-Aldrich Corp. St. Louis, MO), 1% bovine serum albumin (Jackson Immuno Research Laboratories, Inc., West Grove, PA) and 0.25% TritonX-100. After blocking, cells were incubated with antibody against troponin T (TNNT2; Abcam, 1:400 dilution) in PBS with 5% goat serum, 1% bovine serum albumin, and 0.25% TritonX-100 overnight at 4 °C. Cells were then washed with PBS containing 0.25% TritonX-100 and incubated with goat anti-mouse IgG (H+L) secondary antibodies conjugated with (1:1000 dilution) Alexa 555 and 20µg/mL Hoechst 33342 (Thermo Fisher Scientific, Waltham, MA) for 1 hour at room temperature in PBS containing 5% goat serum, 1% bovine serum albumin, and 0.25% TritonX-100. Stained cells were washed with PBS-0.25% TritonX-100 and were mounted with Prolong Diamond Antifade Mountant (Thermo Fisher Scientific, Waltham, MA).

2.13.2 MF-20 Staining of C2C12 myoblasts: C2C12 cells were cultured on gelatinized coverslips in 10 cm culture dishes. Five coverslips were placed on each plate. To induce myogenic differentiation, the general

DMEM media was replaced with differentiation media and allowed to differentiate till Day 4.At the end of day 4, the culture dishes were washed with PBS and fixed with 4% PFA, washed in PBS in 0.25% Triton (PBST) twice and replaced with BSA in PBST containing 5% Goat serum and incubated at room temperature for 30 minutes. The coverslips were removed from the culture dishes and incubated with primary antibody MyHC (Developmental Studies Hybridoma Bank, clone MF-20) overnight and washed twice with PBST and then incubated for one hour with goat anti-mouse IgG (H+L) secondary antibodies conjugated with Alexa 555 (1:500) (Thermo Fisher Scientific, Waltham, MA) along with goat anti-mouse fluorescein-conjugated Alexa Fluor 488 (Thermo Fisher Scientific). After two subsequent PBS washes, the coverslips were mounted on slides with Prolong Antifade medium. 20 µg/mL Hoechst 33342 (1:1000) (Thermo Fisher Scientific, Waltham, MA) was added during the second PBS wash. Cells incubated without primary antibody and with secondary antibody alone were used as staining controls. The myotubes were imaged using Nikon Eclipse Ti-E inverted microscope. The images were analyzed using the Image-J software. Four random fields per coverslip were counted for myotubes, and the average per coverslip was taken to calculated the rated of myotube fusion in both the knockdown and the control C2C12 cell lines. For all staining, sample coverslips incubated without primary antibody, and

only secondary antibody for respective antibodies were used as background controls. A minimum of 4 representative areas was imaged per coverslip. Images taken for each area included snapshots under BF, DAPI filter, and FITC/TRITC. The images were analyzed using the Image-J software. For quantification purposes, three coverslips were used per sample per antibody. Total cells positive for respective target divided by total number of cells (nuclei) for each of the representative areas and then the average per coverslip were taken to calculated the rated of myotube fusion in the respective cell lines of C2C12 such as the control, the sh*Platr14* knockdown and the C2C12 cell line harboring the *Platr14* overexpression vector.

2.13.3 5-Bromo-2'- Deoxyuridine Labeling and detection:

Undifferentiated and Differentiated C2C12 cells (at day 2) were incubated with 5-Bromo-2´-Deoxyuridine, BrdU (10uM) (Invitrogen, Thermo Fischer Scientific, Waltham, MA) for 2 hours at 37 °C. After the PBS wash, twice for 5 minutes; the plates were fixed with 4% PFA for 10 minutes. After the PBS wash, twice for 5 minutes, the DNA denaturation was performed with 1 N HCl in PBS for 30 minutes at 37 °C, followed by neutralization in 0.1 M borate buffer for 10 minutes. The plates were PBS washed twice for 5 minutes, followed by with PBST wash for 5 Minutes. After Blocking with 1X

casein in PBST, for 30 minutes at room temperature, the plates were incubated with anti-BrdU monoclonal antibodies (1:250) (Developmental Studies Hybridoma Bank, clone G3G4) in (1X casein, PBST, 2% Goat Serum) overnight in a humidified chamber. The next day, the plates were washed briefly thrice with PBST followed by incubation with secondary antibody (1:400) AF-546/568 in (1X casein, PBST, 2% Goat Serum) for one hour at room temperature. PBS-washed thrice. 20 µL of DAPI stock solution (1mg/mL) was added to 50 mL of PBS during the first PBS wash. The wells were covered with PBS and stored at 4°C until imaging. Random fields were imaged per well, and the number of BrdU-positive cells over the total number of cell per field were calculated.

2.14 *In vitro* cell migration assay: Undifferentiated and differentiated C2C12 cells (day 2) were maintained at about 100% confluency in general media and differentiated media respectively. Using a 200 µL pipette tip, a scratch was created at the center of the well. The wells were washed and the media replaced. The initial time point was imaged with the Nikon Eclipse Ti. The cells were incubated at 37 °C, 5% CO_2 for 6 hours. After 6 hours the migration rate was imaged. The distance across the scratch was measured using the scale bar from ImageJ software. The migration rate was calculated by using the following formula: The final migrated distance:

The initial distance /the final migrated distance X 100.The experiments were performed in replicates.

2.15 RNA-sequencing:

2.15.1 Library preparation for RNA-sequencing: Assessment of RNA quality using Agilent bioanalyzer with a RNA high sensitivity chip was determined before library preparation. A total of 1 µg of total RNA from the study groups was used for generating the libraries for next-generation sequencing with Illumina's HiSeq3000 platform. Libraries were prepared using the ScriptSeq Complete Gold kit from Epicenter (Lucigen Corporation) according to the manufacturer's recommended procedure for low-input starting material. Libraries prepared were quantified using the pico-green quantification kit from Promega (Promega Corporation, USA), along with Promega's Quantifluor fluorometer.

2.15.2 Data analysis for RNA-sequencing: The raw reads from the RNA-sequence data were aligned to the mouse genome (mm9) by using Bowtie and TopHat (Daehwan K *et al*., 2013). The reads, which mapped to multiple locations of the genome, were discarded before downstream analysis. We generated a comprehensive list of lncRNAs by merging the annotated lncRNAs from the ENCODE project and those listed in the study by Matkovich and colleagues (the combination of annotated lncRNAs from

Noncode 2.0, lncRNAdb, Scripture, fRNAdb, Ensembl, RefSeq, and the UCSC Genome Database). The abundances of lncRNAs, protein coding genes, and other noncoding RNAs were computed by using cufflinks and were expressed as FPKMs (Benham *et al*., 2017). For the C2C12 study, the RNA-Seq data was pre-processed by removing any adaptor contaminates and poor-quality (q30) base calling using cutadapt (Martin, M. 2011). Processed data was aligned to mouse genome (version mm9) using TopHat (Daehwan K *et al*., 2013) and processed through the cufflinks (Trapnell C. *et al*., 2012) standard pipeline using Ensembl GTF (GRCm38.87) reference transcript to generate normalized FPKM for differential analysis.

2.16 RNA Pull Down Assay:

Using the (pCR-Blunt II-TOPO) kit, (Invitrogen, Life Technologies) the Platr14 cDNA was cloned into the pCR™4-TOPO® vector under the T7 promoter. The clones obtained were sequenced to confirm the orientation. AmpliScribeTM T7-FlashTM Biotin-RNA Transcription Kit (Epicenter, Lucigen Corporation) was used to follow the manufacturer's protocol to produce randomly labeled biotinylated RNA. The RNA pull-down was performed with the Streptavidin beads were used to tag the biotin-labeled RNA to properly orient the RNA for the protein binding. 100 mg of cells/100uL of the lysis buffer was used to prepare the lysate. The cell

lysate with a protein concentration of about 100mg/mL was prepared from day 2 post-differentiated C2C12 cell line harboring the *Platr14* overexpression vector. The eGFP empty vector was used as the control. Following the manufacture's protocol, the prepared lysate was allowed to interact with the biotin-labeled RNA to be further eluted out for mass spectrophotometry analysis. Elute with no labeled RNA was used as the control.

2.17 RAP-MS: The RNA Antisense Purification-Mass Spectrophotometry protocol is used to study endogenous RNA: Protein Interactions. Here the protocol according to Guttmann *et al.,* was followed. The biotin labeled RNA probes of 20 nucleotides long were designed that tile the entire *Platr14* nucleotide sequence. The *Platr14* pulled down with the associated proteins were analyzed using mass spectrophotometry to identify the associated proteins.

2.17.1 Sample Preparation for mass spectrometry:

Following lyophilization, protein mixture was reconstituted in 10-20 µL of 25–50% acetonitrile with 100mM ammonium bicarbonate, pH 7.2 based on protein solubility. This was reduced for 15 minutes at 37 °C using 50 mM TCEP in ammonium bicarbonate, pH 7.2. Alkylation was performed without

light exposure for 30 min at room temperature using 50 mM iodoacetamide in ammonium bicarbonate, pH 7.2. The protein sample was diluted to 300 µL with HPLC-grade water, and the proteins were precipitated by chloroform-methanol method. Followed by reconstitution of the precipitated protein in 80 µL of 100 mM ammonium bicarbonate and further digested using 10 µL of a 20 ng/µL stock solution of sequencing-grade trypsin (200 ng, Promega) for 16–24 h at 37 °C. 10 µL of 10% formic acid was added to quench the reaction and the sample was lyophilized to remove ammonium bicarbonate. This was further reconstituted in 20 µL of 1% formic acid for analysis by mass spectrometry. The analysis was performed in triplicate.

2.17.2 LTQ Ion Trap: Liquid chromatography tandem mass spectrometry that has an Agilent 1290 Infinity UPLC system using Solvent A = Water +0.1% formic acid and Solvent B = Methanol +0.1% formic acid. The protocol set up was followed per the protocol of Mali S. *et al*., 2016.The RAW data files were transformed to either mzXML or MGF format using the MSConvert utility program from the ProteoWizard program suite (http://proteowizard.sourceforge.net/tools.shtml). Spectra from the HPLC-MS/MS run were output into MGF format for analysis in X!Tandem (Current GPM 2013.09.07) of the mm9 genome. Each database search used trypsin digestion and default Q-TOF and Ion Trap methods.

2.18 Chromatin Isolation by RNA Purification (ChiRP):

The experiment was carried out as per the protocol of Chu C *et al*. The biotinylated antisense probes were designed tiling the entire *Platr14* nucleotide sequence. About 40 million day 2 post differentiated C2C12 with *Platr14* over-expressed cell lines were used for the study. The *Platr14* retrieval was validated by RT-qPCR to confirm the lncRNA retrieval. RNA input was used as the control. The DNA fraction was extracted from ChIRP samples to be further processed for the high-throughput sequencing libraries per Illumina protocol.

2.19 Bioinformatics: For Gene Ontology studies, in-silico tools such as DAVID Bioinformatics Resources 6.8, GOrilla, STRING Version 10.0 were used. The University of California Santa Cruz (UCSC) Genome Bioinformatics website (http://genome.ucsc.edu), NCBI37/mm9 database, was used to identify the neighboring mesoderm markers of lncRNA, for obtaining the nucleotide sequences in FASTA format, and performing the BLASTN analysis. Noncode 2.0, lncRNAdb, Scripture, fRNAdb, Ensembl, RefSeq databases along with RNA-Sequencing data analysis tools Cutadapt, TopHat, Bowtie Cufflinks were used.

Chapter 3

Results

Long noncoding RNAs in Cardiac and Skeletal muscle differentiation during mouse embryogenesis.

(Some of the results shown here will be included in (Benham *et al*. and George *et al*., In preparation)

3.1 Identification of candidate lncRNAs in the mesoderm lineage:

Even though we have narrow knowledge about the specific functions of lncRNAs in mammalian development and disease, we are aware that lncRNAs are shown to exhibit unique expression patterns across tissues (Derrien et al., 2012; Dinger et al., 2008). Our lab generated transcriptome data from ES cell line UH3, at specific time-points such as undifferentiated ES cells (Day 0) and differentiated stages of cell aggregates known as the Embryoid bodies (EBs) at Days 5,8,10,12 (Figure 3.1A&B). This data with Mesp1 lineage tracing generated from our lab by Soibam and colleagues showed the enrichment for mesoderm developmental processes as well as heart development by the Mesp1-derived cell population. Since lncRNA are seen associated with developmental stages (A Fatica *et al*.,2014), to identify the lncRNA in early mesoderm formation we mined the data for lncRNA profiling. Though 85% to 90% of the annotated transcripts represented mRNA, the rest comprised pseudogenes, antisense RNAs, long noncoding RNAs and small RNAs. A list of 4099 annotated long noncoding RNA (greater than 200 nucleotides) was generated from the sequencing data.

A total of 2916 long non-coding transcripts with FPKM greater than 0.1 were obtained from ES undifferentiated, *Mesp1*-derived and non-*Mesp1*-

derived cell lines. The lncRNA expression profiles of undifferentiated ESCs and populations of *Mesp*1-derived and non–*Mesp1*-derived cells were classified into 16 distinct expression patterns (groups A-P)(Figure 3.1C). Except groups B and K, all other groups showed differential lncRNA expression between the *Mesp1*-derived and non–*Mesp1*-derived cells. Therefore this allowed identifying a set of lncRNAs that were enriched in the mesoderm lineage. A total of 2460 lncRNAs were obtained from the *Mesp1*-derived cell lines. From this selected population 220 lncRNAs were expressed uniquely at distinct time-points. Gene ontology performed showed that this mesoderm related lncRNAs had terms associated with developmental processes heart development, skeletal system development, Kidney development, blood vessel development, vasculature development, cartilage development and mesenchyme development (Table 3.1). From this list we selected 12 lncRNAs shown to have distinct enrichment at specific time points in the differentiated UH3 Mesp1-derived cell population (Figure 3.2).

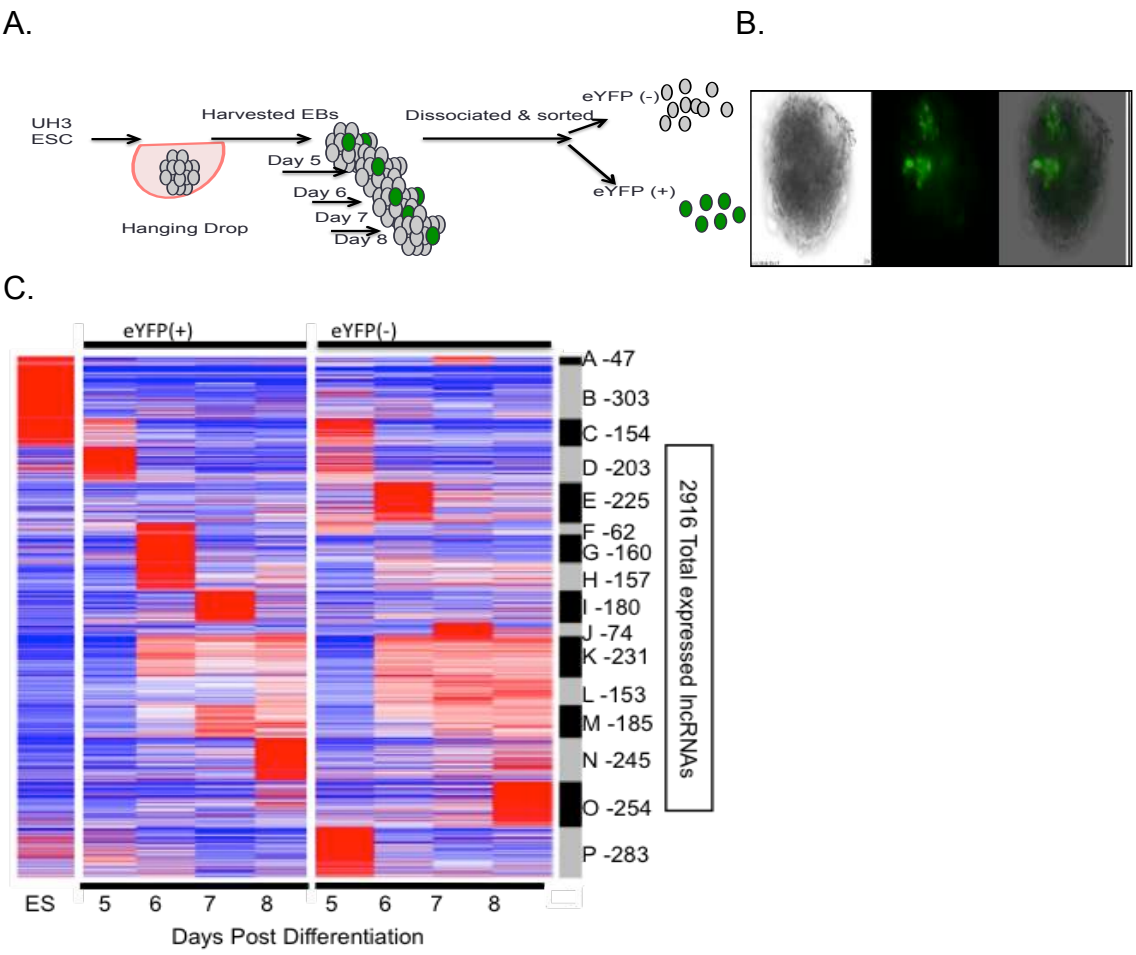


Figure 3.1:LncRNA profiling from UH3 RNA Sequencing data set.

(A) Shows the differentiation of the ESC cell line UH3 to obtain Embryoid Bodies (EBs) constituting *Mesp1* positive and non-*Mesp1*-derived cells. (B).Embryoid Bodies after 6 days in hang drop shown at 20X magnification under Bright Field, Fluorescence YFP filter, and merged, respectively. The YFP+ represents the *Mesp1*-derived cells. (C). Heat map representing the

lncRNAs identified in the *Mesp1* and non-*Mesp1* populations in undifferentiated ESCs and Embryoid bodies across different time points.

Table 3.1: Gene ontology representing the different GO terms associated with the mesodermal markers neighboring the profiled lncRNAs.

GO Terms	P Value
GO:0001501|skeletal system development	5.35E-11
GO:0060348|bone development	2.40E-09
GO:0051216|cartilage development	1.08E-08
GO:0001568|blood vessel development	6.07E-08
GO:0001944|vasculature development	1.15E-07
GO:0001822|kidney development	2.67E-06
GO:0007507|heart development	3.39E-05
GO:0060485|mesenchyme development	4.82E-04

A.

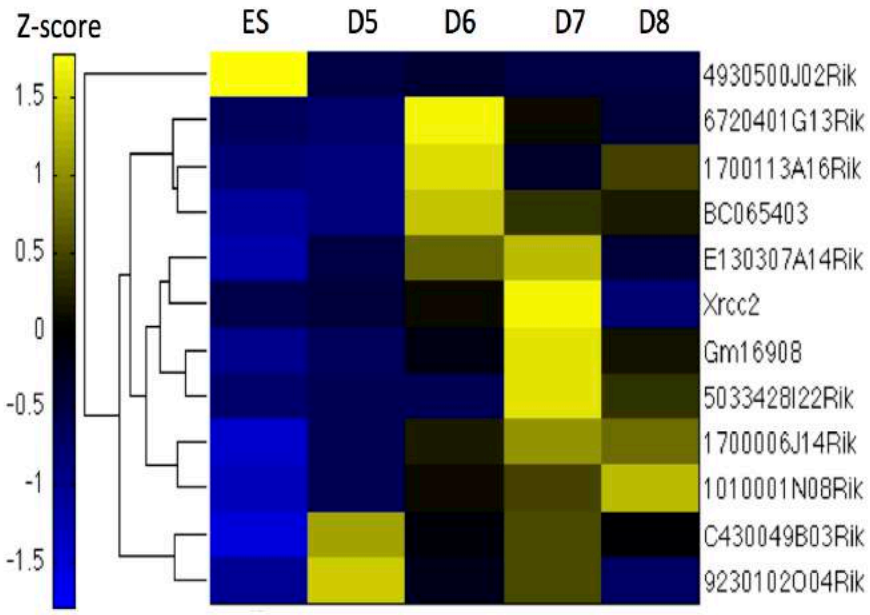
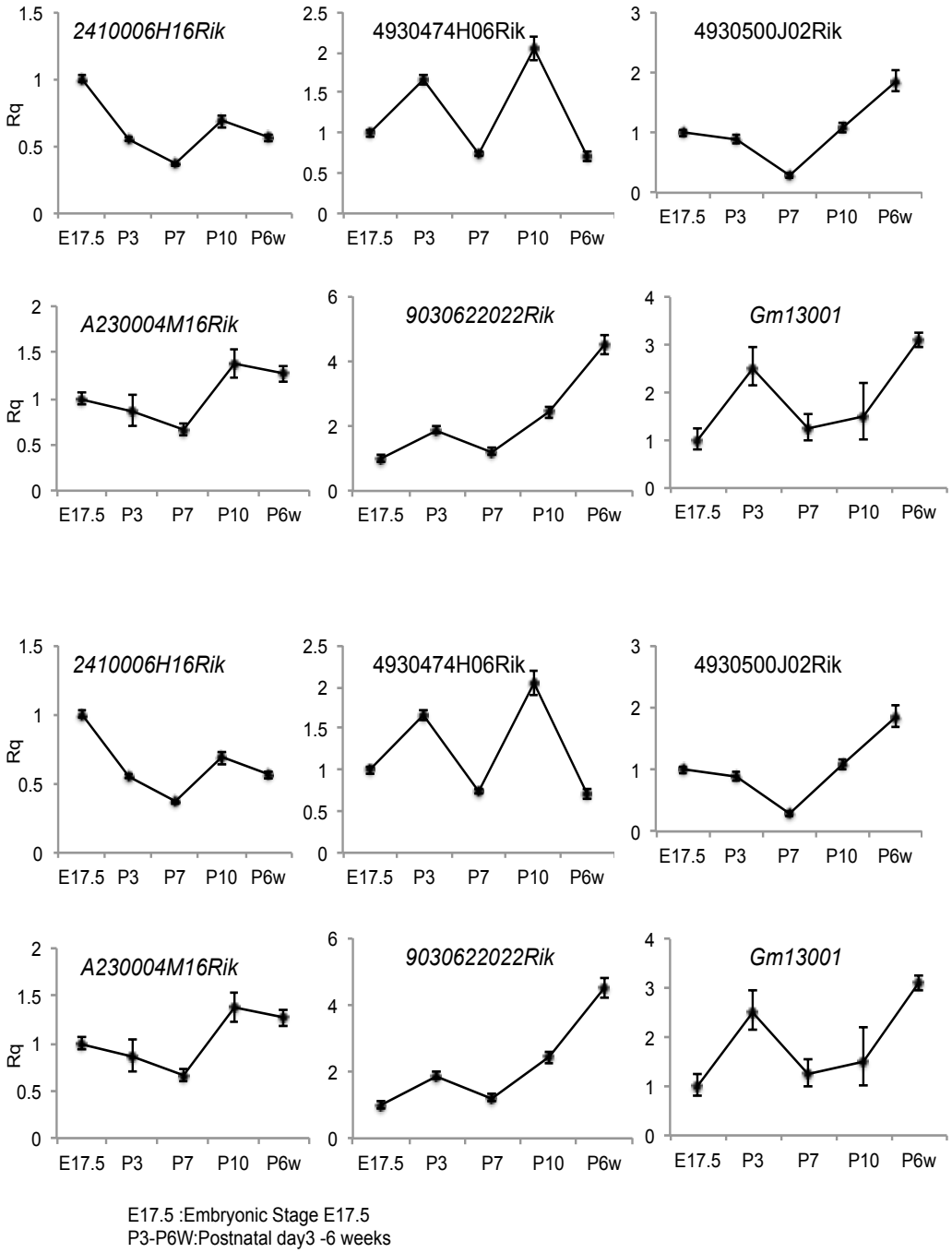


Figure 3.2: Heat map showing day specific expression of the selected lncRNAs for further characterization.

3.2 Identification of candidate lncRNAs in the cardiac lineage:

LncRNAs are known to regulate cardiac formation (S Frank, 2016). Klattenhoff, C *et al.* has shown that expression of lncRNAs in specific cell types during ESC differentiation indicates that they can be candidate regulators of that specific lineage commitment. The 12 lncRNAs (Figure 3.2) we selected for further screening were derived from UH3 cells, a cell line preferably used for the induction of cardiac lineage. As these lncRNAs

were present at a specific time point of differentiation, we established that these would be important at specific stages of differentiation of the cardiac lineage. To further validate its expression in cardiac differentiation *in vivo*, we performed RT-qPCR of wild type mouse tissues procured from the E17.5 embryonic stage as well as postnatal days 5,8,10, and Postnatal 6 weeks (adult stage). Figure 3.3 shows the expression pattern of the selected lncRNAs in embryonic stage as well as the postnatal stages of cardiac formation.



E17.5 :Embryonic Stage E17.5
P3-P6W:Postnatal day3 -6 weeks

Figure: 3.3: In-vivo expression of mesodermal related lncRNAs in heart tissues. Quantitative PCR for lncRNA expression in heart tissue isolated from E17.5, Postnatal Days 3, 7, 10, and 6 weeks.

Polymerase Chain Reaction - PCR; Relative quotient for expression - Rq. All samples have n=3 biological replicates. E17.5 was used as reference. All expressions were normalized to β-actin. All relative expressions were determined using comparative threshold cycle (ΔΔCt) method.

3.2.1 Correlation of Mesoderm related lncRNAs with their neighboring mesoderm and cardiac markers:

LncRNAs can associate with its neighboring genes, (K, W Vance,2014) and can regulate expression of neighboring genes by acting in cis or in trans by diverse mechanisms (Orom and Shiekhattar, 2011; Rinn and Chang, 2012). Based on the UCSC mm9 Genome search we selected few lncRNAs with genomic location closer to neighboring mesodermal markers. RT-qPCR was performed on mouse cardiac tissues obtained from embryonic stages 17.5 and postnatal tissues at day postnatal days 3,7,10 and adult stage P 6 weeks (Figure 3.4). The lncRNAs showed positive and negative correlation. LncRNA *1010001N08Rik* showed a positive correlation with the neighboring *Gata4*. LncRNA *5033428I22Rik* showed a negative trend with *Hand2* except at adult stage where it was up regulated. While lncRNA *5430400D12Rik* followed a similar pattern with

the associated mesoderm marker *Notch1* till the postnatal 10 days, followed by an opposite effect at the adult stage. From this data, we can only conclude that lncRNA may or may not have a direct correlation with its neighboring mesoderm marker.

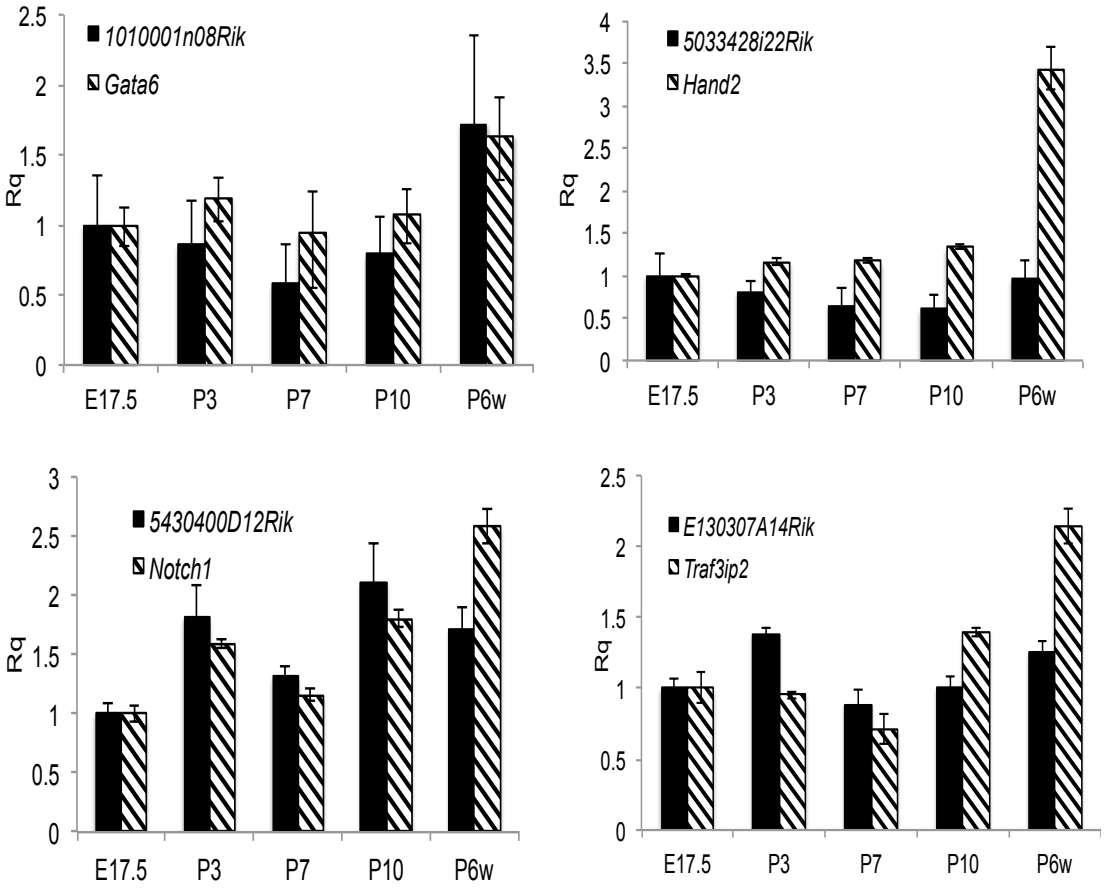


Figure 3.4. *In vivo* correlations of mesodermal related LncRNAs with the neighboring mesoderm markers. Quantitative PCR for lncRNA expression and its neighboring mesodermal marker in heart tissue from isolated hearts from E17.5, Postnatal Days 3, 7, 10, and 6 weeks.

Polymerase Chain Reaction - PCR; Relative quotient for expression - Rq. All samples have n=3 biological replicates. E17.5 heart tissues were used as reference. All expressions were normalized to β-actin. All relative expressions were determined using comparative threshold cycle (ΔΔCt) method.

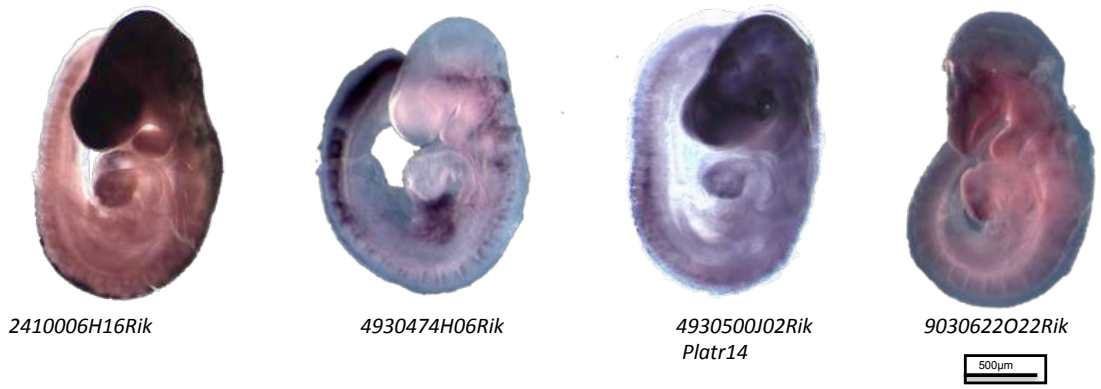
3.2.2 *In vivo* Expression of cardiac related long non-coding RNAs:

In Situ Hybridization studies were performed in E9.5 mouse embryos, to assess the expression of the selected 12 lncrnas in cardiac lineage in vivo. Of the 12 lncRNAs, 4 lncRNAs such as *2410006H16Rik, 4930474H06Rik, 4930500J02Rik (Platr14)* and *9030622O22Rik* showed a positive expression within the primitive ventricle of the developing heart (Figure 3.5A). This data correlated with the *in vivo* expression trend seen in the embryonic (E17.5), the postnatal and adult mouse cardiac tissues (Figure 3.5B).

LncRNAs as *2410006H16Rik, 4930474H06Rik, 4930500J02Rik (Platr14),* and *9030622O22Rik* showed a distinct pattern of expression. *2410006H16Rik* showed an up-regulation in the embryonic stage E17.5, while lncRNA *4930474H06Rik* showed a distinct pattern of expression at

postnatal days 3 and 7. LncRNA *4930500J02Rik (Platr14)* showed an up-regulation at adult stage (P6 weeks) while lncRNA *9030622O22Rik* showed a consistent increase from the embryonic stage to the adult stage.

A.



B.

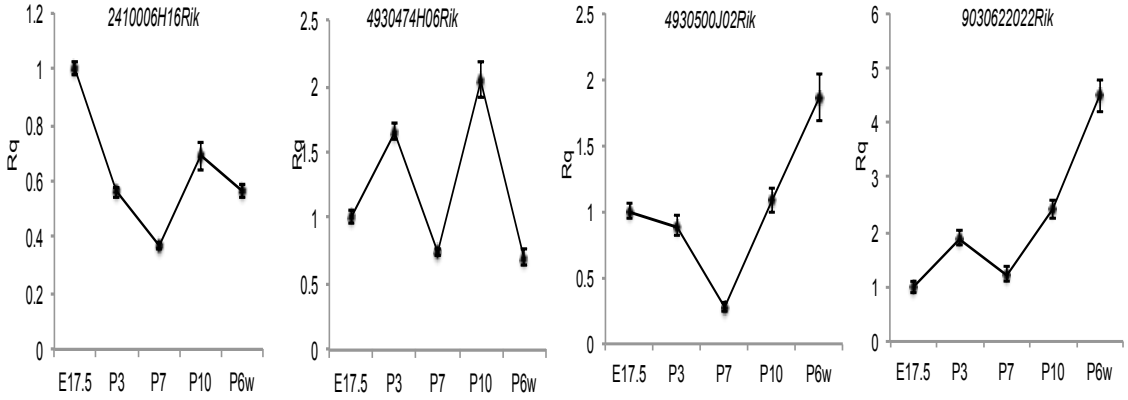


Figure 3.5.Cardiac-related lncRNA expression *in vivo*.

(A). *In-situ* hybridization of E9.5 embryos showing positive expression of lncRNAs in the cardiac region. (B). Quantitative PCR of lncRNA

expression pattern in prenatal and postnatal cardiac tissues. Polymerase Chain Reaction - PCR; Relative quotient for expression – Rq. All samples have n=3 biological replicates. E17.5 was used as reference. All expressions were normalized to β-actin. All relative expressions were determined using comparative thresholdcycle (ΔΔCt) method.

3.3.3 Mesoderm-related LncRNAs in Cardiac Lineage Commitment:

The ESCs have the ability to differentiate into any derivative of the three germ layers including cardiac cell types (Klattenhoff, C *et a*l.). To examine the function of the selected lncRNAs in lineage commitment and its ability to form cardiac tissues, ES cell lines, AB2.2 was used to generate three stably expressed short hairpin RNAs (shRNA) and the cell line with the highest knockdown efficiency was validated by RT-qPCR. The differentiation ability of the ES cells can be visualized as spontaneously contracting EBs during differentiation. All the lncRNA knockdowns performed in AB2.2 cell lines that were allowed to differentiate to form embryoid bodies (EBs). Only 15% of the sh*Platr14* knockdown (Figure 3.6A&B) day11 embryoid bodies showed spontaneous beating as compared to the 25% of beating in control EBs with empty vector (Figure

3.6C). The other lncRNAs knockdown cell lines showed an insignificant change in the EB contractility as compared to the control ES cell line with empty vector pLL3.7 (data not shown). Therefore, it was observed that the knockdown of *Platr14* affects the contractility *in vitro* in embryonic bodies.

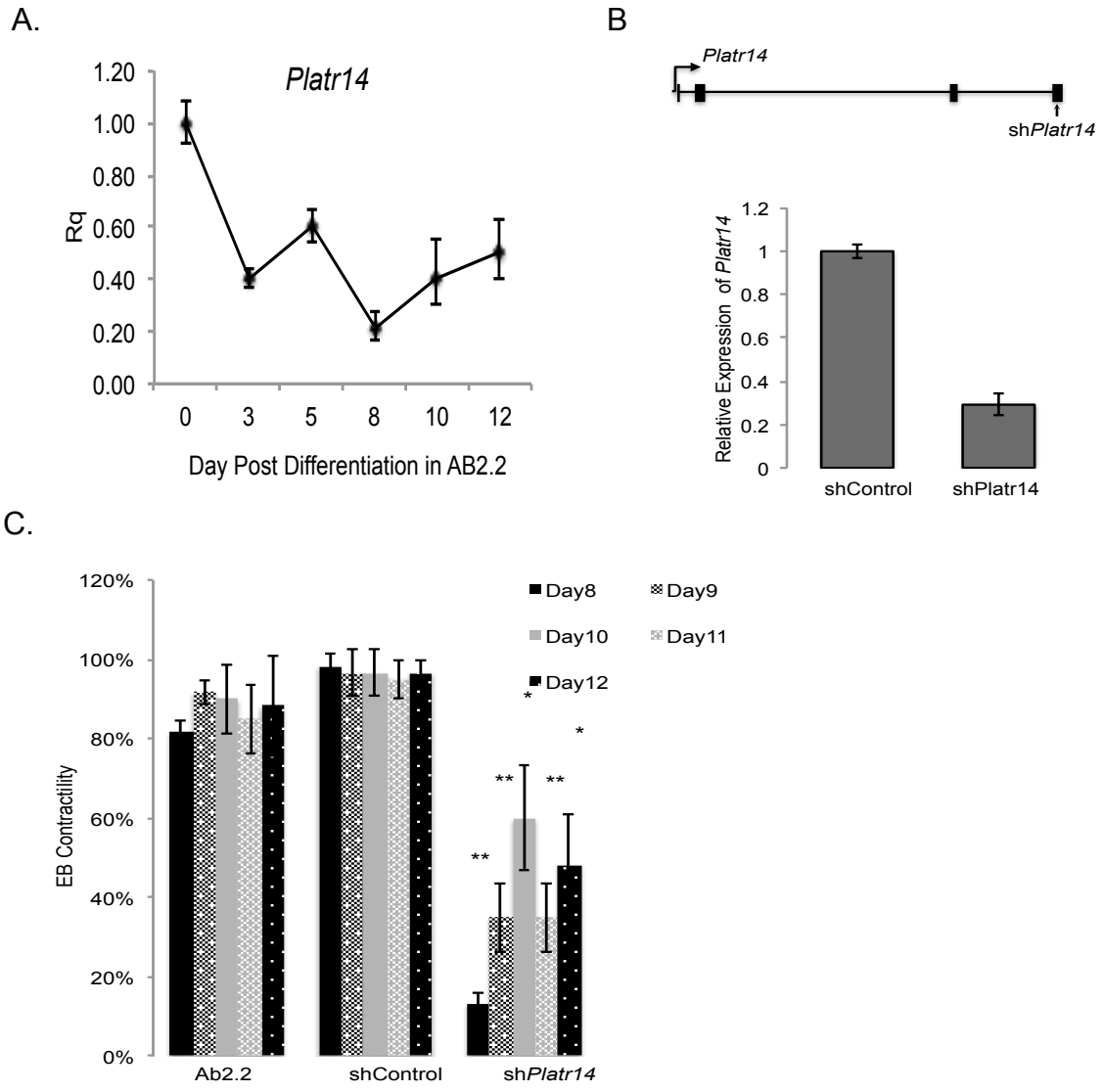


Figure 3.6. Cardiogenic role of *Platr14*. (A) Quantitative PCR for *Platr14* expression in undifferentiated and differentiated AB2.2 ES cells.

(B). Quantitative PCR validation of *Platr14* shRNA knockdown in AB2.2 ES cells. shControl was used as the reference. Polymerase Chain Reaction - PCR; Relative quotient for expression-Rq. All expressions were normalized

to GAPDH. All relative expressions were determined using comparative threshold cycle (ΔΔCt) method. All samples have n=3 biological replicates. (C).Embryoid body contractility assay performed in sh*Platr14* knockdown AB2.2 cells. P-value <0.05(*), P-value <0.01(**).

3.4. Long Non-Coding RNA *Platr14*:

Platr14 is an intergenic lncRNA that is 582 nucleotides long located on chromosome 2 of the mouse genome (mm9) neighboring the gene cleavage stimulation factor, 3' pre-RNA, subunit 3 (*Cstf3).* Bergmann *et al.* has recently reported *4930500J02Rik/ENSMUSG00000086454* as a non-coding transcript having pluripotency function and so the name, *Platr14* Pluripotency associated transcript14.

3.4.1 Knockdown of *Platr14* affects the cardiogenic potential of ESCs *in vivo*:

To further dissect its role in the regulation of cardiac markers at different stages of cardiogenesis we performed RT-qPCR of the RNA isolated from undifferentiated and differentiated AB2.2 ESCs. RT-qPCR validation showed a significant depletion of the *Platr14* transcript in ES AB2.2 cell line with shPlatr14#2 directed against the 582-nucleotide stretch (Figure

3.6B). The expression level of the cardiac markers analyzed in sh*Platr14* AB2.2 EBs at specific time points (Day 0,5,8,10,12) showed a deregulation of the cardiac markers such as *T brachyury, Goosecoid (Gsc), Nkx2-5, Mef2c, Gata4, Tbx5, Tbx20, Myh6,* and *Tnnt2* as compared to the undifferentiated AB2.2 with empty vector as control. The transcription factors *Brachyury (T)* and *Goosecoid (Gsc),* expressed in the primitive streak, have significant roles in mesoderm induction (Costello *et al*., 2011; David *et al.,* 2011). Upon *Platr14* inhibition, mesendoderm markers such as *Goosecoid, T-Brachyury* showed higher levels of expression in undifferentiated ESCs as well as differentiated day 5 EBs by at least 2 fold (Figure 3.7A). Mesoderm/cardiac markers such as *Mef2c* was up regulated by about 1.5 fold at Day 8. *Nkx2.5, Gata4* showed a down-regulation at Day 8 by about 1.5 fold and 2 fold at day 8 and day 10 respectively in sh*Platr14* knockdown AB2.2 EBs (Figure 3.7B). Also the knockdown cell line also exhibited a down-regulation of cardiomyocyte markers such as *Tbx5, Tbx20, Myh6,* and *Tnnt2. Tbx5* and *Tbx20* were down regulated 2 fold and 3 fold respectively at day12 upon inhibition of *Platr14* (Fig 3.7C). While endoderm markers such as *Foxa2* did not show any significant change, *Gata6* showed a down regulation at day10 and day12 in *Platr14* knockdown AB2.2 EBs (Fig 3.7A).

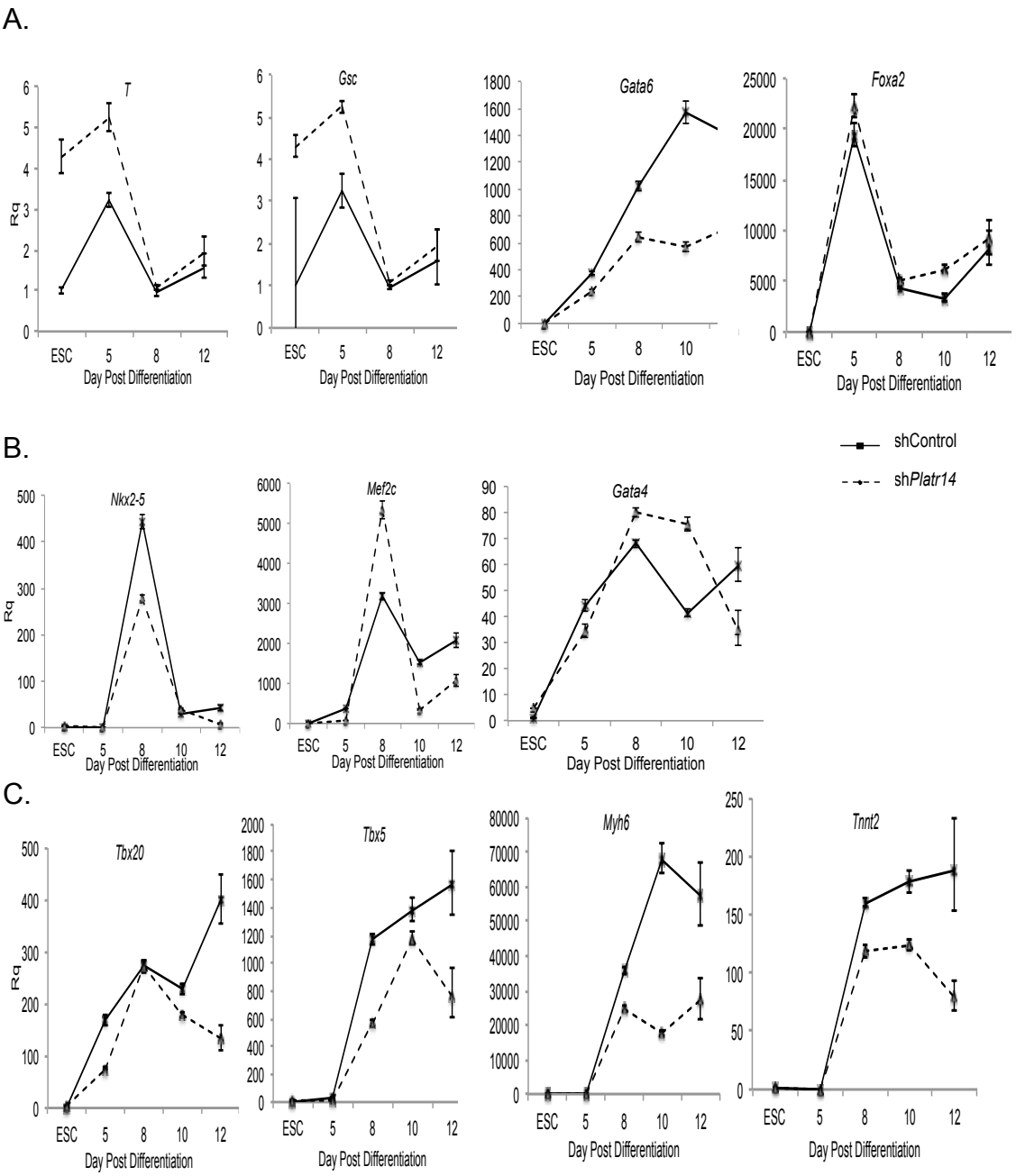


Figure 3.7:Knockdown of Platr14 affects cardiogenic markers.

(A).Quantitative PCR for *Platr14* expression of mesendoderm and endodermmarkers. (B).Cardiac mesoderm markers and (C) Cardiomyocyte

markers, in undifferentiated AB2.2 ES cells (Day 0) as well as differentiated Embryoid bodies at days post differentiation 5, 8, 10 and 12. All expressions were normalized to GAPDH. All relative expressions were determined using comparative threshold cycle (ΔΔCt) method. All samples have n=3 biological replicates. Undifferentiated ESCs day 0 was used as reference. Polymerase Chain Reaction - PCR; Relative quotient - Rq.

3.4.2 Immunocytochemistry studies show a decreased expression of cardiac contractile proteins upon *Platr14* inhibition:

Cardiac troponin is a major component that makes up the contractile unit for the cardiac muscle. Immunocytochemical analysis using cardiac troponin T (*cTnt*) antibody for the *cTnnt2* marker was performed in the *Platr14* knockdown AB2.2 EBs. At day12 post differentiation, sh*Platr14* knockdown AB2.2 EBs showed a reduction in the population of positive *cTnT* cells as compared to control AB2.2 EBs (Figure 3.8A). The population of *cTNNT2* positive cells was relatively reduced as compared to the control with empty vectors (Figure 3.8B). These data helped as to propose that *Platr14* apart from its pluripotency ability it also plays a crucial role in defining the cardiac lineage.

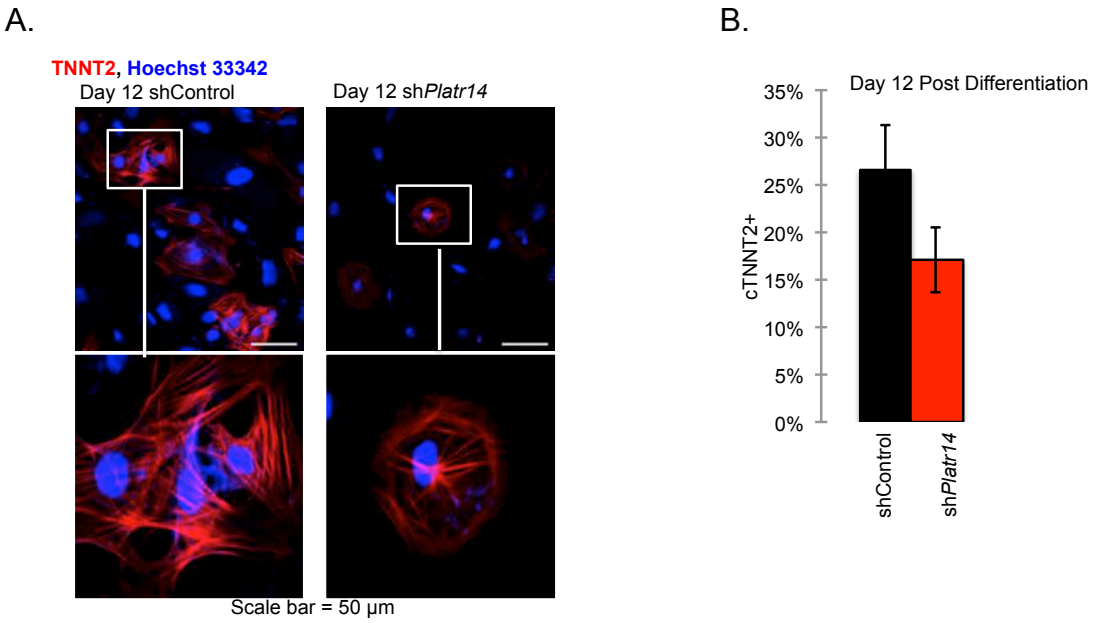


Figure 3.8: Inhibition of *Platr14* shows a decreased expression of Cardiac contractile proteins in day 12 Embryoid bodies.

(A). Immunocytochemistry staining showing the reduced expression of cardiac troponin in sh*Platr14* knockdown. (B). Immunostaining analysis showing a reduction in the *cTNNT2* positive cells in the sh*Platr14* knockdown. Empty vector was used as control.

3.5 Identification of mesoderm related lncRNAs in Skeletal Muscle Formation:

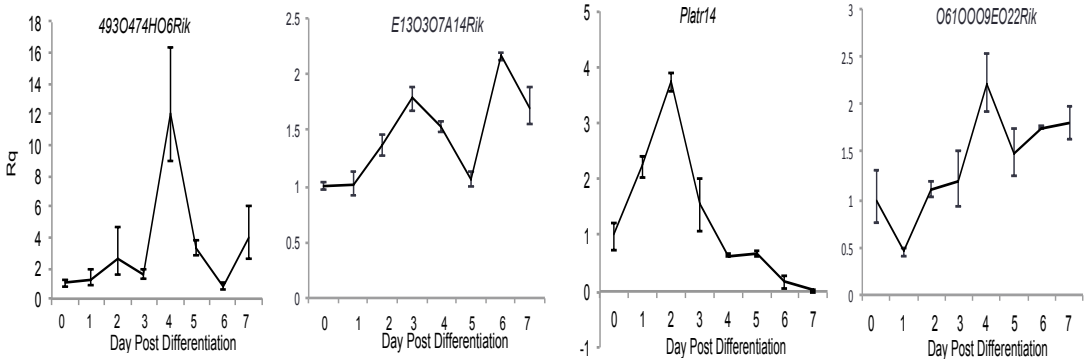
Mesodermal precursors give rise to skeletal muscle formation. Given that the subset of selected 12 lncRNAs are mesoderm related lncRNAs; we

examined its role in myogenesis. We used C2C12 mouse myoblast cell line, which is a widely used cell line for the in-vitro study as it helps to mirror the myogenic changes as in an *in vivo* model. We allowed the differentiation of the C2C12 cell line up to day 7 in differentiation media in the presence and absence of Insulin and performed quantitative RT-qPCR analysis. While no significant expression of the lncRNAs was observed in C2C12 allowed to differentiate in the absence of insulin, four of the selected 12 lncRNAs showed a striking expression pattern. LncRNAs such as *Platr14*, *0610009E022Rik, 493O474HO6Rik,* and *E13O3O7A14Rik* showed a consistent expression pattern. These lncRNAs showed enrichment at a specific time point of differentiation. While lncRNA *0610009E022Rik* was expressed throughout the differentiation stages, *493O474HO6Rik*, *E13O3O7A14Rik*, and *Platr14* strikingly showed elevated levels of expression at specific time points such as day 2, 3 and 4 post differentiation (Figure 3.9A).

This expression trend correlated with the *in-situ* hybridization studies performed in E9.5 embryos. *In-situ* hybridization showed a positive expression of *Platr14, 0610009E022Rik, 493O474HO6Rik,* and *E13O3O7A14Rik* along the dorsal side representing the myotome region of the E9.5 embryos. LncRNA *493O474HO6Rik* showed positive expression in the somites of the head as well as the tail region, specifically

the tail tip. LncRNA *E13O3O7A14Rik* and *Platr14* showed positive expression pattern in somites of the tail, while lncRNA *0610009E022Rik* did not show any distinct expression (Figure 3.9B). Since *Platr14* was seen to show a significant role in mesoderm directed cardiac lineage, we were also interested in the role of *Platr14* in myogenesis.

A.



B.

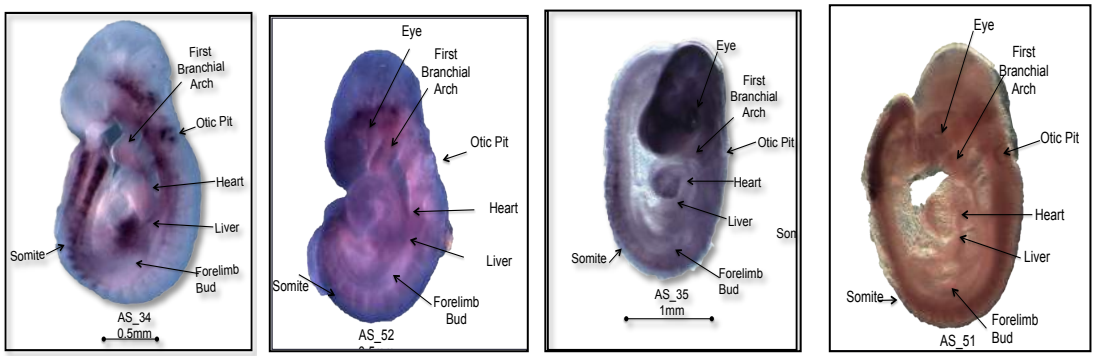


Figure 3.9: Expression pattern of the selected mesodermal LncRNAs in skeletal muscle formation. (A). Quantitative PCR expression pattern in C2C12 cell line. Polymerase Chain Reaction - PCR; Relative quotient for expression-Rq. All expressions were normalized to β-actin. All relative expressions were determined using comparative threshold cycle (ΔΔCt) method. All samples have n=3 biological replicates. (B). Whole Mount In situ hybridization of E9.5 wild type mouse embryos with DIG-labeled antisense RNA probe to a specific lncRNA.

RT- qPCR analysis showed that the expression pattern in embryonic tissues such as the heart, lung, brain, torso, tongue, liver and spleen isolated from E15.5 embryos showed the highest level of *Platr14* expression in the tongue and torso tissues, both constituting the skeletal muscle (Fig 3.10).

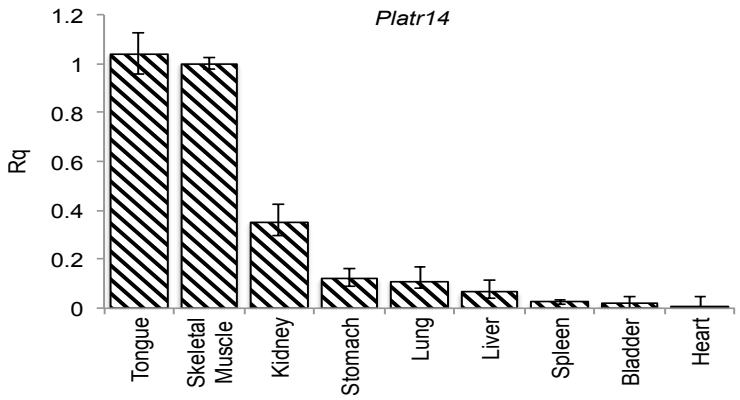


Figure 3.10:*Platr14* expression is up-regulated in Skeletal Muscle. Quantitative PCR for the expression of *Platr14* in wild type mouse embryonic tissues (E15.5). Polymerase Chain Reaction - PCR; Relative quotient for expression – Rq. All expressions were normalized to GAPDH. All relative expressions were determined using comparative threshold cycle (ΔΔCt) method. All samples have n=3 biological replicates.

3.5.1 Knockdown of *Platr14* affects myogenic program *in vivo*:

From the high enrichment observed during the early stages of C2C12 differentiation as well as the correlating *in situ* hybridization data, *Platr14* was selected for further study to establish its role in myogenesis. It was observed from the earlier screening that *Platr14* is up-regulated post-differentiation day 2. Of the 3 shRNAs used for the knockdown of *Platr14* in C2C12, sh*Platr14#2* showed the maximum knockdown and was selected for the future assays (Figure 3.11A). The differentiated sh*Platr14* knockdown C2C12 cell line at specific time points (post differentiation days 0,2,4) showed a deregulation of early and terminal myogenic markers. Inhibition of *Platr14* showed that the genes involved in early stages of myogenesis such as *Cdkna1, Cdknb1, Myod, Mef2c, Myh3* were down regulated in undifferentiated and differentiated day 2 C2C12 cells. Similarly, the terminal markers *Myog, Tnni2, Tnnt3* were down-regulated at Day 4 in the sh*Platr14* knockdown. Undifferentiated C2C12 cell line with empty vector was used as control (Figure 3.11B).

This expression trend of the myogenic markers was further validated with Lipofectamine-mediated transfection with three Dsi RNAs against *Platr14*. After the successful knockdown validation showing about 80% inhibition, the expression levels of the myogenic target genes were validated using RT-qPCR (Figure 3.12A&B). The genes targets showed similar trend as

observed in the sh*Platr14* knockdown. *Cdkna1, Myod, Mef2c, Myog, Tnnt2*, were down-regulated in the early and terminal phases of differentiation of C2C12 cell line.

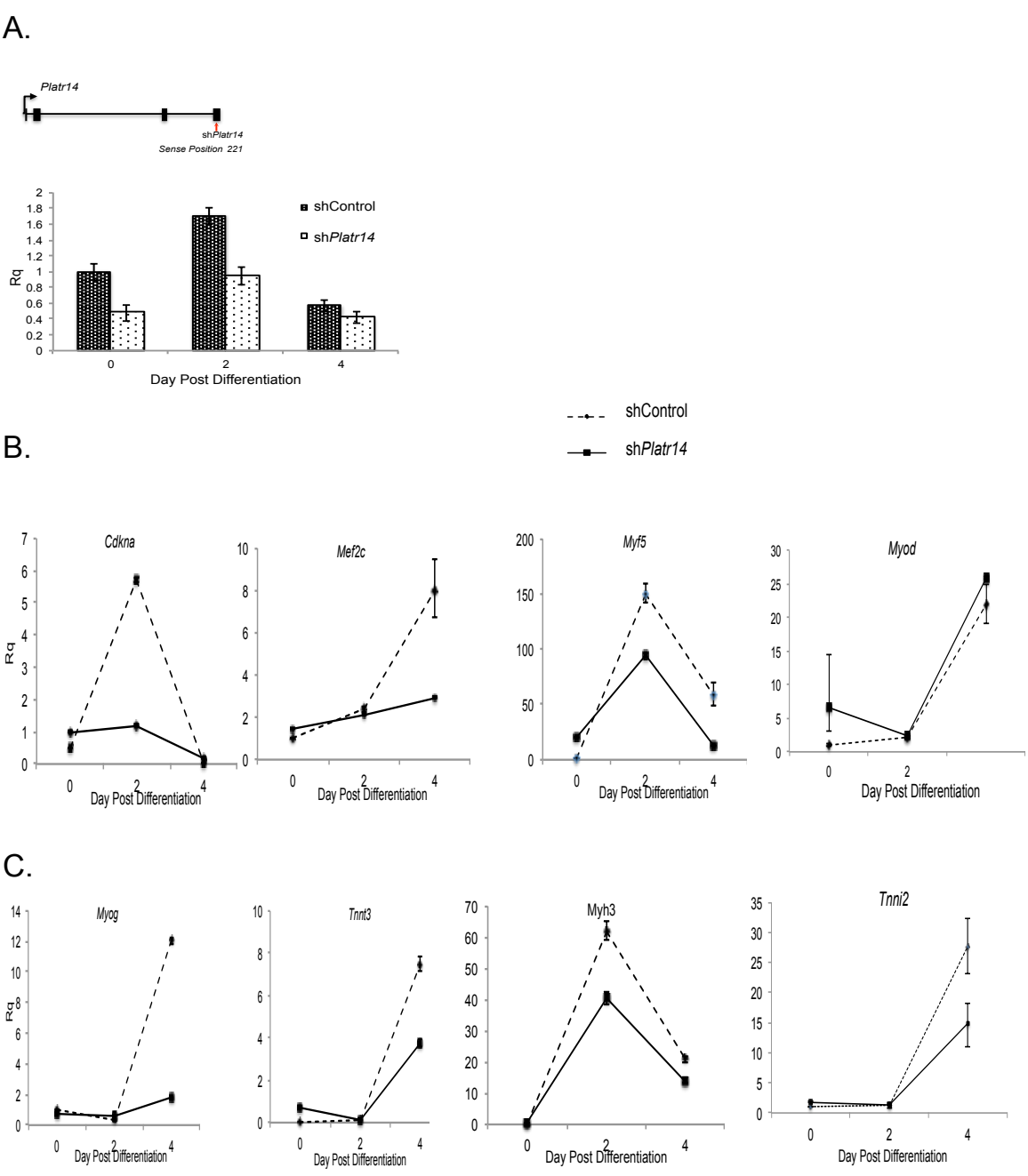
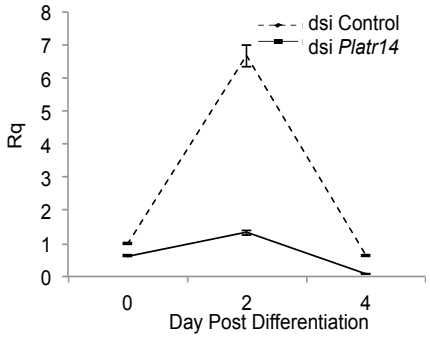


Figure 3.11: *Platr14* is required for myogenic differentiation.

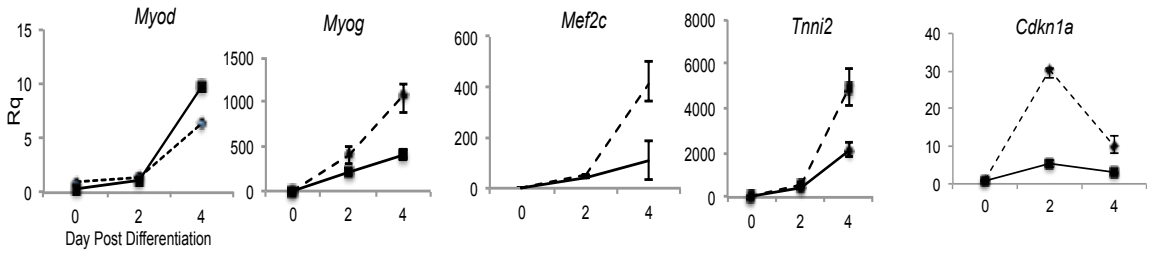
(A). Quantitative PCR for the validation of knockdown of *Platr14* with shRNA(#2) in undifferentiated and differentiated C2C12 cell line.

Quantitative PCR for the expression of (B) early myogenic markers and (C) terminal myogenic markers in sh*Platr14* knockdown in undifferentiated and differentiated C2C12 cell line. Polymerase Chain Reaction - PCR; Relative quotient for expression - Rq. All expressions were normalized to GAPDH. All relative expressions were determined using comparative threshold cycle (ΔΔCt) method. All samples have n=3 biological replicates. Undifferentiated C2C12 myoblasts with empty vector was used as control.

A.



B.



C.

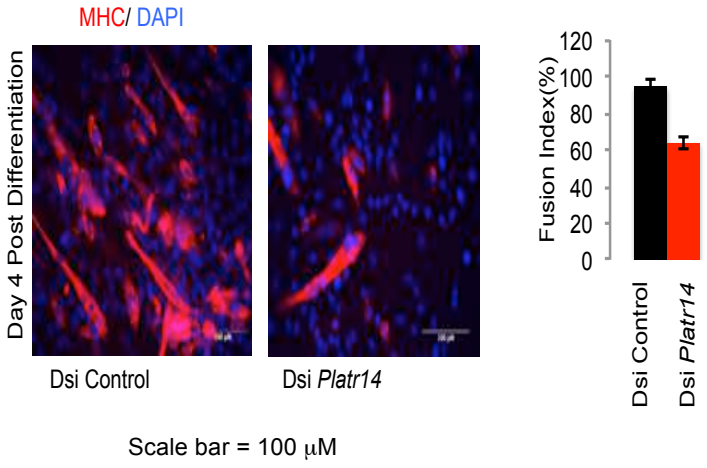


Figure 3.12: Dsi knockdown of *Platr14* affects myogenic differentiation in C2C12 cell line. (A). Quantitative PCR for the validation of knockdown of *Platr14* with 3 Dsi RNA in undifferentiated and differentiated C2C12 cell line. (B). Quantitative PCR for the expression of early myogenic markers

and terminal myogenic markers in Dsi knockdown of *Platr14* in undifferentiated and differentiated C2C12 cell line. Polymerase Chain Reaction - PCR; Relative quotient for expression – Rq. All expressions were normalized to GAPDH. All relative expressions were determined using comparative threshold cycle (ΔΔCt) method. All samples have n=3 biological replicates. Undifferentiated C2C12 myoblasts with scramble Dsi was used as the negative control. (C) Immunocytochemistry staining for myosin heavy chain (MF-20) shows a reduced expression of myosin heavy chain in DsiRNA knockdown differentiated C2C12, day 4 post differentiation.

3.5.2 Immunocytochemistry studies show a decreased expression of Muscle contractile protein upon *Platr14* inhibition:

Immunocytochemistry analysis for the expression of myosin heavy chain (MF-20) in day 4 differentiated cell lines showed that the fusion index was much reduced in the sh*Platr14* knockdown group as compared to the control C2C12 cell line with the empty vector. The Fusion index was calculated from the number of cells with more than 2 nuclei/the total numbers of nuclei in the field of view. The number of myosin heavy chain positive cells also showed a reduction in the *Platr14* knockdown group as

compared to the control (Figure 3.13 A&B). The immunocytochemistry analysis repeated with the Dsi transfection also showed a similar decrease in myosin heavy chain in the *Platr14* knockdown as compared to the C2C12 cell line with scramble Dsi as the negative control. The immunocytochemical analysis performed in Dsi knockdown further confirmed the reduction in the expression of myosin heavy chain protein (Figure 3.12C).

The findings indicate that *Platr14* inhibition affects myogenesis by either slowing or delaying the process of differentiation, as it can be noted that the early myogenic markers catch up with the normal expression trend by day 4 post differentiation.

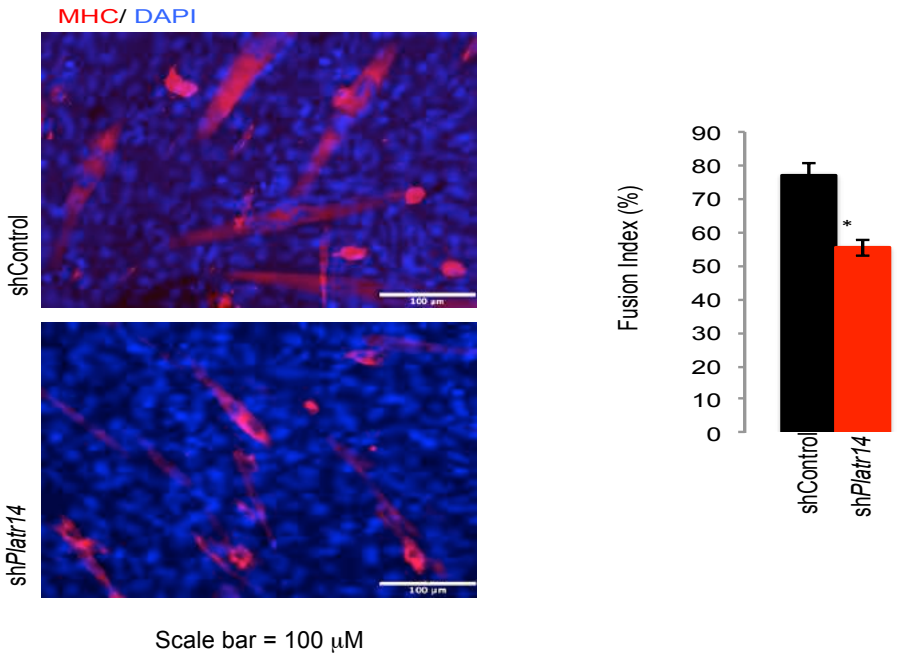


Figure 3.13: Knockdown of *Platr14* shows a decreased expression of myogenic contractile protein, Myosin heavy chain.

(A). Immunocytochemistry staining with MF-20 shows a reduced expression of myosin heavy chain in day 4 post differentiated sh*Platr14* knockdown C21C2 myotubes. (B). A reduction in the fusion index percentage in Platr14 knockdown C21C2 myotubes. P<0.05. Fusion index was calculated by the total number of myotubes with more than 2 nuclei over the total number of nuclei (stained blue=DAPI) in the field area. Scale bar: 100 µm; Fusion Index: P<0.005 (*).

3.5.3 Overexpression of *Platr14* does not accelerate myogenesis:

A gain of function assay by doxycycline induced overexpression of *Platr14* (Figure 3.14A) in C2C12 myoblasts showed an increase in the expression of early myogenic markers such as *Myod* and *Cdkna1* at Day 2 post-differentiation (Figure 3.14B). The terminal markers such as *Myog, Myh3, Tnnt3* (Figure 3.14C) showed an elevated expression at post differentiation day 4. Suprisingly, immunofluorescence staining with MF-20 for myosin heavy chain did not show any significant change in the rate of fusion when compared to the C2C12 cell line control (data not included). This points out that the overexpression of the *Platr14* transcripts does not enhance or quicken differentiation process. Interestingly we also noted that if *Platr14* is induced to express at terminal stages of differentiation such as post differentiation day 4 or day 6, it strongly down-regulates the expression of terminal markers (Figure 3.14 D, E&F). These findings indicate that *Platr14* may have an inhibitory role in terminal differentiation markers and its significance as early pro-myogenic marker similar to other myogenic lncRNAs identified such as *lncMD-1* and *lncDum* (Cesana, Marcella *et a*l., 2011 and Wang, L., 2015).

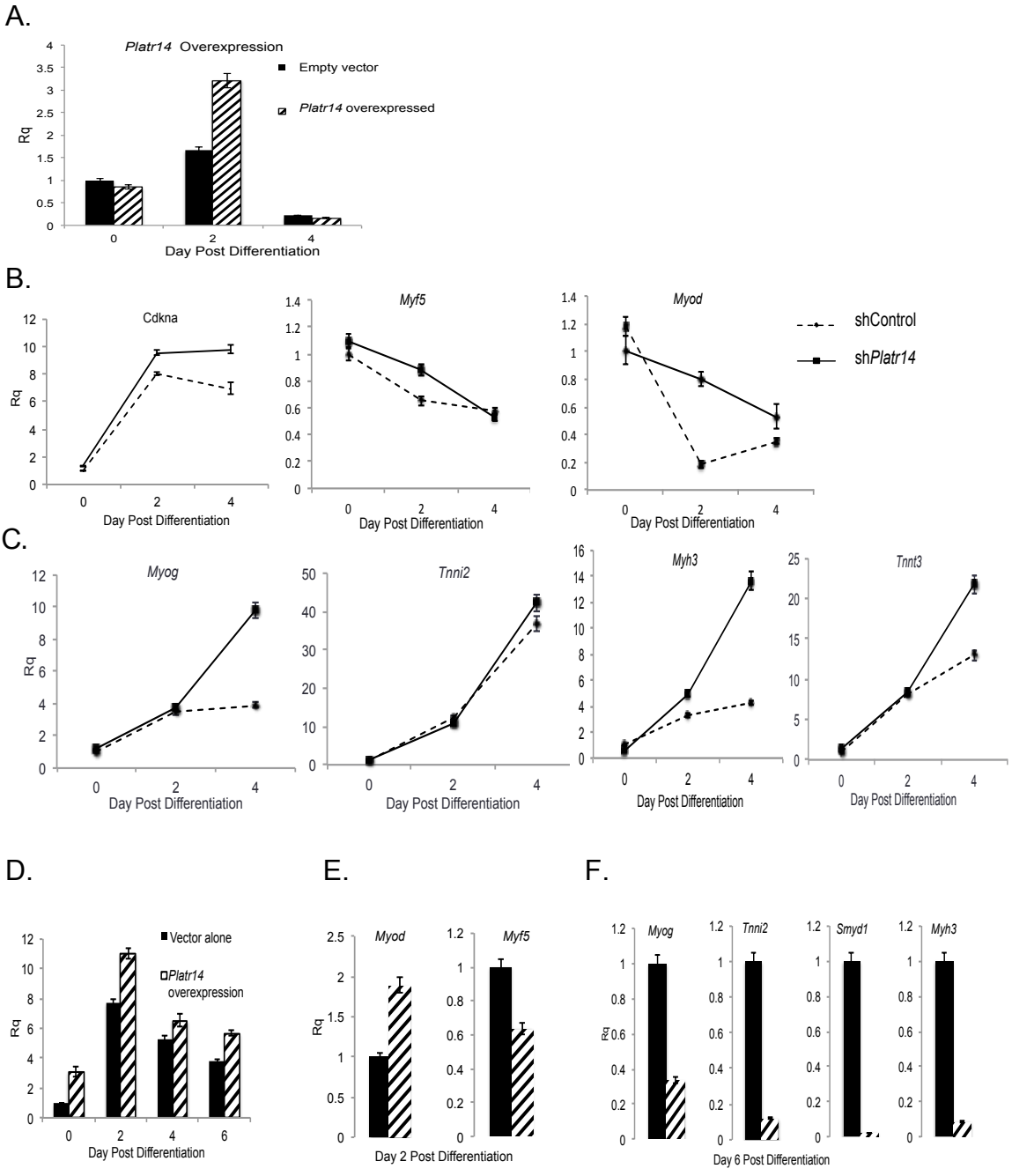


Figure 3.14: Overexpression of *Platr14* does not accelerate myogenesis. (A). Validation of doxycycline induced overexpression of *Platr14* in differentiating C21C2 cell lines at post differentiation day 2.

(B&C). Quantitative PCR for the expression of early and terminal myogenic markers in the *Platr14* overexpressed C2C12 cell lines. Polymerase Chain Reaction - PCR; Relative quotient for expression - Rq. (D). Validation of doxycycline induced overexpression of *Platr14* in differentiating C21C2 cell lines at time points post differentiation day 0,2,4,and 6. (E)&(F). Quantitative PCR for the expression of early and terminal myogenic markers in the *Platr14* overexpressed C2C12 cell lines. All expressions were normalized to GAPDH. All relative expressions were determined using comparative threshold cycle (ΔΔCt) method. All samples have n=3 biological replicates. Undifferentiated C2C12 myoblasts with empty vector was used as the negative control.

3.5.4 The global effect of *Platr14* on mesoderm related lineage processes:

To further study the global gene expression patterns during C2C12 differentiation understand the fusion of *Platr14* we performed next-generation sequencing, to profile the transcriptome of *Platr14* knockdown differentiated C2C12. We compared the post-differentiated day 2 sh*Platr14* knockdown cell line with reference to empty vector. The genes that were expressed at FPKM lesser that 0.5 were classified as down-regulated while

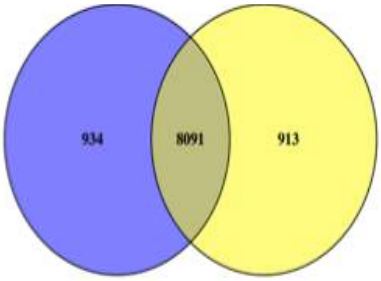
those genes expressed at FPKM greater than 2.0 were classified as up-regulated. We found ~4659 genes were uniquely down-regulated and ~ 913 genes were uniquely up-regulated in the sh*Platr14* knockdown (Figure 3.15A). From the Gene Ontology (GO) (Figure 3.15B) generated it was observed that the enriched terms for genes down-regulated related to processes that are involved in regulation of developmental processes such as cell proliferation, migration, adhesion, metabolic activities, cellular calcium hemostasis and cell fusion as well as formation of contractile proteins. About 50% of the genes down-regulated were associated with biological terms such as cardiac muscle formation, somitogenesis, organelle organization and assembly, cell migration and locomotion, cell-cell signaling and mesenchyme formation.

This demonstrates that *Platr14* may have a direct or an indirect role in driving the undifferentiated myoblast to the differentiated state and can be considered as a pro-myogenic factor.

A.

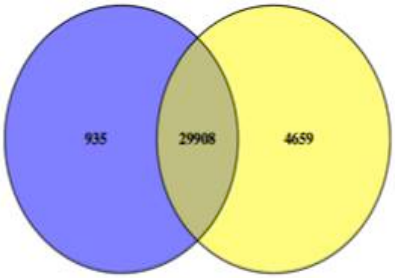
Up regulated gene expression

sh Control sh *Platr14*



Down regulated gene expression

sh Control sh *Platr14*



B. Biological Processes:

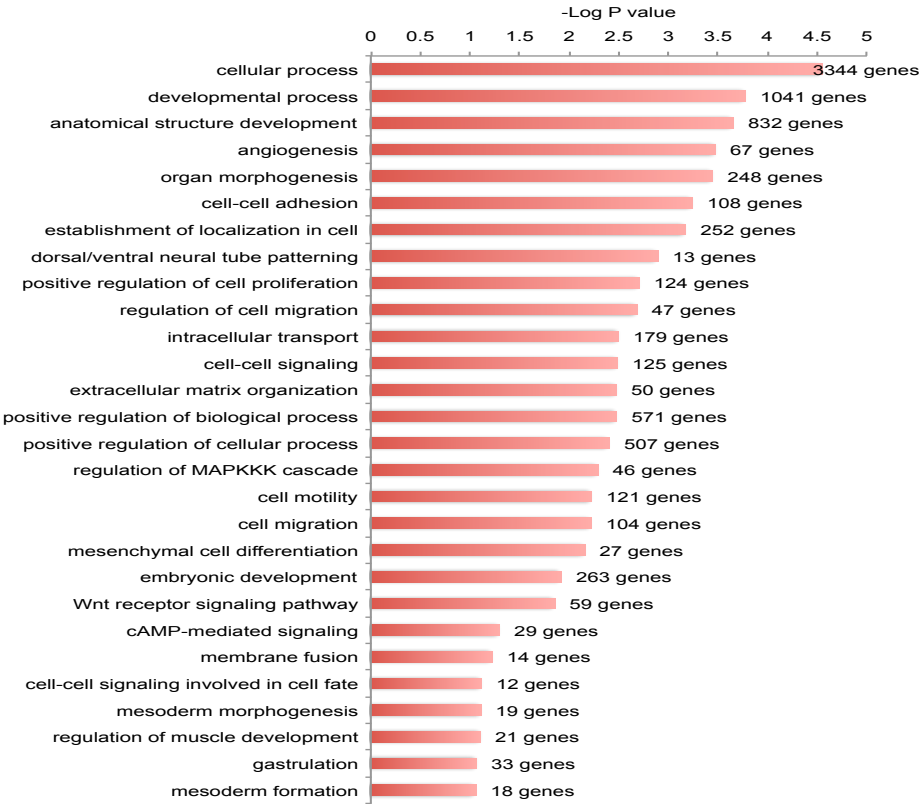


Figure 3.15: *Platr14* is important regulator for the expression of genes associated with the early stages of mesoderm lineage specification.

(A). Venn diagrams representing overlapping and distinct up- or down-regulated genes at day2 post differentiation between C2C12 control group and sh*Platr14* knockdown group. (B). GO terms associated with the down regulated genes in sh*Platr14* knockdown cell line.

3.5.5 Knockdown of *Platr14* affects cell proliferation:

Cell proliferation and differentiation show a remarkable inverse relationship. The proliferating C2C12 cells should exit cell cycle to enter into differentiation (S Ruijtenberg, 2016). Therefore, BrdU assay was performed in replicates to assess the rate of proliferation in the both the sh*Platr14* knockdown. The BrdU-positive cells were quantified by the counting the number of BrdU-positive cells with DAPI positive cells across three random fields. Knockdown sh*Platr14* C2C12 cell line showed an increase in the incorporation of BrdU as compared to the C2C12 cell line with empty vector as control indicating a failure of *Platr14* knockdown C21C2 cell lines to completely exit out of cell cycle to initiate the differentiation process (Figure 3.16 A&B).

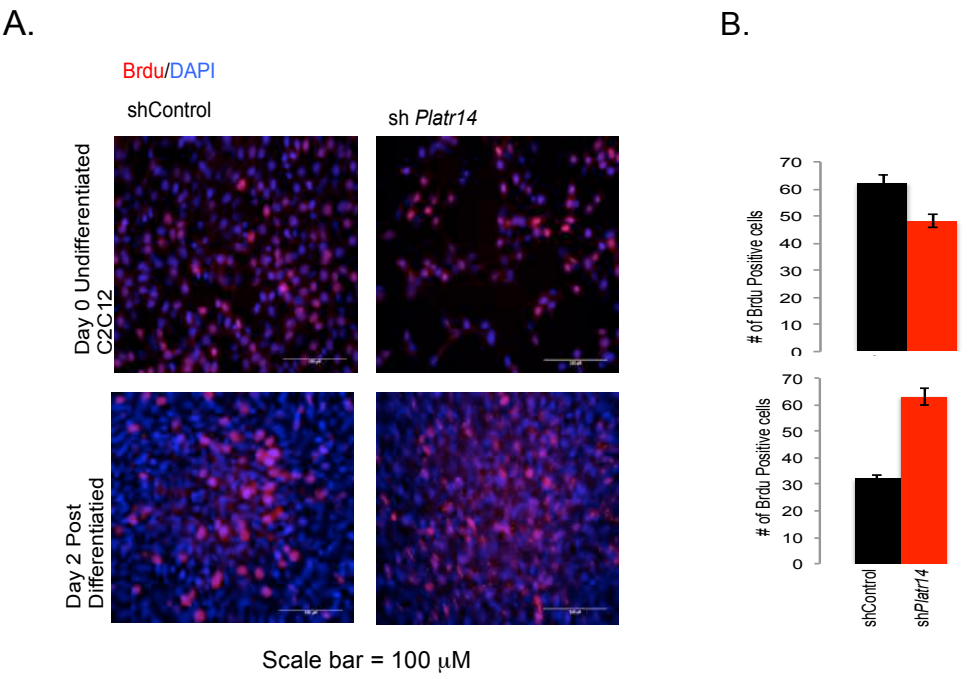


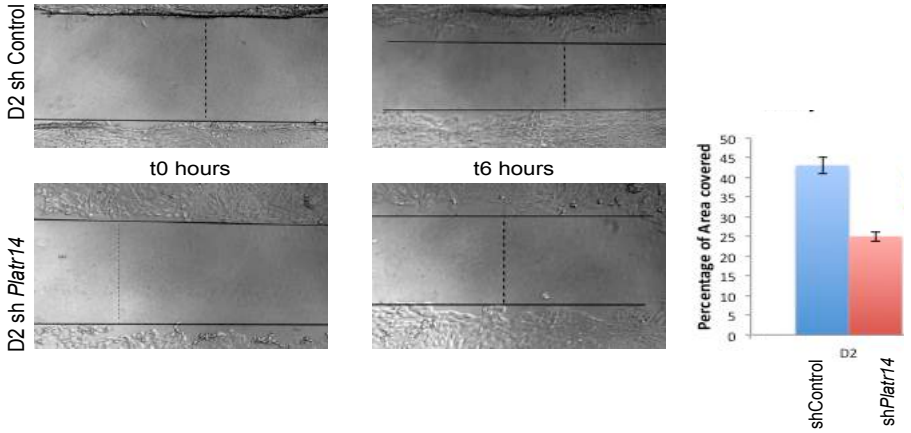
Figure 3.16: *Platr14* knockdown affects cell proliferation. (A & B). BrdU assay showing a reduction in the number of proliferating cells in sh*Platr14* undifferentiated (day0) C2C12 myoblasts and an increase in BrdU incorporation in sh*Platr14* D2 differentiated cells.

3.5.6 Knockdown of *Platr14* affects cell migration:

Myoblast migration is a significant factor in the process of myogenesis and regeneration. It is required for myoblast alignment and fusion to form myotubes (Louis M *et al.*). The two C2C12 cell lines, the knockdown sh*Platr14* and the control with empty vector were allowed to differentiate. A was scratch made along the center of the plate and the cells were allowed

to migrate for 3 and 7 hours. The knockdown sh*Platr14* knockdown showed only a migration rate of 24% in wound coverage after 3 and 6 hours as compared to the migration rate of 43% in the control C2C12 cell line harboring the empty vector pLL3.7 (Figure 3.17A). This shows that the *Platr14* may be required for cell migration. This correlated with the RNA sequencing data in day 2 sh*Platr14* knockdown C2C12 cell line, which showed a down-regulation of genes associated with migration (Figure 3.17B).

A.



B.

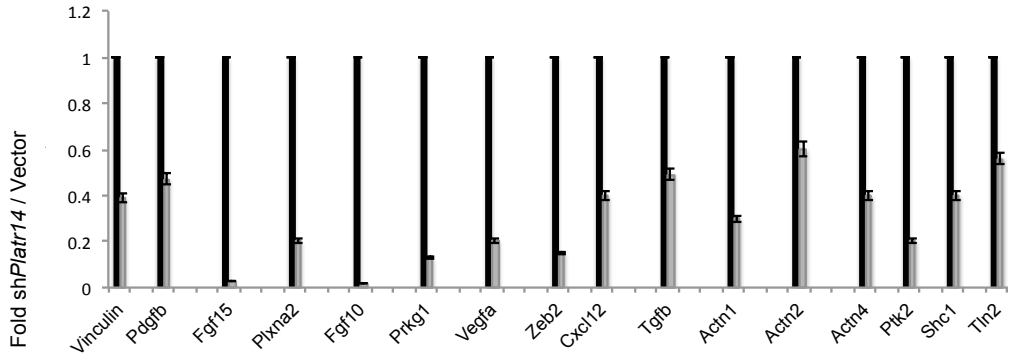


Figure 3.17: *Platr14* may be essential for cell migration. (A). Scratch assay showing the migration rate of C212 cell lines, shPlatr14 knockdown and the control at specific time-points (before and after 6 hours). (B). The genes associated with migration that are down-regulated upon *Platr14* inhibition (from RNA-Sequencing data). The expression is represented as fold change of sh*Platr14*/ Vector.

3.5.7 *Platr14* harbors Repeat Sequences:

LncRNAs are observed to carry repeat sequences or the TE sites within their genome sequence, and these are evolutionarily conserved (R Johnson, 2014). These repeat sequences are found to carry functional domain that can bind to specific protein sequences or nucleotide sequences, both RNA and DNA. *In-silico* analysis revealed repeat sequences at the 3'UTR of *Platr14*. At position 370-500 bp *Platr14* is predicted to have complementarity with about 140 DNA sites in the mouse genome (mm9). This nucleotide stretch on *Platr14* are not conserved across species, while nucleotides from position 510-572 bp are conserved across genomes of chordates such as the mouse, rat, human, orangutan, chimpanzee, guinea pig, and dog. This distinct set of nucleotides may represent the TE sequences randomly interspersed in the lncRNA sequence.

When noted that of the 140 DNA complementary sites, 29 of the targets were differentially expressed upon *Platr14* Knockdown in C2C12 differentiated cells (Figure 3.18). GO analysis of the selected 29 targets showed GO terms associated with positive regulation of developmental processes (Table 3.2).

This indicates a probability of lncRNA association with these predicted DNA sites to epigenetically regulate the transcription of the genes associated with the development.

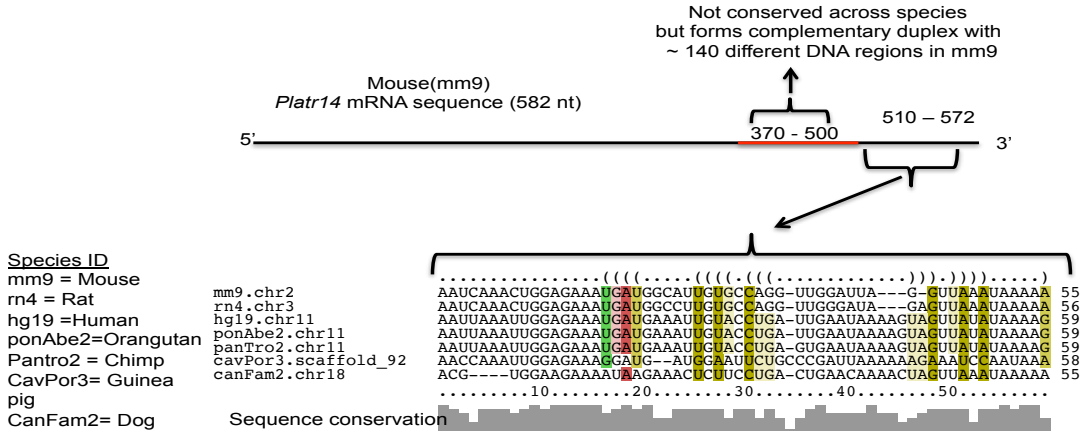


Figure 3.18: *In-silico* analysis of the *Platr14* sequences predicts repeat sequences: *In-silico* prediction of repeat sequences at the 3'UTR of *Platr14.*

Table 3.2: The differentially expressed genes that are predicted to have binding complementarity with the predicted repeat sequences of *Platr14.*

Gene short name	Fold:D2 KD/Vector
Rabgap1l	0.00420167
Vstm4	0.018812552
Ccdc69	0.018923535
Acbd5	0.036039088
Pard3b	0.070348868
Lrrc38	0.086790488
Dcaf8	0.090219933
Oprd1	0.143697775
Syk	0.164651118
Tcf19	0.203678462
Ino80d	0.244346402
Akna	0.286878304
Lypla1	0.370840546
Camk1d	0.404393783
Magi1	0.408476374
Lhx4	0.424406594
Haao	0.513572508
Zc3h8	0.5176092
Dcaf5	0.538493821
Lzts3	0.54920612
Dph1	0.572451663
Atp2a3	0.57623096
Lrrtm2	0.594241621
Slc7a11	0.594241621
Tbc1d22a	0.625822341
Cd8a	0.648561668
Gm5860	0.67662526
Dpp10	0.686628993
Rpl11	0.690657292
Cdc14a	0.698774691

3.5.8 *In vitro* RNA pull down assay shows association of *Platr14* with HnrRNPs and other associated proteins:

Recent findings have demonstrated that lncRNAs have various important regulatory effect on target gene thereby contributing to epigenetic modification, transcription and post-transcriptional processing through specific interactions with proteins and other cellular factors (Lee JT, 2012;

Yoon J, 2013). LncRNA can also act as scaffolds to create discrete protein complexes: lncRNA–RNPs. Under high stringency wash conditions, using *in vitro* RNA pull-down followed by Mass Spectrometry (MS) analysis, we identified that *in vitro*-transcribed biotinylated-*Platr14* sense transcript were associated with proteins linked to developmental process (Table 3.3). Most of the pulled down proteins were found to have been involved in myogenesis. The no RNA control was also associated with some general RNA-binding proteins that were also bound by the beads. When compared with the no RNA control of RNA pull-down experiment, there was a relative abundant association of *Platr14* with heterogeneous nuclear ribonucleoprotein, such as Hnrnpa0, Hnrnpa1, Hnrnpa2b1, Hnrnph1, which have been reported to bind other lncRNAs (K Reich, 2014, X Lan, 2016, M Montes, 2015). It has to be also noted that STRING database generated GO terms for the identified proteins showed its association with pathways involved in cytoskeletal formation, focal adhesion, cell projection indicating its role in myogenic differentiation (Table 3.4). These are preliminary results, and these findings have to be further confirmed with biological replicates. Furthermore, the predicted targets have to be further validated by RNA Immunoprecipitation (RIP) to confirm the lncRNA specific interactions.

Table 3.3 *In vitro* RNA pull down assay shows potential protein-lncRNA interaction.

MS analysis of the proteins pulled out using *in-vitro* RNA pull down.

Rank	log(e)	Description
2	-233.1	Actg1
4	-225.8	Tpm1
6	-186.7	Tpm2
8	-155	Tpm3-rs7
9	-151.9	Tpm3
10	-110	Syncrip (hnRNP Q)
11	-99.3	Actc1
12	-93.6	Actg2
13	-88.5	Coro1c
16	-69.8	Hnrnpa3
17	-64.9	Lima1
20	-54.3	Csda
22	-53.2	Hnrnpd
24	-50.7	Hnrnpab
26	-49.9	Actbl2
27	-41.9	Dhx9
28	-39.4	Rpl12
30	-33.6	Pura
32	-28.5	Fus
34	-23.9	Coro1b
35	-23.1	Tnnt2
36	-20.9	Strap
37	-20.3	Tnni1
40	-18.4	Hnrpdl
41	-16.7	Serbp1
43	-14.7	Gm6180
47	-11	Rbm3
51	-7.2	Gnb2l1
52	-6	Rplp0
53	-5.9	Gm10073
54	-3.9	Eef1a2
55	-3.8	Rbmxrt
56	-2.5	Csde1
57	-2.4	G3bp1
58	-2.3	Tpm4
60	-2.1	Gm10263
64	-2	Flna
61	-2	Hoxd3
62	-2	Tnni2
65	-1.8	Myo1c
66	-1.6	Mrs2
67	-1.5	Hnrnpu
68	-1.4	Nexn
70	-1.3	Zfp64
73	-1.1	Anubl1
71	-1.1	D11Bwg0517e
72	-1.1	Mtap4
74	-1.1	Rab6
75	-1	Syt6

Table 3.4 *In vitro* RNA pull down assay shows potential protein-lncRNA interaction. The STRING database prediction of the GO terms associated with the predicted proteins.

Molecular component (GO):

Pathway ID	Pathway description	Observed gene count	False Discovery Rate
GO.0015629	actin cytoskeleton	13	1.08E-10
GO.0043232	intracellular non-membrane-bounded	24	1.38E-07
GO.0005856	cytoskeleton	18	1.73E-07
GO.0005925	focal adhesion	10	1.73E-07
GO.0030529	ribonucleoprotein complex	12	1.73E-07
GO.0032432	actin filament bundle	6	1.73E-07
GO.0032991	macromolecular complex	26	1.73E-07
GO.0043229	intracellular organelle	38	1.73E-07
GO.0044446	intracellular organelle part	31	2.96E-07
GO.0044430	cytoskeletal part	14	2.14E-06
GO.0001725	stress fiber	5	4.46E-06
GO.0005865	striated muscle thin filament	4	5.14E-06
GO.0042995	cell projection	14	3.58E-05

3.5.9 Identification of endogenous *Platr14*: protein interactions:

Though *in vitro* RNA pull down assay shows lncRNA: protein interaction, the endogenous interaction cannot be established. Therefore, to understand the endogenous protein lncRNA interaction we performed RAP (RNA Antisense Pull down) assay. This allows identifying the RNA associated protein using specific lncRNA antisense probes; followed by

Mass Spectroscopy analysis to identify the proteins associated. 20 nucleotides long antisense probes designed against *Platr14* were able to pull out the endogenous proteins from whole cell lysate. The Mass spectroscopy analysis resulted in protein predictions. Few of the proteins identified such as Zbtb20, Col5a2, and Col4a2 are seen associated with skeletal muscle differentiation (Table 3.4). The experiment needs to be repeated with biological replicates by increasing the sample size to obtain enough starting material for nuclear lysate to validate the protein read outs.

Table 3.4: Identification of endogenous *Platr14*: protein interactions.

RAP-MS shows protein pulled down using RAP probes against *Platr14.*

Protein	Description
Zbtb20	Zinc finger and BTB domain containing 20
Col5a2	Collagen type 5 alpha2
Col4a2	Collagen type 4 alpha2
Dnahc8	Dynein
Anubl1	Zinc Finger, AN1-Type Domain
Actc1	Cardiac Actin
Clcn4-2	Chloride Channel, Voltage-Sensitive 4
Spnb2	Spectrin
Tmem59	Transmembrane Protein 59
Heatr5a	HEAT Repeat Containing 5A
Tmtc4	Transmembrane And Tetratricopeptide Repeat Containing 4
Sass6	SAS6 centriolar assembly protin
Nup133	Nucleoporin
Oas2	Oligoadenylate Synthetase
Zmynd15	Zinc Finger MYND-type containing 15

3.5.10 Outcomes of ChIRP validation to identify Platr14: DNA interactions.

ChIRP-seq is a method using DNA oligonucleotides to capture lncRNAs and their genomic DNA binding sites (Chu *et al.,* 2011). We used probes of 21 nucleotides designed to identify lncRNA-associated proteins in day 2 differentiated Platr14 overexpressed C2C12.Following the protocol of Chu *et al*., 2011, we cross-linked cells extensively with glutaraldehyde, retrieved target RNA with oligonucleotide hybridization. Once we optimized the protocol for *Platr14* pull down, we validated by qRT-PCR the RNA pulled out using the oligonucleotide probes with *LacZ* probes as the control. (Figure 3.19A&B). But we were unable to retrieve sufficient DNA to be used for DNA Sequencing.

A. B.

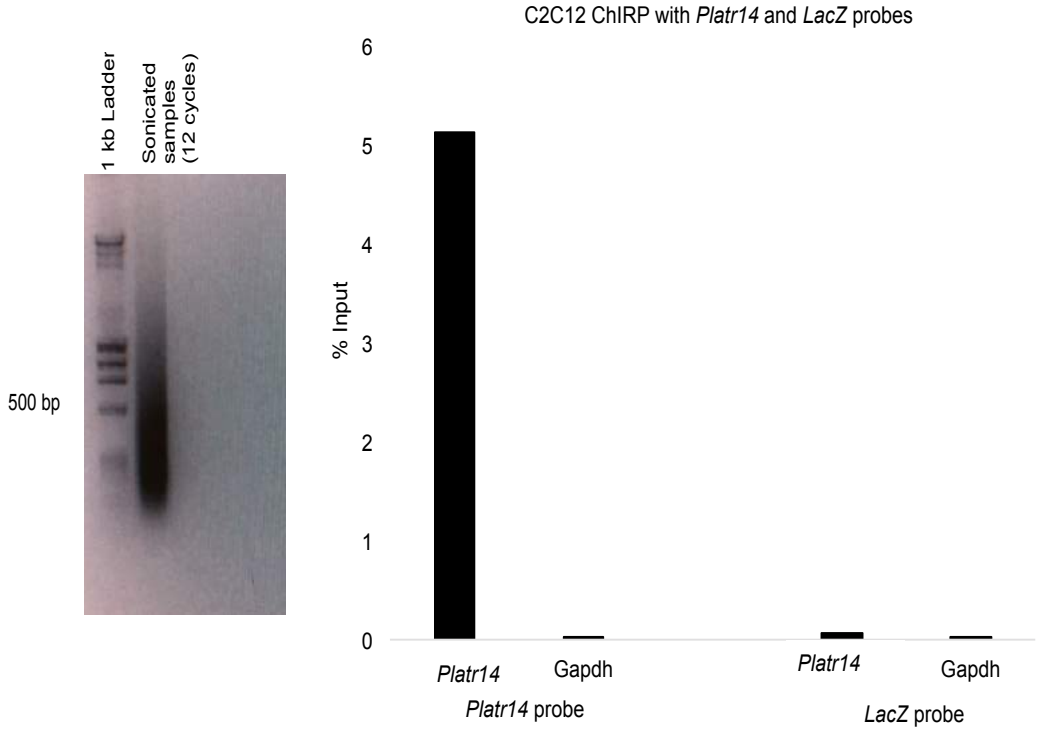


Figure 3.19: Optimization of ChiRP protocol for day 2 differentiated C2C12 cell line. (A). Sonicated samples (12 cycles) of day 2 differentiated *Platr14* overexpressed C2C12 cell line. (B). Optimization of RNA pull down from day 2 differentiated *Platr14* overexpressed C2C12 cell lines. GAPDH was used as Quantitative PCR control. *LacZ* was used as the ChIRP pull down control.

Chapter 4

Discussion

Recent findings of the RNA world have shown that Long non-coding RNAs have diverse roles at all levels of gene regulation from epigenetic mechanisms, nuclear organization to RNA processing and translation. Like microRNAs, long non-coding RNAs are recognized as regulators of embryonic development. As this new layer of gene regulatory network has been defined more lncRNAs have been are seen linked to lineage decisions and differentiation. Various factors control mesoderm specification. But not much understanding has been reached in the role of lncRNAs associated with mesoderm lineage.

In this study, we identified a lineage-specific long non-coding RNA *Platr14.*

From the transcriptome analysis performed using in-vitro ESC differentiation system, UH3 novel lncRNAs expressed in both Mesp1 derived, and non-*Mesp1* derived cell populations were identified. It is understood that lncRNAs are tissue specific and differentiation day specific and *Mesp1* derived cells can differentiate into mesoderm lineage. The subset of mesoderm related lncRNAs we selected for further characterization shows that lncrnas have a distinct expression pattern which indicates its significant role in in the regulation of lineage specification and differentiation.

During the development from fetal to the adult life of the heart, which takes place at postnatal day 7, highly coordinated regulatory networks are required to adapt to the various physiological and pathophysiological changes. Therefore there is a change in regulatory players. The mammalian heart has a remarkable capacity to maintain its regenerative capacity for several days post birth. At the end of postnatal day 7, it quickly looses this capability as the mature cardiomyocytes exit cell cycle. By the end of this, the cardiomyocytes produce binucleated cells. Following which the cell cycle regulators become extremely low in expression and the metabolic genes gets up regulated (Ikenishi A *et al*., 2012). Quantitative PCR studies performed in embryonic and postnatal heart tissues of mouse shows that the expression of the lncrnas studied reduces or shows low expression at postnatal day 7 then up-regulates by adult stage. This may indicate that the lncrnas do not randomly appear at these stages instead may have a direct or indirect regulatory role in the cardiac regulatory network, indicating its lineage specificity.

The ISH performed in embryonic stage 9.5 mouse embryos showed positive expression of only four lncRNAs of the selected subset of 12 lncrnas and this expression correlated with the qRT-PCR analysis. The expression of *9030622O22Rik* and *Platr14* in heart tissues at postnatal

days 3,7 and adult stage shows that these two lncrnas may have a regulatory role throughout the cardiac differentiation and maintenance. *2410006H16Rik* shows a decreased expression in both postnatal and adult heart tissues indicating that this may have ES cell function and may help in the maintenance of ESC characteristics. *4930474H06Rik* did not show a positive expression in the ISH analysis but was enriched in postnatal day 3 and 10 indicating its differentiation stage-specific regulation in the cardiac formation.

Both mesoderm and endoderm lineages are derived from the primitive streak. Murry and Keller, 2008; Kubo *et al.,* 2004; Tada *et al*., 2005 have demonstrated that the mesoderm and endoderm progenitors can co-express *Bra, Foxa2, Gsc,* and *Gata4/6.* Upon *Platr14* knockdown, these mesendoderm markers are up-regulated while the downstream targets of the cardiac lineage such as *Nkx2.5Tbx5, Tbx20, Myf6, Tnnt2* were seen to be down-regulated. *Gata4* is a significant factor in early cardiac differentiation and this factor is regulated directly by *Foxa2* (Rojas, A., 2010). Therefore, the up-regulation of *Gata4* upon the *Platr14* inhibition could be due to increase expression of *Foxa2*. Unlike *Gata4, Gata6* is down regulated since *Nkx2.5* is a direct target of *Gata6* in the early cardiac formation (Molkentin *et al*., 2000). Moreover, the immunocytochemical

analysis revealed that the structural gene such as cardiac Troponin was not activated properly in the sh*Platr14* EBs compared to the controls.

These data indicate that *Platr14* is critical for promoting early gene regulatory network in cardiac development.

***Platr14* in Skeletal Muscle Formation:**

Skeletal muscle is a mesoderm-derived organ system. The mesoderm related lncRNAs screened for the expression in C2C12 cell system showed expression of 4 lncrnas in the differentiation process of C2C12 with an up-regulation at distinct time points. Which correlated with the d.e 9.5 ISH studies. Platr14 showed a distinct expression in cardiac as well as skeletal muscle differentiation. Unlike the up-regulation of *Platr14* in ESC state in ES cells, it was up regulated in day 2, the time point of marked induction of differentiation.

Platr14 knockdown in undifferentiated and differentiated C2C12 showed a deregulation of early myogenic markers such as *Cdkna1, Myod*, and *Myf5*.The exit from cell proliferation is an important stage for the initiation of differentiation. Cell cycle markers such as *Cdkna1* and *Cdknb1* are down regulated by day 2 of differentiation (M Kitzmann, 2001). *Platr14* inhibition showed an increase in *Cdkna1* indicating a failure to exit the cell

cycle. BrdU analysis also indicated an increase in proliferating cells in *Platr14* knockdown as compared to the control cells. Migration is the second step towards myogenesis. It can be noted that upon the knockdown of *Platr14* many genes associated with migration were down-regulated.

The muscle-specific transcription that is suppressed by the polycomb group (PcG) complex is driven by the coordinated effort of *Myod* and Myogenin(*Myog*) (Braun T., 2011). *Myod* is up regulated in the undifferentiated stage as well as upon induction of differentiation followed by down regulated in the terminal stages. But upon *Platr14* knockdown, *Myod* is down regulated indicating an insufficient activation of the myogenic pathway. An important marker of terminal myogenesis, *Myog* was down- regulated at day 4 of C2C12 differentiation. Also the immunocytochemical staining of myosin heavy chain was reduced in *Platr14* inhibited C2C12 cell line as compared to the control. The reduced expression of contractile proteins such as *Tnni2, Tnnt3, and Myh3* in the later stages of differentiation indicates the regulatory role of *Platr14* in the myogenic pathway. Immunostaining of myosin heavy chain showed a reduction in the rate of fusion rate in *Platr14* knockdown day 4 post differentiated C2C12 cells. This indicates that *Platr14* plays a crucial role in

the transition of differentiating myoblasts to differentiated myotubes and a disruption in its expression can either delay or inhibit the formation of primary myofibers.

Knockdown of *Platr14* is also shown to affect the metabolic genes that are involved in cAMP biosynthetic process, ATP synthesis and hydrolysis, fatty acid beta-oxidation, monocarboxylic acid biosynthesis, lipid transport. As pointed out in the literature these are the fundamental metabolic processes involved in driving myogenesis to terminal specification. It has to be further studied to understand if *Platr14* directly affects any of the metabolic genes to regulate the MRFs or vice versa.

The overexpression of *Platr14* is not sufficient for the acceleration of myogenesis. Though overexpression of *Platr14* up-regulated the expression of few of the myogenic markers, there was no significant change in the rate of fusion index in the *Platr14* overexpressed C2C12 cell line. Strikingly, the induced expression of *Platr14* in the later stages of differentiation inhibits the expression of terminal markers indicating a repressor role of *Platr14* in terminal myogenesis. It cannot also be ruled out that *Platr14* may share multiple partners in regulation. This clearly shows the critical role of *Platr14* is a novel lincRNA associated with early stages of skeletal muscle differentiation and would be essential for the

transition of undifferentiated myoblasts to differentiated myocytes.

Analysis of sequence conservation of *Platr14* across species identified homologous conserved sites at the 3'UTR of *Platr14*.The homology of the nucleotides at the 3'UTR of *Platr14* indicates the epigenetic role of *Platr14* in gene regulation. The gene ontology of these predicted *Platr14* binding partners is associated with processes involved in early development stages.

Though further validation is needed for the *Platr14*: protein interactions study; it was seen that *Platr14* could also associate with proteins involved with different biological process involved in differentiation and early development.

Though Bergmann et al. earlier reported *Platr14* to be involved only with pluripotency features; in conclusion, we can say that lincRNA *Platr14* is found to be necessary for driving the mesoderm-derived cardiac and skeletal muscle lineages. *Platr14* has all the hallmarks of a functional lncRNA. 1) *Platr14* is specifically expressed in embryonic cardiac tissues and developing skeletal muscle. 2) There is a transcriptional regulation of Platr14 that is demonstrated by its developmental stage specific expression.3) *Platr14* shows at least some degree of conservation among

species; the conserved sequences being the repeat sequences.4) a deregulation of *Platr14* expression levels results may affect proper development of the tissue.

Therefore directly or indirectly, *Platr14* may act as a co-regulator in the transcriptional regulation of many factors involved in lineage specification during the developmental processes (Figure 4.1).

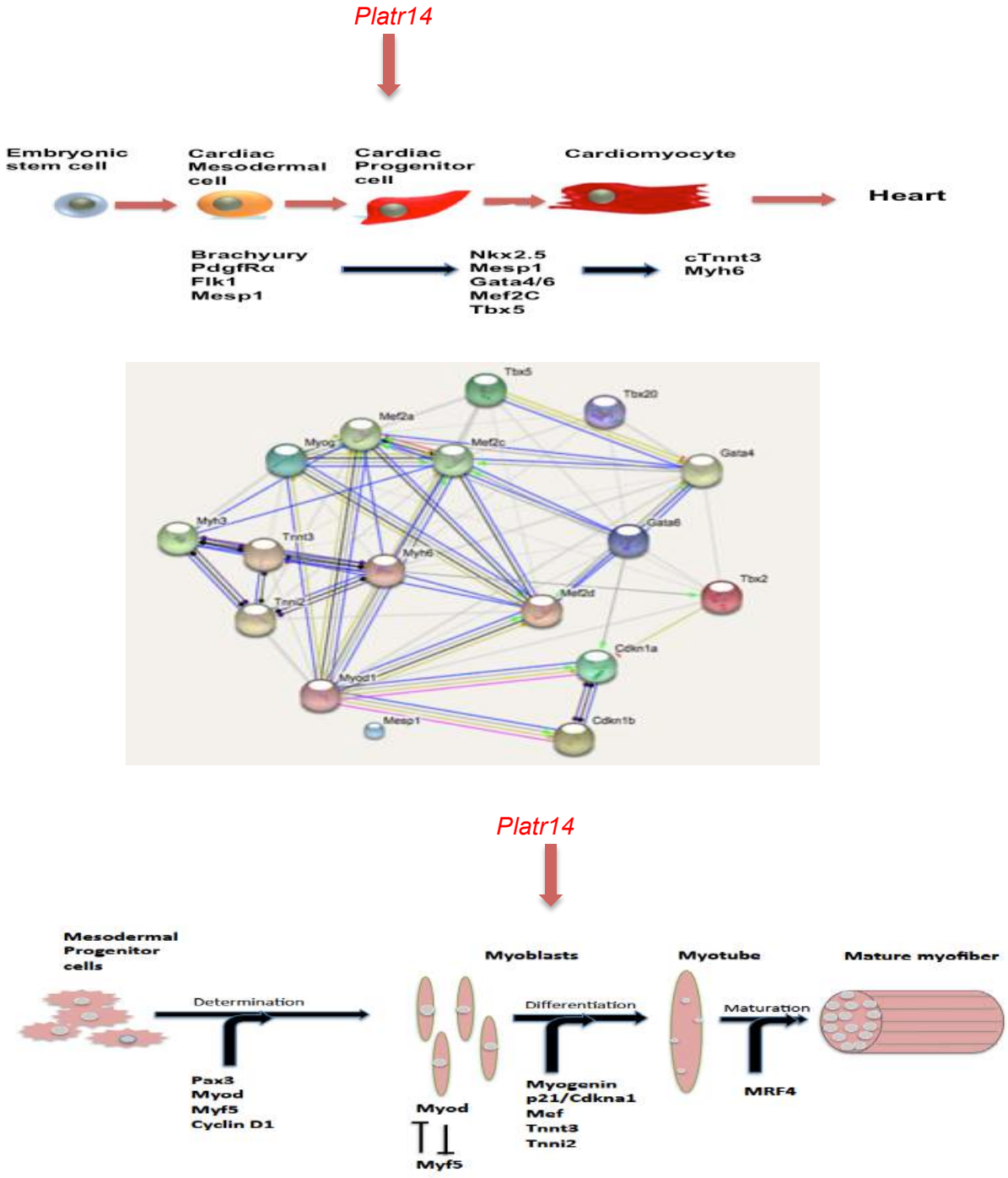


Figure 4.1 Schematic representation of the predicted role of *Platr14* in Cardiac and Skeletal muscle differentiation.

Chapter 5

Concluding remarks, Future Directions and References

5.1 Summary:

In this study, we identified mesoderm related long non-coding RNA associated with cardiac and skeletal muscle differentiation. The enrichment of Platr14 in cardiac and skeletal muscle led us to investigate further its role in early differentiation. Knockdown of *Platr14* in ESCs as well as mesenchymal cell line, C2C12 indicated that this lincRNA is required for the driving the differentiation during cardiogenesis as well as myogenesis. RNA-Seq of the sh*Platr14* showed a significant number of genes were down-regulated that are involved in early stages of myoblast differentiation such cell proliferation and migration. lncRNA: protein interactions studies identified few protein candidates that could be lncRNA partners in driving the myogenic differentiation. We also identified repeat sequences in *Platr14* that are found to be conserved across species and unique DNA binding sites on the mouse genome indicating that *Platr14* can have the ability to bring out a change in the chromosome architecture thereby controlling epigenetic regulation.

5.2 Future Directions:

The lncRNA-protein interactions experiments such as the in vitro RNA pull down, and the RAP MS studies generated few predicted candidate proteins. These findings have to be further validated with RNA Immuno

precipitation (RIP) studies, to confirm the interactions with *Platr14* and the predicted proteins.

The identification of repeat sequences at the 3'UTR of *Platr14* and its homology with other DNA binding sites in the mouse genome predicts probable interactions with these sites. Therefore it can be presumed that this could be accomplished by a conformational change that can occur in the chromatin architecture during the epigenetic regulation of developmental genes. Further experiments such as FISH (Fluorescent In Situ Hybridization) will enable us to validate the complementary DNA sequences with *Platr14*. Advanced techniques such as Assay for Transposase Accessible Chromatin with high-throughput sequencing (ATAC-seq) (Buenrostro et al., 2013, 2015) are needed to identify these interactions to confirm the *in-silico* findings. This assay uses hyperactive Tn5 transposase that can trigger chromatin to cut and ligate adapters and are further processed through high-throughput sequencing in regions that have high accessibility. Thus, it will help in providing a clear picture of the regulatory role of *Platr14* in development.

Therefore the functional role along with the molecular mechanisms will enable Platr14 to be considered as a therapeutic agent in conditions such as cardiac and skeletal muscle dysfunctions.

Significant breakthroughs continue to be achieved in various cardiovascular and skeletal muscular dysfunctions. But not much effort has been given to understanding the gene regulatory mechanism by long non-coding RNA. A clear understanding of the molecular mechanisms involved in lncRNA transcriptome regulation will enable us to exploit these noncoding RNAs as therapeutic tools.

5.2 Conclusion:

The advancement of technologies such as whole genome and transcriptome studies exhibits that protein expression is a highly sophisticated and dynamic process. Emerging evidence points out the significance of long non-coding RNA as regulators at different levels of gene expression governing differentiation and signaling during development and disease conditions. Therefore the identification of *Platr14* in mesoderm lineage specification opens new doors to investigate further its mechanistic role in regulation and to pave the way to clinical application in the future.

5.3 References:

- Aulehla, A., & Pourquié, O. (2010). Signaling Gradients during Paraxial Mesoderm Development. Cold Spring Harbor Perspectives in Biology, **2**(2).

- Bentzinger, C. F., Wang, Y. X., & Rudnicki, M. A. (2012). Building Muscle: Molecular Regulation of Myogenesis. Cold Spring Harbor Perspectives in Biology, **4**(2).

- Bhartiya D Kapoor S Jalali S,*et al*. (2012), Conceptual approaches for lncRNA drug discovery and future strategies. Expert Opin Drug Discov; **7**: 503-13

- Blackwell BJ, Lopez MF, Wang J, Krastins B, Sarracino D, Tollervey JR, Dobke M, Jordan IK, Lunyak VV 2012. Protein interactions with piALU RNA indicates putative participation of retroRNA in the cell cycle, DNA repair and chromatin assembly. Mob Genet Elements **2**: 26–35.

- Britten RJ, Davidson EH 1971. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. Q Rev Biol **46**: 111–138.

- Buske FA, Bauer DC, Mattick JS, Bailey TL 2012. Triplexator: detecting nucleic acid triple helices in genomic and transcriptomic data. Genome Res **22**: 1372–1381.

- C. Niro, J. Demignon, S. Vincent *et al.,* (2010).Six1 and Six4 gene expression is necessary to activate the fast-type muscle gene program in the mouse primary myotome. Developmental Biology **338**:168–182.

- Cabianca, Daphne S. *et al*. 4 (2012). A Long ncRNA Links Copy Number Variation to a Polycomb/Trithorax Epigenetic Switch in FSHD Muscular Dystrophy.Cell **149**: 819–831.

- Campeau E, Ruhl VE, Rodier F, Smith CL, Rahmberg BL, Fuss JO, Campisi J, Yaswen P, Cooper PK, Kaufman PD (2009). A versatile viral system for expression and depletion of proteins in mammalian cells. PLoS One **4**(8): e6529.

- Caretti G, Schiltz RL, Dilworth FJ, Di Padova M, Zhao P, Ogryzko V, Fuller-Pace FV, Hoffman EP, Tapscott SJ, Sartorelli V (2006). The RNA helicases p68/p72 and the noncoding RNA SRA are coregulators of MyoD and skeletal muscle differentiation.Dev Cell **11**: 547-560.

- Carpenter S, Aiello D, Atianand MK, Ricci EP, Gandhi P, Hall LL, Byron M, Monks B, Henry-Bezy M, Lawrence JB, et al. (2013). A long noncoding RNA mediates both activation and repression of immune response genes. Science 341: 789–792.

- Cesana M, Cacchiarelli D, Legnini I, Santini T, Sthandier O, Chinappi M, Tramontano A, Bozzoni I (2011). A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. Cell, **147**: 358-369.

- Chang HY (2011): A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature **472**:120–124.

- Cheng L, Alvares LE, Ahmed MU, El-Hanfy AS, Dietrich S (2004). The epaxial-hypaxial subdivision of the avian somite. Dev Biol. **274**(2): 348–369.

- Cole Trapnell,Adam Roberts,Loyal Goff,Geo Pertea, Daehwan Kim,David R Kelley,Harold Pimentel, Steven L Salzberg,John L RinnLior Pachter (2012).Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.Nature Protocols **7**, 562–578.

- Costello I, Pimeisl IM, Drager S, Bikoff EK, Robertson EJ, Arnold SJ. (2011).The T-box transcription factor Eomesodermin acts upstream of Mesp1 to specify cardiac mesoderm during mouse gastrulation. Nat Cell Biol **13**:1084–1091.

- Daehwan Kim, Geo Pertea, Cole Trapnell, Harold Pimentel, Ryan Kelley, Steven L. Salzberg (2013).TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology Volume 14 (4): 1.

- David R, Jarsch VB, Schwarz F, Nathan P, Gegg M, Lickert H, Franz WM. (2011).Induction of MesP1 by Brachyury(T) generates the common multipotent cardiovascular stem cell. Cardiovasc Res. **92**:115–122.

- Diana Eng, Hsiao-Yen Ma, Michael K. Gross, and Chrissa Kioussi (2013).Gene Networks during Skeletal Myogenesis, ISRN Developmental Biology **2013** (348704).

- E. Berardi, D. Annibali, M. Cassano, S. Crippa, and M. Sampaolesi, 2014. “Molecular and cell-based therapies for muscle degenerations: a road under construction,” Frontiers in Physiology, **5**, article 119,

- E. S. Lander, L. M. Linton, B. Birren, et al., (2001)“Initial sequencing and analysis of the human genome,” Nature **409**:860–921.

- Engreitz, J. M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., … Guttman, M. (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X-chromosome. Science **341**(6147), 1237973.

- Eun B, Sampley ML, Good AL, Gebert CM, Pfeifer K. (2013): Promoter cross-talk via a shared enhancer explains paternally biased expression of Nctc1 at the Igf2/H19/Nctc1 imprinted locus. Nucleic Acids Res, **41**: 817-826.

- F. Relaix, D. Rocancourt, A. Mansouri, and M. Buckingham (2005) A Pax3/Pax7-dependent population of skeletal muscle progenitor cells. Nature, **435**: 948–953.

- Feng J, Bi C, Clark BS, Mady R, Shah P, Kohtz JD (2006): The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultra conserved region and functions as a Dlx-2 transcriptional co-activator. Genes Dev **20**:1470–1484.

- Feng, Y., Hu, X., Zhang, Y., Zhang, D., Li, C., & Zhang, L. (2014). Methods for the Study of Long Noncoding RNA in Cancer Cell Signaling. Methods in Molecular Biology (Clifton, N.J.), **1165**:115–143.

- Frank, S., Aguirre, A., Hescheler, J., & Kurian, L. (2016). A lncRNA Perspective into (Re)Building the Heart. Frontiers in Cell and Developmental Biology **4** :128.

- Gong, C. G. *et al*. (2015). A long non-coding RNA, LncMyoD, regulates skeletal muscle differentiation by blocking IMP2-mediated mRNA translation. Dev. Cell **34**:181–191

- Gong, C., & Maquat, L. E. (2011). lncRNAs transactivate Staufen1-mediated mRNA decay by duplexing with 3'UTRs via Alu elements. Nature, **470**(7333), 284–288.

- Goodier JL, Cheung LE, Kazazian HH Jr (2013). Mapping the LINE1 ORF1 protein interactome reveals associated inhibitors of human retrotransposition. Nucleic Acids Res **41**: 7401–7419.

- Grote P, Wittler L, Hendrix D, Koch F, Wahrisch S, Beisaw A, Macura K, Bläss G, Kellis M, Werber M, *et al*. 2013. The tissue-

specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. Dev Cell **24**: 206–214.

- Gutschner T Diederichs S. 2012. The hallmarks of cancer: a long non-coding RNA point of view. RNA Biol **9** : 703 – 19.

- Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L, *et al*. (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature **477**: 295–300.

- Ikenishi A., Okayama H., Iwamoto N., Yoshitome S., Tane S., Nakamura K., Obayashi T., Hayashi T., Takeuchi T (2012).Cell cycle regulation in mouse heart during embryonic and postnatal stages. Dev. Growth Differ. **54**, 731–738.

- Jeffery D. Molkentina, Christopher Antosb, Brian Mercerb, Tyler Taigena, Joseph M. Mianoc, Eric N. Olson (2000). Direct Activation of a GATA6 Cardiac Enhancer by Nkx2.5: Evidence for a Reinforcing Regulatory Network of Nkx2.5 and GATA Transcription Factors in the Developing Heart, Developmental Biology **217(2**): 301–309.

- Jihong Chen and Qiao Li (2013). Enhancing myogenic differentiation of pluripotent stem cells with small molecule inducers, Cell & Cell & Bioscience **3**:40

- Johnson R, Guigó R (2014).The RIDL hypothesis: transposable elements as functional domains of long noncoding RNAs. RNA **20**(7):959-976.

- Kallen AN, Zhou XB, Xu J, Qiao C, Ma J, Yan L, Lu L, Liu C, Yi JS, Zhang H, Min W, Bennett AM, Gregory RI, Ding Y, Huang Y (2013).The imprinted H19 lncRNA antagonizes let-7 microRNAs. Mol Cell **52**: 101-112.

- L.Kassar-Duchossoy,E.Giacone,B.Gayraud-Morel,A.Jory,D. Gome`s, and S. Tajbakhsh,(2005).Pax3/Pax7 mark a novel population of primitive myogenic cells during development. Genes and Development **19**:1426–1431.

- Li X Wu Z Fu X ,*et al*.(2014). lncRNAs: insights into their function and mechanics in underlying disorders . Mutat Res/Rev Mutat Res; **762** : 1 – 21.

- Lopez-Sanchez C, Garcia-Martinez V, Schoenwolf GC (2001).

Localization of cells of the prospective neural plate, heart and somites within the primitive streak and epiblast of avian embryos at intermediate primitive- streak stages. Cells Tissues Organs **169**:334–346.

- Lu L, Sun K, Chen X, Zhao Y, Wang L, Zhou L, Sun H, Wang H (2013).Genome-wide survey by ChIP-seq reveals YY1 regulation of lincRNAs in skeletal myogenesis. Embo J, **32**: 2575-2588.

- M. K. Gross, L. Moran-Rivard, T. Velasquez, M. N. Nakatsu, K. Jagla, and M. Goulding (2000).Lbx1 is required for muscle precursor migration along a lateral pathway into the limb. Development **127**(2): 413–424.

- M. Lagha, T. Sato, L. Bajard *et al.*(2008).Regulation of skeletal muscle stem cell behavior by Pax3 and Pax7. Cold Spring Harbor Symposia on Quantitative Biology **73**: 307–315.

- M. Murphy and G. Kardon (2011). Origin of vertebrate limb muscle: the role of progenitor and myoblast populations.Current Topics in Developmental Biology **96**:1–32.

- Maamar H, Cabili MN, Rinn J, Raj A (2013). linc-HOXA1 is a

noncoding RNA that represses Hoxa1 transcription in cis.Genes Dev **27**:1260–1271.

- Mackowiak SD, Gregersen LH, Munschauer M, Loewer A, Ziebold U, Landthaler M, Kocks C, le Noble F, Rajewsky N (2013).Circular RNAs are a large class of animal RNAs with regulatory potency. Nature, **495**:333–338.

- Magali Louis, Nadège Zanou, Monique Van Schoor, Philippe Gailly (2008). TRPC1 regulates skeletal myoblast migration and differentiation.Journal of Cell Science **121**: 3951-3959.

- Mao YS, Sunwoo H, Zhang B, Spector DL(2011).Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. Nat Cell Biol **13**:95–101.

- Martianov I, Ramadass A, Serra Barros A, Chow N, Akoulitchev A (2007). Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. Nature **445**:666–670.

- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal, **17**(1), pp. 10-12.

- Mattick JS (2009). Deconstructing the dogma: a new view of the evolution and genetic programming of complex organisms. Ann N Y Acad Sci 1178: 29–46.

- Moulton, Vincent (2005). Tracking down Noncoding RNAs. Proceedings of the National Academy of Sciences of the United States of America 102(7): 2269–2270.

- Mousavi K, Zare H, Dell'orso S, Grontved L, Gutierrez-Cruz G, Derfoul A, Hager GL, Sartorelli V (2013). eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. Mol Cell 51: 606-617.

- Mu Zhu, Jiafan Liu, Jia Xiao, Li Yang, Mingxiang Cai, Hongyu Shen, Xiaojia Chen, Yi Ma, Sumin Hu, Zuolin Wang, An Hong, Yingxian Li, Yao Sun & Xiaogang Wang (2017).Lnc-mg is a long non-coding RNA that promotes myogenesis.Nature Communications 8 :14718

- Myocardial Lineage Development Sylvia M. Evans, Deborah Yelon, Frank L. Conlon and Margaret L. Kirby (2010). Circ Res 107:1428-1444

- Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U,

Baker JC, Grutzner F, Kaessmann H (2014). The evolution of lncRNA repertoires and expression patterns in tetrapods. Nature **505**: 635–640.

- Ng SY, Bogu GK, Soh BS, Stanton LW (2013).The long noncoding RNA RMST interacts with SOX2 to regulate neurogenesis. Mol Cell, **51**:349–359.

- Ng SY, Johnson R, Stanton LW (2012). Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. EMBO J **31**: 522–533.

- P. J. Batista and H. Y. Chang, (2013).Long noncoding RNAs: cellular address codes in development and disease. Cell **152** (6):1298–1307.

- Parameswaran M, Tam PP (1995).Regionalisation of cell fate and morpho- genetic movement of the mesoderm during mouse gastrulation.Dev Genet **17**:16–28.

- Peschiaroli A, Figliola R, Coltella L, Strom A, Valentini A, D'Agnano I, Maione R (2002).MyoD induces apoptosis in the absence of RB

function through a p21WAF1-dependent re-localization of cyclin/cdk complexes to the nucleus. Oncogene **21**:8114-8127.

- Pollard KS, Salama SR, Lambert N, Lambot MA, Coppens S, Pedersen JS, Katzman S, King B, Onodera C, Siepel A, *et al*. 2006. An RNA gene expressed during cortical development evolved rapidly in humans. Nature **443**: 167–172.

- Powell JA, Carrasco MA, Adams DS, Drouet B, Rios J, Müller M, Estrada M, Jaimovich E (2001). IP3 receptor function and localization in myotubes: An unexplored Ca2+ signaling pathway in skeletal muscle. J Cell Sci **114**: 3673–3683

- R. Grifone, J. Demignon, C. Houbron *et al.* (2005). Six1 and Six4 homeoproteins are required for Pax3 and Mrf expression during myogenesis in the mouse embryo. Development **132**(9):2235–2249.

- R. Grifone, J. Demignon, J. Giordani *et al.* (2007). Eya1 and Eya2 pro- teins are required for hypaxial somitic myogenesis in the mouse embryo. Developmental Biology **302**(2): 602–616.

- Rao SS, Chu C, Kohtz DS (1994). Ectopic expression of cyclin D1 prevents activation of gene transcription by myogenic basic helix-

loop-helix regulators. Mol Cell Biol, **14**:5259-5267.

- Rao SS, Kohtz DS (1995). Positive and negative regulation of D-type cyclin expression in skeletal myoblasts by basic fibroblast growth factor and transforming growth factor β: a role for cyclin D1 in control of myoblast differentiation. J Biol Chem **270**:4093-4100.

- Rojas, A., Schachterle, W., Xu, S. -M., Martín, F., & Black, B. L. (2010). FoxA2 binding to an intronic enhancer controls direct transcriptional regulation of Gata4 during early endoderm specification. Developmental Biology **346**(2): 346–355.

- Mali, S. *et al*. (2016). Observations on Different Resin Strategies for Affinity Purification Mass Spectrometry of a Tagged Protein. Anal Biochem **515**: 26-32.

- S. B. P. Chargé and M. A. Rudnicki (2004). Cellular and molecular regulation of muscle regeneration. Physiological Reviews 84(1): 209–238.

- S.R. Eddy (2001).Non–coding RNA genes and the modern RNA world, Nature Reviews Genetics 2:919-929.

- Schmitz, S. U., Grote, P., & Herrmann, B. G. (2016). Mechanisms of

long noncoding RNA function in development and disease. Cellular and Molecular Life Sciences 73:2491–2509.

- Schuster-Gossler K, Bilinski P, Sado T, Ferguson-Smith A, Gossler A (1998). The mouse Gtl2 gene is differentially expressed during embryonic development, encodes multiple alternatively spliced transcripts, and may act as an RNA. Dev Dyn **212**: 214-228.

- Skapek SX, Rhee J, Spicer DB, Lassar AB (1995).Inhibition of myogenic differentiation in proliferating myoblasts by cyclin D1-dependent kinase. Science **267**:1022-1024.

- Sunwoo H, Dinger ME, Wilusz JE, Amaral PP, Mattick JS, Spector DL (2009). MEN epsilon/beta nuclear-retained non-coding RNAs are up regulated upon muscle differentiation and are essential components of paraspeckles. Genome Res **19**: 347-359.

- T. Braun and M. Gautel (2011).Transcriptional mechanisms regulating skeletal muscle differentiation, growth and homeostasis. Nature Reviews Molecular Cell Biology **12**(6): 349–361.

- Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, Freier SM, Bennett CF, Sharma A, Bubulya PA, Blencowe BJ, Prasanth SG,

Prasanth KV (2010). The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol Cell **39**:925–938.

- Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, Shi Y, Segal E, Chang HY (2010). Long noncoding RNA as modular scaffold of histone modification complexes. Science **329**: 689–693.

- Vance, K. W., & Ponting, C. P. (2014). Transcriptional regulatory functions of nuclear long noncoding RNAs. Trends in Genetics **30**(8): 348–355.

- Wamstad JA, Alexander JM, Truty RM, Shrikumar A, Li F, Eilertson KE, Ding H, Wylie JN, Pico AR, Capra JA, Erwin G, Kattman SJ, Keller GM, Srivastava D, Levine SS, Pollard KS, Holloway AK, Boyer LA, Bruneau BG (2012). Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. Cell **151**: 206–220

- Wang J, Gong C, Maquat LE (2013). Control of myogenesis by rodent SINE-containing lncRNAs. Genes Dev, **27**: 793-804.

- Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen

Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA, Wysocka J, Lei M, Dekker J, Helms JA, (2011).A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression Nature **472**, 120–124

- Wang, L., Zhao, Y., Bao, X., Zhu, X., Kwok, Y. K., Sun, K., … Wang, H. (2015). LncRNA Dum interacts with Dnmts to regulate Dppa2 expression during myogenic differentiation and muscle regeneration. Cell Research, **25**(3), 335–350.

- Washietl S, Kellis M, Garber M (2014). Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. Genome Res **24**: 616–628.

- Watts R, Johnsen VL, Shearer J, Hittel DS (2013).Myostatin-induced inhibition of the long noncoding RNA Malat1 is associated with decreased myogenesis. Am J Physiol Cell Physiol, **304**: C995-1001.

- Xing Chen, Chenggang Clarence Yan, Xu Zhang, Zhu-Hong You (2016).Long non-coding RNAs and complex diseases: from experimental results to computational models. Brief Bioinform bbw060.

- Yan L, Yang M, Guo H, *et al*.(2013).Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nat Struct Mol Biol **20**:1131–1139.

- Yang L, Lin C, Liu W, Zhang J, Ohgi KA, Grinstein JD, Dorrestein PC, Rosenfeld MG (2011).ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. Cell **147**:773–788.

- Zabludoff SD, Csete M, Wagner R, Yu X, Wold BJ (1998).p27Kip1 is expressed transiently in developing myotomes and enhances myogenesis. Cell Growth Differ **9**:1-11.

- Zaffran S, Frasch M (2002). Early signals in cardiac development. Circ Res **91**:457–469.

- Zhang P, Wong C, Liu D, Finegold M, Harper JW, Elledge SJ (1999). p21CIP1 and p57KIP2 control muscle differentiation at the myogenin step. Genes Dev **13** : 213-224.

Appendix A.3.1: List of the primers used for real time PCR (qPCR):

Primer	Forward	Reverse
1010001N08Rik	AAACACTCTTGTGGGAGTGGAT	AATCCTGTTCCAGCTGCTAGTC
1700006J14Rik	GCTTCGCTACACGGAAGAAG	GGCCAA TCTCATTCTGCA CT
1700113A16RIK	CTAGGATTCCAGGATCCCTTTC	GCAGAGTGAGTGATGATTTGGA
2310031A07Rik	TCCCTTCTCTCTGCTTCCTG	GAATTGCTTGGCTTCTTTGC
2410006H16Rik	GACAACGTGGAATAGGAGCAG	CTGGAGACTCCTCTGGGAAT
4930474H06Rik	TGGTGGTGTAGTTGCTCTGC	AGACGAGGTCTCCTACGAA
5033428I22Rik	CAAGCACCCTAAAGAAGAGGAA	AAAGGGAAGAATCTCCATCTCC
5430400D12Rik	CATCGAGACTGGCAATTCCT	AAAGGGAAGAATCTCCATCTCC
6720401G13RIK	CTGCATGAGGAAGCAGAGTTTA	GGGAGAGTTTACCTCCCATTTT
9030622O22Rik	CTTTGGGGCACAGAAAGAGA	GATCCACAGGAGCTTGAAGG
A230004M16Rik	AGG AAC AATTGGGGG AAA AC	TCAGTACCTCTGTGCGATCCT
Acta	AGGATGCAGAAGGAGATCACA	CACATCTGCTGGAAGGTAGAC
BC065403	AAATCTTGAGGATGCTGGGATA	AAGCTGTCTCCTGAATGATGTG
C430049B03Rik	CCATTTTCGAGTGGAGTGAGA	CCAGTTTGAGAACGTTTTGTCT
Cdkna1	GTACTTCCTCTGCCTGCTG	AATCTGTCAGGCTGGTCTGC
E130307A14Rik	GACCGCCATGTCACATCTTA	CCTCTTTAGAAGCATCCAGGTG
E130307A14Rik	GACCGCCATGTCACATCTTAA	CCTCTTTAGAAGCATCCAGGTG
G530011O06Rik	GGCCCAAACTGGAATACCTT	TGTCCACACTTTGGTCTTCG
GAPDH	AGGTCGGTGTGAACGGATTTG	TGTAGACCATGTAGTTGAGGTCA
Gata4	CTGTGGCCTGTATCACAAGATG	GTGGTGGTAGTCTGGCAGTTG
Gata6	CTTGCGGGCTCTATATGAAACT	TAGGTTTTCGTTTCCTGGTTTG
Gm13001	CAAGATTTCATGGATTCAGCAG	AAAAAGAGGGGGAGAAGAGATG
GM16908	GACCTATGGGGATATGTGCATT	AGGAAAGGTCTTTGCCTCTTCT
Hand1	CCTTCAAGGCTGAACTCAAAAA	GCGCCCTTTAATCCTCTTCT
Hand2	CTACATCGCCTACCTCATGGAT	TGCTCACTGTGCTTTTCAAGAT
Kis2	CTCCCTTGTTGCTCTGAAGG	AACCAAGGTAGGTCGGGAGT
Mef2c	CATCTCTGTCTGGCTTCAACAC	GTGTGTTGTGGGTATCTCGAAG
Mesp1	GTTCCTGTACGCAGAAACAGC	CAGGTTTCTAGAAGAGCCAGCA
Myf5	CTGTCTGGTCCCGAAAGAAC	GCTCGGATGGCTCTGTAGAC
Myf6	GGCTGGATCAGCAAGAGAAG	CCTGGAATGATCCGAAACAC
Myh3	CACTGTGGAGACAGAAGACAGC	CCATGTCCTCAATCTTGTCAAA
Myh6	CATCAAGGAGCTCACCTACCA	TTGTAGGCCTTCACCTTCAACT
Myod	GGCTACGACACCGCCTACTA	GTGGTGCATCTGCCAAAG
Myog	CAGTGAATGCAACTCCCACA	CTGTCCACGATGGACGTAAG
Nkx2-5	CAAGTGCTCTCCTGCTTTCC	GGCTTTGTCCAGCTCCACT
Notch1	CCACTGTGAACTGCCCTATGT	CTTCACAGTTCTGTCCAGCAAA
Platr14	ACAATGAGGTATTTTGCCTGCT	GAGCCTGGGTTTGTCATATAGC
Smyd1	AGCCTATGCTATTCTCCTGGTG	ATACGCAGCTCCATCTCTGTCT
T	TAACTGGTCTAGCCTCGGAGTG	GTACCATTGCTCACAGACCAGA
Tbx5	TGGAGTTACACAGGATGTCTCG	TACTGGGACCCTGGGTTGGATG
Tnni2	CCAGGCTCCATGTCTGAAGT	TTGCCCCTCAGGTCAAATAG
Tnnt3	AAGCCTCTGAACATTGACCATC	CGAATTTGTCAGTCTCCAGTTG
Traf3ip2	CAAGAAGAGAGACCTGCACAGA	ATTAGATGGTGGGCGGTAGAG

A3.2 List of the shRNA sequences used in the knockdown study:

LncRNA Sites	Sequences
Xrcc2_site1_1	tgGAACTAGAACCCAACCTGTTTGCTGttcaagagaCAGCAAACAGGTTGGGTTCTAGTTGTTttttttc
Xrcc2_site1_2	ctcgaggaaaaaaAACAACTAGAACCCAACCTGTTTGCTGtctcttgaaCAGCAAACAGGTTGGGTTCTAGTCca
Xrcc2_site3_1	tgGAAACAGATAGCTAGGTAAATAGATttcaagagaATCTATTTACCTAGCTATCTGTTTCCTttttttc
Xrcc2_site3_2	ctcgaggaaaaaaAGGAAACAGGATAGCTAGGTAAATAGATtctcttgaaATCTATTTACCTAGCTATCTGTTTCca
Xrcc2_site4_1	tgCAGCCAGTTACTTCAGTAATGCTTCttcaagagaGAAGCATTACTGAAGTAACTGGCTGTCttttttc
Xrcc2_site4_2	ctcgaggaaaaaaGACGCCAGTTACTTCAGTAATGCTTCtctcttgaaGAAGCATTACTGAAGTAACTGGCTGca
4930500j02RIK_ site2_1	tgCTACAAACTGCTTCTGTGTGACTAAttcaagagaTTAGTCACACAGAAGCAGTTTGTAGTCttttttc
4930500j02RIK_ site2_2	ctcgaggaaaaaaGACTACAAACTGCTTCTGTGTGACTAAtctcttgaaTTAGTCACACAGAAGCAGTTTGTAGca
4930500j02RIK_site3_1	tgCAATCACATCCTCCAGAGATTCTGAttcaagagaTCAGAATCTCTGGAGGATGTGATTGGCttttttc
4930500j02RIK_site3_2	ctcgaggaaaaaaGCCAATCACATCCTCCAGAGCTTCTGAtctcttgaaTCAGAATCTCTGGAGGATGTGATTGca
4930500j02RIK_site5_1	tgGAGCCATCTCCCTAGCAATCAAACTttcaagagaAGTTTGATTGCTAGGGAGATGGCTCAGttttttc
4930500j02RIK_site5_2	ctcgaggaaaaaaCTGAGCCATCTCCCTAGCAATCAAACTtctcttgaaAGTTTGATTGCTAGGGAGATGGCTCca
6720401G13RIK_site1_1	tgGCATGAGGAAGCAGAGTTTATTCTTttcaagagaAAGAATAAACTCTGCTTCCTCATGCAGttttttc
6720401G13RIK_site1_2	ctcgaggaaaaaaCTGCATGAGGAAGCAGACTTTATTCTTtctcttgaaAAGAATAAACTCTGCTTCCTCATGCca
6720401G13RIK_site2_1	tgGAATGAGGAAGAACTGGACGATGAAttcaagagaTTCATCGTCCAGTTCTTCCTCATTCCCttttttc
6720401G13RIK_site2_2	ctcgaggaaaaaaGGGAATGAGGTTGAACTGGACGATGAAtctcttgaaTTCATCGTCCAGTTCTTCCTCATTCca
6720401G13RIK_site3_1	tgGGCTGAGTCGCTACCTTGAAATGAAttcaagagaTTCATTTCAAGGTAGCGACTCAGCCAGttttttc
6720401G13RIK_site3_2	ctcgaggaaaaaaCTGGCTGAGTCGCTACCTTGAAATGAAtctcttgaaTTCATTTCAAGGTAGCGACTCAGCCca
1700113A16RIK_site1_1	tgCAAGGAGCTCCCAGCCTCTTAGGAGttcaagagaCTCCTAAGAGGCTGGGAGCTCCTTGTTttttttc
1700113A16RIK_site1_2	ctcgaggaaaaaaAACAAGGAGCTCCCAGCCTCTTAGGAGtctcttgaaCTCCTAAGAGGCTGGGAGCTCCTTGca
1700113A16RIK_site2_1	tgCCAATACGAGGCTAAGGGAATTCCAttcaagagaTGGAATTCCCTTAGCCTCGTATTGGCTttttttc
1700113A16RIK_site2_2	ctcgaggaaaaaaAGCCAATACGAGGCTAAGGGAATTCCAtctcttgaaTGGAATTCCCTTAGCCTCGTATTGGca
1700113A16RIK_site3_1	tgCGTACAGCACGGAAAGAAATAATGAttcaagagaTCATTATTTCTTTCCGTGCTGTACGGTttttttc
1700113A16RIK_site3_2	ctcgaggaaaaaaACGGTACAGCACGGAAAGAAATAATGAtctcttgaaTCATTATTTCTTTCCGTGCTGTACGca
BC065403_ site1_1	tgCAAGCAGAAACCAGTGACAAGTTCTttcaagagaAGAACTTGTCACTGGTTTCTGCTTGTAttttttc
BC065403_ site1_2	ctcgaggaaaaaaTACAAGCAGAAACCAGTGACAAGTTCTtctcttgaaAGAACTTGTCACTGGTTTCTGCTTGca
BC065403 _site4_1	tgAGTTAAGATTGATAAGGGTTGAGAAttcaagagaTTCTCAACCCTTATCAATCTTAACTTTttttttc
BC065403 _site4_2	ctcgaggaaaaaaAAAGATTGATAAGGGTTGAGAAtctcttgaaTTCTCAACCCTTATCAATCTTAACTca
BC065403_site5_1	tgGAGGATGCTGGGATAGTCTGACGTGttcaagagaCACGTCAGACTATCCCAGCATCCTCAAttttttc
BC065403_site5_2	ctcgaggaaaaaaTTGAGGATGCTGGGATAGTCTGACGTGtctcttgaaCACGTCAGACTATCCCAGCATCCTCca

List of the shRNA sequences used in the knockdown study(continued):

LncRNA Sites	Sequences
GM16908_site3_1	tgGATGGACCACCTTCTGTCTAGTTCTttcaagagaAGAACTAGACAGAAGGTGGTCCATCTTtttttc
GM16908_site3_2	ctcgaggaaaaaaAAGATGGACCACCTTCTGTCTAGTTCTtctcttgaaAGAACTAGAGAAGGTGGTCCATCca
GM16908_site4_1	tgCACCCAGAGAGCTAAGAAACTGAGAttcaagagaTCTCAGTTTCTTAGCTCTCTGGGTGTGttttttc
GM16908_site4_2	ctcgaggaaaaaaCACACCCAGAGAGCTAAGAAACTGAGAtctcttgaaTCTCAGTTTCTTAGCTCTCTGGGTGca
GM16908_site5_1	tgGTGCCAGACCCAAGTGAACTTAGTTttcaagagaAACTAAGTTCACTTGGGTCTGGCACCTttttttc
GM16908_site5_2	ctcgaggaaaaaaAGGTGCCAGACCCAAGTGAACTAGTTtctcttgaaAACTAAGTTGAGTTTGGGTCTGGCACca
E130307A14RIK_site1_1	tgCGAACAGCTGCCATGGTAAACTGGCttcaagagaGCCAGTTTACCATGGCAGCTGTTCGGTttttttc
E130307A14RIK_site1_2	ctcgaggaaaaaaACCGAACAGCTGCCATGGTAAACTGGCtctcttgaaGCCAGTTTACCATGGCAGCTGTTCGca
E130307A14RIK_site2_1	tgGAATCACCTGGATGCTTCTAAAGAGttcaagagaCTCTTTAGAAGCATCCAGGTGATTCTTttttttc
E130307A14RIK_site2_2	ctcgaggaaaaaaAAGAATCACCTGGATGCTTCTAAAGACtctcttgaaCTCTTTAGAAGCATCCAGGTGATTCca
E130307A14RIK_site3_1	tgGAGGCAGCTCCAAATGTATTAGGTGttcaagagaCACCTAATACATTTGGAGCTGCCTCCGttttttc
E130307A14RIK_site3_2	ctcgaggaaaaaaCGGAGGCAGCTCCAAATGTATTAGGTGtctcttgaaCACCTAATACATTTGGAGCTGCCTCca
5033428I22RIK_site2_1	tgGCACCCTAAAGAAGAGGAAGAGAGCttcaagagaGCTCTCTTCCTCTTCTTTAGGGTGCTTttttttc
5033428I22RIK_site2_2	ctcgaggaaaaaaAAGCACCCTAAAGAAGAGGAAGAGAGCtctcttgaaGCTCTCTCCTCTTCTTTAGGGTGCca
5033428I22RIK_site3_1	tgCGAGCATCTCCAGCTTGTGCTCCAttcaagagaTGGAGCACAAGCTGGAGACTGCTCGCAttttttc
5033428I22RIK_site3_2	ctcgaggaaaaaaTGCGAGCAGTCTCCAGCTTGTGCTCCAtctcttgaaTGGAGCACAAGCTGGAGATGCTCGca
5033428I22RIK_site5_1	tgCCGTGACTATCTTTCGAGTTAGAGCttcaagagaGCTCTAACTCGAAAGATAGTCACGGTGttttttc
5033428I22RIK_site5_2	ctcgaggaaaaaaCACCGTGACTATCTTTCGAGTTAGAGCtctcttgaaGCTCTAACTCGAAAGATAGTCACGGca
1700006J14RIK_site2_1	tgAGATCTCAATGCTGAGCATTTCTTGttcaagagaCAAGAAATGCTCAGCATTGAGATCTCTttttttc
1700006J14RIK_site2_2	ctcgaggaaaaaaAGAGATCTCAATGCTGAGCATTTCTTGtctcttgaaCAAGAAATGCTCAGCATTGAGATCTca
1700006J14RIK_site3_1	tgGGTCCAGCCACCTGACCCAGCCTGAttcaagagaTCAGGCTGGGTCAGGTGGCTGGACCGAttttttc
1700006J14RIK_site3_2	ctcgaggaaaaaaTCGGTCCAGCCACCTGACCCAGCCTGAtctcttgaaTCAGGCTGGGTCAGGTGGCTGGACCca
1700006J14RIK_site5_1	tgCTGATGCTGAGGTAGAAATTGAATGttcaagagaCATTCAATTTCTACCTCAGCATCAGTTttttttc
1700006J14RIK_site5_2	ctcgaggaaaaaaAACTGATGCTGAGGTAGAAATTGAATGtctcttgaaCATTCAATTTCTACCTCAGCATCAGca
1010001N08RIK_site1_1	tgGAAACACTCAGGATGAAGATGCTGGttcaagagaCCAGCATCTTCATCCTGAGTGTTTCCTttttttc
1010001N08RIK_site1_2	ctcgaggaaaaaaAGGAAACACTCAGGATGAAGATGCTGGtctcttgaaCCAGCATCTTCATCCTGAGTGTTTCca
1010001N08RIK_site2_1	tgAGATCTCAATGCTGAGCATTTCTTGttcaagagaCAAGAAATGCTCAGCATTGAGATCTCTttttttc
1010001N08RIK_site2_2	ctcgaggaaaaaaAGAGATCTCAATGCTGAGCATTTCTTGtctcttgaaCAAAGAAATGCTCAGCATTGAGATCTca
1010001N08RIK_site3_1	tgGGTCCAGCCACCTGACCCAGCCTGAttcaagagaTCAGGCTGGGTCAGGTGGCTGGACCGAttttttc
1010001N08RIK_site3_2	ctcgaggaaaaaaTCGGTCCAGCCACCTGACCCAGCCTGAtctcttgaaTCAGGCTGGGTCAGGTGGCTGGACCca
1010001N08RIK_site5_1	tgCTGATGCTGAGGTAGAAATTGAATGttcaagagaCATTCAATTTCTACCTCAGCATCAGTTttttttc
1010001N08RIK_site5_2	ctcgaggaaaaaaAACTGATGCTGAGGTAGAAATTGAATGtctcttgaaCATTCAATTTCTACCTCAGCATCAGca
C430049B03RIK_site2_1	tgCCACTGCCTGGGTAAGTCTGGGTCTttcaagagaAGACCCAGACTTACCCAGGCAGTGGAAttttttc
C430049B03RIK_site2_2	ctcgaggaaaaaaTTCCACTGCCTGGGTAAGTCTGGGTCTtctcttgaaAGACCCAGACTTACCCAGGCAGTGGca
C430049B03RIK_site3_1	tgGGCTTAAGAAGGAGTGATTTGTATTttcaagagaAATACAAATCACTCCTTCTTAAGCCACttttttc
C430049B03RIK_site3_2	ctcgaggaaaaaaGTGGCTTAAGAAGGAGTGATTTGTATTtctcttgaaAATACAAATCACTCCTTCTTAAGCCca
C430049B03RIK_site5_1	tgGGAGAAGAGTGTTAAACTTCTGGAGttcaagagaCTCCAGAAGTTTAACACTCTTCTCCAGttttttc
C430049B03RIK_site5_2	ctcgaggaaaaaaCTGGAGAAGAGTGTTAAACTTCTGGAGtctcttgaaCTCCAGAAGTTTAACACTCTTCTCCca
9230102O04RIK_site3_1	tgACAGCTTACTGGAGAGTGTGACTCTttcaagagaAGAGTCACACTCTCCAGTAAGCTGTGAttttttc
9230102O04RIK_site3_2	ctcgaggaaaaaaTCACAGCTTACTGGAGAGTGTGACTCTtctcttgaaAGAGTCACACTCTCCAGTAAGCTGTca
9230102O04RIK_site4_1	tgCAGCGTGCACGGAATCGAATTAAACttcaagagaGTTTAATTCGATTCCGTGCACGCTGCAttttttc
9230102O04RIK_site4_2	ctcgaggaaaaaaTGCAGCGTGCACGGAATCGAATTAAACtctcttgaaGTTTAATTCGATTCCGTGCACGCTGca
9230102O04RIK_site5_1	tgCATGGTGAAAGCTAAGGCACTGTGGttcaagagaCCACAGTGCCTTAGCTTTCACCATGTCttttttc
9230102O04RIK_site5_2	ctcgaggaaaaaaGACATGGTGAAAGCTAAGGCACTGTGGtctcttgaaCCACAGTGCCTTAGCTTTCACCATGca

A3.3 List of the Dsi RNA sequences used in the Knockdown study:

Platr14 sites	**Dsi Sequences**
Platr14 site 1	GGAGCUCUAGAAUUUCAC
	CACCUUUGUGAAUUCUA
Platr14 site 2	GUGGAAUCAGCUCAAAUC
	UCAUGCUGAUUUGAGCUG
Platr14 site 3	AGGAGCUCUAGAAUUUCA
	ACCUUUGUGAAAUUCUAG

A3.4 The list of probes used in RAP-MS and ChIRP Assays:

Forward sequence *Platr14*

1tttggaatgg ggacagcata tatggttcag actacaaact gcttctgtgt gactaagaga
61agaagaatca gagtaaagag aaacatactc tcttaatctc cgaagattaa cggcaggagg
121agctctagaa tttcacaaag gtgctgtcag gacacatgaa aatctagaga aaaagaccag
181cttgaagaga ccaaaaagaa gaaagcctgc caatcacatc ctccagagat tctgagacac
241acacacccct ggtgtagggg gacaccatct gaaaggtgac acccattcta cggaatataa
301gtaccagcta aagaggacac acacttgaag aaaaaggaga caatgaggta ttttgcctgc
361tcatatgtct atgcctggtg cccccagagg ccagaagagg acatcaggag ttaaagacag
421ttgtgaccca ccatgtgcta tatgacaaac ccaggctctc tggaagagca gccagtgagt
481gctcgtgaat gctgagccat ctccctagca atcaaactgg agaaatgatg gcattgtgcc
541aggttggatt aggttaaata aaaatcgtaa tcattggcta gg

The probe sequences used:

1. TGTCCTGACAGCACCTTT
2. CTGGAGGATGTGATTGGC
3. GGGTGTCACCTTTCAGAT
4. GAGAGCCTGGGTTTGTCATA
5. TGCTAGGGAGATGGCTCA
6. CCTAATCCAACCTGGCAC