Doctorate Dissertation

博士論文

Study on regulatory mechanisms governing the expression of transcription factor genes during mouse heart development

（マウス心臓発生過程における転写因子遺伝子の制御メカニズムに関する研究）

A Dissertation Submitted for Degree of Doctor of Philosophy

February 2017

平成 29 年 2 月博士（理学）申請

Department of Biological Sciences, Graduate School of Science,

The University of Tokyo

東京大学大学院理学系研究科生物科学専攻

Yutaro Hori

堀　優太郎

Abstract

Congenital heart diseases occur frequently among newborns, indicating that heart development is a complex process that requires precise control of differentiation and growth of cardiac cells and also precise patterning of these cells. Although the mechanism underlying heart development is still not fully understood, previous studies have suggested that dose-sensitive characteristics of cardiac transcription factors play critical roles for heart development. Many of the cardiac transcription factor genes require both alleles for proper heart development and their mutations result in highly variable phenotypes even among inbred strain mice.

Recent studies about regulatory mechanism of gene expression in mammals have revealed that long non-coding RNAs (lncRNAs) are an important modulator at the transcriptional and translational levels. Based on the hypothesis that lncRNAs also play important roles in mouse heart development, I attempted to comprehensively identify lncRNAs by comparing embryonic and adult mouse hearts.

RNA-seq analysis of the ventricles of E10.5, E13.5 and 8 week mice identified 787 lncRNA candidates. By comparing the expression of these candidates in mouse brain RNA-seq data, I identified 316 lncRNA candidates among them as non-ubiquitous and heart-selective ones. Next, I examined the distribution of lncRNAs in the mouse genome. Gene ontology analysis revealed that many heart-selective lncRNAs are present near genes important for the heart development. Importantly, many of them are transcribed from the promoters of neighboring genes in a head-to-head divergent manner. I next tried to address significance of the enrichment of genes with bidirectional lncRNAs among haploinsufficient genes. Based on the report that identified haploinsufficient genes and RefSeq transcript annotation database, I showed that the enrichment is observed not only in the heart but also in many other tissues.

In this study, I focused on the lncRNA divergently transcribed from *Tbx5*, since it is

evolutionarily well conserved and *Tbx5* is also a haploinsufficient gene. Quantitative RT-PCR (qRT-PCR) analysis demonstrated that the expression pattern of the lncRNA is slightly different from that of *Tbx5*, indicating that they are regulated separately, though they share the same promoter. Using ES cells, I knocked down the lncRNA by inserting strong transcription stop signals into the second exon using the Crispr/Cas9 system. Knockdown (KD) mice derived from the ES cells were embryonic lethal and exhibited severe hypoplasia of ventricle. The expression pattern of *Tbx5* mRNA was unchanged and the RNA-seq of KD mice suggested normal differentiation of cardiomyocytes. Thus, although the precise mechanism is still unsolved, this transcript is likely to function in the morphogenesis of the mouse heart.

During the study of Chapter I, I found that *Tbx5* has multiple promoters and the expression pattern of Tbx5 protein could be altered in the lncRNA KD mice. Based on these findings, I tried to understand the translational regulation of *Tbx5* in Chapter II. *Tbx5* has been shown to be expressed at the left-side of the ventricle and crucial for the proper formation of the ventricular septum. Although disturbance of this left-right gradient is well-known to cause univentricular heart completely lacking the left-right identity, the molecular mechanism underlying the formation of this gradient is yet to be understood.

First, by conducting qRT-PCR and immunohistochemistry on*Tbx5*, I found a distinct expression pattern at the mRNA and protein levels in the ventricle. Even in postnatal mice that lack Tbx5 protein signal, the mRNA level was comparable to that in embryos. I reanalyzed the RNA-seq data I used in Chapter I and found that there are three *Tbx5* promoters and one of them, which I call promoter A, showed the embryo-specific expression pattern. I compared the expression levels of these alternative promoter isoforms by qRT-PCR and revealed that the expression pattern of promoter A isoform was consistent with that of Tbx5 protein. The spatial expression pattern of promoter A isoform in the ventricle of embryos was also remarkably consistent with the protein expression pattern. Furthermore, by luciferase assay, the 5’ UTR of the highly-expressed isoform with an inconsistent expression pattern exhibited strong translational repression activity. However, knockout (KO) of isoform A turned out to have little impact on Tbx5 protein production in ESC differentiation system, indicating that *Tbx5* is subject to a complex post-transcriptional control.

In conclusion, these findings suggested that lncRNAs, especially bidirectionally transcribed ones, might play a role in the precise expression regulation of dose-sensitive transcription factor genes that are close to them. In addition, the study in Chapter II suggested various potential strategies to achieve complex expression patterns of these genes.

Table of Contents

Abbreviations

CHD: congenital heart disease

gRNA: guide RNA

ESC: embryonic stem cell

FDR: false discovery rate

HOS: Holt-Oram Syndrome

IHC: immunohistochemistry

ISH: *in situ* hybridization

IVS: interventricular septum

KD: knockdown

KO: knockout

LA: left atrium

lncRNA: long non-coding RNA

LV: left ventricle

ON: overnight

PRC: polycomb repressive complex

qRT-PCR: quantitative real-time polymerase chain reaction

RA: right atrium

RNA-seq: RNA sequencing

RT: room temperature

RV: right ventricle

Tbx5ua: *Tbx5 upstream antisense product*

uORF: upstream open reading frame

UTR: untranslated region

VSD: ventricular septal defect

WB: western blot

WT: wildtype

General Introduction

Regulation of gene expression level is critical for development. In fact, polysomy is a condition in which an organism has one or more chromosome than normal and human trisomies having one extrachromosome often results in miscarriage rather than live birth. Except trisomies of chromosome 18 (Edwards syndrome) and 21 (Down syndrome), most of trisomies are embryonic lethal. Turner syndrome exhibits various abnormalities and is a condition in which a female is partly or completely missing an X chromosome. These abnormalities associated with alteration of chromosome number indicate importance of gene expression levels in development. It is well-known that expression levels of transcription factor genes are particularly important as their alteration often cause abnormalities even when only one of the two copies are mutated (haploinsufficiency). Moreover, mutations of many of these haploinsufficient genes lead to highly variable phenotypes (Zlotogora, 2003).

Although the expression levels of genes are crucial in development, they are highly variable at the single cell-level. Advent of imaging technologies using fluorescent proteins have made it possible to observe fluctuations of gene expression in living cells in real-time. Fluctuations of gene expression at the transcriptional and translational levels have been studied extensively and it has been revealed that gene expression is intrinsically variable, leading to high degree of diversity in cell populations (Blake et al., 2003; Elowitz et al., 2002; Raser and O'Shea, 2004; Wu et al., 2016). Although prokaryotic cells are found to have much higher degree of fluctuation, to the level that the number of mRNA and protein are not correlated at all (Taniguchi et al., 2011), eukaryotic cells also show probabilistic behavior in transcription and translation, and it can even lead to phenotypic variations in single-celled organisms (Blake et al., 2006). In multicellular organisms, however, high levels of phenotypic homogeneity are recognized. In particular, the phenotypic similarities between identical twins are extremely high, indicating so-called genetic determinism. How can the dichotomy between this remarkable reproducibility of development and huge noise at the cell level be resolved? This kind of phenotypic stability is a manifestation of robustness in development and is one of the biggest issues in developmental biology (Kitano, 2004).

The heart is an organ with the highest occurrence of congenital abnormalities, suggesting that its developmental processes are finely regulated and fragile. Many transcription factor genes involved in the heart development are known to be dose-sensitive, exhibiting lower penetrance and haploinsufficiency (Fahed et al., 2013). The transcription factor, *Tbx5,* which I studied in this thesis paper, is one of such genes. Mutation of *Tbx5* is the cause of Holt-Oram syndrome characterized by ventricular and/or atrial septal defects and forelimb abnormalities (Mori and Bruneau, 2004). Holt-Oram syndrome model mice show highly variable phenotypes, from completely normal to severe loss of cardiac walls and forelimbs (Bruneau et al., 2001). Interestingly, *Tbx5* is also a haploinsufficient gene, i.e. both alleles are required for normal development. There are two kinds of mechanisms that underlying haploinsufficiency. One kind is that an allele cannot produce a sufficient amount of proteins and the other kind is that the gene precisely requires a certain expression level (Deutschbauer et al., 2005). For haploinsufficient transcription factor genes, the latter case is dominant. Indeed, several studies indicated that overexpression of some of these cardiac transcription factor genes lead to developmental abnormalities (Breckenridge et al., 2009; Espinoza-Lewis et al., 2011; Gove et al., 1997; Liberatore et al., 2000). Thus, the heart development requires very precise and sensitive regulatory mechanism of gene expression, which is likely to be a cause of developmental fragility. These characteristics make the heart an interesting model for gene regulation studies.

Furthermore, the heart is an interesting organ in terms of evolution. Cardiac structures among vertebrates are quite diverse, fish have one atrium and one ventricle, amphibians have two atria and one ventricle, and birds and mammals have two atria and two ventricles. It is still an interesting unsolved issue how the septation between cardiac chambers has evolved. *Tbx5* has been suggested to play a crucial role in the evolution of ventricular walls since non-septated animals express *Tbx5* uniformly in the heart while animals with complete walls show a steep gradient of *Tbx5* expression in the ventricle. The loss of this gradient leads to a single ventricle indicating that the spatial expression pattern is essential (Koshiba-Takeuchi et al., 2009). Other genes such as *Hand1* and *Hand2* show such restricted expression patterns, possibly contributing to the evolution of four cardiac chambers (Olson, 2006).

In this study, to understand the molecular mechanisms underlying the precise control of gene expression in terms of dosage and space, I investigated long non-coding RNAs (lncRNAs), which have recently attracted much attention as important gene expression modulators. Unlike small non-coding RNAs such as miRNA, lncRNAs are yet to be sufficiently functionally classified, while many of them are suggested to work as transcriptional modifiers by recruiting epigenetic factors to specific loci (Wang and Chang, 2011). In chapter I, I tried to find clues to understand what regulates the expression levels of cardiac transcription factors by focusing on lncRNAs. By performing RNA-seq analysis on developing mouse ventricles I found that lncRNAs are enriched near cardiac transcription factors. Furthermore, I showed that lncRNAs transcribed from bidirectional promoters are enriched among haploinsufficient genes not only in the heart but in other tissues. I focused on the lncRNA that is divergently transcribed from the *Tbx5* promoter and showed that knockdown of this transcript results in embryonic lethality with hypoplastic ventricle. In Chapter II, I found that the expression pattern of *Tbx5* at the mRNA and protein levels appear to differ. By examining RNA-seq analysis, I found that the expression pattern of one of the isoforms of *Tbx5* is similar to that of protein. Examination on these isoforms confirmed that the expression pattern of this isoform is quite similar to that of Tbx5 protein both spatially and temporally. Furthermore, a relatively highly expressed promoter with inconsistent expression pattern is indicated to have very low translational ability. However, KO of the isoform with consistent expression pattern in ES cells did not result in the depletion of Tbx5 protein. Although the exact mechanism of Tbx5 protein regulation is not still understood, I was able to show that *Tbx5* is under complex transcriptional and translational control.

Chapter I

Comprehensive identification of long non-coding RNAs in mouse heart development and the analysis of the divergent lncRNA accompanying a cardiac transcription factor gene *Tbx5*

Introduction

Morphogenesis is a complex process in which appropriate cell types are differentiated and positioned at the right place and at the proper timing. The surprising reproducibility of developmental processes is underpinned by the robustness of genetic program (Bateson and Gluckman, 2012). However, in spite of the high robustness under intact genetic condition, the program can be easily collapsed under genetic abnormalities, since some genes require both alleles for proper function (i.e. haploinsufficiency) (Seidman and Seidman, 2002). In the heart, even slight failures of the program lead to congenital heart diseases (CHDs), which occur frequently as high as around one in a hundred births (Hoffman and Kaplan, 2002). Genetic studies have shown that many of the transcription factor genes in heart are regulated in a highly spatiotemporal manner and show haploinsufficiency (Srivastava, 2006). However, it has not been well understood how such intricate control of gene expression in terms of expression pattern and dosage is achieved.

Comparative genomics have shown that the complexity of the body plan and the proportion of non-coding region of genome have clear positive correlation evolutionarily (Taft and Mattick, 2003). It is now generally understood that the vast broad "junk" area is necessary for gene regulation in a fine and complicated way (Pennisi, 2012). Many evo-devo studies support this view by suggesting that evolution of multicellular organisms is largely driven by the adjustments in transcriptional regulators such as enhancer elements, but not by functional evolution of protein-coding genes (Carroll, 2008). Recent advancements in genomics and transcriptomics have demonstrated that nearly a half of the mammalian genome is actually transcribed into RNAs (Carninci et al., 2005). Long non-coding RNA (lncRNA) is an emerging class of RNA that is generally defined by lacking the ability to produce functional proteins and being longer than 200 nucleotides. Many of these molecules have been demonstrated to work as transcriptional or translational regulators (Wang and Chang, 2011). Some lncRNAs are known to recruit epigenetic regulators to specific loci in the genome to modulate transcription. For example, the classical lncRNA *Xist* recruits polycomb repressive complex 2 (PRC2) to the X chromosome in *cis* to inactivate one of the two X chromosomes to achieve dosage compensation (Brockdorff, 2013). Many lncRNAs studied so far have been indicated to bind to epigenetic factors and recruit them to defined genomic loci. Moreover, a large proportion of lncRNAs were suggested to bind to PRC proteins, possibly epigenetically silencing define loci (Khalil et al., 2009). Other lncRNAs function as post-transcriptional modulators of gene expression through the formation of duplexes with mRNA to inhibit translation by RNAi (i.e. antisense transcripts) (Faghihi and Wahlestedt, 2009) or through the inhibition of miRNAs by working as so-called sponges (Ebert and Sharp, 2010) or through controlling splicing (Bardou et al., 2014). Although much attention has been paid to lncRNAs recently, the low conservation of sequence among orthologs and the difficulty of determining three-dimensional structures make it difficult to classify these molecules functionally and evolutionally. Their biochemical characters (e.g. strong nonspecific binding to proteins) also make it difficult to dissect their precise molecular functions (Davidovich et al., 2015; Novikova et al., 2013). Many lncRNAs show stage- and tissue-specific expression patterns, suggesting their roles in development (Gloss and Dinger, 2015).

Several lncRNAs that function in the mammalian heart development have been reported, but the identification and characterization of lncRNAs in the mammalian heart is still insufficient (Anderson et al., 2016; Grote et al., 2013; Klattenhoff et al., 2013). Considering the regulative nature of lncRNAs, they are thought to be the key components in solving the aforementioned problems regarding the developmental fragility in mammalian hearts.

Here in chapter I, I report that key cardiac transcription factors possess lncRNAs in close proximity, particularly as bidirectional promoter transcripts and that one lncRNA near *Tbx5*, namely *Tbx5ua*, is required for heart development. *Tbx5ua* knockdown mice showed abnormally thin ventricular walls and were embryonic lethal.

Materials and Methods

RNA-seq

Total RNAs from embryonic and adult mice were extracted using Sepasol-RNA I Super G (Nacalai #09379-55). The cDNA libraries for paired-end RNA-seq for the screening of lncRNAs were prepared from 1ug of RNAs with Truseq Stranded Total RNA Library Prep Kit (Illumina #RS-122-2201) according to Illumina's instructions. The cDNA libraries for tetraploid chimeric mice were prepared by Smart-Seq2 protocol according to the original paper (Picelli et al., 2013) with 12 cycles of preamplification and 9 cycles of enrichment PCR.

qRT-PCR

Total RNA was extracted with Sepasol-RNA I Super G (Nacalai #09379-55). cDNA samples were prepare using RevaTra Ace qPCR RT Master Mix with gDNA remover (Toyobo #FSQ-301). Real-time PCR was performed with SYBR Premix EX Taq II (Takara #RR820). The PCR conditions were as follows: 95°C for 30 s followed by 50 cycles of 95°C for 5 s and 60°C for 30 s, and subsequent dissociation curve measurement. I used *Gapdh* as internal control. Gene-specific primers are listed below.

Gapdh

5'-TGTGTCCGTCGTGGAT-3'

5' -TTGCTGTTGAAGTCGCAGGAG -3'

Tbx5

5'-GCCTGGAAACCTTGCTTCGATA-3'

5'-ACGTGTAAGCCGGGAGCTTG-3'

Tbx5 (bidirectional promoter isoform)

5'-AGCTACCTCGCCTCAGTGAG-3'

5’-TTCGTGGAACTTCAGCCACAG-3’

Tbx5ua

5’-AAAGAGAGCTGCCACTCCTG-3’

5’-TCTGTCACATCCAACACCAA-3’

Culture and genome editing of ES cells

ES cells were cultured on MEF feeder in ES culture medium (i.e. Knockout DMEM (Gibco #10829018), 20% Knockout Serum Replacement (Gibco #10828028), 1 * GlutaMAX (Gibco #10566016), 1 * NEAA (Sigma #M7145), 1mM sodium pyruvate (Gibco # 11360070), 10^{-4}M 2-Mercaptoethanol, 1000 U/ml LiF (Wako #198-15781)

The gRNA target sequence to induce double strand break was 5’-GTCACTGCCGCTCCAATCCTCGG-3’.

We designed the gRNA with Cas-OFFinder (http://www.rgenome.net/cas-offinder/) to minimize the number of off-target sites. Homology directed repair donors were constructed so that the *Neo*^{R} or *EGFP* expressing cassette was flanked by ~1000 bp 5’ and 3’ homologous arms cloned from genomic DNA. ES cells were transfected with Cas9 expressing plasmid, gRNA expressing plasmid and the donor plasmid along with non-gRNA expressing negative control. After two days, the ES cells were passaged onto SNL feeder cells and cultured for 8 days with 250 µg/ml G418 (Nacalai #16513-26) and surviving EGFP-positive colonies were manually picked up. After one more cycle of single colony pick up to ensure that the ES cells are clonal, they were subjected to cell permeable Cre treatment (Münst et al., 2009) to remove the selection cassettes, and then EGFP-negative colonies were picked up to obtain cells without selection cassettes. Finally, the ES cells from each colony were genotyped and karyotyped.

Generation of tetraploid chimeric mice

Generation of chimeric mice was performed as described previously (Tanimoto et al., 2008) in collaboration with Laboratory Animal Resource Center, University of Tsukuba and Tokyo Medical and Dental University.

Histology

Immunohistochemistry for Tbx5 were performed as follows. Antigen retrieval was performed by microwaving the sections in 10mM citrate acid pH 6.0. Then they were permeabilized for 10 minutes in 0.2 % Triton X in PBS at RT. Blocking was performed with 10% Blocking One (Nacalai #03953-95) in PBST. Tbx5 antibody (Santa Cruz Biotechnology #sc-17866) was diluted 1/100 in 5% Blocking One/PBT and second antibody (Invitrogen #A-11037) was diluted 1/200.

In situ hybridization of *Tbx5* gene was performed as follows. First, cryosections were permeabilized in 0.2 N HCl for 15 minutes. After washing with PBT three times, the sections were re-fixed with freshly made 4% PFA for 15 minutes. After washing, the sections were hybridized with DIG labeled probes at 70ºC for ON. The next day, the sections were washed with 0.2* SSC three times. After blocking the sections with 10 % sheep serum for 1 hour, 1/1000 diluted anti-DIG-AP Fab fragment (Roche #11093274910) were added and incubated for an hour at RT. After washing with TBST, the sections were washed with NTMT and colored with BM purple.

Statistics

For Welch's *t*-test, we confirmed that each group is approximately normally distributed by Shapiro-Wilk test.

Results

To identify novel lncRNAs that are specifically expressed during heart development in mice, I extracted total RNA from the ventricles of embryonic day (E) 10.5 and E13.5 and 8 weeks-old mice and prepared cDNA libraries, that were subjected to paired-end 2 * 100bp RNA-seq. The resulting read count was approximately 40 M reads for each sample. The obtained reads were mapped to the mouse genome (mm10) with Tophat2 (Kim et al., 2013), and the mapped reads were assembled using Cufflinks (Trapnell et al., 2010) with and without UCSC transcript annotations. Because many of the currently known functional lncRNAs are spliced and because it is difficult to confirm the existence of non-spliced transcripts unless they are expressed at very high levels, I focused on spliced lncRNA candidates in my analysis. I set the lower limit of expression at fragments per kilobase of exon per million mapped fragments (fpkm) of 1, because above that level, the accuracy of the reconstruction of known transcripts without the transcript reference was sufficiently high (Figure 1). I also checked if exons of known genes were incorrectly annotated as lncRNAs. I found that the direction of a majority of the lncRNAs that are located within 10,000 bp from known genes are in the opposite direction from them (225 vs 86), suggesting that such mis-annotations are rare.

From the assembled transcripts, already known mRNAs or functional RNAs that are not generally classified as lncRNAs (e.g., snoRNA and tRNA) were removed, and I also omitted RNAs that have CDS longer than 1/3 of their total length according to the standard of Ensembl, since they are potentially protein-coding transcripts. As a result, I was able to identify 787 candidates of spliced lncRNAs. To omit lncRNAs that are ubiquitously expressed without tissue specificity, I examined the expression of the obtained candidates in the mouse brain. Because the brain is an organ that diverges from the heart at a very early developmental stage and originates from the ectoderm, whereas the heart originates from the mesoderm, I used the brain as a reference organ. I here just wanted to exclude lncRNAs that are expressed with no tissue specificity and did not intended to find lncRNAs that are exclusively expressed in the heart since many genes are known to function differently according to the context of the tissues. The comparison revealed that 316 of the identified spliced lncRNA candidates were selectively expressed in the heart (Figure 2). I checked the expression of these genes in the kidney and the liver and found that only 34 were expressed in both of them, and 213 of them were expressed only in the heart. I found that some lncRNA candidates were expressed in a stage-specific manner, suggesting that they may have roles in heart development or maturation.

Many of the cardiac transcription factor genes have neighboring lncRNAs

First, I plotted the distribution of the expression levels of the obtained lncRNAs at E10.5 along with that of mRNAs. Consistent with the previous reports, the expression levels of lncRNAs were much lower than those of mRNAs. Interestingly, almost no heart-selective lncRNAs had fpkm values higher than 10 (Figure 3). Since many lncRNAs are known to modulate the transcription of neighboring genes in *cis*, I tried to identify the neighboring genes of the identified lncRNAs. The distribution of the distances from the transcriptional start site (TSS) of lncRNAs to the nearest genes was examined (Figure 4). Overall, the distance distribution of all obtained lncRNAs seemed to be similar to that of mRNAs. However, heart-selective lncRNAs were unexpectedly found to be at greater distances to protein-coding genes. The median distances were 12,626, 12,024 and 22,522 for mRNAs, all lncRNAs and heart-selective lncRNAs, respectively ($p \approx 3.9 * 10^{-1}$ for all lncRNAs vs. mRNAs; and $p \approx 8.4 * 10^{-8}$ for heart-selective lncRNAs vs. mRNAs, Mann-Whitney U test). Next, I examined what types of genes were enriched among the genes closest to lncRNAs. To this end, I conducted a gene enrichment analysis on such protein coding genes using the DAVID bioinformatics tool (http://david.ncifcrf.gov/) (Huang et al., 2009) and found that transcription factor genes were enriched among genes near lncRNAs in the heart. I also found that the genes associated with heart development were more strongly enriched among the genes near heart-selective lncRNAs when compared to the genes near lncRNAs lacking tissue specificity (Table 1).

I next tried to identify antisense lncRNAs and lncRNAs from bidirectional promoters. Bidirectional promoters produce two transcripts in a head-to-head divergent manner and attract a lot of attention as important sources of lncRNAs. Preceding studies have revealed that many of them regulate the genes with which they share promoters. I evaluated lncRNAs that had their TSS within 3,000 bp from the promoter of protein coding genes as lncRNAs driven by bidirectional promoters. Consistent with the result that the distance between heart-selective lncRNAs and their neighboring genes is generally greater than the distance between all lncRNAs and their neighboring genes, both antisense lncRNA and bidirectional lncRNA were enriched among lncRNAs that are expressed both in the heart and the brain (Figure 5). Some of the lncRNAs and neighboring genes were judged to be both antisense and bidirectional because of alternative promoter isoforms.

Next, in order to clarify the relationship between mRNAs and their bidirectional lncRNAs, I calculated Pearson correlation coefficients between the log2-transformed expression levels of the bidirectional promoter pairs over the course of development. The distribution of the correlation coefficients is plotted in Figure 6. Many gene pairs clearly show positive or negative correlation, and the positive correlation appears to be dominant (Figure 7).

By searching the protein coding genes that are close to lncRNAs, I found many transcription factor genes that have critical functions for heart development (i.e., *Tbx5, Tbx20, Nkx2-5, Gata4, Gata6, Sall4, Hand1, Hand2, Wt1, Nr2f1, Irx3* and *Irx5*). Notably many of these lncRNAs were bidirectional lncRNAs (i.e., *Tbx5, Tbx20, Nkx2-5, Gata6, Sall4, Hand1, Hand2, Wt1, Nr2f1, Irx3* and *Irx5*). Some of these lncRNAs (e.g., those divergent to *Irx5, Gata6 and Wt1*) are expressed in the kidney or in the liver, and in such cases divergent genes are also expressed, suggesting that the expression of bidirectional pairs are correlated not only temporally but spatially. I examined the conservation of these lncRNAs near transcription factors by searching the RefSeq database and found that at least some lncRNAs were conserved in the human genome (*Tbx5, Nkx2-5, Hand2, Gata6, Wt1* and *Nr2f1*) (Figure 8) and that the bidirectional lncRNA to *Tbx5* (*Lnc125*) was even conserved in chicken, which diverged from mammals 400 million years ago. Here, I judged bidirectional lncRNAs to be conserved solely based on the existence of transcripts at the corresponding loci, since the sequences of lncRNAs are known to evolve rapidly.

Because haploinsufficient transcription factor genes seem to be highly enriched among the genes that are in close proximity to divergent lncRNAs, I determined whether the enrichment was limited to the heart or whether it was more generally true (Jay et al., 2005; Moskowitz et al., 2004). Using the mouse RefSeq transcript database (GRCm38.p3) and the paper that comprehensively identified haploinsufficient genes (Dang et al., 2008), I tried to determine the proportion of genes with bidirectional lncRNAs among all genes and among haploinsufficient genes. I indeed found that haploinsufficient genes were significantly more enriched among genes with bidirectional lncRNAs ($p = 3.4 * 10^{-5}$ based on hypergeometric distribution) (Figure 9). To exclude the possibility that the tissue specificity of bidirectional lncRNAs and haploinsufficient genes generates pseudo-correlations, I calculated the proportion of housekeeping genes among all genes and among haploinsufficient genes and showed that the proportions were not significantly different (Figure 10) (Eisenberg and Levanon, 2013).

Generally, the conservation of lncRNAs across species is very low compared to protein-coding transcripts. However, the *Tbx5*-divergent lncRNA is observed among a wide range of species. *Tbx5* is also a dosage-sensitive gene (Moskowitz et al., 2004). These findings prompted me to examine the function of the *Tbx5*-divergent lncRNA.

Analysis of the *Tbx5*-divergent lncRNA

Tbx5 is a transcription factor that is known to be essential for the development of the heart and forelimb. Holt-Oram syndrome is a dominant disorder caused by a single-allele mutation of *TBX5* and is characterized by hypoplasia of the forelimb, abnormalities in the thumb, and atrial and/or ventricular septal defects (Bruneau et al., 2001; Holt and Oram, 1960; Li et al., 1997). Importantly, the phenotypes of Holt-Oram syndrome show a high degree of variance, indicating that the dose of *TBX5* is crucial in normal heart development (Mori and Bruneau, 2004).

Hereafter I will call this lncRNA as *Tbx5 upstream antisense product* (*Tbx5ua*). *Tbx5ua* homolog is present in human genome annotation and it is named *TBX5-AS1* (Figure 8). When compared with humans *TBX5-AS1* the sequence of *Tbx5ua* is relatively well conserved at the 5' region, although it is hard to judge if this conservation is the consequence of functional demand since both the promoter and enhancer elements also exhibit a high degree of conservation. *Tbx5ua* is transcribed from one of the promoters of *Tbx5* in the opposite direction and overlaps with the intron of one of the *Tbx5* isoforms (Figure 11, RefSeq: XM_006530282.3, isoform 1). RNA-seq data suggest that *Tbx5ua* is alternatively spliced, producing several isoforms (Figure 11). In Figure 11, I labeled isoforms that were identified in our RNA-seq experiment in at least one stage. Reanalysis of previously published intact/nuclear RNA-seq of cardiomyocytes revealed that *Tbx5ua* is not clearly localized (Figure 12) (Preissl et al., 2015). Previous study reports that quite a few lncRNAs actually exhibit this type of non-localized expression patterns (Cabili et al., 2015).

I first quantified the expression level of the transcript in the heart ventricle, atrium and forelimb during normal development by quantitative RT-PCR (Figure 13). I found that the expression level of *Tbx5ua* was increased in the ventricle as development progressed, which was inconsistent with the expression pattern of *Tbx5*. I also examined the expression level of the *Tbx5* isoform that is also transcribed from the bidirectional promoter (Isoform 2, RefSeq: XM_006530280.1). The expression level of that isoform was stable during the entire developmental process, which was also different from the expression pattern of *Tbx5ua* (Figure 14). Next, I compared the expression level of the lncRNA in both of the ventricles at E11.5 because it is well-known that the expression level of *Tbx5* is higher in the left ventricle than in the right ventricle and that the steep gradient is crucial for establishing a proper ventricular septum (Koshiba-Takeuchi et al., 2009; Takeuchi et al., 2003). I observed that *Tbx5ua* expression was almost the same between the left and right ventricles at E11.5, while I confirmed the differential expression level of *Tbx5* (Figure 15). These results suggest that *Tbx5ua* is not just a byproduct of *Tbx5* and is regulated separately as a different product.

Tbx5ua-knockdown (KD) mice were embryonic lethal with severe abnormalities in the heart

To determine the function of *Tbx5ua*, I knocked down both alleles of *Tbx5ua* by inserting three tandem copies of bovine growth hormone polyadenylation site (3xpA) at the second exon to prematurely stop transcription in C57BL/6J-derived ES cells using the CRISPR/Cas9 system (Figure 16) (Tanimoto et al., 2008). By tetraploid complementation, I obtained completely ES cell-derived mouse embryos from two ES cell lines and their phenotypes were consistent between lines. The expression level of *Tbx5ua* in E9.5 KD mice was strongly repressed to approximately 1/10 of that in control embryos, showing successful knockdown (Figure 17). Although the expression levels of *Tbx5* and *Tbx5ua* seemed to be anticorrelated in the heart during development (Figure 13), KD of *Tbx5ua* did not result in the increase of *Tbx5* expression level. I also showed that the expression levels of the different *Tbx5* isoforms that are transcribed from all three promoters were not significantly changed (Figure 18).

At E9.5, chimeric KD embryos appeared to be normal except the heart. KD mice obtained from the first line of ESC were slightly larger than WT mice, but that was not the case in the second line (Figure 19). It is common that chimeric embryos obtained from tetraploid complementation vary in their stages, which makes analyzing their phenotype difficult at E9.5. When I investigate the phenotype of the KD hearts, the right ventricle was hypoplastic at E9.5 (Figure 20A, 21A). Hematoxylin and eosin (HE) staining of the cryosections indicated that the ventricular walls of E9.5 KD mice were irregular and the pre-ventricular septal region appeared to be thinner, which was not verified statistically but at least consistent between the two ESC lines (Figure 20B, 21B, 22). The number of pictures for whole embryo and that for section were different because the sections of some embryo were not properly prepared due to my mishandling. None of the embryos showed a visible abnormality in the forelimbs, which is observed in *Tbx5*-deficient embryos. By E13.5, all of the KD embryos were dead with a pale body. One of the WT embryos had an abnormal outflow tract, which was likely to be caused by tetraploid complementation. Nonetheless, the overall phenotype was markedly different between WT and KD. The hearts of KD embryos showed severe ventricular hypoplasia (Figure 23), which was probably the cause of the lethality. The forelimbs seemed completely normal even at this stage, which was a significant difference between the phenotype of the *Tbx5ua* KD mice and that of the mouse model of Holt-Oram syndrome (i.e., *Tbx5* heterozygous knockout) (Figure 23). The phenotypes among KD embryos seemed to be heart-specific, suggesting that they are attributed to genomic modification.

In situ hybridization of *Tbx5* revealed normal mRNA expression in the KD ventricle (Figure 24A). *In situ* hybridization of *Nppa*, which often exhibits altered expression pattern in embryos with abnormal morphogenesis, showed an expanded expression around the pre-ventricular septal region of KD embryos (Figure 24B).

To comprehensively investigate the genes affected by *Tbx5ua* knockdown, I performed RNA-seq with the RNAs extracted from the ventricles of tetraploid chimeric embryos derived from either KD or WT ES cells. I used three embryos for each group and used the Smart-Seq2 protocol to generate libraries from the small amount of RNA. By gene ontology analysis, I found that the genes involved in heart development were significantly enriched among the genes that were determined to be significantly changed (False Discovery Rate; FDR < 0.10, Figure 25A). However, none of the structural genes that are important for cardiomyocyte contraction were changed (Figure 25C), suggesting the possibility that *Tbx5ua* has a critical role in morphogenesis rather than in cell differentiation. Finally, I conducted principal component analysis (PCA) on the RNA-seq data (Figure 25D). The two groups were evidently distinguished only by considering the first principal component.

Discussion

This study revealed that many cardiac transcription factor genes had neighboring lncRNAs, especially bidirectional ones. The clear correlation of the expression level seen in some bidirectional pairs suggests their regulatory roles. Haploinsufficiency among many cardiac transcription factors indicates the possibility that these neighboring lncRNAs are important in stabilizing the expression level of the neighboring genes. Since the expression level of transcription factors is generally low, even a small change could lead to severe consequences. The bidirectional lncRNAs might function to precisely control the expression level of those genes encoding expression level-sensitive protein product. An alternative hypothesis is that these transcription factors may form an optimal transcriptional environment for lncRNAs to evolve. As some transcription factor genes can cause the direct lineage reprogramming, they are thought to define the cell types. Thus, making use of these preexisting transcriptional environments might be a cost-efficient way to evolve cell type specific lncRNAs. Some studies have demonstrated that bidirectional transcription is inevitable and so-called transcription ripple effect exists (Almada et al., 2013; Ebisuya et al., 2008). These findings also support my idea by showing that the preexisting transcriptional environment offers precursor transcripts that potentially evolve into defined, functional ones. In summary, active transcription factor genes could provide an environment for lncRNAs to evolve by offering cell type-specific and active epigenetic environment.

I identified *Tbx5ua* that is conserved from mammals to birds. Comparison of the sequence of *Tbx5ua* between mouse and chicken showed less similarity, but it does not mean that the function is not conserved as the previous examples, i.e. the exact conservation at the sequence level is not necessarily required for functional conservation of lncRNAs (Okamoto et al., 2011; Tripathi et al., 2013). *Tbx5ua* was not found in NCBI genomic annotations of reptiles, amphibians or fish at the corresponding loci. In fact, by conducting the reanalysis on the publicly available RNA-seq data

(GSE41338) (Rabbow et al., 2012) that include RNA-seq of the adult heart of chicken, anole and frog, I could confirm that *Tbx5ua* is expressed only in chicken among these species at the adult stages (Figure 19). It is interesting that *Tbx5ua* is conserved in the two-ventricle animals, in which complete ventricular septum exists, but not in non-septated animals. There is a possibility that acquisition of *Tbx5ua* might have contributed to the evolution of complete ventricular septum.

I showed that *Tbx5ua* lncRNA is required for proper heart development. Since I knocked down *Tbx5ua* by prematurely terminating the transcription, the formation of transcription complex at the transcription start site is not inhibited. Thus, if the transcription of *Tbx5ua* itself is important for altering the local transcriptional environment, my KD scheme is not sufficient to assess the true function of *Tbx5ua*. Although preliminary, my data suggested that the expression pattern of Tbx5 protein is altered in the KD mice (Figure 27). While I do not have any evidences supporting the direct roles of *Tbx5ua* on *Tbx5*, the function of *Tbx5ua* might be atypical for a divergent lncRNA since many of such lncRNAs like *Upperhand* were shown to alter the transcription of neighboring genes as repeatedly stated. How the left-sided expression of *Tbx5* is regulated is an unsolved important issue to understand the molecular mechanism of heart development (Smemo et al., 2012).

In conclusion, this study has revealed that many genes, particularly transcription factor genes, involved in heart development possess lncRNAs in their close proximity. Furthermore, many of these lncRNAs were transcribed from the promoter of protein-coding genes divergently. The fact that bidirectional lncRNAs are enriched among haploinsufficient genes indicated their functional roles for the regulation of dose-sensitive genes.

Chapter II

Differential expression pattern and translational ability among *Tbx5* isoforms underlie the left-sided expression of Tbx5 protein during mouse heart development

Introduction

Development of mammalian hearts starts with the differentiation of endocardial tubes and their fusion. After tubular hearts are formed, they bend to differentiate into distinct regions (Srivastava, 2006). The formation of ventricular and atrial septum is a functionally important yet error-prone process. Indeed, ventricular and atrial septal defects (VSD and ASD respectively) comprise a large fraction of congenital heart disease (CHD) occurrence (Minette and Sahn, 2006). Although many genes are known to cause VSD when mutated, most CHD cases including VSD cannot be explained by such mutations in coding regions, indicating the importance of non-coding regulatory elements (Postma et al., 2015; Wamstad et al., 2014). Actually, many genes involved in the heart development, especially transcription factor genes, are dose-sensitive and require precise patterning of expression (Bruneau, 2008; Olson, 2006; Seidman and Seidman, 2002).

Although the exact molecular mechanism underlying the development of ventricular septum is still elusive, it is well-known that *TBX5* plays a central role in the process. *TBX5* is a transcription factor gene, of which mutations cause a dominant disorder Holt-Oram syndrome (HOS), which is characterized by ventricular and/or atrial septal defects and the hypoplasia of forelimb (Bruneau et al., 2001; Holt and Oram, 1960; Li et al., 1997; Mori and Bruneau, 2004). As a single allele mutation of *TBX5* causes HOS, *TBX5* is a haploinsufficient, dose-sensitive gene like many other cardiac transcription factors. Previous studies have shown that the left-sided expression of *TBX5* is required for the proper formation of ventricular septum (Takeuchi et al., 2003). Indeed, both knockout and overexpression of *Tbx5* in the entire ventricle resulted in the complete loss of the ventricular septum, showing the importance of the left-sided expression (Koshiba-Takeuchi et al., 2009). Moreover, the expression patterns of *Tbx5* among species without ventricular walls show no left-right difference, apparently explaining the evolution of the ventricular wall among vertebrates (Jensen et al., 2013; Koshiba-Takeuchi et al., 2009). In spite of all the studies, however, it still remains unknown how the left-right difference in the ventricular expression of *Tbx5* is regulated at the molecular level (Smemo et al., 2012).

I found that the protein expression and the mRNA expression of *Tbx5* during development are apparently different. Through reanalysis of the RNA-seq data, I found that *Tbx5* has three promoters and one of them, which I call promoter A, was likely to contribute to the majority of Tbx5 protein expression in the ventricle. This distal promoter is located around 40,000 bp upstream of the second exon and has not attracted much attention until now. I found that the promoter A isoform expression level was clearly correlated with the protein expression level of *Tbx5*, while the other promoter isoforms did not. This result led to the hypothesis that the *Tbx5* mRNA isoform from the promoter previously thought to be the “main” promoter has a quite low translational ability. Indeed, the 5’UTR of this isoform was demonstrated to have strong ability to repress translation. Thus, of the two highly-expressed isoforms of *Tbx5*, one with quite low translational ability exhibits the expression pattern different from the protein expression pattern, providing an explanation for the inconsistency between transcription and translation of *Tbx5*.

Materials and Methods

Immunohistochemistry of Tbx5

For tissue sections, IHC was performed as follows. Antigen retrieval was performed by microwaving the sections in 10mM citrate acid pH 6.0. Blocking was performed with 10% Blocking One (Nacalai #03953-95) in PBST and Tbx5 antibody (Santa Cruz Biotechnology, sc-17866) was diluted 1/100 in 5% Blocking One/PBT and second antibody (Invitrogen, A-11037) was diluted 1/200. For cell culture, cells were fixed 10 min with 4% paraformaldehyde. Tbx5 antibody was diluted 1/100 and cTnT antibody (Santa Cruz Biotechnology, sc-20025) was diluted 1/500.

Western blot

Total proteins were extracted by lysing tissues in RIPA buffer. Blocking was performed with Blocking one and Tbx5 antibody (Santa Cruz Biotechnology #sc-17866) was diluted 1/200 and Gapdh antibody (Chemicon #MAB374) was diluted 1/1,000 in 5% Blocking one in TBST. Second antibody was diluted 1/10,000. Finally, luminescent signal was detected with ECL Western blotting detection reagents (Amersham #RPN2109) using ImageQuant LAS4000mini (GE lifescience).

qRT-PCR

Total RNA was extracted with Sepasol-RNA I Super G (Nacalai #09379-55). cDNA samples were prepared by using RevaTra Ace qPCR RT Master Mix with gDNA remover (Toyobo #FSQ-301). Real-time PCR was performed with SYBR Premix EX Taq II (Takara #RR820). I used *Gapdh* as internal control. The primers are listed below.

Gapdh

5’-TGTGTCCGTCGTGGAT-3’

5’ -TTGCTGTTGAAGTCGCAGGAG -3’

Tbx5

5’-ATGGCCGATACAGATGAGGG-3’

5’-TTCGTGGAACTTCAGCCACAG-3’

Tbx5 (promoter A isoform)

5’-GTCCAGTGTTCATCCGGTCA-3’

5’-TTCGTGGAACTTCAGCCACAG-3’

Tbx5 (promoter A isoform full exon cloning)

5’-GTCCAGTGTTCATCCGGTCA-3’

5’-CTCCGTGCTGGAACATTCCTC-3’

Tbx5 (promoter B isoform)

5’-AGCTACCTCGCCTCAGTGAG-3’

5’-TTCGTGGAACTTCAGCCACAG-3’

Tbx5 (promoter C isoform)

5’-GAATGCATCCCCCTGT-3’

5’-TTCGTGGAACTTCAGCCACAG-3’

Renilla luciferase

5’- AAGAGCGAAGAGGGCGAGAA-3’

5’- TGCGGACAATCTGGACGAC-3’

Firefly luciferase

5’- CAACTGCATAAGGCTATGAAGAGA-3’

5’- ATTTGTATTCAGCCCATATCGTTT-3’

Luciferase assay

Hek293T cells were passaged into 48 well plates the day before transfection. 28.5 ng of modified firefly luciferase plasmids and 28.5 ng of renilla luciferase plasmids as internal control were transfected with Lipofectamine 2000 (Invitrogen #11668019). After 36 hours, the cells were lysed and the luciferase activities were measured with Dual-Luciferase Reporter Assay System (Promega #E1980).

Results

Protein and mRNA expression patterns of *Tbx5* in the mouse ventricle appear to be different

Since the expression pattern of Tbx5 protein has not been well-characterized, I conducted immunohistochemistry of Tbx5 at several stages and confirmed the graded expression at E9.5 when the left-right difference in the ventricle is known to appear (Figure 28A). At E13.5, the expression of Tbx5 was detected in the trabeculae whereas much weaker signals were observed in the ventricular walls and in the ventricular septum(Franco et al., 1998). It is well known that the cells in the wall and in the trabeculae express different genes. Unexpectedly, many cells in the right ventricle expressed Tbx5 at this stage, though the number of Tbx5-positive cells was apparently smaller than that in the left ventricle (Figure 28B). At P2, Tbx5 signal was detected in the atrium but not in the ventricle (Figure 28C). Interestingly, the expression level of Tbx5 in the atrium was different between the left and the right. This asymmetrical expression might contribute to the functional and structural differences of the right and left atrium. To confirm the low expression level of Tbx5 in the ventricle of the postnatal mice, I performed western blot and indeed confirmed that the expression of Tbx5 protein was at an undetectable level, supporting my histological results (Figure 29).

I next examined the expression pattern of *Tbx5* at the mRNA level by quantitative reverse transcription PCR (qRT-PCR). At E9.5, the expression level of *Tbx5* in the left ventricle was about twice as that in the right ventricle (Figure 30A). Moreover, qRT-PCR over development revealed that the ventricular expression level of *Tbx5* mRNA in postnatal mice was comparable to that in embryos (Figure 30B). Thus, the expression levels of protein and mRNA of *Tbx5* appeared to be distinct.

Three isoforms of *Tbx5* mRNA show different expression patterns spatiotemporally

I hypothesized that different isoforms of *Tbx5* could explain the seeming discrepancy between mRNA and protein levels. To investigate isoforms of *Tbx5,* I analyzed RNA-seq data on total RNAs extracted from the ventricles of E10.5, E13.5 and adult mice (GEO: GSE93324). The analysis revealed that *Tbx5* has three promoters (Figure 31A, B). The one nearest to the second exon has been commonly regarded as the "main" promoter. I hereafter call this promoter, promoter C, and the exon transcribed from this promoter, exon C. Just 2,000 bp upstream of this promoter there is another promoter, which I call promoter B. Although some genomic annotations on mouse genome suggest that exon B is spliced to form a junction with exon C, my RNA-seq data did not include such a read. The third promoter, promoter A, was located about 38,000 bp upstream of promoter C. RNA-seq data detected this isoform only in embryos (Figure 31A). Therefore, the expression pattern of the promoter A isoform appeared to be similar to that of Tbx5 protein, which was also undetectable in adult ventricles. In fact, promoter A is located around the region that was identified as one of the enhancers that shows an expression pattern close to that of *Tbx5* in Smemo *et al* (Smemo et al., 2012). The start codon of *Tbx5* is located in the second exon, so these isoforms only differ in their 5' UTRs, but not in their protein coding sequences. I examined the number of reads that contain splice junctions to estimate the relative expression level among three isoforms. Reads containing "exon B-exon 2" splice junction was around 1/10 of all reads containing "exon 1-exon2" junctions in E10.5, indicating low contribution of this isoform (Figure 32).

All the three promoters also exist in the human and chicken genomic annotations, but not in zebrafish and frog. Furthermore, reanalysis on publicly available RNA-seq data on human embryonic and adult ventricles (GEO: GSE78567) showed that the transcription of promoter A is restricted to embryos, whereas the other two isoforms are still expressed in adults, consistent with the case in mice (Figure 31A). These findings based on RNA-seq analysis suggested that the differences among the alternative promoter isoforms underlie the Tbx5 protein expression pattern.

I then performed qRT-PCR on the three isoforms in the ventricle during heart development.

The expression level of the promoter A isoform dropped drastically as the heart matures and was below the detection limit of qRT-PCR in postnatal mice, confirming the RNA-seq data. On the other hand, the expression levels of other isoforms were roughly constant during development, which contradicts the very low protein expression level in postnatal ventricles (Figure 33). Because RNA-seq analysis on the adult ventricles showed that the full-length *Tbx5* was actually transcribed, this raises the possibility that the promoter C isoform, which comprise the majority of the *Tbx5* transcription in postnatal mice, has a very low translational ability. In the postnatal atrium, where Tbx5 protein is detectable, I observed that all the three isoforms are expressed (Figure 34).

I next compared the mRNA expression levels in the left and right ventricle. As is well-known and I showed by IHC, Tbx5 protein is much highly expressed in the left ventricle. qRT-PCR on E9.5 mice demonstrated that the expression level of the promoter A product was much higher in the left ventricle while the promoter B and C isoforms did not show such left-right difference (Figure 35).

Thus, both the temporal and spatial expression pattern of promoter A isoform is in good accordance with the protein expression pattern of Tbx5 while the expression level of promoter C isoform seems to be constant irrespective of stage and region, suggesting the primary role of promoter A isoform in Tbx5 protein production. To check if promoter A is capable of producing transcripts coding full-length protein, I conducted PCR cloning and confirmed that full-exon transcripts are indeed obtained from promoter A (Figure36).

Low translational ability of the promoter C isoform is attributed to its 5’ UTR *in vitro*

To fully understand the regulation of Tbx5 protein expression, it is important to reveal the difference among three isoforms. I examined if 5’ UTRs of these three isoforms affect translational ability of the downstream ORF by using dual-reporter luciferase assay with HEK293T cells. 5’ UTR of each isoform was added to the firefly luciferase gene and the renilla luciferase gene was used as an internal control. The addition of exon C resulted in severe reduction of firefly luciferase activity while 5’ UTR of isoform A and B also inhibited luciferase translation to some degree (Figure 37). qRT-PCR on luciferase genes revealed that the addition of 5’ UTRs had a much smaller impact on the mRNA abundancies (Figure 38), suggesting that the translational activity is strongly impaired by the addition of exon C 5’ UTR.

To unravel what causes the translational inhibition, I compared the sequences of first exons of three isoforms and found that exon C has 3 AUGs before the correct translation start site, whereas exon A has 1 and exon B has 2. All the coding sequences from these upstream AUGs had stop codons before the correct AUG or shifted codon frames, prohibiting coding Tbx5 protein. This kind of missense coding frames in 5’ UTR are called upstream open reading frames (uORFs) and they are generally thought to down-regulate protein expression by inhibiting translation from the correct start codons. 5’ UTR of human *TBX5* also have multiple missense ATGs.

I examined the effect of these uORFs in exon C on protein expression by mutating all three AUGs to UUGs and conducted luciferase assay. The disruption of uORFs recovered the luciferase activity by about 5-fold but it did not result in full recovery of expression (Figure 41). Therefore, the uORFs in exon C partly explain low translational ability of exon C 5’ UTR, but there must be other unidentified factors.

Tbx5 protein expression was observed in isoform A-knockout ES cells

Since the expression pattern of promoter A isoform in the ventricle is quite similar to that of Tbx5 protein, I hypothesized that this isoform accounts for most of Tbx5 protein production. I knocked out promoter A isoform by deleting the whole exon A using the Crispr/Cas9 system in ES cells (Figure 42). I differentiated KO and WT ES cells *in vitro* by hanging drop method. Immunohistochemistry revealed that differentiated cardiomyocytes derived from promoter A KO ESCs expressed Tbx5 protein just as WT ESCs did (Figure 43). Although it is difficult to draw a definitive conclusion from my experiment, which produces different types of cardiomyocytes at an arbitrary ratio, it was shown that isoform A is not the only isoform that can produce a decent amount of Tbx5 protein as I first hypothesized.

Discussion

In chapter II, I first revealed that the expression patterns of *Tbx5* mRNA and protein are clearly different. The expression pattern of *Tbx5* mRNA is relatively homogeneous while that of protein is highly stage-and-region specific. I identified three highly expressed isoforms of *Tbx5* mRNA and found that their expression patterns are different. Furthermore, the expression pattern of isoform A is quite similar to that of protein in the ventricle. Isoform A has never been studied and characterized before. In fact, it was not even in RefSeq until June 2016. Preceding studies (Smemo et al., 2012) and public genomic annotations such as Refseq and Ensembl have considered promoter C as the main driver of *Tbx5.* Consequently, previous efforts to unravel the regulation of *Tbx5* focused on elements around promoter C. However, I demonstrated that the expression pattern of isoform C is clearly different from that of Tbx5 protein and that the 5' UTR of this isoform inhibits translation in HEK293 cells. These results of isoform C, the low expression level of isoform B in the ventricle and the similarity of the expression pattern of isoform A to that of Tbx5 protein strongly indicated that the major source of Tbx5 protein is isoform A. However, knockout of A isoform did not result in the loss of Tbx5 protein expression *in vitro*. Based on these results, I propose that the regulation of Tbx5 proteins is regulated both at the transcriptional and translational level in a very complex manner depending on the stage and region. Interestingly, promoter A is only observed among animals with complete ventricular septum (i.e., mammals and birds), suggesting that it played an important role during evolution. Although the function of each isoform is still elusive, they are evolutionarily conserved among mammals and birds, suggesting that they have specific functions. Indeed, the sequences of these three exons are well-conserved between humans and mice. To understand their functions, knockout studies using mice would be necessary.

Another unsolved issue is how exon C 5' UTR affects translational activity *in vitro*. I showed that uORFs affect translational ability, but the disruption of uAUGs did not result in the full recovery of luciferase activity. Repression through ubiquitous small RNAs or RNA biding proteins may be responsible, but it was hard to judge from the sequence if such mechanisms are involved (Araujo et al., 2012). It is possible that the de-repression of promoter C isoform by genetic mutations is a cause of previously uncharacterized VSD cases since the loss of Tbx5 gradient by overexpression is also known to lead to VSD (Koshiba-Takeuchi et al., 2009). However, isoform C can still account for a major amount of protein production in cardiomyocytes since the experiment was performed using HEK293 cells. If so, translational adjustment should be the key mechanism that realizes the strict spatiotemporal modulation of Tbx5 protein expression.

After the establishment of the ventricular wall, Tbx5 protein expression is weak in the ventricular wall and ventricular septum and strong in the trabeculae in the both ventricles. A higher expression level is also observed in the right atrium of postnatal mice (Figure 29C). Of the two atria, only the right atrium develops protruding muscle structures. Therefore, it is possible that the expression of Tbx5 protein at a later stage of development is linked to the development of protruding structures inside chambers.

Conclusion

In my Doctoral course studies, I tried to understand the molecular basis of robustness of development by focusing on the regulation of transcription factors, especially dose-sensitive ones. To this end, I first focused on lncRNAs in the mouse heart and showed that they are present near cardiac transcription factors, especially as bidirectional products. Since bidirectional promoter lncRNAs can theoretically monitor the activity of their own promoters, it is possible that they fine-tune the expression levels of the paired transcription factor genes. I picked up the lncRNA divergent to *Tbx5* and analyzed its function by making knockdown chimeric mice since *Tbx5* is a transcription factor gene which shows strong dose-sensitivity and the regulation of *Tbx5* is an unsolved important issue for understanding heart morphogenesis. Although the exact molecular function of this lncRNA was elusive, I showed that the lncRNA is necessary for normal development of the heart. The fact that many important cardiac transcription factor genes accompany bidirectional lncRNAs suggested that they play important roles in heart development. However, it remains unknown if they share general function. In fact, since the correlation between bidirectional gene-lncRNA pairs show both positive and negative tendencies (Figure 6), I guess that the function of such divergent lncRNAs is not universal. It is also unsure if these bidirectional pairs are really obtained from bidirectional promoters because if they really share the promoter, negative correlation at the transcriptional level is unlikely. Furthermore, many divergent lncRNAs like *Fendrr* and *Upperhand* are known to be necessary for the transcription of the paired genes, but it is unnatural because the transcription of the divergent pairs should be regulated simultaneously. One possible explanation is that bidirectional lncRNAs amplify and stabilize the expression of its promoter but is still is unsolved why some genes require such mechanism while others do not. These observations suggest that the true significance of divergent lncRNAs is yet to be understood and more detailed study on different lncRNAs is needed.

In chapter II, I tried to elucidate the regulation of *Tbx5* at the protein level. I showed that the expression pattern of *Tbx5* mRNA and protein are different, strongly suggesting that post-transcriptional control is involved in its regulation. Based on the finding that *Tbx5* has an embryo-specific promoter, I examined the expression patterns and translational abilities among different isoforms. The expression pattern of the embryo-specific promoter isoform is similar to that of Tbx5 protein, whereas the isoform with an expression pattern different from Tbx5 protein is suggested to have a very low translational ability. Interestingly, many cardiac transcription factor genes including *Tbx5* have uORFs in their 5’ UTRs, which possibly makes their half-life shorter and thus allows swift turn-off of gene expression when necessary. In spite of all the results supporting the view that isoform A is the main source of Tbx5 protein, KO of exon A resulted in normal Tbx5 production among differentiated ESCs. In conclusion, it is certain that Tbx5 protein expression in the ventricle is regulated both transcriptionally and translationally, but it is still not understood when and where each isoform contributes to proteins expression at what level.

The results of Chapter II also provide insights into the results of KD mouse experiment in Chapter I. When I quantified *Tbx5* in the whole ventricle, the expression levels of the three isoforms were not significantly affected. However, because the expression levels of them are low and the sample number is not large enough, the results of qRT-PCR and RNA-seq are not sufficient to make definitive conclusion as to detect the local change of their expression patterns. *Tbx5 in situ* hybridization is also not sensitive or quantitative enough to distinguish isoforms either. Thus, my results in Chapter I does not exclude the possibility that the expression pattern of *Tbx5* is altered in at least one isoform and the expression pattern of protein is affected.

In summary, these studies suggested that cardiac transcription factor genes develop various strategies such as neighboring lncRNAs and alternative isoforms to achieve precise and complex regulation. It would be important to understand the exact mechanism and significance of these strategies at the molecular level and to what extent these phenomena are general.

Acknowledgements

I would like to express my deepest gratitude to Dr. Atsushi Miyajima for giving me the opportunity in his lab to complete this work. I owe a lot to Dr. Akiyama, Dr. Tabata and other faculty members of Institute of Molecular and Cellular Biosciences at the University of Tokyo for their generous support for my study. I would like to thank Satoru Takahashi, Yoko Tanimoto, Tetsushi Furukawa, Jun K Takeuchi, Kazuko Koshiba-Takeuchi for their collaboration on my study. I am also grateful for my present and previous colleagues including ones in Dr. Taira's lab for their fruitful advices and discussions. My work was supported by Grant-in-aid from Japan Society for the Promotion of Science for Young Scientists and Graduate Program for Leaders in Life Innovation of the University of Tokyo. Lastly, I would like to show my greatest appreciation to my parents, family and friends for their warm support.

References

Almada, A.E., Wu, X., Kriz, A.J., Burge, C.B., and Sharp, P. a (2013). Promoter directionality is controlled by U1 snRNP and polyadenylation signals. Nature *499*, 360–363.

Anderson, K.M., Anderson, D.M., McAnally, J.R., Shelton, J.M., Bassel-Duby, R., and Olson, E.N. (2016). Transcription of the non-coding RNA upperhand controls Hand2 expression and heart development. Nature *539*, 433–436.

Araujo, P.R., Yoon, K., Ko, D., Smith, A.D., Qiao, M., Suresh, U., Burns, S.C., and Penalva, L.O.F. (2012). Before it gets started: Regulating translation at the 5' UTR. Comp. Funct. Genomics *2012*.

Bardou, F., Ariel, F., Simpson, C.G., Romero-Barrios, N., Laporte, P., Balzergue, S., Brown, J.W.S., and Crespi, M. (2014). Long Noncoding RNA Modulates Alternative Splicing Regulators in Arabidopsis. Dev. Cell *30*, 166–176.

Bateson, P., and Gluckman, P. (2012). Plasticity and robustness in development and evolution. Int. J. Epidemiol. *41*, 219–223.

Blake, W.J., KAErn, M., Cantor, C.R., and Collins, J.J. (2003). Noise in eukaryotic gene expression. Nature *422*, 633–637.

Blake, W.J., Balázsi, G., Kohanski, M. a, Isaacs, F.J., Murphy, K.F., Kuang, Y., Cantor, C.R., Walt, D.R., and Collins, J.J. (2006). Phenotypic consequences of promoter-mediated transcriptional noise. Mol. Cell *24*, 853–865.

Breckenridge, R.A., Zuberi, Z., Gomes, J., Orford, R., Dupays, L., Felkin, L.E., Clark, J.E., Magee, A.I., Ehler, E., Birks, E.J., et al. (2009). Overexpression of the transcription factor Hand1 causes predisposition towards arrhythmia in mice. J. Mol. Cell. Cardiol. *47*, 133–141.

Brockdorff, N. (2013). Noncoding RNA and Polycomb recruitment. RNA 429–442.

Bruneau, B.G. (2008). The developmental genetics of congenital heart disease. Nature *451*, 943–948.

Bruneau, B.G., Nemer, G., Schmitt, J.P., Charron, F., Robitaille, L., Caron, S., Conner, D.A., Gessler, M., Nemer, M., Seidman, C.E., et al. (2001). A murine model of Holt-Oram syndrome defines roles of the T-Box transcription factor Tbx5 in cardiogenesis and disease. Cell *106*, 709–721.

Cabili, M.N., Dunagin, M.C., McClanahan, P.D., Biaesch, A., Padovan-Merhar, O., Regev, A., Rinn, J.L., and Raj, A. (2015). Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. Genome Biol. *16*, 20.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., et al. (2005). The transcriptional landscape of the mammalian genome. Science *309*, 1559–1563.

Carroll, S.B. (2008). Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. Cell *134*, 25–36.

Dang, V.T., Kassahn, K.S., Marcos, A.E., and Ragan, M.A. (2008). Identification of human haploinsufficient genes and their genomic proximity to segmental duplications. Eur. J. Hum. Genet. *16111*, 1350–1357.

Davidovich, C., Wang, X., Cifuentes-Rojas, C., Goodrich, K.J., Gooding, A.R., Lee, J.T., and Cech, T.R. (2015). Toward a consensus on the binding specificity and promiscuity of PRC2 for RNA. Mol. Cell *57*, 552–559.

Deutschbauer, A.M., Jaramillo, D.F., Proctor, M., Kumm, J., Hillenmeyer, M.E., Davis, R.W., Nislow, C., and Giaever, G. (2005). Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. Genetics *169*, 1915–1925.

Ebert, M.S., and Sharp, P.A. (2010). Emerging roles for natural microRNA sponges. Curr. Biol. *20*, R858–R861.

Ebisuya, M., Yamamoto, T., Nakajima, M., and Nishida, E. (2008). Ripples from neighbouring transcription. Nat. Cell Biol. *10*, 1106–1113.

Eisenberg, E., and Levanon, E.Y. (2013). Human housekeeping genes, revisited. Trends Genet. *29*, 569–574.

Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic Gene Expression in a Single Cell. Science (80-.). *297*, 1183–1186.

Espinoza-Lewis, R.A., Liu, H., Sun, C., Chen, C., Jiao, K., and Chen, Y. (2011). Ectopic expression of Nkx2.5 suppresses the formation of the sinoatrial node in mice. Dev. Biol. *356*, 359–369.

Faghihi, M., and Wahlestedt, C. (2009). Regulatory roles of natural antisense transcripts. Nat. Rev. Mol. Cell Biol. *10*, 637–643.

Fahed, A.C., Gelb, B.D., Seidman, J.G., and Seidman, C.E. (2013). Genetics of congenital heart disease: The glass half empty. Circ. Res. *112*, 707–720.

Franco, D., Lamers, W.H., and Moorman, A.F.M. (1998). Patterns of expression in the developing myocardium : towards a morphologically integrated transcriptional model. Cardiovasc. Res. 38 *38*, 25–53.

Gloss, B.S., and Dinger, M.E. (2015). The specificity of long noncoding RNA expression. Biochim. Biophys. Acta - Gene Regul. Mech. *1859*, 16–22.

Gove, C., Walmsley, M., Nijjar, S., Bertwistle, D., Guille, M., Partington, G., Bomford, A., and Patient, R. (1997). Over-expression of GATA-6 in Xenopus embryos blocks differentiation of heart precursors. EMBO J. *16*, 355–368.

Grote, P., Wittler, L., Hendrix, D., Koch, F., Währisch, 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.

Hoffman, J.I.E., and Kaplan, S. (2002). The incidence of congenital heart disease. J. Am. Coll. Cardiol. *39*, 1890–1900.

Holt, M., and Oram, S. (1960). Familial heart disease with skeletal malformations. Br. Heart J. *22*, 236–

242.

Huang, D.W., Lempicki, R. a, and Sherman, B.T. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Jay, P.Y., Rozhitskaya, O., Tarnavski, O., Sherwood, M.C., Dorfman, A.L., Lu, Y., Ueyama, T., and Izumo, S. (2005). Haploinsufficiency of the cardiac transcription factor Nkx2-5 variably affects the expression of putative target genes. FASEB J. *19*, 1495–1497.

Jensen, B., Wang, T., Christoffels, V.M., and Moorman, A.F.M. (2013). Evolution and development of the building plan of the vertebrate heart. Biochim. Biophys. Acta J. *1833*, 783–794.

Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B.E., van Oudenaarden, A., et al. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc. Natl. Acad. Sci. U. S. A. *106*, 11667–11672.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

Kitano, H. (2004). Biological robustness. Nat. Rev. Genet. *5*, 826–837.

Klattenhoff, C. a, Scheuermann, J.C., Surface, L.E., Bradley, R.K., Fields, P. a, Steinhauser, M.L., Ding, H., Butty, V.L., Torrey, L., Haas, S., et al. (2013). Braveheart, a Long Noncoding RNA Required for Cardiovascular Lineage Commitment. Cell *152*, 570–583.

Koshiba-Takeuchi, K., Mori, A.D., Kaynak, B.L., Cebra-Thomas, J., Sukonnik, T., Georges, R.O., Latham, S., Beck, L., Henkelman, R.M., Black, B.L., et al. (2009). Reptilian heart development and the molecular basis of cardiac chamber evolution. Nature *461*, 95–98.

Li, Q.Y., Newbury-Ecob, R. a, Terrett, J. a, Wilson, D.I., Curtis, a R., Yi, C.H., Gebuhr, T., Bullen, P.J., Robson, S.C., Strachan, T., et al. (1997). Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family. Nat. Genet. *15*, 21–29.

Liberatore, C.M., Searcy-Schrick, R.D., and Yutzey, K.E. (2000). Ventricular expression of tbx5 inhibits normal heart chamber development. Dev. Biol. *223*, 169–180.

Minette, M.S., and Sahn, D.J. (2006). Ventricular septal defects. Circulation *114*, 2190–2197.

Mori, A.D., and Bruneau, B.G. (2004). TBX5 mutations and congenital heart disease: Holt-Oram syndrome revealed. Curr. Opin. Cardiol. *19*, 211–215.

Moskowitz, I.P.G., Pizard, A., Patel, V. V, Bruneau, B.G., Kim, J.B., Kupershmidt, S., Roden, D., Berul, C.I., Seidman, C.E., and Seidman, J.G. (2004). The T-Box transcription factor Tbx5 is required for the patterning and maturation of the murine cardiac conduction system. Development *131*, 4107–4116.

Münst, B., Patsch, C., and Edenhofer, F. (2009). Engineering Cell-permeable Protein. J. Vis. Exp. e1627.

Novikova, I. V., Hennelly, S.P., and Sanbonmatsu, K.Y. (2013). Tackling structures of long noncoding RNAs. Int. J. Mol. Sci. *14*, 23672–23684.

Okamoto, I., Patrat, C., Thépot, D., Peynot, N., Fauque, P., Daniel, N., Diabangouaya, P., Wolf, J.-P., Renard, J.-P., Duranthon, V., et al. (2011). Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. Nature *472*, 370–374.

Olson, E.N. (2006). Gene regulatory networks in the evolution and development of the heart. Science *313*, 1922–1927.

Pennisi, E. (2012). ENCODE Project Writes Eulogy For Junk DNA. Science (80-.). *337*, 1159–1161.

Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat. Methods *10*, 1096–1098.

Postma, A. V, Bezzina, C.R., and Christoffels, V.M. (2015). Genetics of congenital heart disease: the contribution of the noncoding regulatory genome. J Hum Genet *61*, 13–19.

Preissl, S., Schwaderer, M., Raulf, A., Hesse, M., Grüning, B.A., Köbele, C., Backofen, R., Fleischmann, B.K., Hein, L., and Gilsbach, R. (2015). Deciphering the Epigenetic Code of Cardiac Myocyte

Transcription. Circ. Res. *117*, 413–423.

Rabbow, E., Rettberg, P., Barczyk, S., Bohmeier, M., Panitz, C., Horneck, G., Heise-rotenburg, R. Von, Hoppenbrouwers, T., Willnecker, R., Baglioni, P., et al. (2012). The Evolutionary Landscape of Alternative Splicing in Vertebrate. Science (80-.). *12*, 374–387.

Raser, J.M., and O'Shea, E.K. (2004). Control of stochasticity in eukaryotic gene expression. Science *304*, 1811–1814.

Seidman, J.G., and Seidman, C. (2002). Transcription factor haploinsufficiency: When half a loaf is not enough. J. Clin. Invest. *109*, 451–455.

Smemo, S., Campos, L.C., Moskowitz, I.P., Krieger, J.E., Pereira, A.C., and Nobrega, M.A. (2012). Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. Hum. Mol. Genet. *21*, 3255–3263.

Srivastava, D. (2006). Making or breaking the heart: from lineage determination to morphogenesis. Cell *126*, 1037–1048.

Taft, R.J., and Mattick, J.S. (2003). Increasing biological complexity is positively correlated with the relative genome-wide expansion of non-protein-coding DNA sequences. Genome Biol. *5*, P1.

Takeuchi, J.K., Ohgi, M., Koshiba-Takeuchi, K., Shiratori, H., Sakaki, I., Ogura, K., Saijoh, Y., and Ogura, T. (2003). Tbx5 specifies the left/right ventricles and ventricular septum position during cardiogenesis. Development *130*, 5953–5964.

Taniguchi, Y., Choi, P.J., Li, G., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X.S. (2011). Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. Science *329*, 533–539.

Tanimoto, Y., Iijima, S., Hasegawa, Y., Suzuki, Y., Daitoku, Y., Mizuno, S., Ishige, T., Kudo, T., Takahashi, S., Kunita, S., et al. (2008). Embryonic stem cells derived from C57BL/6J and C57BL/6N mice. Comp. Med. *58*, 347–352.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

Tripathi, V., Shen, Z., Chakraborty, A., Giri, S., Freier, S.M., Wu, X., Zhang, Y., Gorospe, M., Prasanth, S.G., Lal, A., et al. (2013). Long Noncoding RNA MALAT1 Controls Cell Cycle Progression by Regulating the Expression of Oncogenic Transcription Factor B-MYB. PLoS Genet. *9*.

Wamstad, J.A., Wang, X., Demuren, O.O., and Boyer, L.A. (2014). Distal enhancers: New insights into heart development and disease. Trends Cell Biol. *24*, 294–302.

Wang, K.C., and Chang, H.Y. (2011). Molecular mechanisms of long noncoding RNAs. Mol. Cell *43*, 904–914.

Wu, B., Eliscovich, C., Yoon, Y.J., and Singer, R.H. (2016). Translation dynamics of single mRNAs in live cells and neurons. Science (80-.). *1084*, aaf1084.

Zlotogora, J. (2003). Penetrance and expressivity in the molecular age. Genet. Med. *5*, 347–352.

Figures

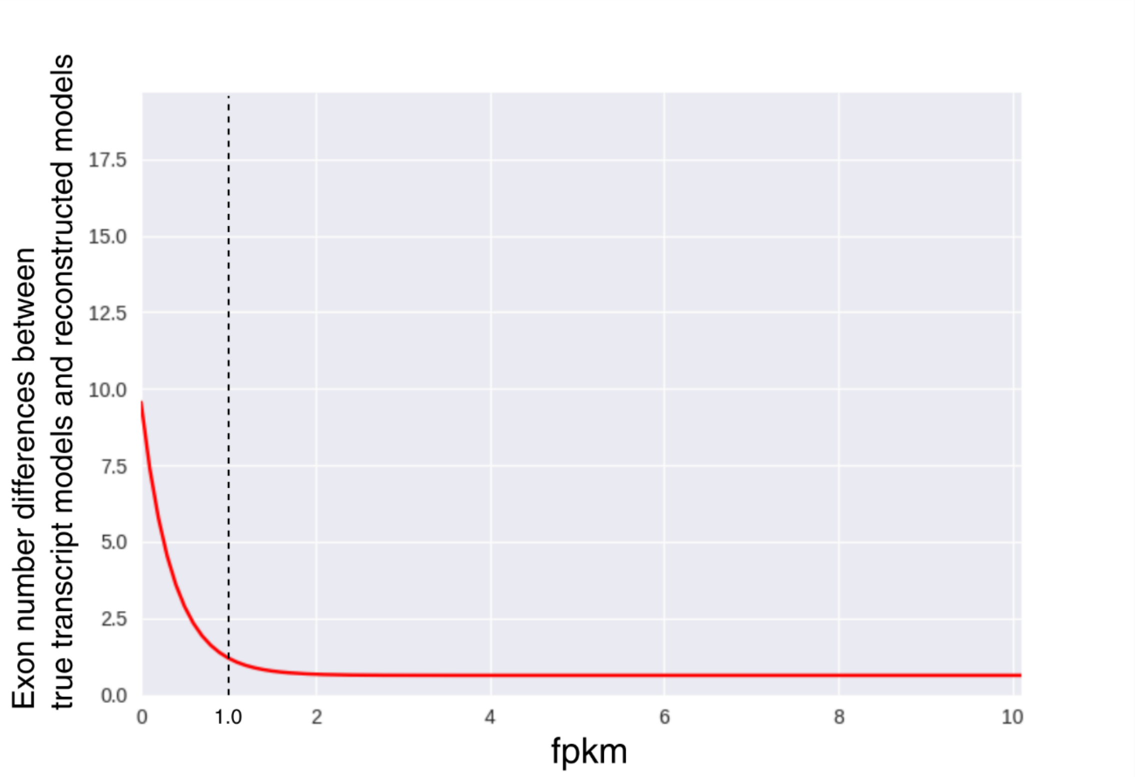


Figure 1: Validation of minimum fpkm value in the reconstruction of transcripts

We counted the exon numbers of reconstructed transcripts and compared them with known exon numbers. The exon numbers were determined based on the maximum of alternative transcripts for each gene. We only took into account genes with their exon number 12 or less since the exon numbers of more than 98.5% of known lncRNAs expressed in the heart fall under the category. The relation between exon number differenced and fpkm was fitted with an exponential curve. This result demonstrates that 1.0 fpkm is sufficient to infer gene models in our RNA-seq experiment.

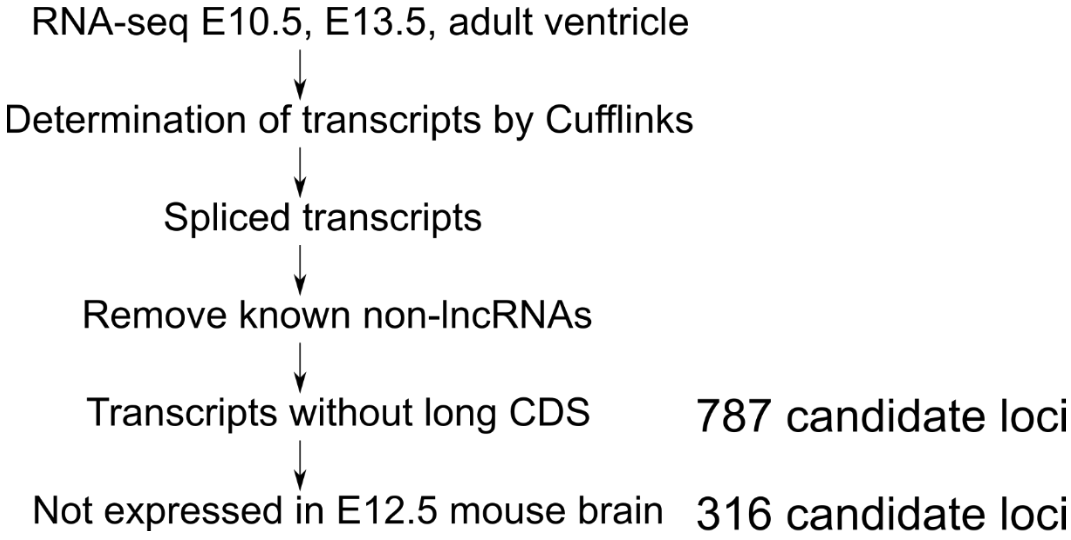

RNA-seq E10.5, E13.5, adult ventricle
Determination of transcripts by Cufflinks
Spliced transcripts
Remove known non-lncRNAs
Transcripts without long CDS
787 candidate loci
Not expressed in E12.5 mouse brain
316 candidate loci

Figure 2: The screening procedure of lncRNAs

I extracted total RNA from E10.5, E13.5 and 8w ventricles. The obtained reads were mapped to the mouse genome (mm10) with Tophat2 and transcript models were determined with Cufflinks. From the determined models, known non-lncRNAs were removed based on UCSC genome annotation and Refseq. As a result, 787 candidate loci were determined and 316 of them were not expressed in E12.5 mouse brain.

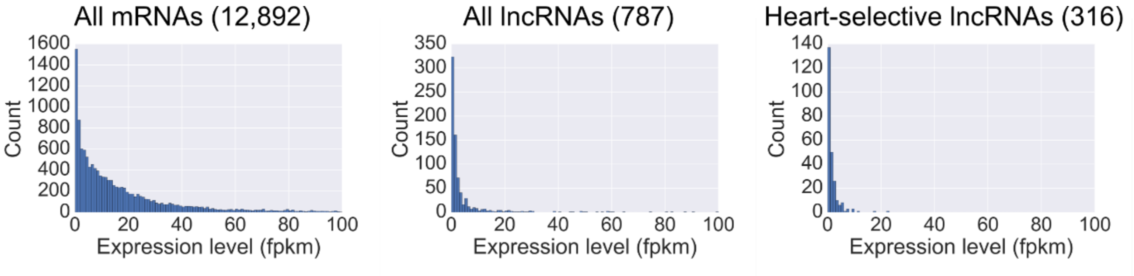


Figure 3: The histograms of expression levels of the lncRNAs at E10.5

The expression levels (fpkm) of genes in each category at E10.5 were plotted. The expression levels of lncRNAs were generally much lower than those of mRNAs. In particular, almost no lncRNAs with tissue specificity had fpkm higher than 10. The number of genes that are expressed at E10.5 in each category is shown in the parenthesis.

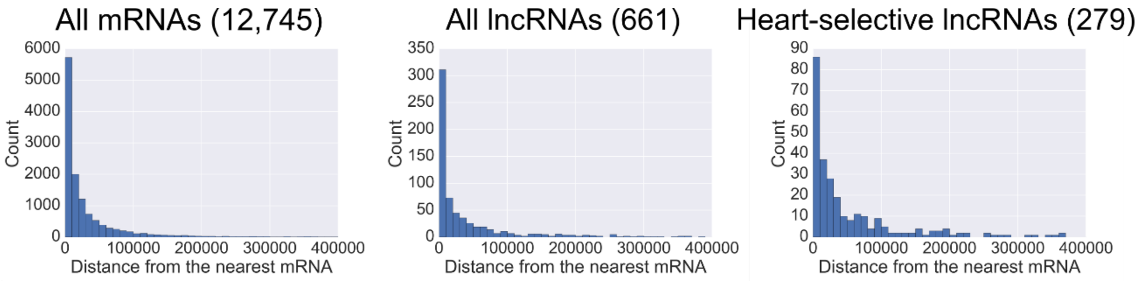


Figure 4: The distances from lncRNAs to their nearest protein-coding genes

Heart-selective lncRNAs were generally at greater distances from protein-coding genes. Genes with TSS that is overlapped with other genes were omitted. The number of genes in each category is shown in the parenthesis.

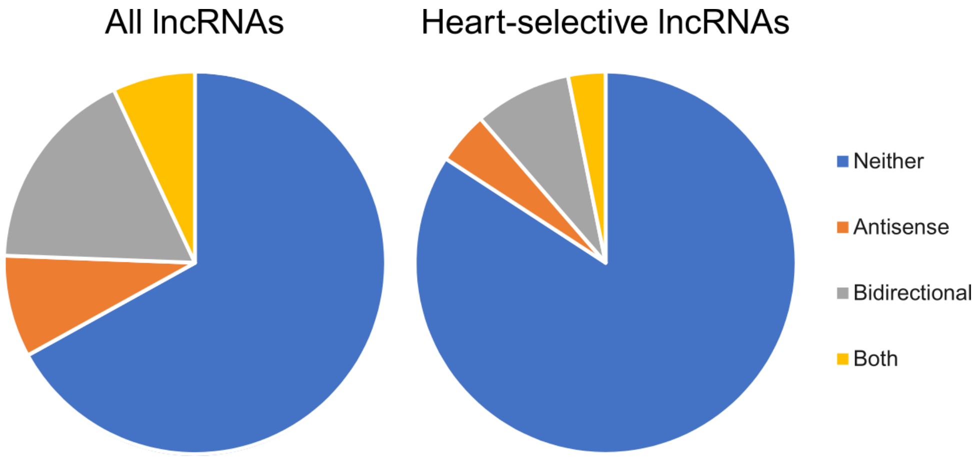


Figure 5: Classification of lncRNA candidates found in the screening

The number of lncRNAs that are antisense or bidirectional to protein-coding genes were determined. Heart-selective lncRNAs are less likely to be bidirectional or antisense lncRNAs. Some lncRNAs were judged to be both antisense and bidirectional since they or neighboring genes have multiple isoforms.

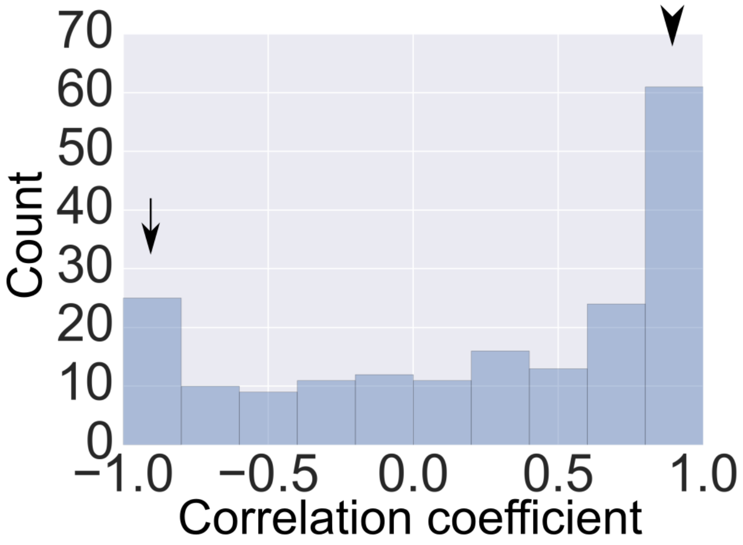


Figure 6: The distribution of Pearson correlation coefficients between bidirectional promoter pairs over development

I plotted Pearson correlation coefficients between the log2-transformed expression levels of bidirectional pairs based on E10.5, E13.5 and 8 week RNA-seq data. If such bidirectional pairs do not have correlation, the correlation coefficients should distribute uniformly. The arrow indicates negative correlation and the arrowhead indicates positive correlation. Quite a few pairs show strong correlation.

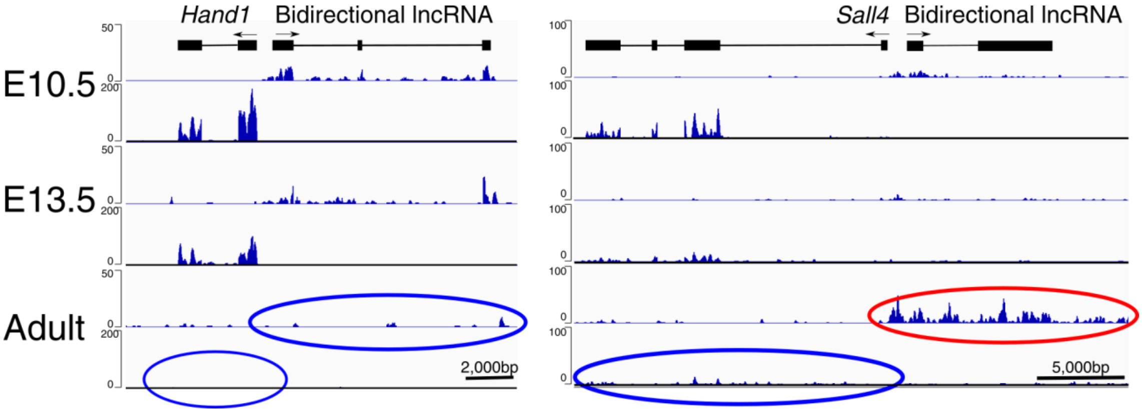


Figure 7: Examples of bidirectional gene pairs with positive and negative correlation

Hand1 (left) shows positive correlation while *Sall4* (right) shows negative correlation with the bidirectional transcript during development.

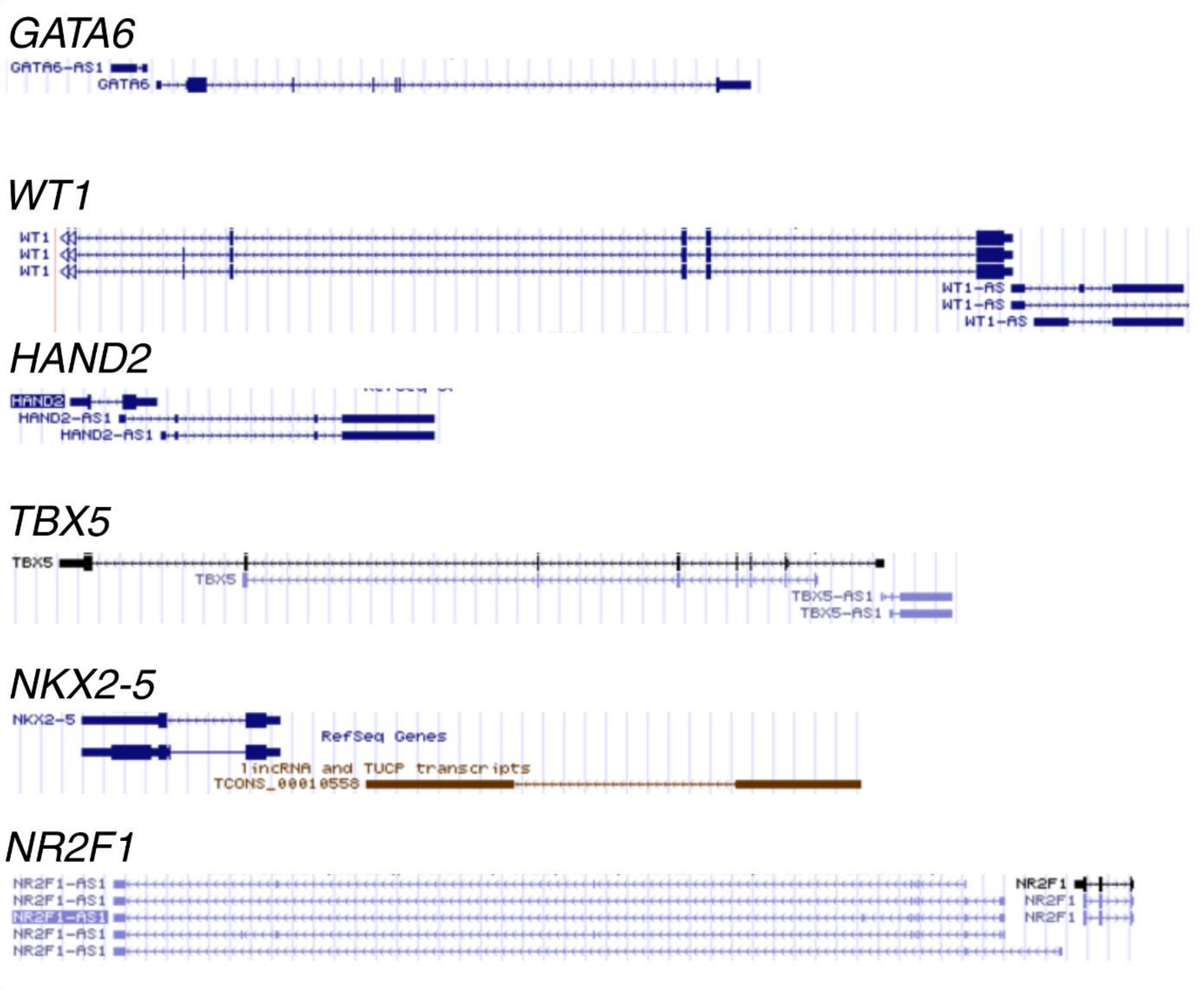


Figure 8: Conserved bidirectional lncRNAs in human

Conserved bidirectional lncRNAs in human USCS genome annotation are shown. Many important cardiac transcription factor genes have conserved bidirectional lncRNAs.

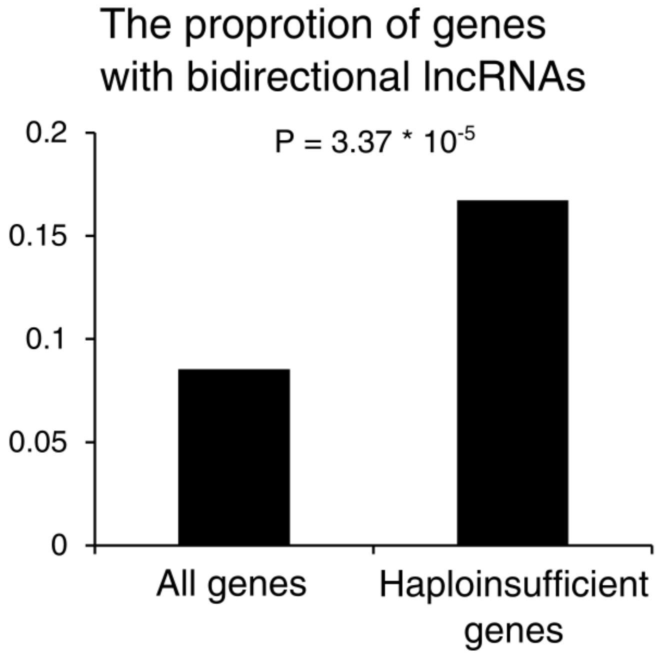


Figure 9: The proportion of protein-coding genes that possess bidirectional lncRNAs in each category

The proportion was calculated based on the paper that comprehensively identified haploinsufficient genes and Refseq database (GRCm38.p3). Bidirectional lncRNAs were significantly enriched among haploinsufficient genes not only in the heart.

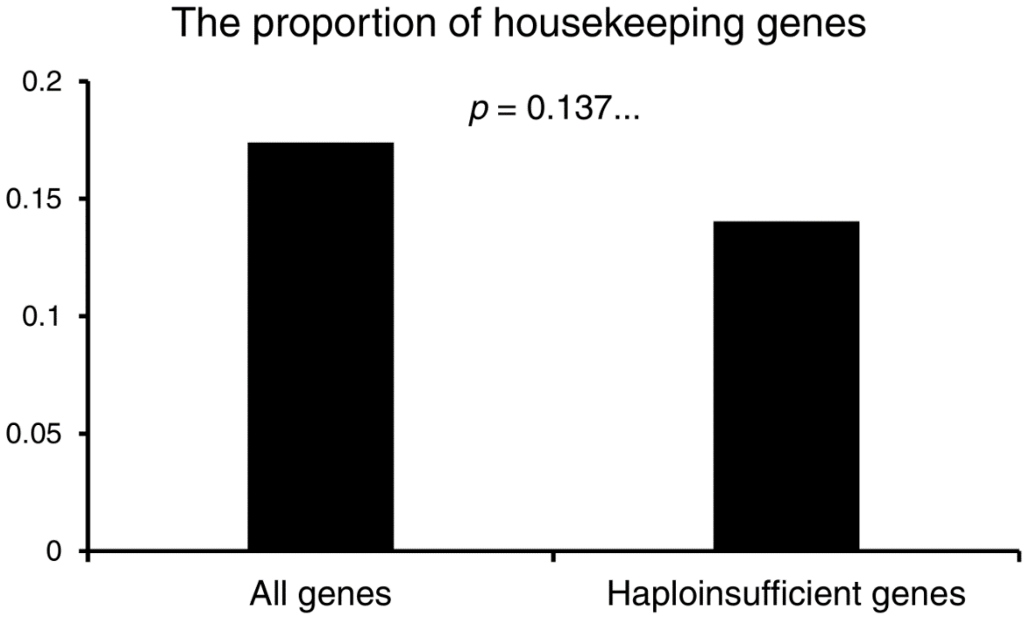


Figure 10: The proportion of housekeeping genes among all genes and haploinsufficient genes

The proportion of housekeeping genes among all genes and among haploinsufficient genes was calculated and it was not found to be significantly correlated. This result eliminates the possibility that the enrichment of genes with bidirectional lncRNAs among haploinsufficient genes is due to the pseudo-correlation generated through housekeeping-haploinsufficient correlation.

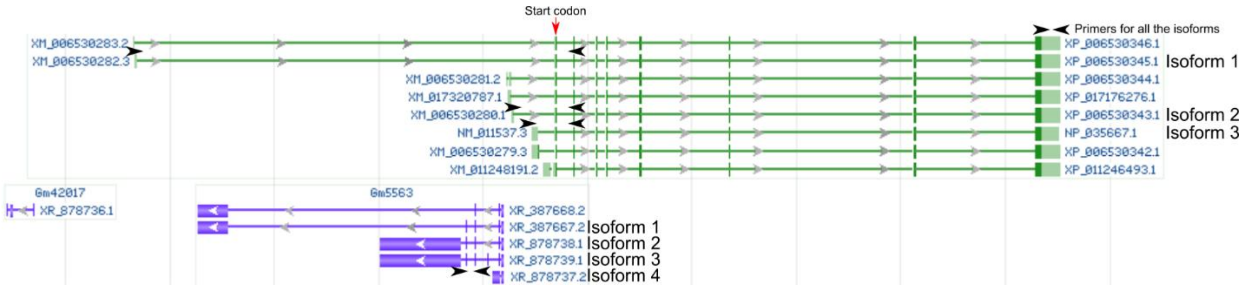


Figure 11: Genomic annotation of *Tbx5ua* and *Tbx5*

RefSeq genome annotation of the mouse *Tbx5* locus is presented. The isoforms of *Tbx5* and *Tbx5ua* that were identified in our RNA-seq analysis were labeled with isoform numbers. The qPCR primers used to quantify *Tbx5* and *Tbx5ua* are indicated as arrowheads.

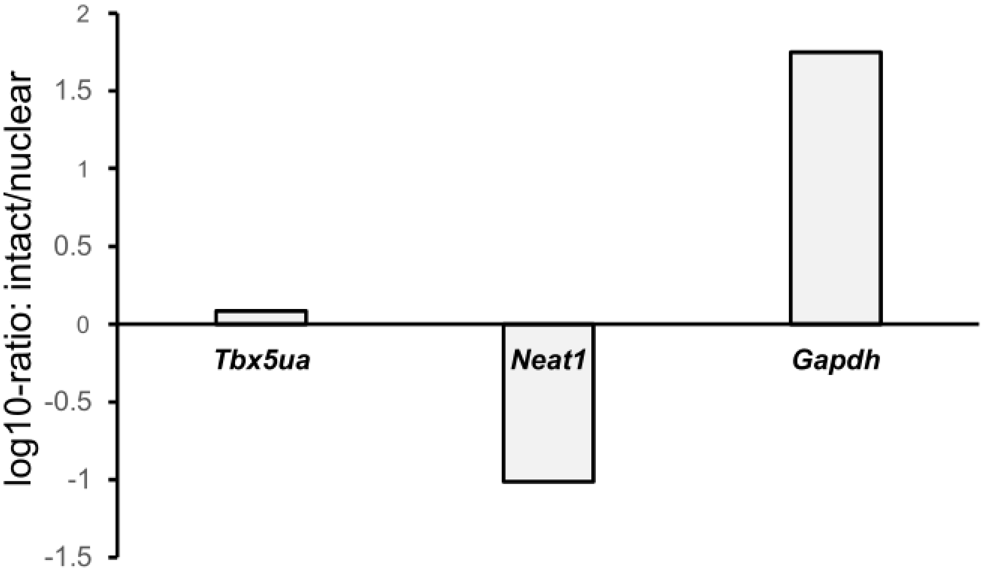


Figure 12: Subcellular localization of *Tbx5ua*

The log10-ratios of intact/nuclear RNA abundances were determined based on a publicly available RNA-seq data. *Gapdh* and *Neat1* serve as controls for cytoplasmic and nuclear localized RNA, respectively. *Tbx5ua* does not show clear nuclear/cytoplasmic localization, which is observed in many other well-known lncRNAs.

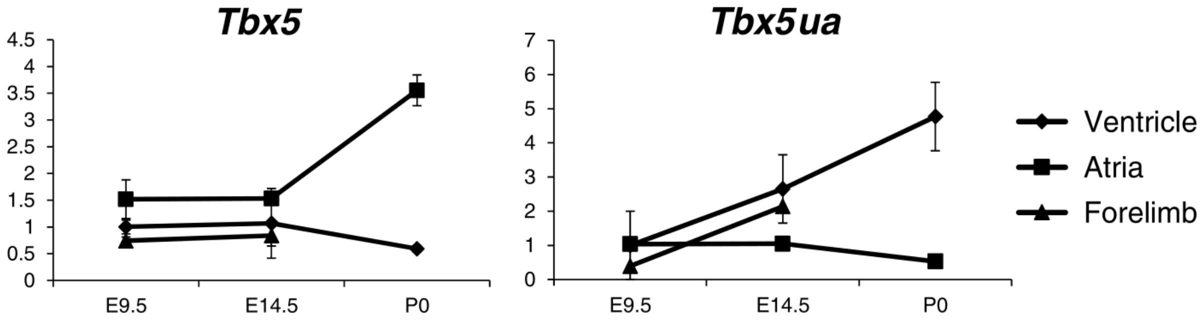


Figure 13: qRT-PCR analysis of the expression pattern of *Tbx5* and *Tbx5ua* in development

The expression levels of *Tbx5* and *Tbx5ua* during development as determined by qRT-PCR. In the ventricle, the expression level of *Tbx5ua* is increased with the progression of development while that of *Tbx5* is decreased (n=3). Error bars indicate SEM.

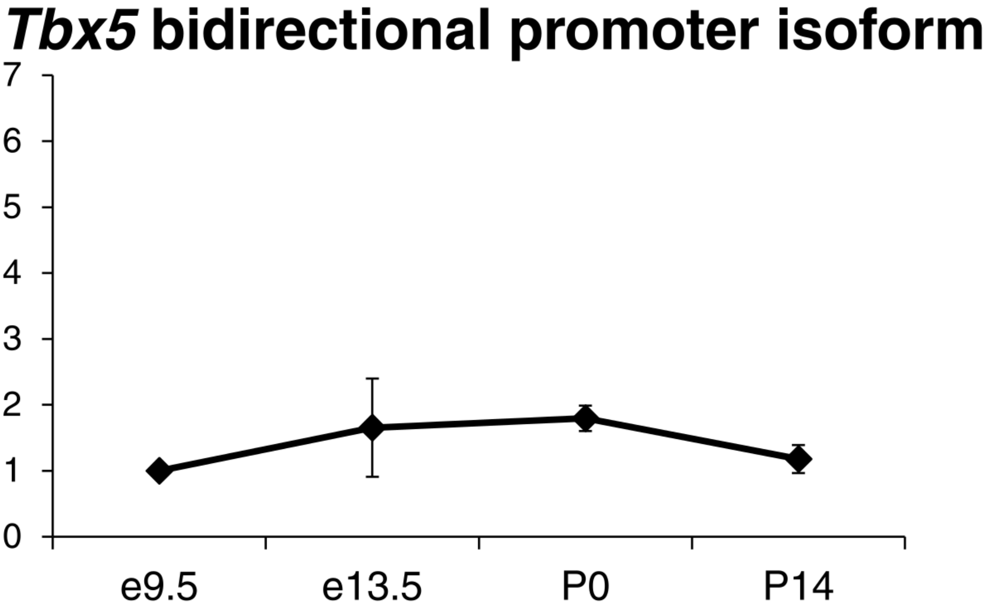


Figure 14: qRT-PCR analysis of a Tbx5 isoform that is transcribed from the promoter that also produces Tbx5ua (isoform 2)

The expression pattern of this isoform over development is also inconsistent with that of *Tbx5ua*, indicating that they are post-transcriptionally modulated or the directional of transcription is somehow controlled (n=3).

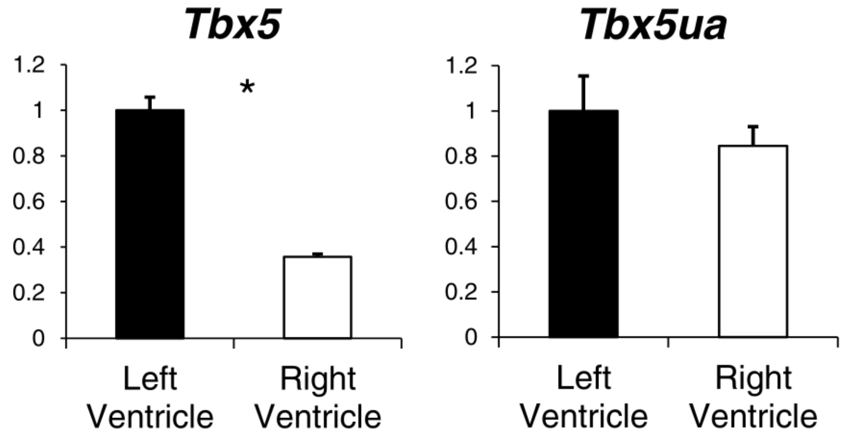


Figure 15: The expression levels of *Tbx5* and *Tbx5ua* in the left and right ventricle at E11.5 by qRT-PCR

The expression level of *Tbx5* shows a well-known left-right difference while that of *Tbx5ua* in the right ventricle is comparable to that in the left ventricles (n=3, *: p < 0.05, Welch's *t*-test). These results suggest that although they share the promoter, they are regulated separately.

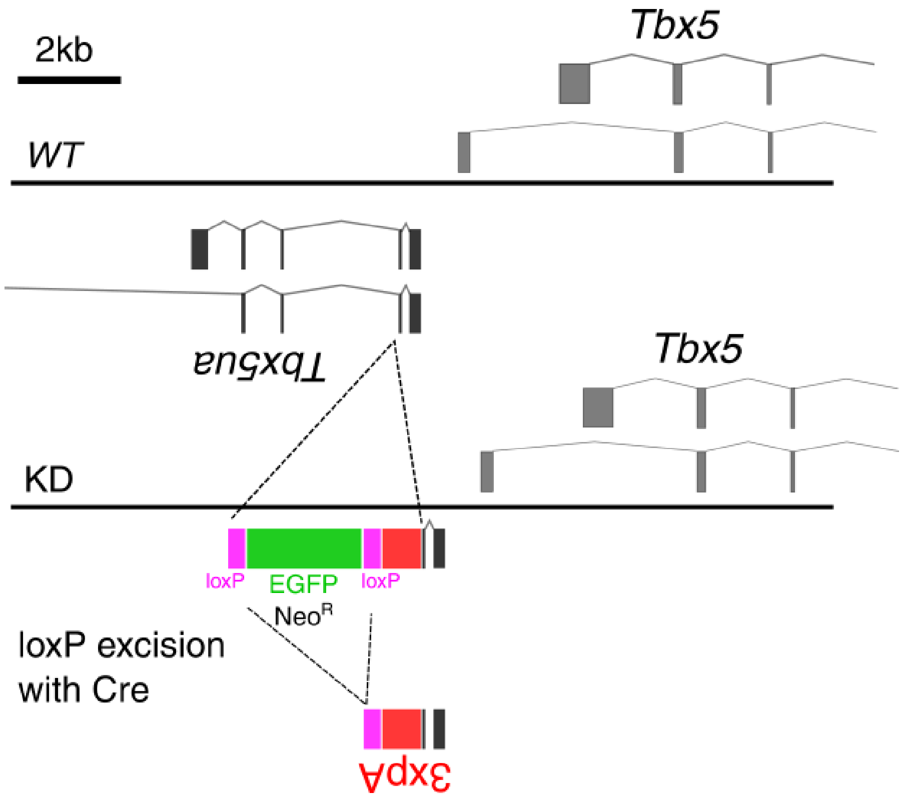


Figure 16: Schematic diagram of *Tbx5ua* knockdown experiment

Three tandem copies of bovine growth hormone polyadenylation signals were inserted along with neomycin resistance gene (*Neo*^{R}) and *EGFP*. The selection markers were subsequently removed with cell-permeable Cre recombinase.

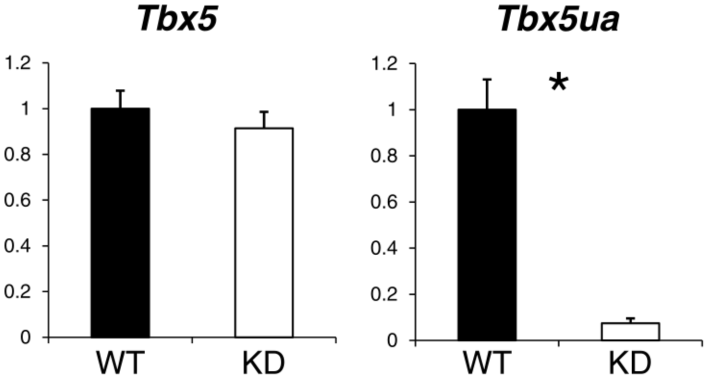


Figure 17: Knockdown of *Tbx5ua* was confirmed by qRT-PCR

Chimeric mice were made by tetraploid complementation assay from the ES cells. Total RNA was extracted from ventricles of E9.5 chimeric mice and qRT-PCR was performed. While the expression level of *Tbx5* was not changed, *Tbx5ua* was knocked down as expected (n=4, *: $p < 0.05$, Welch's *t*-test).

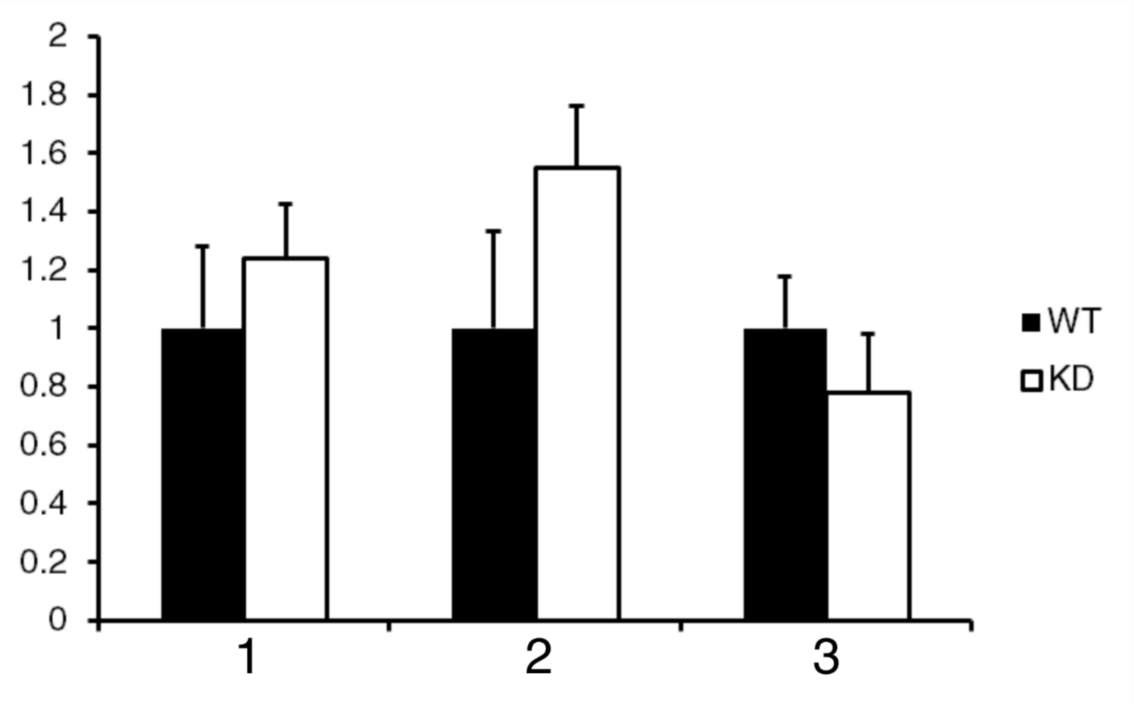


Figure 18: qRT-PCR of different isoforms of *Tbx5* in KD mouse

qRT-PCR analysis of KD and WT mice for all the *Tbx5* isoforms detected in my RNA-seq analysis. The expression levels are not significantly changed even when I look at each major isoform. The isoform numbers are indicated in Figure 3A. (n=4, *: $p < 0.05$, Welch's *t*-test).

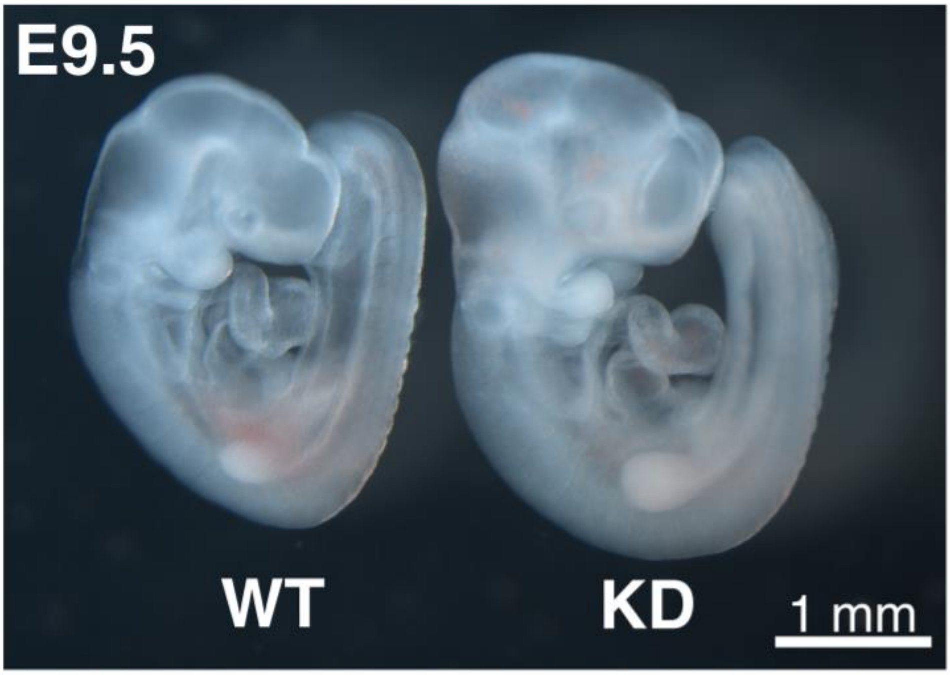


Figure 19: The morphology of chimeric mice at E9.5

The body size and forelimb looked completely normal. In fact, KD mice were slightly bigger at E9.5 in the first round of experiment (left), while in the second round such tendency was not observed (right).

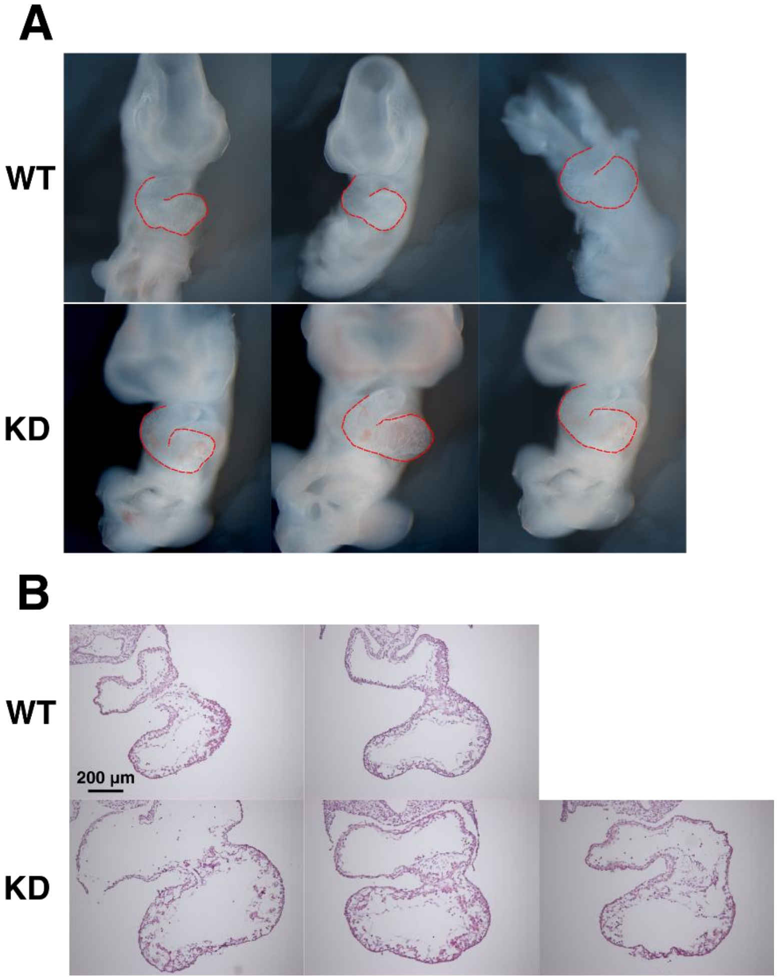


Figure 20: The heart of chimeric mice at E9.5

(A) The right ventricle of KD mice shows hypoplasticity as shown by the outlines of ventricle and the arrowheads. The right panel of WT shows an embryo that lacks its head due to my mishandling. (B) Hematoxylin-eosin staining of sections. The ventricular wall seems to be irregular, especially around the interventricular zone.

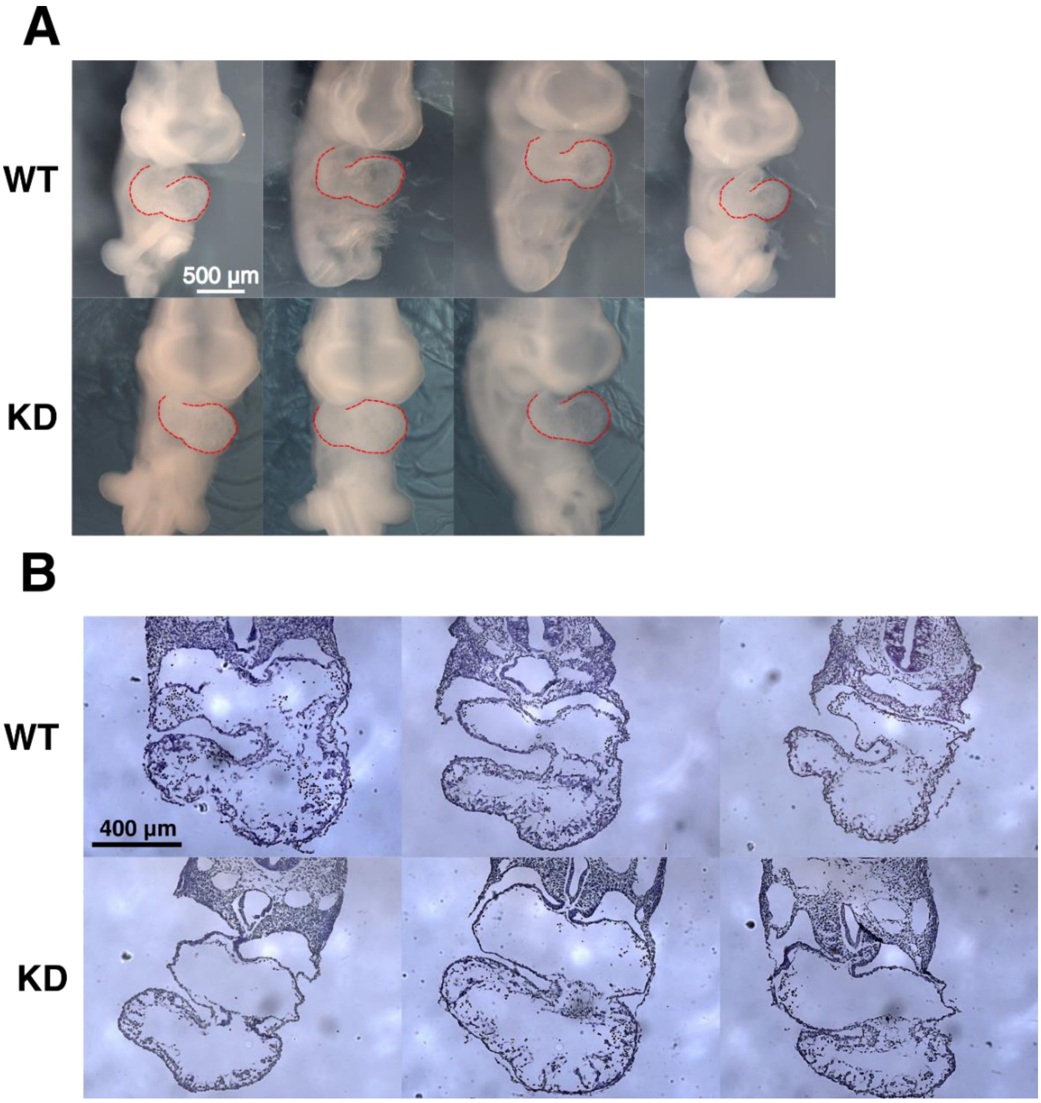


Figure 21: E9.5 chimeric mice obtained from another line of KD ESCs

Morphological phenotype of *Tbx5ua* KD embryos derived from another ESC line is shown and is consistent with our first ESC line.

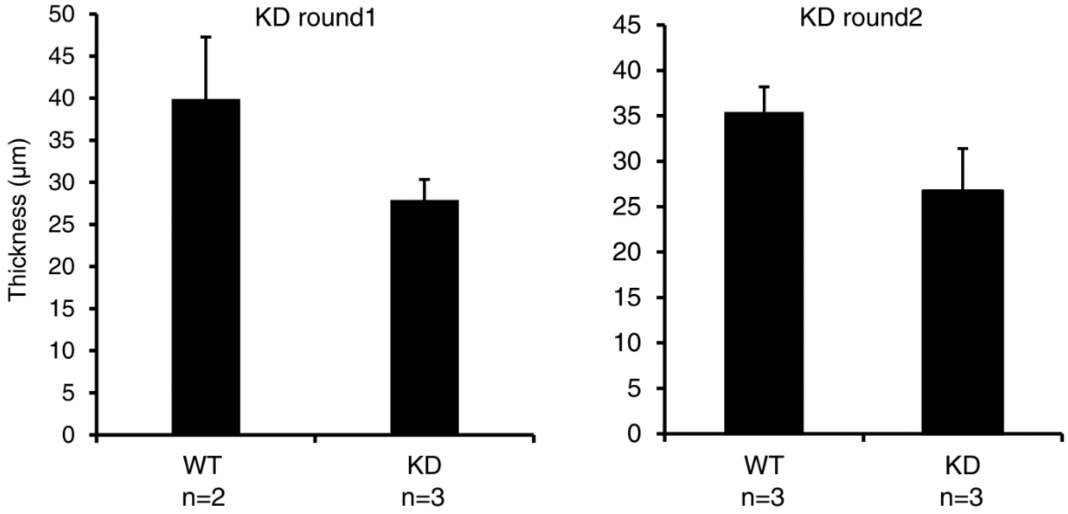


Figure 22: Measurement of ventricular wall thickness in WT and KD mice

The thickness of the ventricular wall around the interventricular zone was measured for WT and KD embryos. Although not statistically verified, KD embryos tended to have thinner wall.

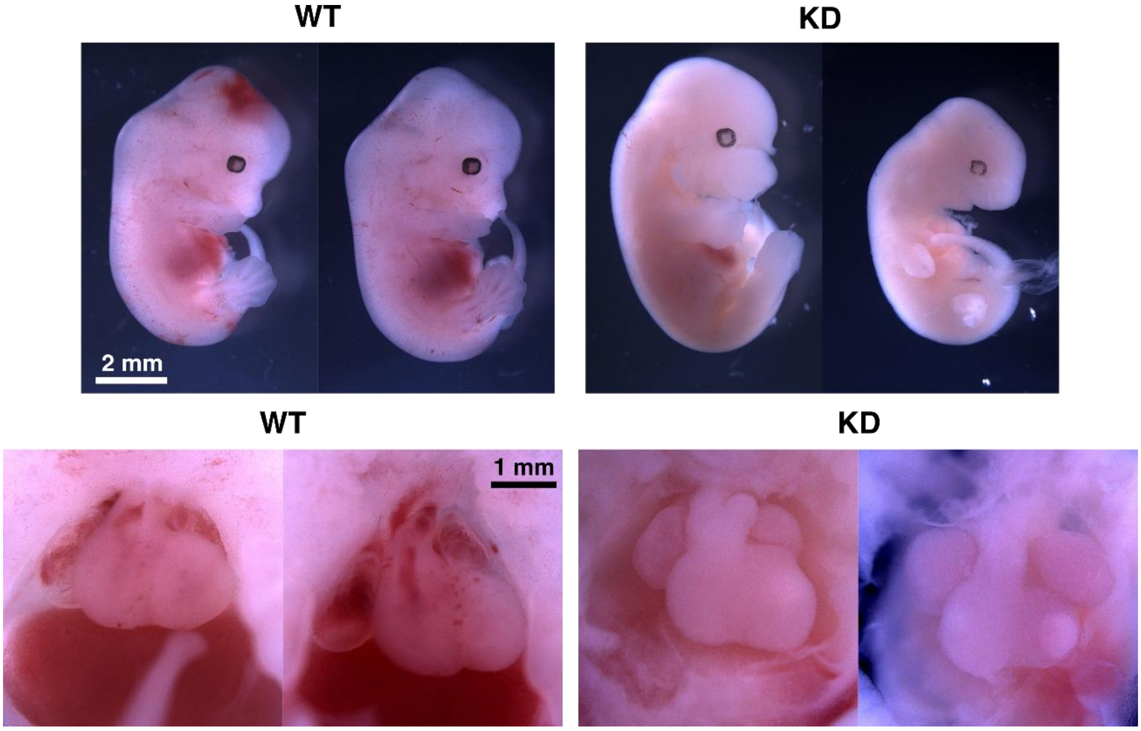


Figure 23: The morphology of E13.5 chimeric embryos

KD chimeric embryos were dead with pale body and severe hypoplasticity in the ventricle and atrium at E13.5. Even at this stage, forelimbs looked completely normal.

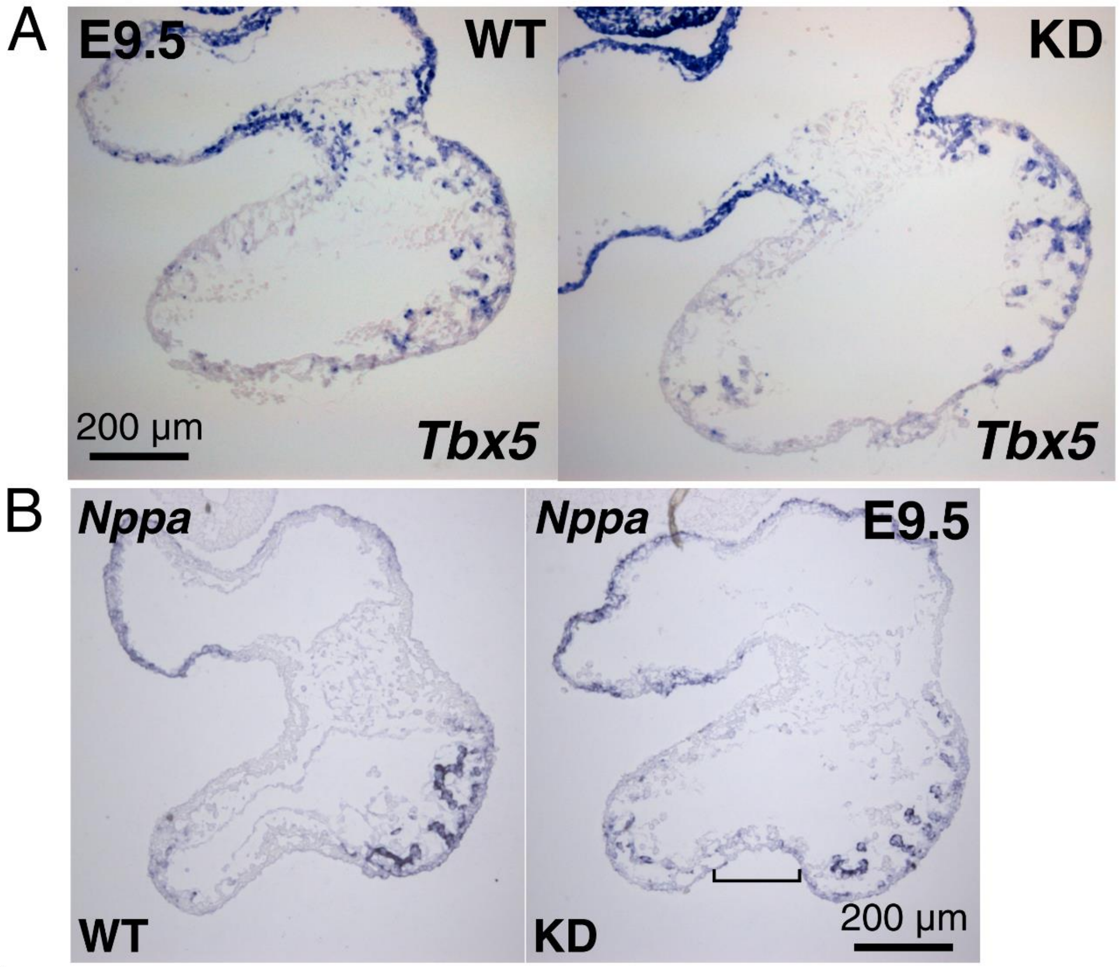
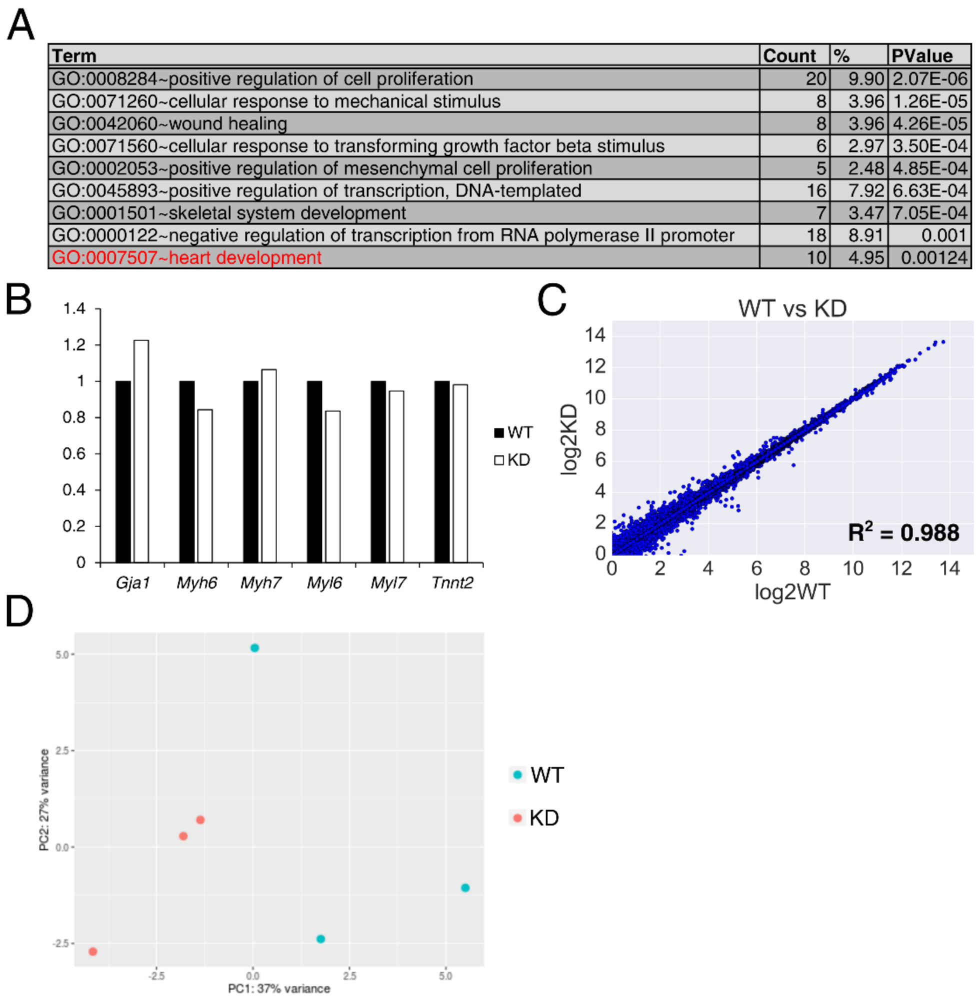


Figure 24: *In situ* hybridization of *Tbx5* and *Nppa* of E9.5 chimeric mice

(A) The expression pattern of *Tbx5* at the mRNA level appeared to be not changed. (B) KD embryos showed an ectopic expression of *Nppa* around the pre-ventricular septal region, which is frequently observed among embryos with abnormal development of ventricular septum.



Term	Count	%	PValue
GO:0008284~positive regulation of cell proliferation	20	9.90	2.07E-06
GO:0071260~cellular response to mechanical stimulus	8	3.96	1.26E-05
GO:0042060~wound healing	8	3.96	4.26E-05
GO:0071560~cellular response to transforming growth factor beta stimulus	6	2.97	3.50E-04
GO:0002053~positive regulation of mesenchymal cell proliferation	5	2.48	4.85E-04
GO:0045893~positive regulation of transcription, DNA-templated	16	7.92	6.63E-04
GO:0001501~skeletal system development	7	3.47	7.05E-04
GO:0000122~negative regulation of transcription from RNA polymerase II promoter	18	8.91	0.001
GO:0007507~heart development	10	4.95	0.00124

Figure 25: Representative results of RNA-seq on the ventricles of E9.5 chimeric mice

I performed SmartSeq2 RNA-seq to comprehensively identify genes that were changed in the knockdown mice at E9.5 (n=3). (A) Gene ontology analysis included terms related to heart development. (B) Structural genes important for heart contraction were not affected, suggesting that the differentiation of cardiomyocytes was not affected in a major way. (C) The scatter plot of log2-transformed expression levels shows that the expression pattern of KD embryos did not change drastically. (D) Principal component analysis on the RNA-seq analysis. WT and KD mice are distinguishable only by the first component.

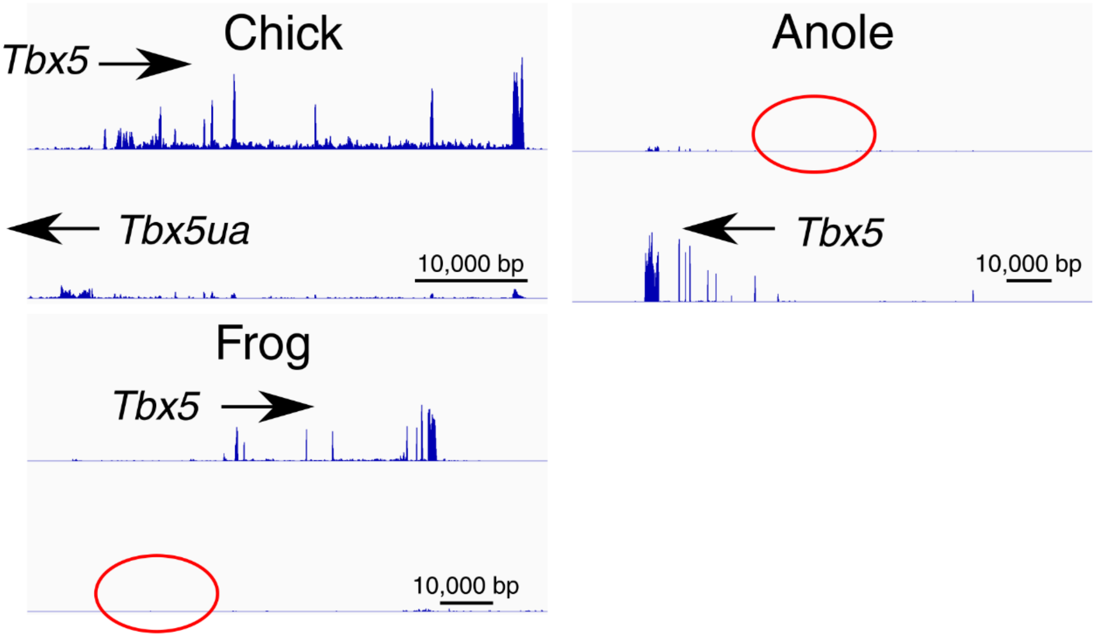


Figure 26: *Tbx5ua* is conserved in chicken, but not in lizard or frog

Reanalysis of RNA-seq data from chicken, anole and zebrafish. *Tbx5ua* is conserved only in chicken, which possesses a complete ventricular septum, among these species.

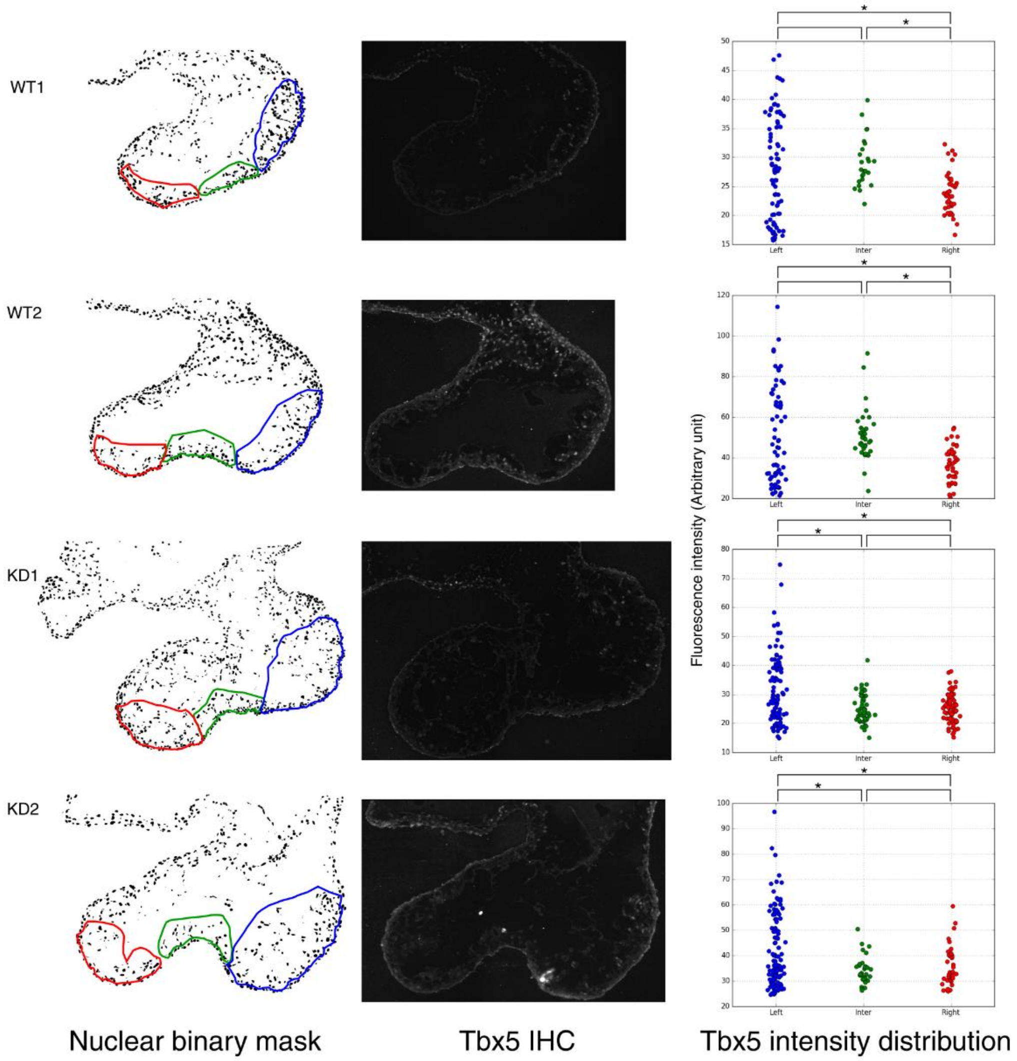


Figure 27: Tbx5 immunohistochemistry suggests altered protein expression in KD

Tbx5 IHC of WT and KD embryos were quantified. The ventricle was divided into three regions and the staining intensity in each nuclear was measured using ImageJ. Nuclear binary masks were produced from DAPI staining. Note that we only quantified cells that are not in the outermost layer of the ventricle because speckle-like background was observed in the region. Mann-Whitney U test was performed for each sample with multiple testing correction with Holm method (*: $p < 0.05$). Tbx5 expression is vanished in the interventricular zone and diminished in the left ventricle in KD embryos.

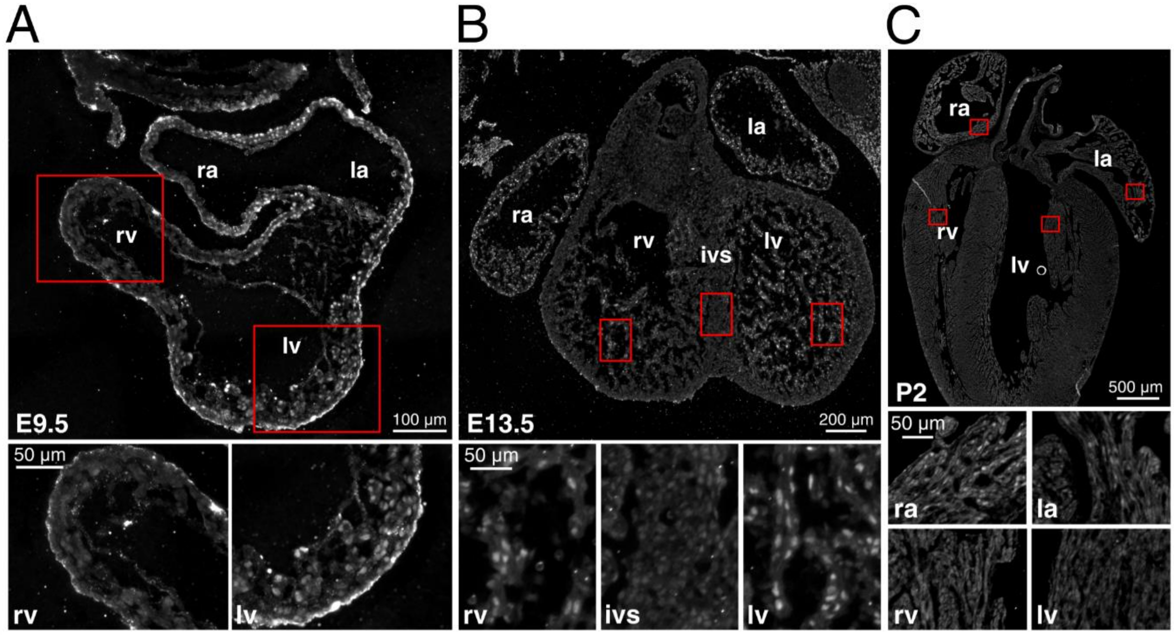


Figure 28: Immunohistochemistry of Tbx5

(A) At E9.5, clear gradient in the ventricle is observed. (B) At E13.5, strong signals were detected in the trabeculae, but not in the ventricular walls or in the ventricular septum. (C) At P2, the number of Tbx5-positive cells were very small in the ventricle.

a: atrium, lv: left ventricle, rv: right ventricle, la: left atrium, ra: right atrium, ivs: interventricular septum

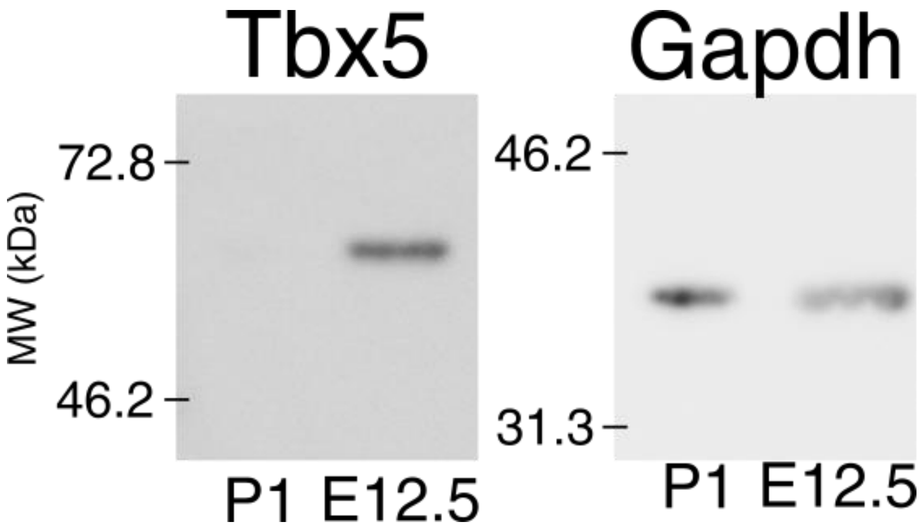


Figure 29: Western blot of Tbx5 and Gapdh

Western blot of Tbx5 on proteins extracted from ventricles of P1 and E12.5 mice. Tbx5 was not detected at P1, confirming the result of immunohistochemistry.

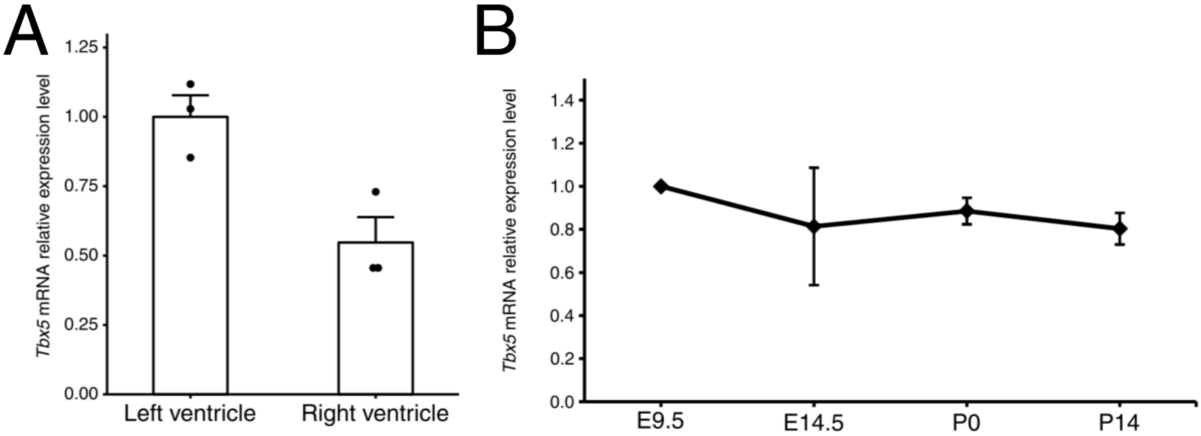


Figure 30: The expression pattern of *Tbx5* mRNA in the ventricle by qRT-PCR

The expression level of *Tbx5* mRNA in the left is just about twice that in the right at E9.5 (left). The expression level of *Tbx5* does not show a strong decrease over development despite the results of Tbx5 protein. These findings show that the expression pattern of *Tbx5* mRNA and protein are different. (n=3)

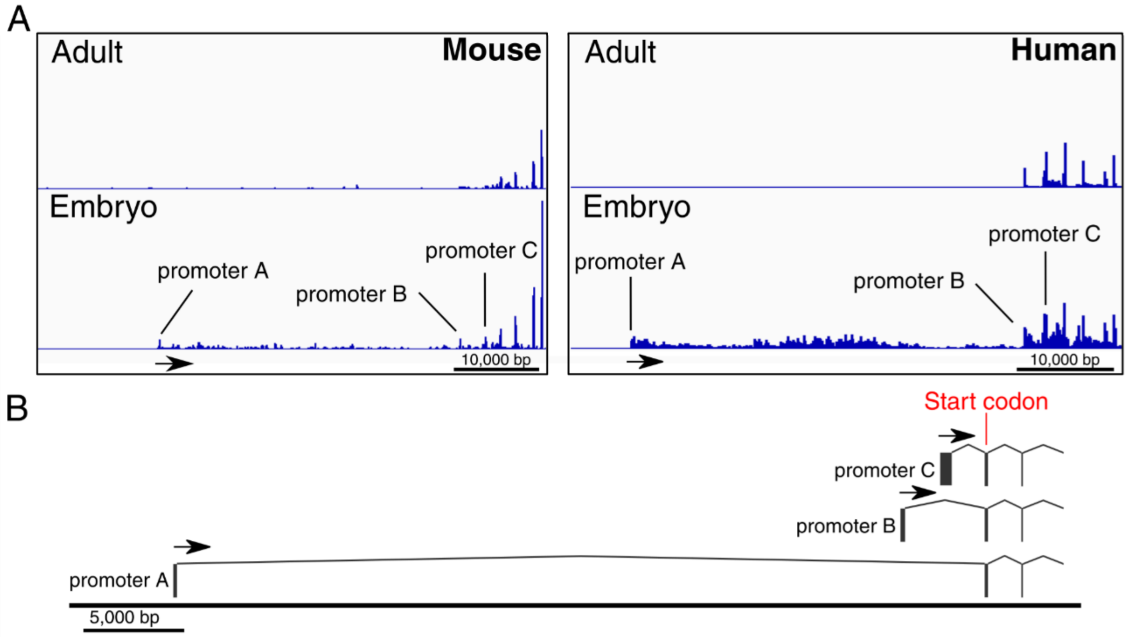


Figure 31: Alternative promoters of *Tbx5* gene

I reanalyzed the RNA-seq data to investigate alternative isoforms of *Tbx5* and found that *Tbx5* actually have three promoters. The expression level of promoter B is apparently lower than that of promoter A and C. The transcription of promoter A appeared to be limited to embryos (top). *Tbx5* has three promoters and the start codon for proper translation is in the second exon so all the three alternative promoter isoforms are different only in their 5' UTRs (bottom).

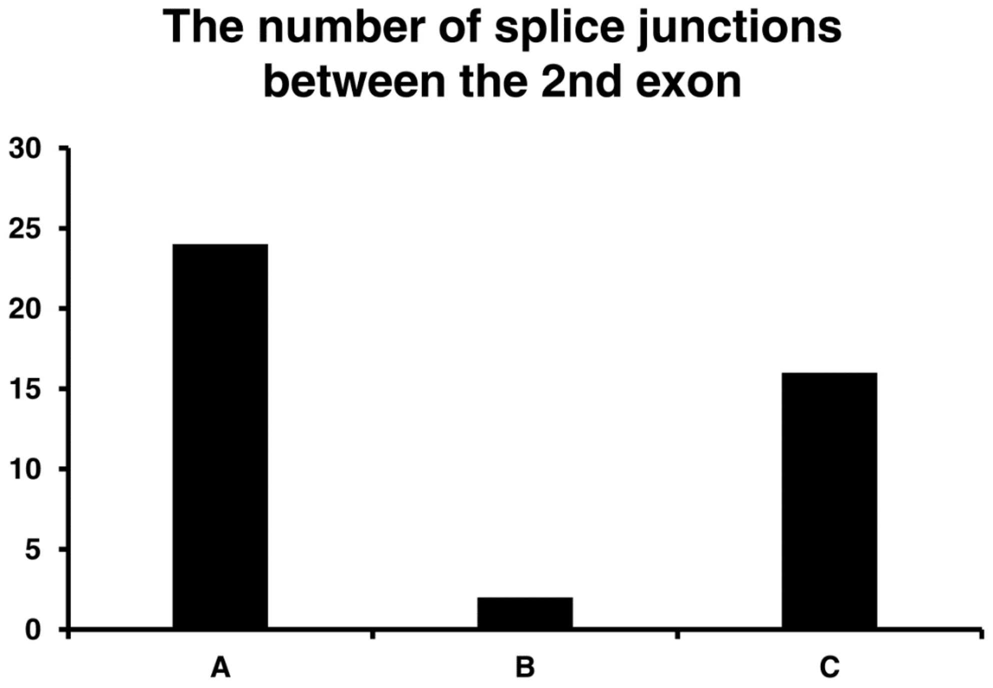
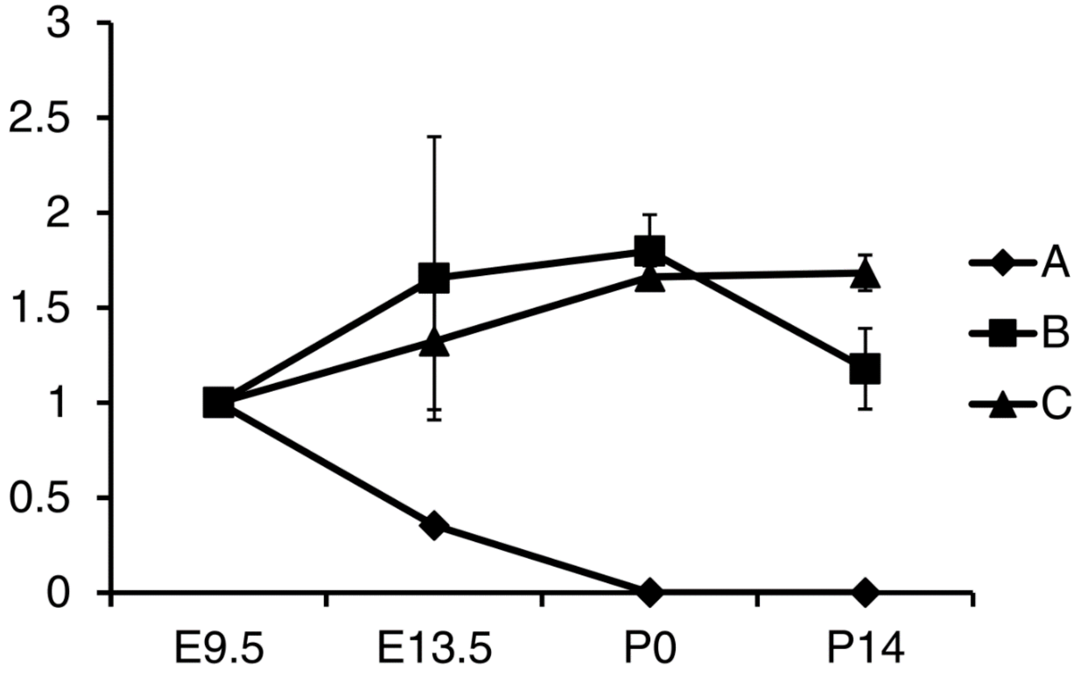


Figure 32: Comparison of the expression levels of three isoforms of *Tbx5* inferred from the RNA-seq data

Using E10.5 RNA-seq data, we counted the number of “1st exon-2nd exon” junctions of each isoform to get the rough estimation of the abundance ratio of Tbx5 isoforms. Of the three, B seemed to be expressed at a very low level.



.

Figure 33: The expression patterns of three isoforms in the ventricle over development by qRT-PCR

The expression level of promoterA isoform decreases rapidly over development and lower than the detection limit of qRT-PCR in postnatal mice. The expression levels of the other two promoters are roughly stable over development. (n=3)

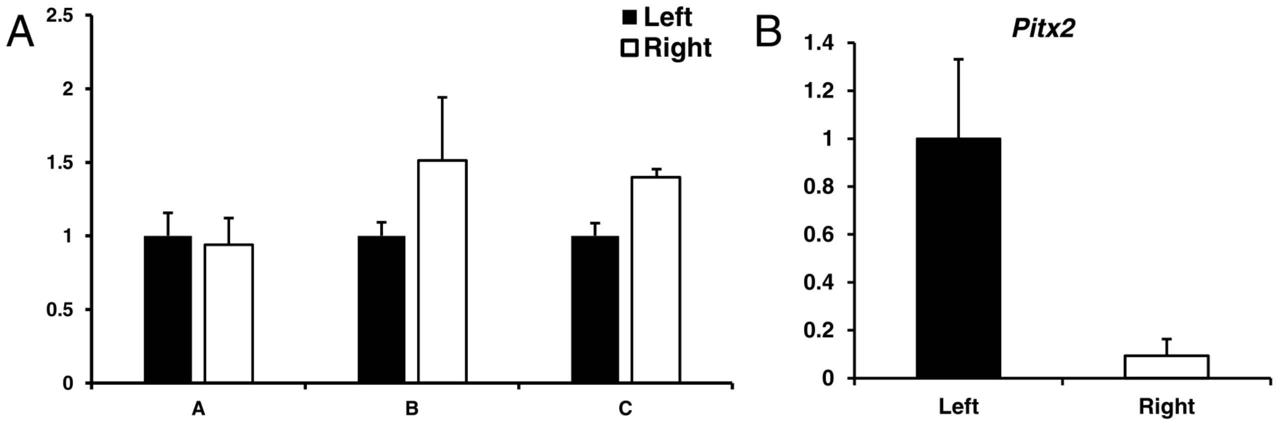


Figure 34: qRT-PCR of three isoforms of *Tbx5* in the atrium

(A) qRT-PCR of *Tbx5* was performed for the left and right atria of P1 mice. All the three isoforms were detected in both the left and right atrium and the left-right difference at the transcript level was small. (B) Pitx2 was used to confirm that the left and right atria were correctly isolated (n=3).

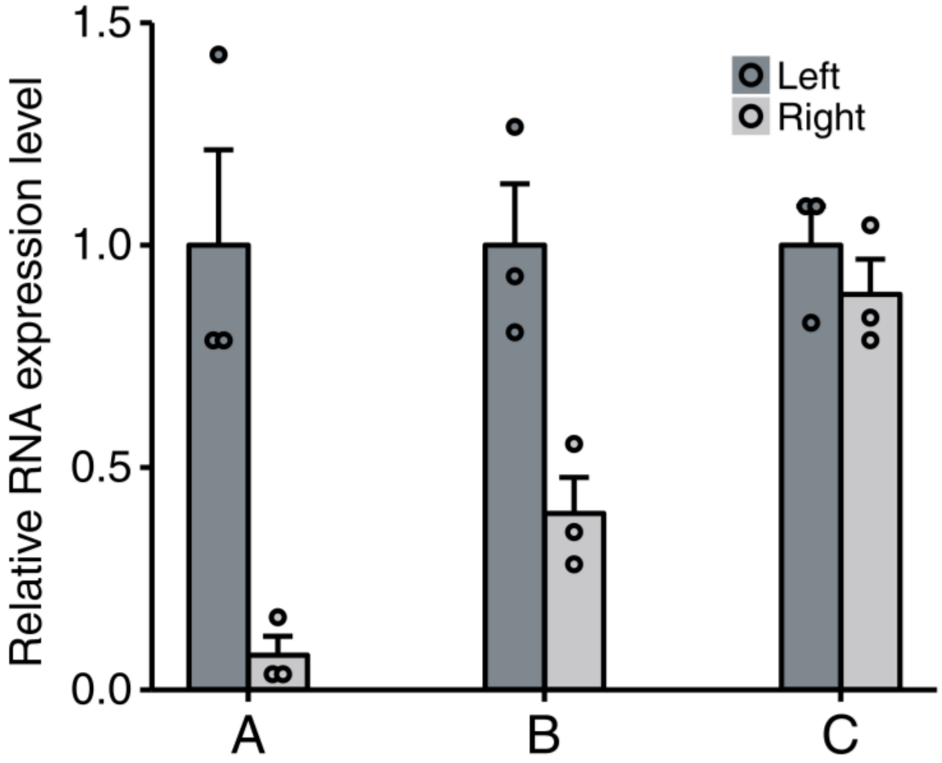


Figure 35: qRT-PCR of three isoforms in the left and right ventricle at E9.5

Promoter A shows a clear left-sided expression while the expression level of promoter C is almost the same between the left and the right. The expression pattern of promoter A isoform is similar to that of Tbx5 protein (n=3).

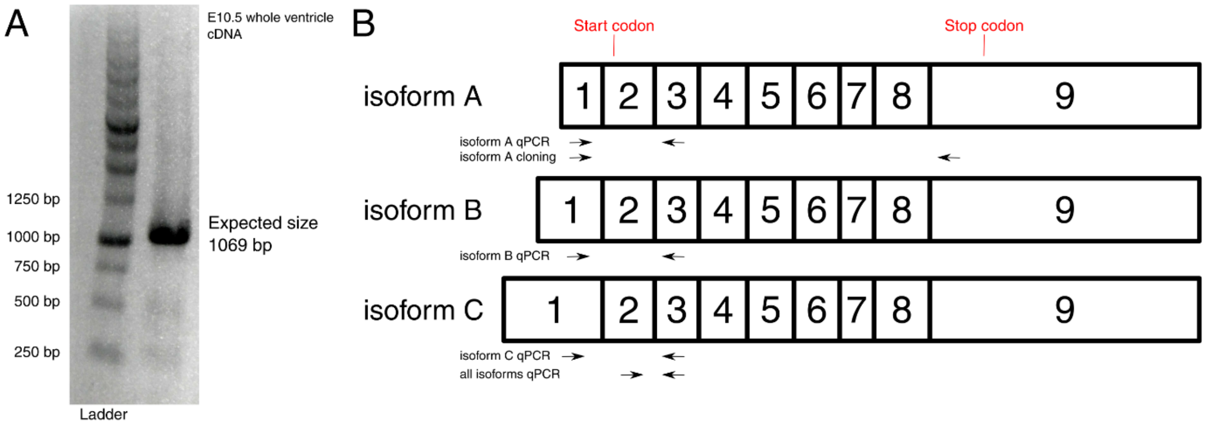


Figure 36: Cloning of isoform A and the primers used for PCR of this study

(A) to confirm the existence of full-exon isoform A, we performed PCR using cDNA derived from the E10.5 whole ventricle. A band was observed at the expected size and sequencing confirmed that this amplified product was indeed Tbx5 including exon A. (B) the structures of three isoforms of Tbx5. The only difference among them is the first exon and their protein coding sequences are the same. The locations of PCR primers used in this study and the structure of luciferase assay construct for isoform C were depicted.

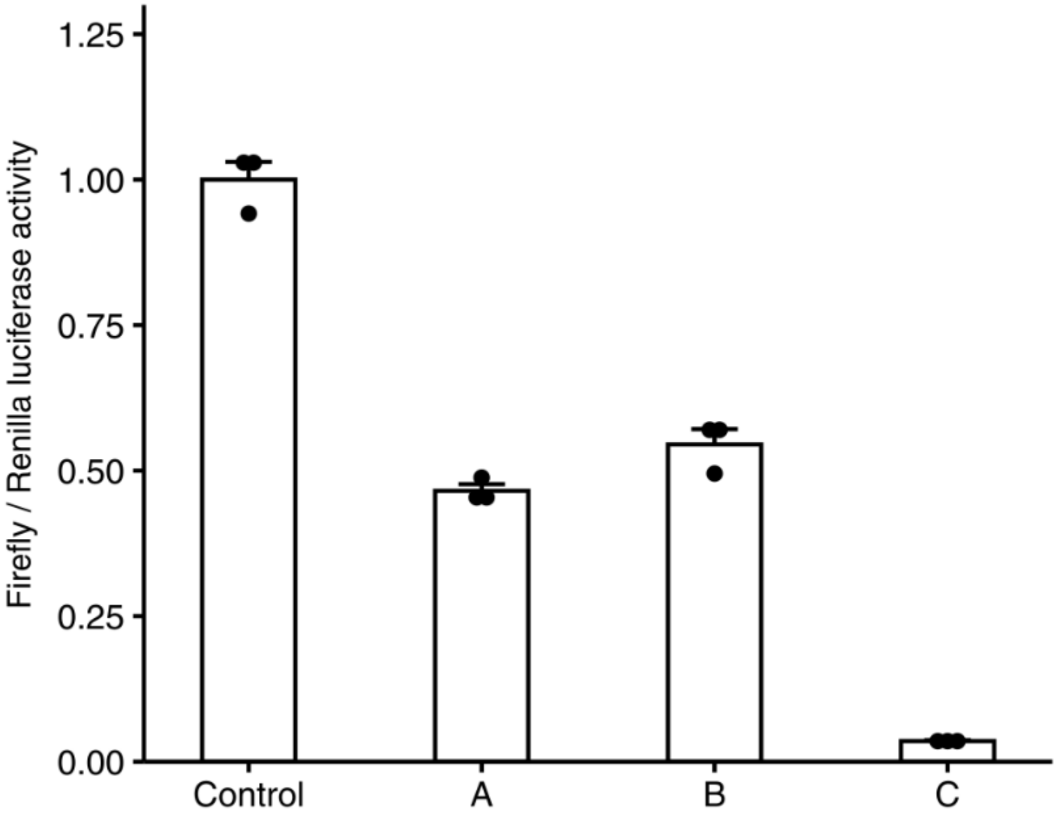


Figure 37: The effect of 5’ UTR of each isoforms on protein expression examined by dual-luciferase assay

Dual-reporter luciferase assay was conducted to investigate the translational effect of 5’ UTR of three *Tbx5* isoforms. I added 5’ UTRs to the firefly luciferase gene and used renilla luciferase as an internal control. The addition of promoter C isoform 5’ UTR resulted in a strong repression of luciferase activity. Control indicates unmodified firefly luciferase (n=3).

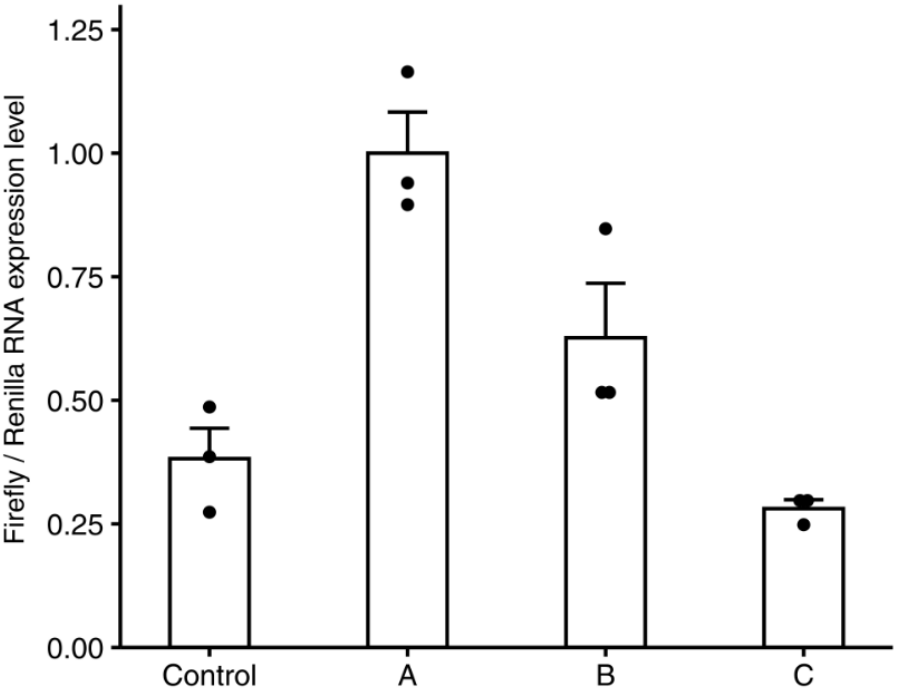


Figure 38: qRT-PCR of luciferase genes

qRT-PCR was performed against firefly and renilla luciferase gene. The addition of 5’ UTR of promoter A and B isoforms up-regulated mRNA expression levels while the addition of isoform C 5’ UTR had little effect on transcription, indicating that the decrease in the luciferase activity is primarily attributed to translational inhibition (n=3).

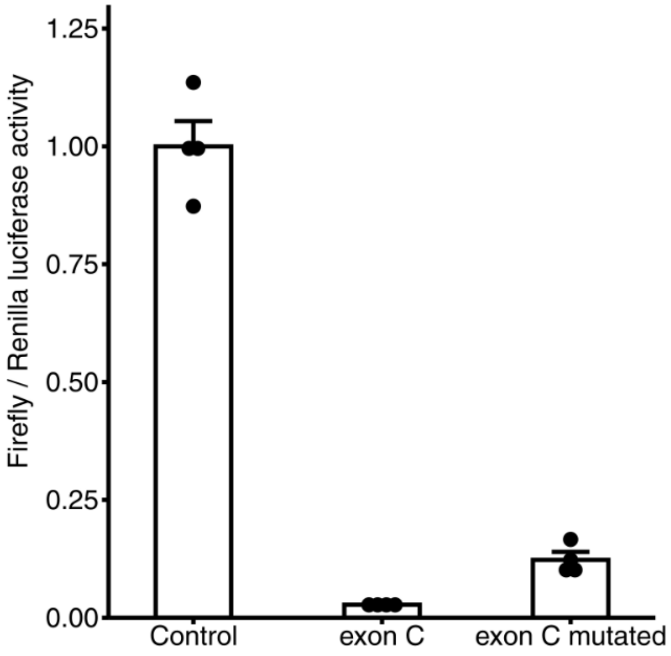


Figure 39: Mutation of uAUGs in exon C resulted in partial recovery of luciferase activity

Since upstream ORFs are known to have strong translational repression ability, I mutated AUGs of exon C to UUGs and examined if luciferase activity is recovered. The mutation of AUGs increased the activity by about 5-folds but the recovery was far from complete. Therefore, other strong repressive elements need to be considered to fully understand the nature of the repression (n=4).

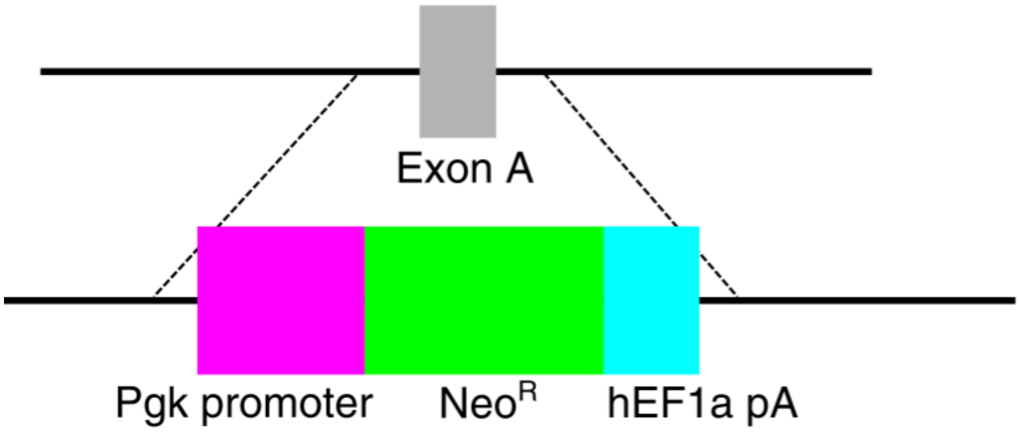


Figure 40: Schematic diagram of isoform A KO

Whole exon A was biallelically removed by knock-in using the Crispr/Cas9 system by replacing the first exon with Neomycin resistant gene cassette.

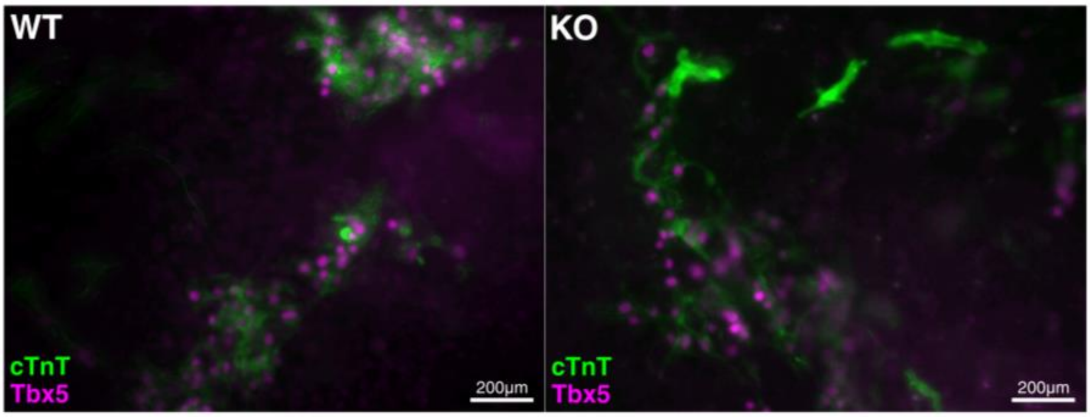
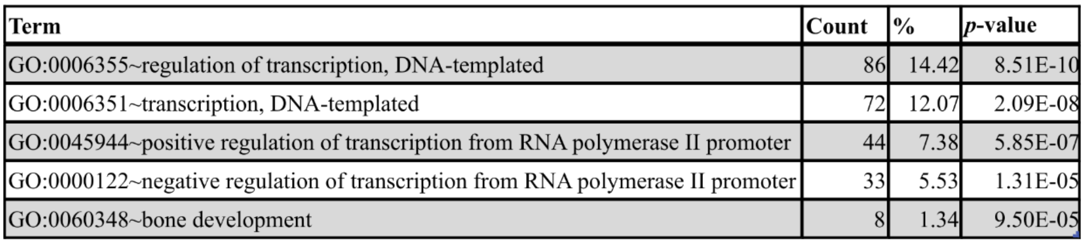


Figure 41: Antibody staining of WT and KO differentiated ESCs

Differentiated ESCs were stained with antibodies against Tbx5 and cTnT, a cardiomyocyte marker. Tbx5 protein expression was observed even among cardiomyocytes derived from promoter A KO ESCs, indicating that isoform A is not necessary for Tbx5 protein production.

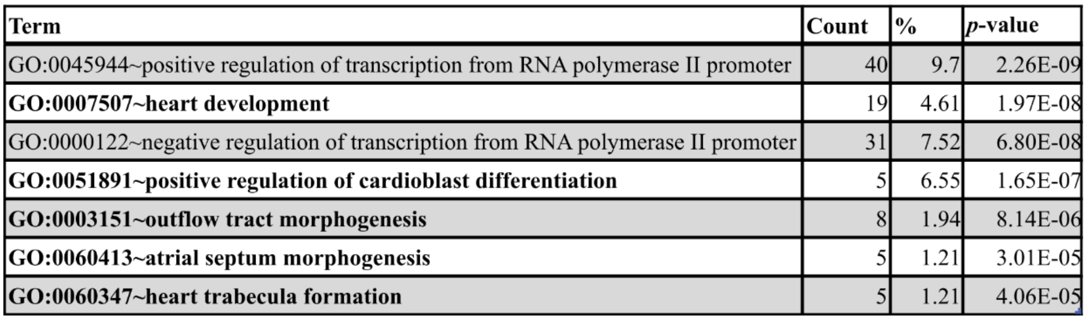
Tables

The nearest genes of lncRNAs expressed both in the heart and the brain



Term	Count	%	*p*-value
GO:0006355~regulation of transcription, DNA-templated	86	14.42	8.51E-10
GO:0006351~transcription, DNA-templated	72	12.07	2.09E-08
GO:0045944~positive regulation of transcription from RNA polymerase II promoter	44	7.38	5.85E-07
GO:0000122~negative regulation of transcription from RNA polymerase II promoter	33	5.53	1.31E-05
GO:0060348~bone development	8	1.34	9.50E-05

The nearest genes of heart-selective lncRNAs



Term	Count	%	*p*-value
GO:0045944~positive regulation of transcription from RNA polymerase II promoter	40	9.7	2.26E-09
GO:0007507~heart development	19	4.61	1.97E-08
GO:0000122~negative regulation of transcription from RNA polymerase II promoter	31	7.52	6.80E-08
GO:0051891~positive regulation of cardioblast differentiation	5	6.55	1.65E-07
GO:0003151~outflow tract morphogenesis	8	1.94	8.14E-06
GO:0060413~atrial septum morphogenesis	5	1.21	3.01E-05
GO:0060347~heart trabecula formation	5	1.21	4.06E-05

Table 1: Gene ontology analysis on the nearest genes of lncRNAs

In order to known what kind of genes is close to lncRNAs, I conducted gene enrichment analysis on the gene nearest to lncRNAs. Heart-selective lncRNAs were significantly enriched near genes important for heart development while such tendency was not observed in the nearest genes of the lncRNAs expressed both in the heart and the brain.