Decoding the transcriptional landscape of Nkx2-5

in heart development and disease

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Ann-Kristin Alteköster

aus Herford

Köln, 2019

The work described in this thesis was conducted during January 2015 and January 2019 under supervision of Prof. Dr. Richard Harvey at the Victor Chang Cardiac Research Institute, 405 Liverpool Street, Sydney 2010, NSW, Australia and Prof. Dr. Thorsten Hoppe at the Institute for Genetics at the University of Cologne, Zülpicher Straße 47a, D-50674 Cologne, Germany.

Die Arbeiten, die in dieser Dissertation beschrieben sind, wurden durchgeführt zwischen Januar 2015 und Januar 2019 unter der Anteitung von Prof. Dr. Richard Harvey am Victor Chang Cardiac Research Institute, 405 Liverpool Street, Sydney 2010, NSW, Australien und Prof. Dr. Thorsten Hoppe am Institut für Genetik der Universität zu Köln, Zülpicher Straße 47a, D-50674 Köln, Deutschland.

Table of Contents

Table of Contents

List of Figures

List of Tables

List of Abbreviations

All abbreviations used in this work are defined at the first use in each chapter. Some of the mostly commonly used abbreviations are given below.

bp	basepair
°C	degree Celsius
cDNA	complementary DNA
DAPI	4'6-diamidino-2-phenylindole
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide
E. coli	Escherichia coli
EDTA	ethylendiaminetetraacetic acid
g	gramm
kb	kilobases
kDA	kilodalton
min	minute
ml	millileter
mM	millimolar
mRNA	messenger RNA
ms	milliseconds
n	sample size
p-value	probability value
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PFA	paraformaldehyde
RNA	ribonucleic acid
rpm	revolutions per minute
RT	room temperature
s	seconds
SDS-PAGE	SDS polyacrylamide gel electrophoresis
SEM	standard error of the mean
SS	sodium dodecyl sulfate
TAE	tris base, acetic acid, EDTA buffer
TBS	tris buffered saline
TRIS	tris(hydroxymethyl)aminomethane
UV	ultraviolet light
V	Volt
µg	microgram
µl	microliter
Wnt	Wingless-related integration site

Abstract

A significant amount of the mammalian genome previously found to not code for proteins and considered 'junk', is actually pervasively transcribed and specifies a dynamic network of regulatory RNAs, the most abundant of which are long non-coding RNAs (lncRNAs). lncRNAs are critical players in many biological processes including epigenetic, transcriptional and post-transcriptional control as well as protein metabolism. Many exhibit cell-type or developmentally specific expression patterns and are associated with human disease. Interestingly, the majority of single nucleotide polymorphisms (SNPs) identified through genome wide association studies (GWAS) map to non-coding regions of the genome, which could potentially affect ncRNAs, given that ~80% of the human genome is transcribed.

Here, we show that the key cardiac transcriptional regulator, *Nkx2-5,* is, in humans and mouse, surrounded by two previously uncharacterised, long cardiac transcripts, which we named *NkxDS* and *NkxUS*. In humans, NKX2-5 mutations are commonly associated with congenital heart disease and a variety of specific cardiac malformations. Several lncRNAs have also been identified to be crucial for cardiac function as well as cardiac lineage commitment. At least 100 lncRNAs have been found to be cardiac specific, however, for most, their functional mechanisms, especially *in vivo,* remain unknown. This work aimed to characterise the two novel transcripts in order to explain whether *NkxUS* and *NkxDS* regulate *Nkx2-5* itself or other processes important for heart function, development and/or disease.

This work shows that both mouse transcripts are nuclear and cardiac-enriched, expressed in the heart throughout development and are also expressed specifically in the human heart. *NkxDS* is a novel long isoform of *Nkx2-5* as it contains the *Nkx2-5* protein-coding sequence. Therefore, *Nkx2-5* may have a protein-coding function as well as a non-coding role within the nucleus. In contrast, *NkxUS* is an independent transcript to *Nkx2-5*. Downregulation of mouse *NkxUS* expression in the HL-1 cardiac cell line displayed an upregulation of transcripts involved in cardiac contraction and conduction. Moreover, *NkxUS* interacting partners included proteins localising to the nucleus and mitochondria, which are involved in metabolism and in the regulation of transcription. Interestingly, a GWAS SNP associated with an increased resting heart rate lies within and disrupts a 600nt RNA structure conserved between mouse and human *NkxUS*. Analysis of mice harboring a partial deletion of this structure, generated by CRISPR/Cas9 technology, revealed no developmental defects, nor did the deleted structure have any effect on *Nkx2-5* expression. However, remarkably,

a similar phenotype as described for humans harboring the GWAS SNP was observed, with homozygous mice exhibiting an increased resting heart rate. The phenotype was shown to result from an elevated pacemaker activity, most likely due to enhanced signaling through the cAMP pathway caused possibly by an upregulation of ion channels.

In conclusion, this work has discovered *NkxDS* as a new isoform of the master cardiac transcription factor *Nkx2-5* and has shown that a conserved RNA structure within *NkxUS* plays an important role in the regulation of resting heart rate. These results provide new insights into the functional importance of lncRNA secondary structures and their involvement in GWAS SNPs associated traits and diseases.

Zusammenfassung

Ein erheblicher Teil des Säugetiergenoms, von dem zuvor festgestellt wurde, dass er nicht für Proteine kodiert und deshalb als "Müll" bezeichnet wurde, wird tatsächlich größtenteils transkribiert und bezeichnet ein dynamisches Netzwerk von regulatorischen RNAs, von denen lange nicht-kodierende RNAs (lncRNAs) die am weitverbreitetsten sind. LncRNAs spielen bei vielen biologischen Prozessen eine entscheidende Rolle, wie zum Beispiel bei der Kontrolle von epigenetischen, transkriptionellen und posttranskriptionellen Vorgängen sowie des Proteinstoffwechsels. Viele weisen zellart- oder entwicklungsspezifische Expressionsmuster auf und sind mit menschlichen Erkrankungen assoziiert. Interessanterweise befindet sich die Mehrheit der durch genomweite Assoziationsstudien (GWAS) identifizierten Einzelnukleotid-polymorphismen (SNPs) in nicht-kodierenden Regionen des Genoms. Diese SNPs könnten daher möglicherweise ncRNAs beeinflussen, da etwa 80% des menschlichen Genoms transkribiert wird.

Mit dieser Arbeit zeigen wir, dass der wichtige kardiale Transkriptionsfaktor, *Nkx2-5*, bei Mensch und Maus von zwei zuvor nicht charakterisierten langen herzspezifischen Transkripten umgeben ist, die wir als *NkxDS* und *NkxUS* bezeichnet haben. Beim Menschen sind NKX2-5 Mutationen häufig mit angeborenen Herzerkrankungen und einer Vielzahl spezifischer Herzfehlbildungen verbunden. Es wurde festgestellt, dass mehrere lncRNAs für die Herzfunktion sowie die Herzentwicklung eine wichtige Rolle spielen. Mindestens 100 lncRNAs sind herzspezifisch, jedoch sind die Funktionsmechanismen, besondere *in vivo*, für die Meisten noch unbekannt. Diese Arbeit zielte darauf ab, die beiden neuen Transkripte zu charakterisieren, um herauszufinden, ob *NkxUS* und *NkxDS Nkx2-5* selbst oder andere Prozesse, die für die Herzfunktion, die Herzentwicklung oder Krankheiten eine wichtige Rolle spielen, regulieren.

Diese Arbeit zeigt, dass beide Maus Transkripte im Zellkern und im Herzen angereichert sind, während der gesamten Entwicklung im Herzen und auch im menschlichen Herzen spezifisch exprimiert sind. *NkxDS* ist eine neue lange Isoform von *Nkx2-5*, da es die Protein-kodierende Sequenz von *Nkx2-5* enthält. Daher könnte *Nkx2-5* eine Protein-kodierende sowie eine nicht-kodierende Rolle innerhalb des Zellkerns ausführen. Im Gegensatz dazu ist *NkxUS* ein unabhängiges Transkript zu *Nkx2-5*. Die Verminderung der *NkxUS* Expression in der Maus HL-1 Herzzelllinie führte zu einer erhöhten Expression von Transkripten, die an der Herzkontraktion und -leitung beteiligt sind. Darüber hinaus interagiert *NkxUS* mit Proteinen, die im Zellkern und in den Mitochondrien vorhanden und am Metabolismus und an der Regulation der Transkription beteiligt sind.

Interessanterweise liegt ein GWAS-SNP, der mit einer erhöhten Ruheherzfrequenz assoziiert ist, innerhalb einer 600-nt RNA Struktur, die zwischen Maus und menschlichem *NkxUS* konserviert ist. Die Analyse von Mäusen, welche durch die CRISPR/Cas9-Technologie erzeugt wurde und Teile dieser Struktur fehlen, zeigte keine Entwicklungsdefekte und keine Änderung in der Expression von *Nkx2-5.* Bemerkenswerterweise wurde jedoch ein ähnlicher Phänotyp wie für Menschen, die den GWAS SNP enthalten, beobachtet, da homozygote Mäuse eine erhöhte Ruheherzfrequenz aufwiesen. Es konnte gezeigt werden, dass der Phänotyp aus einer erhöhten Schrittmacheraktivität resultiert, höchstwahrscheinlich aufgrund einer verstärkten Signalisierung des cAMP-Signalwegs, möglicherweise durch eine erhöhte Expression von Ionenkanälen.

Zusammenfassend hat diese Arbeit *NkxDS* als eine neue Isoform des Transkriptionfaktors *Nkx2-5* entdeckt und gezeigt, dass eine konservierte RNA Struktur in *NkxUS* eine wichtige Rolle bei der Regulierung der Ruheherzfrequenz spielt. Diese Ergebnisse liefern neue Einblicke in die funktionale Bedeutung von lncRNA Sekundärstrukturen und ihre Auswirkungen auf GWAS-SNPs, die mit Erkrankungen assoziiert sind.

1 Introduction

1.1 The heart and its physiology

The heart is the first organ to form and function in the vertebrate embryo [1]. It is a highly modified muscular pump, regulating and maintaining homeostasis by distributing oxygenated blood throughout the body. Blood is not only essential for providing oxygen but also nutrients to all tissues of the body and removing unwanted metabolic waste. Without a functioning heart, the organism will die within minutes.

1.1.1 Adult heart anatomy

Anatomically, the human heart consists of a left and right side, each with an atria and ventricle, four chambers in total [2-4] (Figure 1A). The right atrium receives oxygen-poor blood from the systemic veins and pumps it to the right ventricle, which in turn distributes the blood to the lungs via the pulmonary arteries (Figure 1A). Oxygen-rich blood from the lungs is received by the left atrium, then pumped into the left ventricle and via the aorta to the body's tissues. To prevent blood regurgitation, the chambers are separated from each other and the great vessels by valves. Atria and ventricles are separated by tricuspid (on the left) and mitral valve (on the right), which prevent back flow of blood during ventricular contraction. Ventricles and great vessels are separated by the pulmonary (on the right) and aortic valve (on the left), which prevent regurgitation after ventricular contraction [5].

1.1.2 The cardiac cycle

One cardiac cycle describes the event of one heart beat beginning with the atria contracting and ending with the relaxation of the ventricles [6]. It has two phases, diastole and systole. Diastole is the phase of relaxation during which the chambers relax and refill with blood, while systole is the phase of contraction and blood ejection out of the chambers. Both atria and ventricles undergo systole and diastole. The atrioventricular valves between the atria and ventricles are open in atrial diastole and systole ensuring that blood can flow into the ventricles. During ventricular systole, the atrioventricular valves close and the aortic and pulmonary valves open to allow blood flow into the aorta and pulmonary artery, respectively. At this point no blood enters the ventricles, only the

atria. The cardiac cycle begins with atrial systole, then ventricular systole, atrial diastole and ends with ventricular diastole until the cycle repeats itself.

1.1.3 Layers of the heart

The cardiac chamber wall is divided into three layers, the endocardium, myocardium and epicardium [5] (Figure 1B). The endocardium forms the endothelial lining of the heart and is the innermost and thinnest layer. It has a smooth surface, which prevents blood from adhering to the walls and therefore prevents potential damage of the heart and formation of blood clots. The myocardium is the middle and thickest layer. It consists mostly of cardiomyocytes, the cardiac muscle cells, which are able to contract. The epicardium is the mesothelium forming the outermost layer. It functions as a protective layer against infections and tumor propagation and provides lubrication for frictionless movement of the heart.

The thickness of the myocardium varies across the chambers. Atrial myocardium is thinner than that of the ventricles as less force is needed to pump blood into the neighboring chamber compared to distant organs. Additionally, the myocardium is thicker in the left than right ventricle as it has to generate more power to distribute blood throughout the body.

In numbers, the heart consists of about 30% cardiomyocytes [7]. Of the non-cardiomyocyte fraction, endothelial cells account for about 64%, resident mesenchymal cells including fibroblast for about 27% and leukocytes for 9% [7]. Inter-cellular crosstalk between these cell types and layers is essential not only during cardiac development but also for normal heart function. Small molecules, peptides and proteins for example can be secreted from endothelial cells and affect contractility of the myocardium or even induce an increase in cardiomyocyte cell size [8, 9].
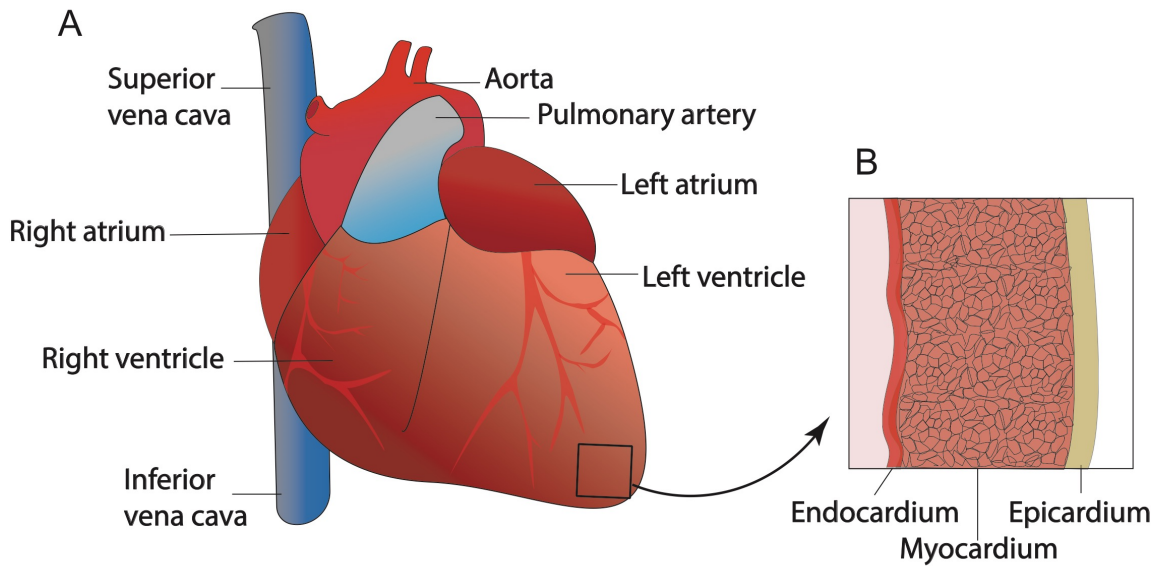
Figure 1 The adult mammalian heart. A) Structure of the adult heart. B) The three layers of the chamber walls.

1.1.4 Cardiac conduction system

The beating of the heart provides the powerful force required for pumping blood throughout the body and depends upon the proper activation and recovery of electrical excitation of the chamber myocardium, which is regulated by the cardiac conduction system. The cardiac conduction system consists of five elements, the sinoatrial node (SAN), the atrioventricular node (AVN), the bundles of His, the bundle branches and the Purkinje fibres [10]. Electrical excitation is mediated by different ion currents across the plasma membrane that result in action potentials, which are then propagated from cell to cell via gap junctions [11, 12]. The heartbeat is initiated by electrical impulses in the SAN, the pacemaker of the heart, which lies at the junction of the right atrium and superior vena cava [13, 14]. The SAN is a heterogeneous population of cells consisting of mostly pacemaker cells, which are specialised cardiomyocytes, but also of non-pacemaker cells like endothelial cells, fibroblasts and atrial myocytes [13, 15]. Adult pacemaker cells originate from and resemble embryonic cardiomyocytes by having a weaker contractile force and the ability to depolarise spontaneously [16, 17]. They are characterised by lacking a true resting membrane potential and being able to generate an action potential without external stimuli. Action potentials generated in the SAN are propagated to the atria causing atrial contraction. From the atria the impulse is conducted to the AVN, which is located in the interatrial septum, a wall of tissue

separating both atria [10]. The AVN delays the stimuli briefly giving the atria enough time to fully empty its blood into the ventricles. The valves close and the atria refill with blood. From the AVN the signal is propagated to the bundles of His and then via its left and right branches to the Purkinje fibre network on each side of the heart to activate the ventricular cardiomyocytes causing ventricular contraction [10].

Dysfunction in any of the five elements can lead to increased, decreased or irregular heart rates (cardiac arrhythmia) [18].

The propagation of action potentials can be visualised using an electrocardiogram (ECG), in which the P wave represents atrial depolarisation, the P-R interval the time it takes for the impulse to reach the ventricles through the AVN, the QRS-complex the ventricular depolarisation and the T-wave ventricular repolarisation (Figure 2).
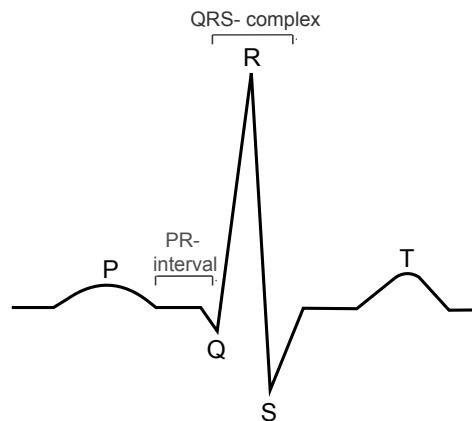
Figure 2 Schematic of an ECG of a human heart in normal sinus rhythm. The P-wave illustrates atrial contraction, the PR-interval reflects atrial and atrioventricular nodal conduction, the QRS-complex represents ventricular contraction and the T-wave depicts ventricular recovery.

1.1.4.1 The mechanism of sinoatrial node automaticity

A normal healthy human heart beats on average about 42 Million times per year, all due to the unique ability of the SAN to spontaneously depolarise and to generate regular action potentials to stimulate contraction.

The SAN action potentials are distinct from other cardiomyocytes and can be divided into three phases (Figure 3A and B). The action potentials start with Phase 4, the hyperpolarisation phase, in which the SAN specific hyperpolarisation-activated cyclic nucleotide-gated potassium (HCN) channels are activated by hyperpolarisation and conduct a slow, inward depolarising sodium

current (Figure 3A) [19, 20]. This current is often also referred to as the "funny current", as it has several unusual features such as activation by hyperpolarisation rather than depolarisation, permeability to both sodium and potassium and modulation by cyclic adenosine monophosphate (cAMP) [21]. Once a certain threshold is reached, low voltage gated T-Type calcium channels open followed by the high voltage gated L-Type calcium channels causing further depolarisation (Phase 0, depolarisation phase, Figure 3A) [22-25]. After a delay the potassium channels open, allowing efflux of potassium ions, which causes repolarisation of the cell (Phase 3, repolarisation phase, Figure 3A) [26]. Once the cell is completely repolarised, the cycle repeats itself spontaneously without a rest.

Action potentials of cardiomyocytes are quite different in that they have a true resting potential, a rapid depolarisation phase, a plateau phase and do not spontaneously depolarise (Figure 3B). Cardiomyocytes have to be activated by action potentials from adjacent cells to reach the threshold for calcium channel activation.

The mechanism of spontaneous depolarisation of SAN cells is mediated by two oscillators, the membrane and calcium clocks, which work in concert with each other (Figure 3C) [27-29]. The membrane clock resides in the plasma membrane and as mentioned above generates depolarisation mainly through the pacemaker current via HCN channels and also the voltage gated T- and L-Type calcium channels (Figure 3C). The calcium clock initiates spontaneous depolarisation by a rhythmic calcium release from the sarcoplasmic reticulum (SR), the main intracellular calcium store, mediated by the ryraonide receptor (RYR) (Figure 3C) [30, 31]. This calcium release activates the sodium calcium exchanger (NCX) in the plasma membrane, which exchanges three sodium ions for one calcium ion leading to depolarisation of the cell (Figure 3C) [30, 32]. Upon reaching the threshold to activate the voltage gated calcium channels, both clocks are in synchrony. After excitation, relaxation is facilitated by cytosolic calcium removal mainly via calcium uptake back into the SR by the sarcoplasmic/endoplasmic reticulum calcium ATPase (SERCA) (Figure 3C) [28]. Moreover, this uptake refills the calcium store and primes the cell for the next cycle [28].

Although action potentials are spontaneously generated by SAN cells, the rate of this activity can be modulated by external factors such as the autonomic nervous system and drugs through β-adrenergic and muscarinic stimulation, which modulates the activity of cAMP-dependent protein kinases that phosphorylate the illustrated channels in Figure 3C, except NCX [33-35].
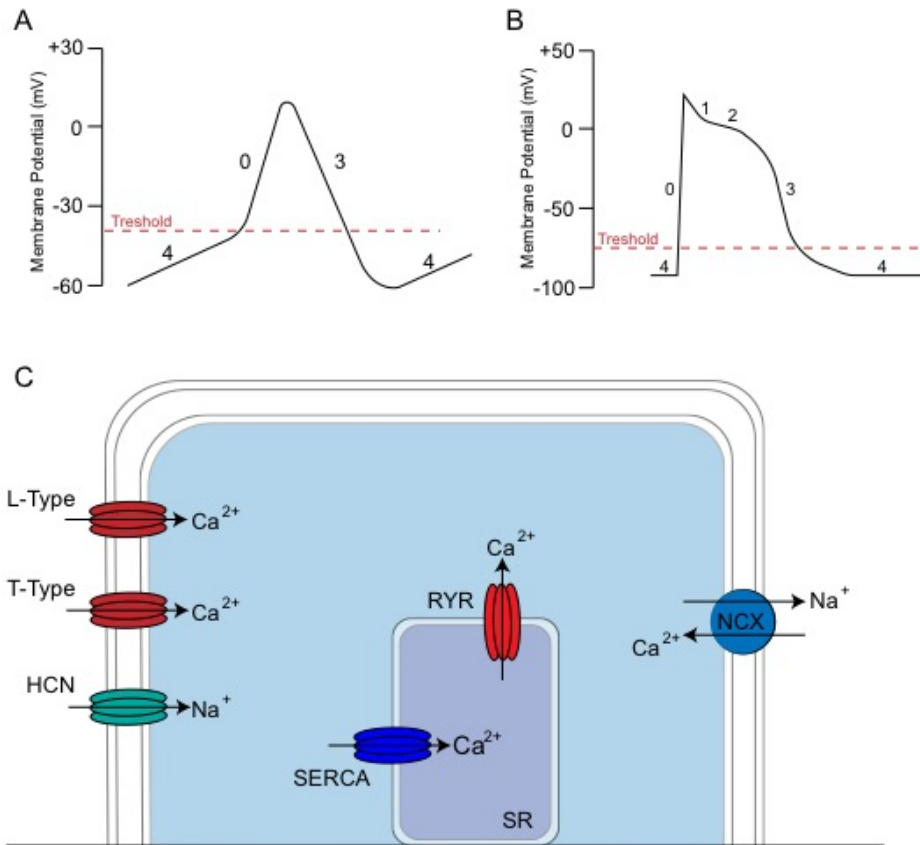
Figure 3 Sinoatrial node automaticity. A) Action potential from SAN cells. Phase 4 is the hyperpolarisation phase, Phase 0 the depolarisation and Phase 3 the repolarisation phase. In Phase 4, HCN channels are activated by hyperpolarisation upon reaching a membrane potential of ~-60 mV. These channels conduct a slow, inward depolarising sodium current. Upon reaching a threshold value of ~-50 mV T-type calcium channels open followed by the L-type calcium channel at ~-40mV further depolarising the cell (Phase 0). Repolarisation occurs when potassium channels open and calcium channels close (Phase 3). B) The cardiomyocyte action potential is quite different to SAN action potentials as they have a true resting potential (Phase 4), which stays constant at ~-90 mV, depolarise much more rapidly (Phase 0) and exhibit an initial repolarisation phase (Phase 1) and a plateau phase (Phase 2), which SAN cells do not possess. Further, these cells do not spontaneously depolarise and have to be activated by an action potential from an adjacent cell. C) The mechanism of SAN automaticity is provided by the membrane (HCN channels, L-type and T-type calcium channels) and the calcium (RYR, NCX, L-type calcium channel and SERCA) clocks. HCN channels are activated by hyperpolarisation and conduct an inward sodium current, which slowly depolarises the cell and upon reaching a threshold value activates the L-type and T-type calcium channels leading to further depolarisation. The calcium clock facilitates spontaneous depolarisation via calcium release from the SR through RYR, which activates NCX. NCX exchanges one sodium for three calcium further depolarising the cell. Upon reaching the threshold value L-type calcium channels open and both clocks are in synchrony. Calcium is taken up into the SR by SERCA to prepare the cell for the next depolarisation.

1.1.5 Adult heart disease

Cardiovascular disease (CVD) is the world's leading cause of death and disability and includes diseases affecting the heart or blood vessels. Globally, about 17.9 Million people die yearly from CVD representing 31% of all deaths (source WHO).

The most common types of adult cardiovascular disease are stroke, heart failure and coronary artery disease with heart attack and stroke accounting for about 85% of CVD deaths worldwide (source WHO). Smoking, obesity, increased age, diabetes, high cholesterol, hypertension, higher resting heart rate, physical inactivity and a family history of cardiovascular disease are well known risk factors for CVD. A few risk factors can be reduced by adopting a healthy lifestyle including a healthy diet, increasing physical activity, loosing weight and stopping smoking. However, given the high incidence of CVD, it is still surprising how little is known about the underlying causes.

1.1.5.1 Types of adult heart disease

There are several types of CVD which involve the adult heart including heart failure, myocardial infarction, cardiomyopathy and cardiac arrhythmias.

In heart failure, the pumping efficiency of the heart is decreased, thus it can no longer meet the body's requirement for blood supply and the organs as well as heart muscle do not get enough oxygen to work effectively [36].

Myocardial infarction (MI), commonly known as heart attack, occurs due to blockage of a coronary artery [37]. Coronary arteries supply the myocardium with oxygen-rich blood needed to meet its high metabolic demand. A decrease in oxygen supply causes death of cardiomyocytes and other cell types and ultimately results in the loss of functional myocardium [37]. In response to MI, the heart undergoes remodelling to try and adapt to the changes. Cardiac remodelling includes scar formation, infarct expansion, ventricular wall thinning and chamber dilatation, resulting eventually in reduced cardiac function and heart failure [38].

Cardiomyopathy refers to diseases affecting the myocardium, which causes enlargement and stiffness, leading to decreased contractility, thus inadequate pumping of the heart [39, 40]. It is a group of conditions that alter the structure of the myocardium, leading to severe cardiac

dysfunction and heart failure. Some are inherited while others are the result of viral infections or heart attacks [40-44].

Cardiac arrhythmias are caused by disrupted electrical signaling in the heart, leading to an abnormal heart beat [18, 45]. The heart is either beating too fast (tachycardia), too slow (bradycardia) or irregular (fibrillation) independent of exercise or stress [18]. The underlying mechanism responsible for cardiac arrhythmias can either be due to abnormalities in impulse conduction or formation in the SAN [46]. For example, sinus tachycardia is described as an elevated resting heart rate with no or minimal physical activity and results from either enhanced pacemaker automaticity, disordered autonomic nervous system activation or both [18, 46]. An elevated resting heart rate increases the oxygen demand and metabolic load on the heart and is associated with an increased risk for cardiovascular disease and all-cause mortality independent of blood pressure and physical fitness [47, 48]. Bradycardia results from a decreased intrinsic SAN function or a block in conduction, most commonly within the AV node or the His-Purkinje system [46]. Furthermore, disruption of the SAN membrane clock can result in heart rates alternating between high and low, termed tachycardia – bradycardia syndrome [18, 49]. An irregular heart rate can also occur due to premature depolarisation before full repolarisation or delayed after-depolarisation [18]. Both of which can occur under conditions of intracellular calcium overload, which can be caused by random and spontaneous calcium release from the sarcoplasmic reticulum, which in turn results in depolarisation of the cell [18, 46].

1.2   Heart development

The human heart begins to develop at conception, starts beating after about three weeks as a linear tube and is completely formed by 8 weeks into pregnancy. Cardiac development into a four chambered heart is a highly complex and dynamic process involving the migration and integration of several cell lineages (Figure 4) [50]. Briefly, the heart is formed from two progenitor cell populations, termed the first and second heart field, which exhibit different behaviors and lineage fates [51-53]. Embryonic development of the heart can be divided into four major phases: Cardiac crescent formation, formation of the linear heart tube, cardiac looping and chamber specification, and chamber septation and maturation.

During gastrulation, cardiac progenitors derived from paired regions of the lateral mesoderm extend across the midline to form the cardiac crescent or first heart field [1, 51, 53] (Figure 4A and B). Next, the progenitor cells move ventrally to form the linear heart tube, which consists of an inner endothelial and outer myocardial layer and is able to beat continuously (Figure 4C) [50, 51, 54]. Progenitor cells from the second heart field, derived from pharyngeal and splanchnic mesoderm, contribute progressively to both poles of the elongating linear heart tube, while it undergoes a process termed cardiac looping [51, 55]. The cells from the second heart field were shown to contribute to the outflow tract, right ventricle and both atria, while those of the first heart field contribute to the left ventricle and both atria [1, 52, 55]. During cardiac looping, the heart tube lengthens and folds to the right into an S-shaped form, which shows distinct anatomical features for ventricular and atrial components as well as the outflow tract (Figure 4D) [51]. With ongoing looping morphogenesis, the chambers and major vessels are placed in their correct alignment with the future atria and outflow tract being positioned above the future ventricle and the myocardium further expands with the chambers becoming more distinct [51]. Further, endocardial cushions form at the outflow tract, which later give rise to the different valves. During the next phase, atria and ventricles are separated into left and right by septa formation and the outflow tract is transformed into the ascending aorta and the pulmonary trunk [56]. During further growth and differentiation, the cardiac conduction system and coronary circulation are established to give rise to the adult four-chambered heart (Figure 4E).
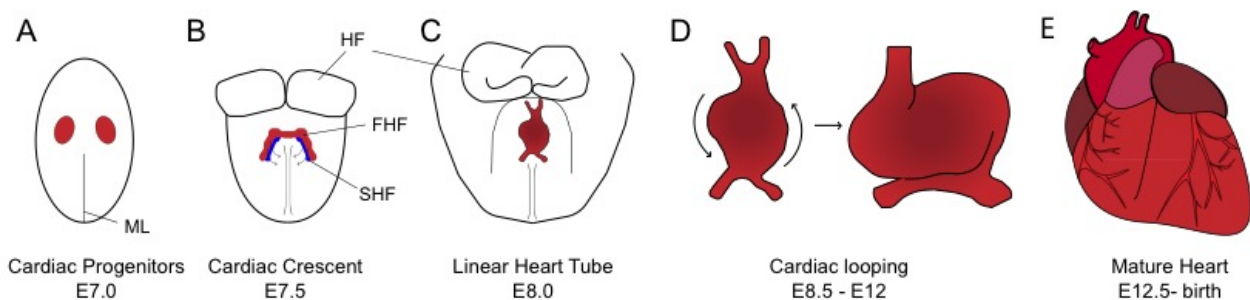
Figure 4 The embryonic development of the heart. Schematic representing mouse heart development with staging in days of embryonic development (E) including the first cardiac progenitors which migrate over the midline (A) to form the cardiac crescent and first heart field (B), second heart field formation at the cardiac crescent stage (B), the formation of the linear heart tube (C), which undergoes cardiac looping (D), followed by septation and maturation of the heart (E). Arrows indicate the direction of movement. ML= midline, HF= head folds, FHF= first heart field, SHF= second heart field.

1.2.1 Congenital heart disease (CHD)

The development of the heart is a complex and complicated process, requiring tight and accurate regulation of related genes over time and space as described above. If one of these crucial steps does not happen at the right time, the heart does not develop properly leading to congenital heart disease (CHD). CHD is defined as a defect in the structure of the newborn's heart and is worldwide the most common birth defect, affecting up to 1% of newborns. CHD can affect the walls between the atria or ventricle, the valves, blood vessels, or the chambers themselves [57]. While some defects cause no or few symptoms, others need treatments directly after birth including in the most severe cases multiple surgeries. Others are lethal even before birth. Many different types of CHD exist, likely arising due to defects in the early stages of heart development [50, 58]. The cause of disease is largely still unknown; therefore it is important to understand the genetic and developmental pathways that shape the heart [50, 58].

1.2.1.1 Types of CHD

The many different types of CHD can be grouped based on anatomic and physiologic features.

The most common types are septal defects, in which the wall between the atria or ventricles fails to close completely during development [59]. These defects allow blood flow between the right and left chambers, thereby reducing cardiac function. Depending on the size of the lesion, the defects can cause mild to severe outcomes.

One of the most serious defects is hypoplasia of the heart, in which one of the chambers and/or aorta are severely underdeveloped and not able to pump blood efficiently [60].

Other defects can be group together based on a blocked or narrowed vein, artery or valve leading to an increased resistance for the heart in order to pump blood to the body's tissue, which may cause hypertension and cardiac hypertrophy.

Further, cyanotic defects are a group of CHD's, in which newborns exhibit bluish skin due to lack of oxygen. Several birth defects can result in cyanotic defects including abnormalities in any of the large blood vessels, which lead to or from the heart, or the valves [59]. Classically this type of CHD is not due to a single abnormality, rather a combination of defects which results in

deoxygenated blood bypassing the lungs and entering the aorta. The most common cyanotic defect is tetralogy of Fallot which is due to a combination of four defects [61]:

- a ventricular defect
- an overriding aorta, meaning that the aorta is localised directly over the ventricular septal defect, thus allowing deoxygenated blood to enter the left ventricle and be pumped throughout the body
- pulmonary stenosis, in which the pulmonary artery is constricted thus decreasing the amount of blood going to the lungs and increasing the resistance to blood flow, which leads to fourth defect
- right ventricular hypertrophy, in which the myocardium of the right ventricle is abnormally thick

1.3   Cardiac gene regulatory networks

A gene regulatory network (GRN) consists of a set of "controller" and "responder" genes spread across the genome, which interact with each other to coordinate and regulate cellular functions important for development, differentiation and environmental adaptation.

1.3.1 Transcription factors

Development and function of the mammalian heart are regulated by complex gene networks that orchestrate morphogenesis, physiology and adaptation [50, 58, 62, 63].

The network includes several evolutionary conserved core transcription factors, also referred to as kernels (Figure 5). The most critical kernels include NK2 Homeobox 5 (NKX2-5), myocyte enhancer factor 2c (MEF2c), Heart and Neural Crest Derivatives Expressed 1/2 (HAND1/2), T-box 5/20 (TBX5/20), Insulin Gene Enhancer Protein (ISL1), Serum Response Factor (SRF) and zinc-finger transcription factor GATA4/6 (Figure 5) [62, 63]. All of these are expressed in the cardiac progenitors and differentiating cells [62]. Components of the network physically interact and function synergistically with one another and with an array of other transcription factors to control

cardiogenesis [62, 63]. Not surprisingly, disruption of the network has catastrophic consequences for development with mutations in at least 25 transcription factors identified in CHD patients [64].

Understanding and determination of the components of the cardiac GRN and its subcircuits, as well as regulators of these core transcription factors could give further insight into heart development and thus the underlying cause of heart disease.

A summary of the network is shown in Figure 5.
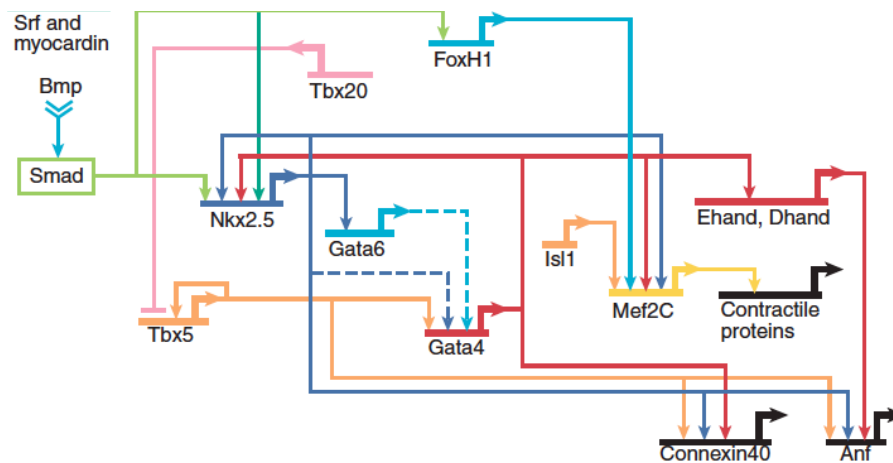
Figure 5 Simplified cardiac gene regulatory network of the core transcription factors involved in vertebrate heart development (from [65]). NKX2-5 sits at the center of the network interacting with several other kernels.

1.3.2 The transcription factor NKX2-5

One of the key transcriptional regulators of the heart and earliest known marker of vertebrate cardiac lineage is the homeodomain transcription factor NKX2-5 [66-68]. The abbreviation NK refers to Nirenberg and Kim, the researches that identified this class of homeodomain containing proteins in Drosophila [69]. Originally identified as *tinman*, a gene essential for heart and visceral muscle development in *Drosophila*, *Nkx2-5* is highly conserved and expressed in cardiac precursor cells from *Drosophila* to humans [68, 70].

NKX2-5 is a central player of the cardiac GRN and known to regulate a number of other genes involved in this network. In the mouse, NKX2-5 is expressed in the forming myocardium from E7.0 onwards, is expressed in both heart fields and its expression in the heart is maintained throughout development [68, 71]. Expression was also detected in the developing foregut, spleen,

tongue, thyroid and stomach [68]. After birth, NKX2-5 continues to be expressed in the heart, however, with much lower expression. In mouse, targeted disruptions of NKX2-5 have catastrophic outcomes for development, with early embryonic lethality by E10.5 due to the absence of cardiac looping, abnormal chamber formation and growth retardation [72-74]. Heterozygous and conditional *Nkx2-5* mouse models show milder abnormalities leading to septal defects, conduction disorders and a propensity for lethal arrhythmias [74-77]. In summary, studies in mice have shown that *Nkx2-5* is important for the formation and maintenance of chamber myocardium as well as the cardiac conduction system [72, 73, 78-80]. Based on its essential role in cardiac development it does not come as a surprise that mutations within NKX2-5 are associated with a collection of congenital heart disorders. In humans, about 40 mutations within NKX2-5 were linked to structural malformations as well as conduction, alignment and compaction defects of the heart, most commonly atrial and ventricular septal defects and atrioventricular conduction block as well as tetralogy of Fallot [76, 81-83].

1.3.2.1 Transcriptional regulation of *Nkx2-5*

The NKX2-5 gene in mouse as well as humans consists of two main exons. Studies in the mouse identified a complex regulatory region of about 27kb surrounding NKX2-5, consisting of two additional upstream exons, termed exon 1a and 1b, as well as a variety of independent enhancer and negative regulatory elements which drive *Nkx2-5* expression in different tissues and compartments of the heart (Figure 6) [73, 84-87]. Several alternative *Nkx2-5* isoforms exist with exon 1a and 1b also being able to be spliced to *Nkx2-5* coding region [84, 88]. So far, nine activating and three inhibitory elements were identified providing tissue specificity by driving *Nkx2-5* expression in cardiac and non-cardiac tissue, but none alone could account for the complete *Nkx2-5* expression pattern [85].

In total, seven activating regions were identified which recapitulate *Nkx2-5* expression in cardiac tissues (Figure 6). Activating region 1 (AR1) drives early *Nkx2-5* expression in cardiac progenitor cells throughout heart development until the chambers form after which its expression becomes strictly localised to the right ventricle (Figure 6) [87]. AR2 recapitulate *Nkx2-5* expression in the cardiac crescent and linear heart tube, and later becomes restricted to the outflow tract and right ventricle (Figure 6) [86]. A third cardiac enhancer (AR3) is responsible for expression in the

outflow tract (Figure 6). These three enhancers can be repressed by inhibitory regions (IR), which are located immediately next to the individual enhancers (Figure 6) [84, 87]. Additional enhancers drive expression in both ventricles (AR4), right ventricle (AR5) and both atria (Upstream Homology [UH] region 5) [88, 89]. An additional upstream enhancer composed of clustered repeats of Smad and GATA DNA binding sites, termed the Gata-Smad-responsive enhancer (G-S), recapitulates expression in the cardiac crescent (Figure 6) [90].

Lastly, enhancer elements for non-cardiac tissues included AR6, which is responsible for thyroid expression, AR7 for expression in thyroid, spleen, stomach and pharynx and UH4 for expression in tongue (Figure 6) [88, 89].

Further, it was shown that transcription factors interact with some of these enhancers to participate in regulating *Nkx2-5* expression [86, 89-91]. A network interference model predicted binding of the transcription factor Tead1 (TEA Domain Transcription Factor 1), GATA4, MSX2 (Msh Homeobox 2) and TGIF1 (TGFB Induced Factor Homeobox) to AR2 and EGR1 (Early Growth Response 1) and SOX11 (SRY-related HMG-box) to AR1 [86, 87, 91].

In summary, *Nkx2-5* is a tightly regulated gene with a set of upstream enhancers directing *Nkx2-5* expression during cardiac progenitor commitment and specification and another set driving expression during chamber specification and septation [89].
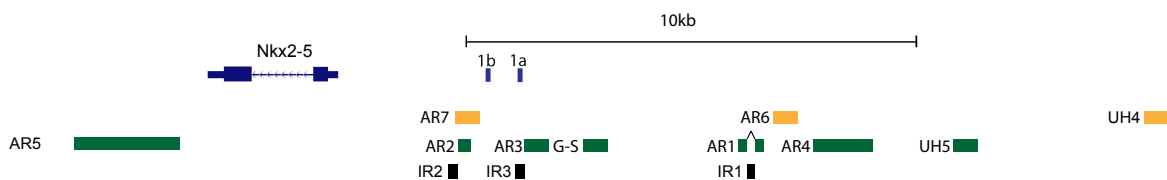
Figure 6 Mouse *Nkx2-5* genomic region with locations of regulatory elements. Diagram showing *Nkx2-5* with its upstream exons, enhancers and inhibitory elements. Cardiac enhancers are indicated in green and non-cardiac in yellow. Inhibitory regions are highlighted in black. AR= activating region, G-S= Gata-Smad-responsive enhancer, UH= upstream homology region, IR= inhibitory region.

1.3.3 Long non-coding RNAs

Recent years showed that GRNs are not only controlled by transcription factors but also by long non-coding RNAs (lncRNA). Advances and widespread application of high-throughput sequencing methods revealed that the vast majority of the human genome does not code for genes that

produce proteins. Only around 2% of the human genome actually consists of protein-coding genes. Prior to the completion of the human genome project, it was assumed that human complexity was due to an increased number of protein-coding genes. However, the results showed that humans have a similar number of protein-coding genes to, for example, nematode worms, suggesting that the number of protein-coding genes cannot explain human complexity [92, 93]. However, the proportion of non-coding DNA was shown to be positively correlated with increased organismal complexity. Humans have the highest amount of non-coding DNA, indicating that this might be a more accurate measurement of complexity [94, 95].

About 80% of our DNA is transcribed, mostly into non-coding RNAs (ncRNAs), which do not contain a defined open reading frame [96]. They can be divided into short and long ncRNAs based on their transcript size. Short ncRNAs include mainly well-validated classes of regulatory RNAs like microRNAs (miRNAs), which will not be further discussed in this thesis.

Transcripts with a length greater than 200 nucleotides are arbitrarily defined as lncRNAs. LncRNAs are a rapidly growing class with until now up to 96,308 and 17,216 described human lncRNA genes and transcripts, respectively (Source: NONCODE). Until recently it was thought that GRNs were comprised exclusively of protein-coding genes mediating control of the genome and that disease resulted as a consequence of dysfunction of this network. However, the interest in lncRNAs has produced a major shift in our understanding of gene regulation with lncRNAs playing critical roles within GRNs as structural and regulatory elements.

1.3.3.1 Structure

Intriguingly, lncRNAs share many processing properties with protein-coding RNAs in that many are transcribed by RNA Polymerase II (or III), spliced, have a 5'cap and a poly(A) tail [97, 98]. One of the only differences lies within the absence of an open reading frame that can be translated [99]. Examination of their genomic location revealed that many protein-coding genes are in fact surrounded by a host of lncRNAs transcripts [100-108]. These can be expressed in sense or antisense orientation and can be intronic, overlapping or intergenic relative to the location of nearby protein-coding genes [109-111].

1.3.3.2 Localisation

Many lncRNAs display highly specific sub-cellular localisation patterns providing key steps towards elucidating their possible regulatory roles [98, 112]. The patterns can range from specifically nuclear, exclusively cytoplasmic or a combination of both [113]. A major fraction of lncRNAs are retained in the nucleus with some localising to specific sub-nuclear organelles [112]. *NEAT1* (Nuclear enriched abundant transcript 1) and *MALAT1* (Metastasis Associated Lung Adenocarcinoma Transcript 1), for example, localise to specific nuclear bodies, the paraspeckles and nuclear speckles, respectively, two compartments involved in splicing [114, 115]. *NEAT1* plays an essential role in paraspeckle formation and maintenance, while *MALAT1* is required for optimal splicing [114-116].

1.3.3.3 Expression

Analysis of lncRNAs revealed that many exhibit exceptional cell-type, tissue and developmental specific expression patterns [117-121]. One major concern for many years regarding the functionality of lncRNAs has been their low abundance in whole tissue RNA sequencing data. However, the emergence of single cell RNA sequencing revealed that rather than being distributed at low expression over thousands of cells, lncRNAs expression was highly precise and dynamic [121-123]. In individual cells, lncRNAs were expressed at similar levels to mRNAs, whereas in pooled cells expression appeared lower [122, 123]. Therefore, even though some lncRNAs exhibit low expression in bulk tissues, they may still play essential biological roles and should not be ignored. Further, since lncRNAs may not produce a protein, their function is tied to the specific tissue or cell-types in which they are expressed. Hence, it is useful to determine their cellular and tissue specific expression as a first step towards exploring their function.

1.3.3.4 Conservation

In contrast to mRNAs, most lncRNAs lack or have low sequence conservation, which for many years stimulated a heated debate regarding their biological relevance. However, non-conservation agrees with the observation that lncRNAs emerged only recently in evolution and that the sequence of

lncRNAs is not constrained by the need to encode amino acids. In contrast to sequence conservation, splice sites, promoters or chromatin signatures of many lncRNAs have been found to be evolutionary well conserved [124-127]. Further, computational approaches identified that secondary structures within lncRNAs are often conserved in vertebrates [128-130]. These conserved structures frequently show overlap in expression between vertebrates and/or co-localisation with binding sites for the same RNA-binding proteins, indicating that knowledge of the nature and conservation of RNA structures might imply biological relevance and give insights into their functionality and mechanisms of action [119, 128, 129].

1.3.3.5 Function

Unlike miRNAs, which post-transcriptionally regulate gene expression, lncRNAs have more diverse functions. LncRNAs can form specific RNA-RNA, RNA-DNA and RNA-protein interactions and thereby can regulate genetic output at almost every stage of a gene's life cycle, from epigenetic, transcriptional and post-transcriptional control to protein metabolism [117, 131-135]. Over the past decade, there have been many studies linking lncRNAs with a variety of functions within cellular processes including regulation of chromatin remodelling, pre-mRNA processing, mRNA stability and splicing as well as protein stability, activity or localisation [112, 136-140]. They commonly interact with chromatin-modifying proteins including the repressive PRC2 (polycomb repressive complex 2) and the activating MLL/TrxG (Mixed-Lineage Leukemia/Trithorax group) complex as well as histone-modifying enzymes and transcription factors to control gene expression [97, 121, 134, 139, 141]. LncRNAs have various mechanisms of function. They may act as a guide to recruit proteins or enzymes to their specific site of action, as decoys to prevent target site binding or even as modular scaffolds to aid in the assembly of relevant complexes (Table 1) [142, 143]. In fact, many lncRNAs were identified to directly interact with either PRC2 or MLL/TrxG to inhibit or activate target gene expression [144, 145]. Others act as miRNA sponges, competing for binding with miRNA targets, thus trapping the miRNA and leading to the activation of its target genes instead of silencing (Table 1) [146-149]. LncRNAs can act in *cis* and/or *trans* and thus are able to control gene expression from neighbouring to very distant genomic loci [105, 150]. Additionally, lncRNAs can indirectly control expression of neighboring genes through mechanisms of their biogenesis rather than the lncRNA molecule itself [104]. When terminating the transcription start site of the cardiac lncRNA *Uph*

(Upperhand), which is transcribed from a divergent promoter shared with the cardiac transcription factor *Hand2, Hand2* expression was lost [104]. In contrast, knockdown of the transcript had no effect [104].

Overall, thousands of lncRNAs still remain un-characterised and a major challenge lies in defining their function. Regarding the heart, several lncRNAs have been identified to be involved in the regulation of cardiac pathways [101, 139, 148, 149, 151, 152]. Some are involved in metabolism and cardiac conduction, others altered during heart development as well as disease and some showing cardioprotective properties (Table 1) [101, 102, 105, 151-156]. Considering the still unknown function of thousands of lncRNAs, many more are likely to be discovered as important regulators of heart development and/or function and may represent an additional layer of regulation to the cardiac transcription factor composed GRN.

1.3.3.5.1 Role of LncRNAs in heart development

Analysis of the transcriptome of embryonic (E13.5) and adult mouse hearts, revealed hundreds of differentially expressed lncRNAs with about 117 being cardiac enriched, indicating that many lncRNAs likely exhibit specific functions in cardiac development [157]. Interestingly, the same study compared healthy adult with hypertrophic hearts and found far less differentially expressed lncRNAs [157]. Another RNA sequencing study identified 1378 lncRNAs expressed in E8.25 hearts with about 200 being specifically expressed in the heart at this stage [158].

Examples of lncRNAs altered during development include *Braveheart*, *Fendrr (*Fetal-lethal non-coding developmental regulatory RNA), *ALIEN* and *CARMEN* (Cardiac mesoderm enhancer-associated noncoding RNA), all of which are expressed in cardiac progenitor cells and, except *ALIEN*, found to interact with PRC2, either as a decoy (*Braveheart*) or guide (*Fendrr*) (Table 1) [102, 103, 152, 154]. The mechanisms of how *CARMEN* regulates gene expression through PRC2 remains unknown [154]. Interestingly, *Fendrr* is also able to interact with the activating MLL/TrxG complex [103]. Additionally, *Fendrr* is one of the very few lncRNAs involved in heart development that have been functionally interrogated *in vivo* in mice [103]. Consistent with its early cardiac expression, *Fendrr* loss-of-function resulted in disturbed cardiac morphogenesis and embryonic lethality at E13.75, most likely due to myocardial dysfunction [103].

However, most of the lncRNAs involved cardiac development lack *in vivo* proof of function and were mostly studied *in vitro* in differentiated stem cells, which can provide important insights into their function, however, has to be interpreted with caution.

1.3.3.5.2 Relevance of lncRNAs in cardiovascular disease

Besides their involvement in cardiac development, lncRNAs are also important regulators of homeostasis in adult hearts and their disruption is linked to a variety of human cardiovascular diseases, including heart attack, heart failure, hypertension and cardiac arrhythmias. Annotation of the cardiac transcriptome after cardiac disease identified hundreds of deregulated lncRNAs and together with their often very specific expression, they have great potential to be used as biomarkers for cardiac form and function as well as therapeutic targets in the future [101, 139, 156, 159].

Deep RNA sequencing of failing and non-failing human heart samples before and after mechanical support with an assist device revealed dynamic regulation of myocardial lncRNAs [156]. About 10% of the lncRNAs, deregulated between failing and non-failing hearts, showed improved expression after mechanical circulatory support, suggesting that lncRNAs play biochemical roles in reverse remodeling observed with mechanical support [156]. Further, the expression profile of lncRNAs was reported to be more predictive than the profile of mRNAs or miRNAs to discriminate failing from non-failing hearts, indicating an important pathophysiological role of lncRNAs in the cardiovascular system [156].

Examples of cardiovascular disease associated lncRNAs are summarised in Table 1. Some lncRNAs, including *ANRIL* (Antisense noncoding RNA in the INK4 locus) and *MIAT* (Myocardial infarction associated transcript), were identified in genome-wide association studies (GWAS), in which a single nucleotide polymorphism (SNP) associated with an increased risk for coronary artery disease and myocardial infarction was identified within the *ANRIL* and *MIAT* locus, respectively [108, 160]. In fact, an increasing number of the trait- and disease-associated GWAS SNPs map to non-coding, intergenic regions of the genome with many found to be within transcriptionally active regions, suggesting that the causative SNPs may affect ncRNA's expression and/or functions and highlighting their potential as candidate biomarkers [161-165].

Of note, lncRNAs may also be cardioprotective. During cardiac stress, including myocardial infarction and cardiac hypertrophy, *Mhrt* (Myosin heavy chain associated RNA transcript) expression is repressed and restoring expression was reported to provide cardioprotection [153].

Moreover, lncRNAs were shown to predict cardiac disease development and survival. By analsying plasma samples of heart failure patients, expression of the mitochondrial transcript *LIPCAR* was found to increase significantly during the process of cardiac remodeling after myocardial infarction and during chronic heart failure, thus *LIPCAR* expression levels may serve as a prognostic marker for heart failure [106, 166].

Taken together, the highly specific expression patterns and powerful regulatory functions of lncRNAs in cardiac disease onset and progression as well as association to genetic risk factors supports their potential as biomarkers for diagnosis and prognosis as well as promising targets for new therapeutic approaches. Several tools are available to modulate lncRNAs to either inhibit or overexpress lncRNAs. LncRNA expression can be reduced during disease by using antisense-based approaches, or restored by reintroducing its activity using for example viral delivery [167, 168]. But also here, challenges in improving delivery and reducing toxicity still exist. Furthermore, the field of lncRNA research is still relatively new and to date detailed mechanism of function is only described for a few lncRNAs, with even fewer studies providing functional validation *in vivo*. Thereby, emphasising the need for detailed characterisation and systematic functional testing of lncRNAs to allow the development of highly targeted therapeutic approaches, which would provide a considerable benefit to human healthcare.

Table 1 Summary of functional roles and mechanisms of known cardiac lncRNAs

Name	Role	Mechanism	References
ALIEN	Cardiac development	Unknown	[152]
Braveheart	Cardiac development	Decoy	[102]
CARMEN	Cardiac development & homeostasis	Unknown	[154]
Fendrr	Cardiac development	Guide	[103]
ANRIL	Genetic risk factor for coronary artery disease	Scaffold	[160]
MIAT	Genetic risk factor for myocardial infarction	miRNA sponge	[107, 108, 169]

TINCR	Inhibits cardiac hypertrophy	Guide	[170]
Mhrt	Inhibits cardiac hypertrophy	Decoy	[153]
CARL	Inhibits cardiac apoptosis	miRNA sponge	[148]
CHRF	Promotes cardiac hypertrophy	miRNA sponge	[149]
CHAST	Promotes cardiac hypertrophy	Unknown	[171]
LIPCAR	Predicts onset and survival in heart failure patients	Unknown	[106]
SENCR	Promotes smooth muscle cell contraction	Decoy	[172]

1.4   Discovery of two novel transcripts surrounding mouse *Nkx2-5*

In mice, as well as in humans, *Nkx2-5* is an isolated protein-coding gene, having no neighbouring protein-coding genes located 50-70kb up- and downstream. However, while scanning available ENCODE RNA data of adult and embryonic mouse hearts, Dr. Nicole Schonrock (previous Harvey group member, Victor Chang Cardiac Research Institute, Australia) identified the area surrounding *Nkx2-5* to be very transciptomically rich. Highly expressed, long RNAs were expressed directly downstream as well as on the opposite strand upstream of mouse *Nkx2-5* (Figure 7) [96]. This expression seems to be heart specific as it was not found in ENCODE transcriptome data of any other tissue type (Figure 7). Based on their genomic location relative to *Nkx2-5*, the transcripts were termed *NkxDS* for Nkx2-5 DOWNSTREAM and *NkxUS* for Nkx2-5 UPSTREAM.
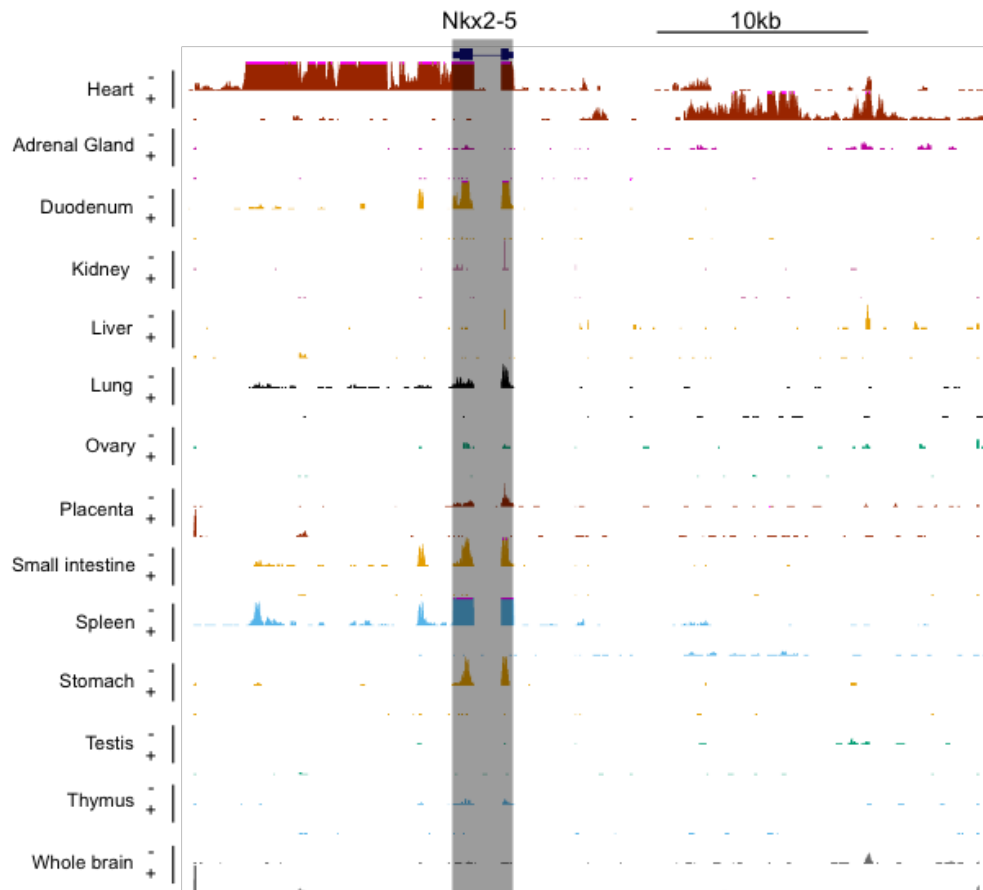
Figure 7 Transcriptional landscape of mouse *Nkx2-5*. Overview of the genomic location of *Nkx2-5* (in gray) and ENCODE transcriptome data on different mouse tissues showing that *Nkx2-5* is surrounded by heart specific transcripts.

1.5 Aim of this thesis

Nkx2-5 has been studied for about two decades to analyse its contribution to heart development and CHD. However, very little is known about its biochemical function and regulation, or how it participates in guiding the cardiac GRN.

Based on the recently discovered cardiac enriched transcripts and their close proximity to *Nkx2-5*, we hypothesised that *NkxUS* and *NkxDS* are either involved in the regulation of *Nkx2-5* itself or other genes important for heart development and/or function. The aim of this thesis was to characterise *NkxUS* and *NkxDS* to determine their potential role within the cardiac GRN. Specifically, my experimental approaches are summarised below in Table 2:

Table 2 Summary of thesis aims with experimental approach and objectives.

Feature Analysed	Experimental Approach	Objective
Structure	Northern blot	Analysis of RNA length
	5' and 3'RACE	Determination of RNA boundaries
	FANTOM CAGE data analysis	Investigation of RNA start
Expression	Real Time-PCR	Overall tissue specificity
	In-situ hybridisation	High resolution cardiac sub-structure analysis
Localisation	Cell-fractionation	Sub-cellular localisation
	RNA-Fluorescent *In Situ* Hybridisation (FISH)	Sub-nuclear localisation
Coding Potential	Polysome profiling	Translation potential
	Open reading frame analysis	Coding potential prediction
Conservation	Real Time-PCR	Expression in human samples
	In silico structural analysis	RNA structure conservation
Relevance in disease	GWAS association	Identification of genetic risk factors
Function	Knockdown in cardiac cells	Insights into biological function *in vitro*
	Chromatin Isolation by RNA Purfication (ChIRP)	Identification DNA and protein interacting partners
	CRISPR mouse model	Investigation of *in vivo* function

Thereby, this work provides a detailed analysis of two novel transcripts and has the potential to increase our understanding of *Nkx2-5* genetic regulation as well as deepen our knowledge of lncRNA function and regulatory mechanisms in the heart.

2 Results

2.1 Structure

2.1.1 Mouse *NkxUS* and *NkxDS* are long, polyadenylated transcripts

Elucidating the structure, including length and RNA boundaries, of the cardiac enriched transcripts surrounding *Nkx2-5* was the first aim of this work and was performed by Dr. Nicole Schonrock (former Victor Chang Cardiac Research Institute, Australia).

To identify the transcript length, northern blotting was performed on 5µg and 10µg of adult mouse heart RNA with probes for *Nkx2-5, NkxDS* and *NkxUS* (Figure 8A). The blot results showed that both are long transcripts, *NkxDS* is ~11kb in length and *NkxUS* ~10kb in length. Interestingly, a ~11kb transcript was detected with both, *Nkx2-5* mRNA and *NkxDS*-specific probes, indicating that *NkxDS* possibly contains the *Nkx2-5* probe sequence and that probably the important cardiac transcription factor *Nkx2-5*, may have an uncharacterised long isoform (Figure 8A). To analyse if *NkxDS* and *Nkx2-5* share the same transcription start site, we mapped the 5'end of *NkxDS* using FANTOM Cap Analysis of Gene Expression (CAGE) [173, 174]. FANTOM CAGE is a catalogue of 5'ends of capped transcripts and available for investigation on the University of California, Santa Cruz (UCSC) genome browser (Figure 8C) [175]. The FANTOM CAGE data showed one prominent peak at the annotated *Nkx2-5* start, suggesting that *Nkx2-5* and *NkxDS* share a similar transcription start site (Figure 8C, blue CAGE counts). *NkxDS* could thus have been generated via read-through of the *Nkx2-5* primary transcript. We performed several Real-Time (RT)-PCRs on adult mouse heart tissue with different set of primers, six forward and three reverse primers, across the recognised *Nkx2-5* mRNA polyadenyation site, including the 5'untranslated region (UTR) and both exons (Figure 8B) of *Nkx2-5*. As a control for genomic DNA contamination, we performed each RT-PCR with two cDNA samples (with and without the reverse transcriptase enzyme) (Figure 8B). The RT-PCR results showed expression for all the tested primer pairs and only for those using the cDNA samples with the enzyme, indicating that *NkxDS* can be continuous with the *Nkx2-5* mRNA and can contain both exons and the 5'UTR of *Nkx2-5* (Figure 8B). The exact 3'end of *NkxDS* was mapped using 3'rapid amplification of cDNA ends (RACE) with a Poly(dT) primer, additionally also identifying a canonical AAUAAA polyadenylation signal, which is a signal for RNA polyadenylation as well as transcription termination (Figure 8C) [176]. This long, potentially non-coding, tail may contain miRNA or protein binding sites that could regulate *Nkx2-5* mRNA stability, sub-cellular localisation or protein translation.

NkxUS, on the other hand, is produced several kilobases upstream and on the opposite strand to *Nkx2-5*. To identify the start site of *NkxUS*, 5'RACE was performed on adult mouse heart RNA. The results of two clones revealed that *NkxUS* is comprised of two exons, one short one (157bp), which sits very close to the second longer exon (~10kb) and is also present in the ENCODE heart RNA sequencing data (Figure 8C). As for *NkxDS*, the exact 3'end was mapped using 3'RACE, thus also revealing a canonical polyadenylation site (AAUAAAA) (Figure 8C).

In summary, *NkxUS* is longer than 200nt, spliced and polyadenylated, fulfilling all features defining typical lncRNAs [99, 109]. With respect to the location to the nearest protein-coding gene, *Nkx2-5*, *NkxUS* would be defined as an intergenic RNA [177]. Moreover, we may have discovered a potential novel isoform of *Nkx2-5, NkxDS.*
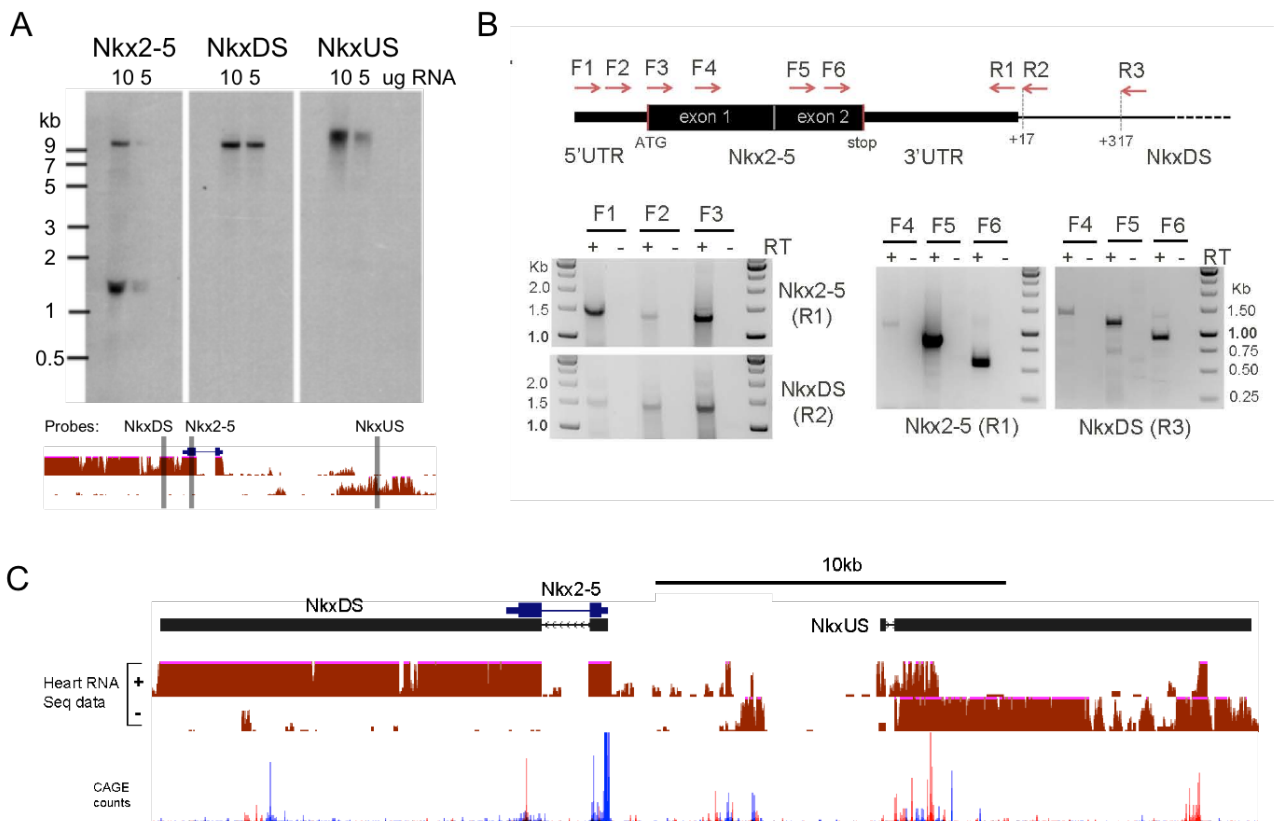
Figure 8 Mouse *NkxDS* and *NkxUS* are long transcripts. A) Northern blot using specific probes targeting *Nkx2-5, NkxDS* and *NkxUS*. For each probe 5µg and 10µg of adult mouse heart RNA was used. The locations of the probes are indicated in the lower image. B) RT-PCR on mouse heart cDNA with six different forward primers (F1-F6) and three different reverse primers (R1-3) across the proposed *Nkx2-5/NkxDS* boundary suggests that *NkxDS* is continuous with *Nkx2-5* mRNA and that *NkxDS* can contain both exons and the 5'UTR of *Nkx2-5*. The primer locations are shown in the upper image. The gel showed bands for all RT-PCRs and only for the samples containing reverse transcriptase (+). C) Map of the genomic location of *Nkx2-5* and the surrounding transcripts, *NkxUS* and *NkxDS*, showing ENCODE mouse heart RNA sequencing data from the plus (+) and minus (-)

strand and FANTOM CAGE counts, with blue indicating counts on the minus strand and red on the plus strand (adapted from UCSC browser [178]).

2.2 Expression

2.2.1 Mouse *NkxUS* and *NkxDS* are highly enriched in cardiac tissues

Previously characterised lncRNAs were found to function in a tissue-, cell-type or developmental specific manner [98, 117, 119]. Therefore, as a first step towards exploring the potential function of *NkxDS* and *NkxUS*, we analysed their expression levels during development and in different mouse tissues.

RT-PCR analysis was performed on 12 adult mouse tissues as well as E9.5 heart tissue and HL-1 cells, a mouse atrial cardiomyocyte cell line, with primers for *Nkx2-5, NkxDS* and *NkxUS* (Figure 9A) [179]. Primers for *Nkx2-5* measure both *Nkx2-5* as well as *NkxDS* expression, since the transcripts overlap we were unable to design *Nkx2-5* specific primers (Figure 9A). The RT-PCR results showed that all transcripts are highly enriched in adult as well as embryonic cardiac tissues (Figure 9A). As previously shown, *Nkx2-5* is mainly expressed in the heart, both embryonic and adult, but additionally showed some low-level expression in spleen and tongue [68]. *NkxDS* mimicked *Nkx2-5* mRNA expression, whereas *NkxUS* was exclusively expressed in cardiac tissue in adult mice (Figure 9A). Relative to *Nkx2-5, NkxUS* and *NkxDS* were more lowly expressed in HL-1 cells as well as adult mouse heart, which is a common feature of lncRNAs (Figure 9B and C) [97].

In situ hybridisation performed by Dr. Gonzalo del Monte Nieto (Victor Chang Cardiac Research Institute, Australia) on embryonic and adult mouse tissues, using the same probes as used for northern blotting, confirmed the RT-PCR results showing that both transcripts are cardiac-enriched (Figure 9D-H). Here showing data of sections from E8.5 (Figure 9D), E17.0 (Figure 9E-G) and adult (Figure 9H). However, the entire experiment was performed on sections from embryos at E8.5, E9.5, E10.5, E17.0, postnatal day 10 and adult and the complete Figure can be found in Appendix I (Figure 49). The results showed that *NkxUS* and *NkxDS* are expressed in the heart throughout development (Figure 9D-H). *NkxDS* and *Nkx2-5* share a similar expression pattern, whilst the expression pattern of *NkxUS* is markedly different. From E8.5 to adulthood, *NkxDS* and *Nkx2-5* are cardiac expressed and predominantly localised to the myocardium (Figure 10D, F and H, arrows). In contrast, *NkxUS* is mainly expressed in the endocardium and to a lesser extend in the

myocardium (Figure 10D, F and H, arrowhead). Furthermore, *NkxUS* was not confined to the heart during early development and was also expressed in other tissues such as the lungs and neural tube (Figure 10G), but became strictly cardiac after birth (Figure 9A). Interestingly, *NkxUS* is strongly expressed in the endocardium surrounding the valve leaflets at E17.0 (Figure 9F, arrow) and in the sinoatrial node (Figure 9G, inset), whereas no expression was observed for *NkxDS* and *Nkx2-5* in these tissues (Figure 9F, arrowheads and 9G). However, after birth and in adult mice all three transcripts are expressed in the valve endocardium (Appendix Figure 49E[3] and F[3]). Upon close inspection, *NkxUS* and *NkxDS* showed a more speckled, potentially nuclear staining, whilst *Nkx2-5* displayed a more diffuse, potentially cytoplasmic staining (Appendix Figure 49D[4]).

In summary, *NkxUS* and *NkxDS* are expressed in the heart throughout development, are cardiac enriched in adult mouse tissues and predominately expressed in distinct layers of the heart wall, suggesting that both transcripts do not function in a developmentally specific but perhaps in a cell- or tissue specific manner.
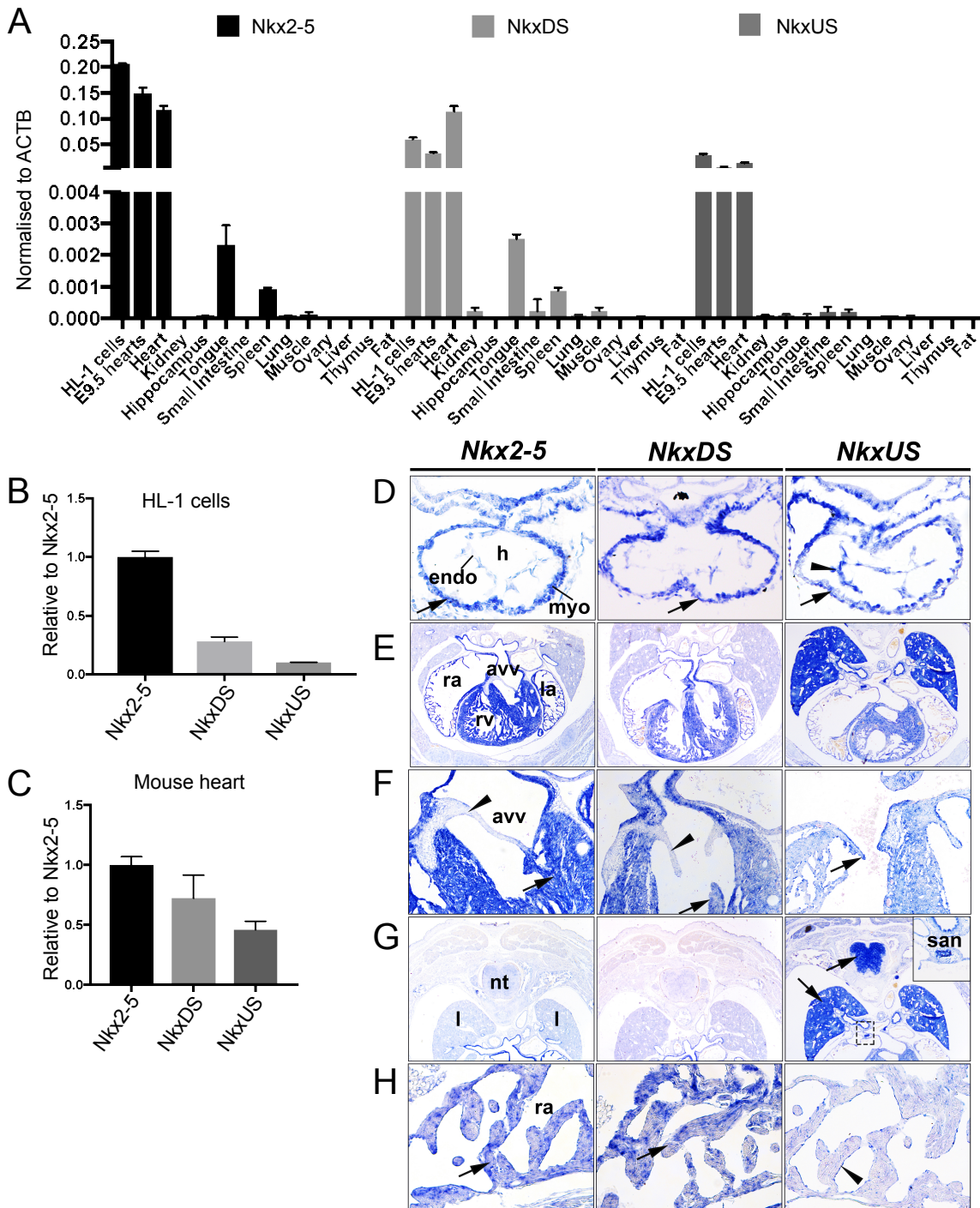
Figure 9 Expression in mouse tissues. A) RT-PCR on different adult mouse tissues as well as E9.5 mouse heart and the cardiac cell line HL-1 showed that both transcripts are highly enriched in cardiac tissues. *Nkx2-5* and *NkxDS* share a similar expression pattern with low expression in some non-cardiac tissues (spleen and tongue). *NkxUS* is strictly cardiac enriched in adult mouse tissues. Data presented as mean ± SD, n=3 B-C) RT-PCR on HL-1 cells (B) and adult mouse heart (C) showed that *NkxUS* and *NkxDS* are relative to *Nkx2-5* much more lowly expressed. Data presented as mean ± SD, n=3. D-H) *Nkx2-5* (left panels), *NkxDS* (middle panels) and *NkxUS* (right panels) expression pattern by *in situ* hybridisation on sections from embryos at E8.5 (D), E17.0 (E-G) and adult (H) showing that both transcripts are expressed in the heart throughout development. Heart (h), endocardium (endo), myocardium (myo), left ventricle (lv), right atrium (ra), left atrium (la), right

ventricle (rv), left ventricle (lv), atrioventricular valve (avv), neural tube (nt), lungs (l), sinoatrial node (san).

2.2.2 *NkxUS* and *NkxDS* expression in differentiating embryoid bodies is similar to *Nkx2-5*

In addition to their development, tissue- and cell-type specificity, lncRNAs have also been shown to be dynamically regulated during differentiation [180]. In order to examine this, we analysed expression levels of *NkxUS* and *NkxDS* in a differentiation time course of mouse embryonic stem cell (mESC) into embryoid bodies (EBs). mESCs are immortal cells, derived from the inner cell mass of a blastocyst during gastrulation and are able to differentiate into ecto-, meso- and endoderm *in vitro* [181-183]. Using conventional EB differentiation protocols, only a minor proportion of mESCs spontaneously differentiate into cardiomyocytes. However, the addition of ascorbic acid was shown to enhance cardiomyocyte differentiation [184]. mESCs were differentiated into EBs enriched for cardiomyocytes by Hananeh Fonoudi following a published method [185]. RNA was extracted at day 0, 4, 7, 10, 14, 17, 21, 24 and 30 and expression levels of *Nkx2-5, NkxDS* and *NkxUS* were analysed via RT-PCR using the same primers as in section 2.2.1. Expression was normalised and analysed relative to day 7. As previously observed, *Nkx2-5* expression levels followed a very dynamic and stage-specific expression pattern with low expression at day 0 (undifferentiated mESCs) and day 4 (differentiating mESCs), highest expression at day 7 and then a drop in expression, which remained low until the end of the time course (Figure 10A) [186]. *NkxUS* and *NkxDS* shared a very similar dynamic expression pattern to *Nkx2-5* (Figure 10A). This was consistent with previous cardiomyocyte differentiation studies, which identified that lncRNAs are significantly correlated in expression with their neighboring genes compared to randomly selected ones [186]. The observed dynamic and stage-specific expression of *Nkx2-5* highlights how it might activate specific gene programs or pathways in a time dependent manner to induce cardiomyocyte differentiation. The fact that all three transcripts are expressed at the same stage suggests that they may act in similar processes during cardiac development. Expression analysis of *NkxUS* and *NkxDS* at day 7 relative to *Nkx2-5* showed, as before for whole heart and HL-1 cells (Figure 9 B and C), a much lower expression level for both transcripts (Figure 10B).
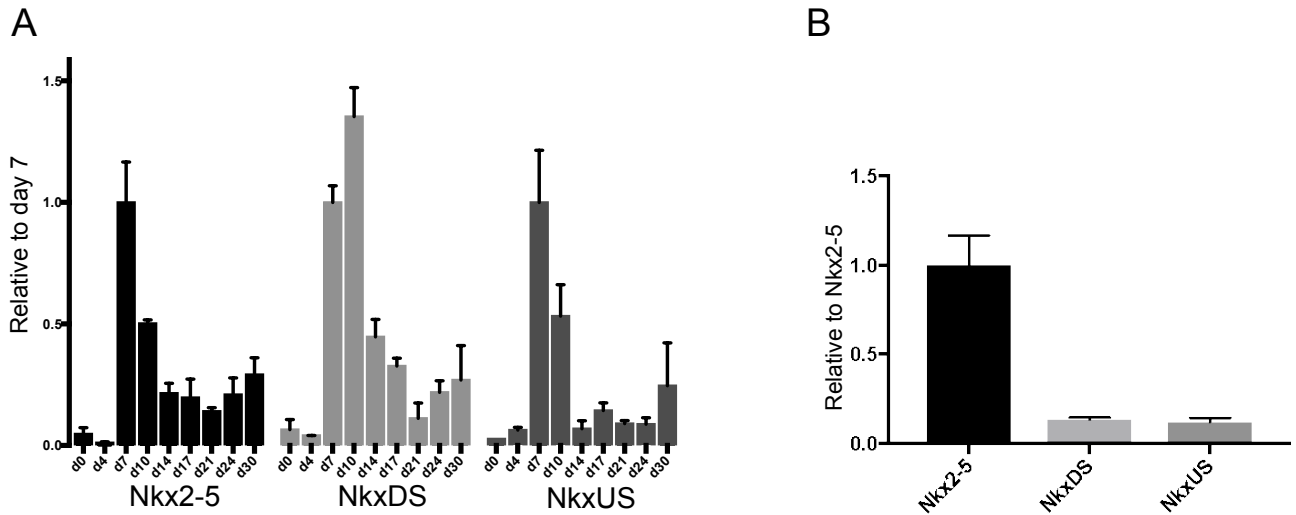
Figure 10 *NkxUS* and *NkxDS* follow the expression pattern of *Nkx2-5* in a time course of mouse embryonic stem cells into embryoid bodies enriched for cardiomyocytes. A) RT-PCR on different timepoints of the time course measuring *Nkx2-5, NkxDS* and *NkxUS* expression relative to day (d) 7 showed that all three transcripts are expressed at the same timepoints. B) Expression at day 7 of *NkxUS* and *NkxDS* relative to *Nkx2-5* showed that *NkxDS* and *NkxUS* are around 80% lower expressed compared to *Nkx2-5*. Data presented as mean ± SD, n=3.

2.2.3 *NkxUS* and *NkxDS* are expressed in purified cardiomyocytes at different postnatal stages

To test for cell-type specific expression in mature cells, we interrogated RNA sequencing data from mouse cardiomyocytes at different postnatal stages, which were freshly isolated and purified by Prof. Robert Graham's group (unpublished).

Cardiomyocytes were harvested at four different time points, postnatal (P) days 2, 10, 13 and 73. Unlike in HL-1 cells, whole heart and the mESC timecourse (Figure 9B, C and 10B), *Nkx2-5* was only marginally more highly expressed than *NkxDS,* but *NkxDS* and *Nkx2-5* mRNA were more highly expressed than *NkxUS* (Figure 11). This was perhaps not surprising since *in situ* hybridisation showed that *NkxUS* is predominantly expressed in the endocardium (Figure 9D). Interestingly, the expression dynamics of both transcripts is similar to *Nkx2-5*, with a decrease in expression with ongoing postnatal development, which was already previously described for *Nkx2-5* [187].

Analysis of expression levels in cells and tissues showed that both transcripts are cardiac enriched, expressed in the heart throughout cardiac development and that *NkxUS* is to some extend also expressed in adult mouse cardiomyocytes. Both RNAs showed a similar dynamic

developmental expression pattern to *Nkx2-5*, indicating that all three transcripts may function at similar timepoints and potentially in related cardiac pathways.
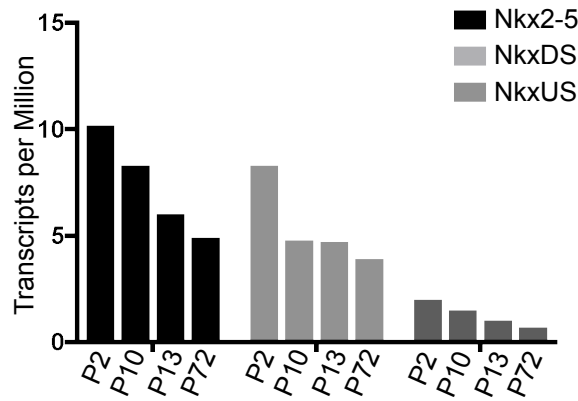
Figure 11 *NkxUS* and *NkxDS* follow the expression pattern of *Nkx2-5* in purified cardiomyocytes. RNA sequencing data of purified mouse cardiomyocytes harvested at different postnatal (P) time points obtained from the Robert Graham group (unpublished) revealed that both transcripts exhibit lower expression with ongoing development similar to *Nkx2-5*. *NkxDS* is almost as abundant as *Nkx2-5* at all the postnatal stages, whereas *NkxUS* is much more lowly expressed in purified cardiomyocytes at any postnatal stage. Results shown in transcripts per million; n=1.

2.3 Subcellular localisation

2.3.1 Both transcripts are retained in the nucleus

Identification of the subcellular localisation of lncRNAs can give significant insight into their function, as nuclear retained lncRNAs have very distinct roles from cytoplasmic enriched lncRNAs [112, 113].

Nuclear and cytoplasmic HL-1 cell fractionations followed by RT-PCRs were performed by Dr. Nicole Schonrock with *NEAT1* as a control for nuclear localisation and *GAPDH* as a cytoplasm-enriched control. The fractionations showed enrichment for *NkxDS* and *NkxUS* in the nucleus (Figure 12A), whereas *Nkx2-5* was slightly enriched in the cytoplasm (Figure 12A). Similar to *GAPDH, Nkx2-5* showed also some nuclear expression (Figure 12A), potentially due to *Nkx2-5* primers also measuring *NkxDS* expression, which was mostly nuclear (Figure 12A).

To validate the nuclear expression of *NkxDS* and *NkxUS* and gain further insights into their subnuclear localisation, we performed RNA-fluorescence *in situ* hybridisation (RNA-FISH) with

probes targeting *NkxDS, NkxUS* and *NEAT1*. Using RNA-FISH, individual RNA molecules can be simultaneously detected, systematically quantified and analysed for subcellular localisation patterns. A set of 48 oligonucleotide probes, each 20 bases long, with a Quasar 670 fluorescent label and a different sequence to cover the entire transcript were designed for *NkxUS* and *NkxDS*. The *Nkx2-5* protein-coding region was excluded for *NkxDS* probe design. For *NEAT1* a commercially available probe set (Biosearch Technologies Inc.) with a Quasar 570 fluorescent label was used. The probes from one set collectively hybridise to the same target RNA, which results in a concentration of fluorophores at a single location producing a distinct signal, which is detectable using fluorescent microscopy [113].

For the experiment, fixed HL-1 and C2C12 cells were used. C2C12, which is an immortalised mouse myoblast cell line, was used as a non-cardiac cell line and served as a negative specificity control since *NkxUS* and *NkxDS* are not expressed in these cells [188].

RNA-FISH results showed that *NkxUS* and *NkxDS* are exclusively expressed in the nucleus of cardiac cells (Figure 12B), which were stained with DAPI. Whilst transcripts appeared to occur throughout the nucleus, some cells showed one to three brighter foci (Figure 12B), which could represent transcription start sites and hence accumulation of nascent transcript. Furthermore, both transcripts seem to be excluded from the nucleolus, visible as DAPI-negative "holes" (Figure 12B). Since *NEAT1* localises to paraspeckles and is expressed in most cell types, it localised as bright discrete foci in cardiac and non-cardiac cells (Figure 12B) [114, 189].

Taken together, the RNA-FISH results confirmed the results from the cell fractionation and *in situ* hybridisation experiments showing that both transcripts are retained in the cell nucleus. Furthermore, neither transcript seemed to localise exclusively to a sub-nuclear structure as signal was observed throughout the nucleus, suggesting that its biological function is within the nucleus, but most likely not associated with specific sub-nuclear organelles.

A

B
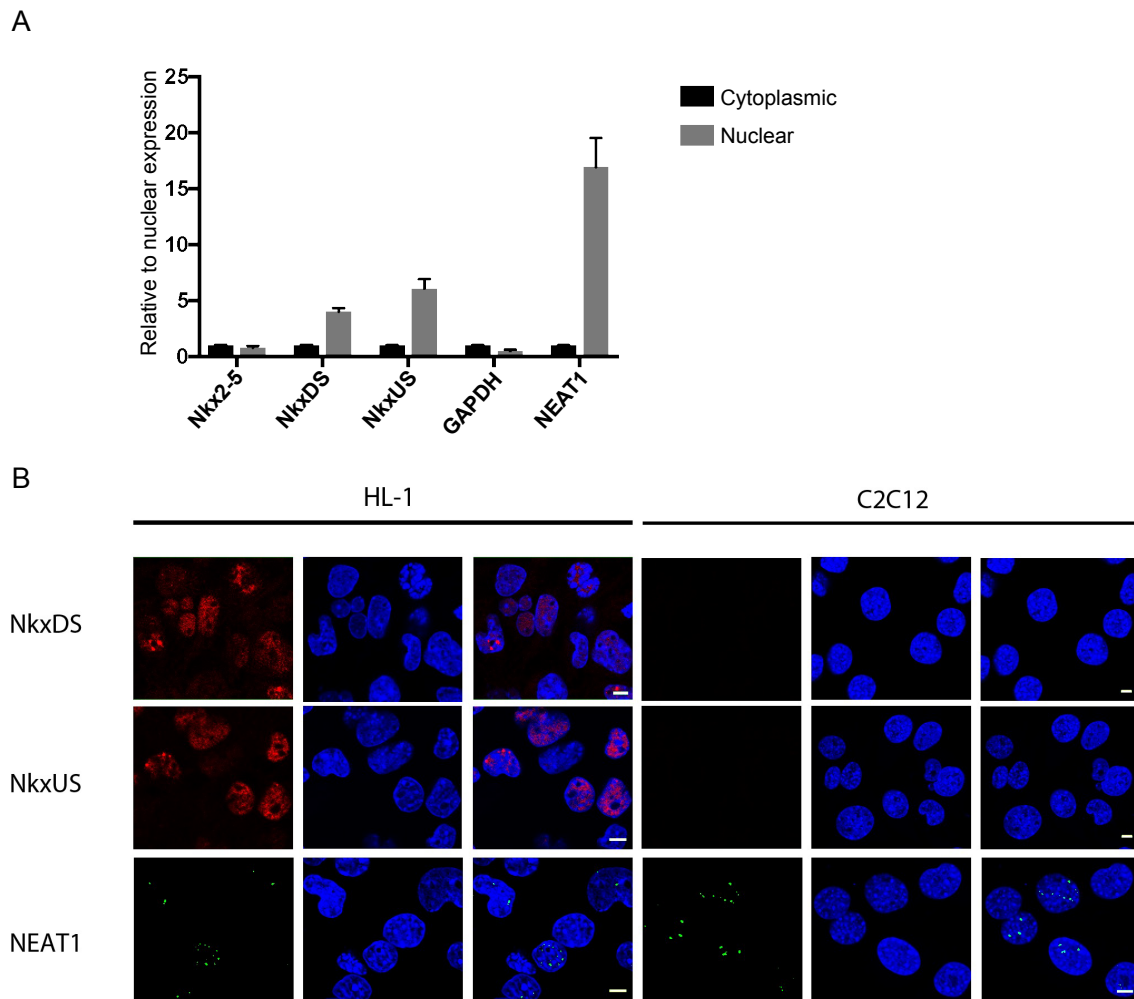
Figure 12 *NkxUS* and *NkxDS* localise to the nucleus in cardiac cells. A) Nuclear and cytoplasmic fractionation of HL-1 cells showing enrichment of *NkxDS* and *NkxUS* in the nucleus. Data presented as mean ± SD, n=3 B) RNA-FISH showed exclusive nuclear localisation with no specific sub-nuclear localisation for *NkxDS* or *NkxUS* in cardiac cells (HL-1), whereas no expression was observed in non-cardiac cells (C2C12). *NEAT1* showed sub-nuclear localisation in both cell types. DAPI is shown in blue. *NkxUS* and *NkxDS* specific probes are shown in red. *NEAT1* specific probes are shown in green. Scale bar = 5µm.

2.4 Coding potential of *NkxUS* and *NkxDS*

To clarify the coding potential of *NkxUS* and *NkxDS*, we used both experimental and computational approaches.

2.4.1 Polysome gradient profiling resulted in minimal association of *NkxUS* and *NkxDS* with polysomes

We performed an RNA polysome gradient profile in HL-1 cells, in collaboration with Prof. Thomas Preiss (Australian National University, Australia). A polysome gradient can give insight into the degree of association of an RNA with ribosomes as actively translated RNAs are generally bound by several ribosomes (polysomes) [190]. To obtain the polysome profile of a cell, they are treated with cycloheximide, which inhibits translation and immobilises the ribosome on the RNA, hence keeping them attached to the RNA. Afterwards, the cell lysate is layered on a sucrose gradient to separate RNAs according to the number of bound ribosomes. Well-translated RNAs have more ribosomes bound and therefore sediment in the denser (later) fractions of the gradient. However, there is a possibility that RNAs which sediment in the denser fractions are actually not bound to ribosomes but merely associate with other protein complexes. Therefore, in a second experiment, puromycin, which detaches ribosomes from RNAs, was added to the cells. Thus, if an RNA was truly bound to polysomes it should now sediment in the less dense fractions, while RNAs which were bound to other proteins should still sediment in the denser fractions.

In the experiment, *Nkx2-5* served as a control for a well- translated RNA. In normal ("cycloheximide-treated") HL-1 cells, virtually all *Nkx2-5* mRNA sedimented in the dense-polysome fraction (Fraction 9) as expected for a well-translated mRNA (Figure 13A). With puromycin treatment, *Nkx2-5* sedimented in Fractions 4-6, which suggests that *Nkx2-5* does indeed associate with polysomes and is a well-translated mRNA (Figure 13B).

Like *Nkx2-5* mRNA, *NkxDS* sedimented in the denser fractions (Figure 13A), although this persisted after puromycin treatment (Figure 13B), suggesting that *NkxDS* may associate with a non-polysome protein complex that also sediments in these denser fractions.

The majority of *NkxUS* transcripts sedimented in Fraction 2 in both normal- and puromycin-treated HL-1 cells (Figure 13A and B), suggesting that, consistent with their nuclear localisation, the majority of *NkxUS* RNAs are not translated. There did appear to be a minor peak in Fractions 8-9 in both conditions, which suggests that *NkxUS* may associate with another non-polysomal protein complex. There was also a small peak appearing in Fractions 4-6 in puromycin-treated HL-1, indicating that a minority fraction of *NkxUS* might associate with polysomes. However, the puromycin experiment was not performed in replicates and thus the results have to be interpreted with caution.

In summary, the polysome profiling revealed no or very minimal ribosome association of *NkxUS*. The minimal presence of *NkxUS* in denser polysome fractions may represent an association of *NkxUS* with other ribosomal associated RNAs, much the same way as miRNAs are also detected in late polysome fractions although they are not actually translated [191, 192]. Interestingly, *NkxDS* and *Nkx2-5* behaved similar in the cycloheximide but different in the puromycin experiment, which suggests that *NkxDS* potentially associates with a non-polysomal protein complex and not with polysomes even though it contains the *Nkx2-5* coding region.
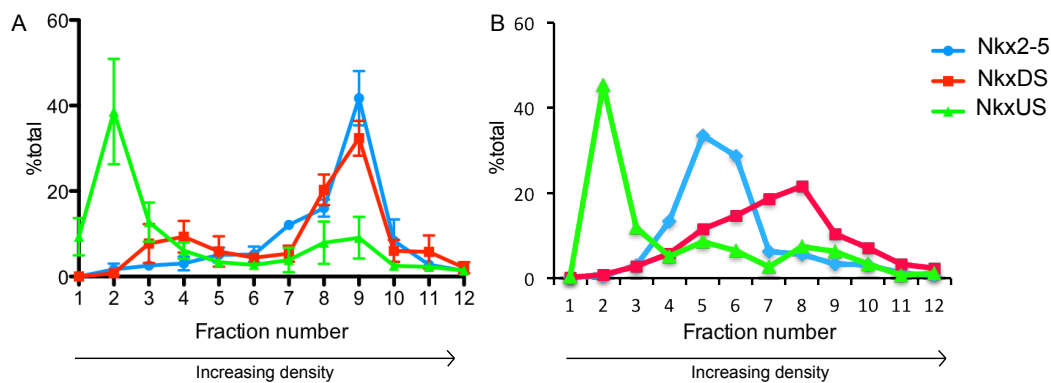
Figure 13 Polysome gradient profiling showed no or little association of *NkxUS* and *NkxDS* with ribosomes. A) Cycloheximide treated HL-1 cells, in which the RNA is kept in place with the polysomes, showed that *Nkx2-5* and *NkxDS* sediment in the later polysome fraction whereas *NkxUS* sediments in the early ones, indicating that *NkxUS* does not associate with polysomes but *NkxDS* might. *Nkx2-5* was as expected detected in the late fractions as it is a well translated RNA and thus associates with polysomes. Data presented as mean ± SD, n=3 B) Puromycin treated HL-1 cells, in which the polysomes are detached from the RNA, showed that *NkxDS* still sediments in the late fractions, indicating that it might associate with a non-polysomal protein complex rather than ribosomes. *Nkx2-5* sediments as expected in earlier fractions compared to A) as it was detached from ribosomes. *NkxUS* still sedimented in the early fractions indicating that it does not associate with polysomes. n=1.

2.4.2 Open reading frame analysis revealed no coding potential for the *NkxDS* tail and *NkxUS*

As a computational approach, we determined the likelihood of open reading frames (ORFs) contained in both transcripts encoding proteins using freely available tools. ORFs are defined as a stretch of codons containing no stop codon and thus have the ability to be translated. Identification of long ORFs may indicate protein-coding potential as long ORFs are thought to be unlikely to occur

by chance in lncRNAs. ORF analysis of mouse *NkxDS* and *NkxUS* was performed using the Coding-Potential Assessment Tool (CPAT), which evaluates characteristic features of coding RNA sequences, including maximum ORF size, codon and hexamer usage bias as well as dinucleotide usage [193]. Since the majority of lncRNAs lack sequence conservation, this method is based on an alignment-free approach. CPAT calculates four features:

- the maximum ORF length
- ORF coverage, which is the ratio of ORF to transcript length (longer RNAs are expected to have longer ORFs by chance)
- Fickett score, which measures the combinatorial effect of nucleotide composition and codon usage bias, hence to which degree each base is favored in one codon position compared to another [194]
- hexamer score, which determines the relative degree of hexamer usage bias in a particular sequence (positive values suggests protein-coding)

From these features a coding probability is calculated.

Analysis of *NkxDS*, excluding the *Nkx2-5* mRNA sequence, revealed a maximum ORF length of 306 nucleotides, a Fickett score of 0.6 and negative hexamer score, resulting in low coding probability (0.09), indicating that, according to CPAT, this ORF is not protein-coding. By comparison, when using the *Nkx2-5* mRNA sequence, we get a Fickett score of 1.2, a positive hexamer score (0.5) and high coding probability (0.99), as expected for a protein-coding sequence. Analysis of *NkxUS*, revealed a maximum ORF length of 294 nucleotides, a Fickett score of 0.8 and negative hexamer score (-0.25), resulting in low coding probability (0.05). Altogether, according to CPAT, the longest ORFs in *NkxUS* and *NkxDS* are predicted to be non-coding.

However, recent studies identified a novel class of short regulatory polypeptides (less than 100 amino acids in length), termed micropeptides [195-197]. Due to their translation from small ORFs, this emerging class has been missed in conventional genome annotation. There is increasing evidence that functionally important micropeptides can be hidden in transcripts annotated as lncRNAs [196, 198-200]. These micropeptides were found to be conserved over large evolutionary distances and can thereby be discovered using phylogenetic conservation analysis [195, 201, 202]. Based on the idea that evolutionary conservation indicates functionality, a study identified about 100 small ORFs within annotated ncRNAs or untranslated mRNA regions, which show strong conservation and are indicated to be translated when overlapping with ribosome profiling data

[202]. Once potential conserved small ORFs are detected, other techniques can help to provide evidence for the translation into micropeptides, such as ribosome footprinting and mass spectrometry [195]. Thus, many lncRNAs might have been mis-annotated and/or could possess both protein-coding and non-coding functions [97]. Using CPAT, several smaller ORFs were identified for *NkxDS* and *NkxUS*, which, according to CPAT, did not have any coding potential.

We further scanned *NkxDS* and *NkxUS* for potential ORFs of any length conserved across multiple species using PhyloCSF, a method which was previously used to identify the micropeptide DWORF [199, 203]. The phyloCSF measures codon substitution frequencies (CSF) and compares nucleotide sequence alignments across multiple species to analyse the likelihood of representing a protein-coding region. If the alignment is likely to be a conserved protein-coding region the phyloCSF score will be positive. PhyloCSF has been integrated into the UCSC browser and revealed negative, non-coding scores for mouse *NkxUS* and *NkxDS*, whereas clear positive, protein-coding scores were observed for the *Nkx2-5* (Figure 14). Regions where no PhyloCSF score was observed indicate that PhyloCSF did not have sufficient statistical power to calculate an accurate score.

Taken together, the polysome profiling and coding potential analysis data along with the exclusively nuclear localisation suggests that both transcripts are unlikely to be translated. At this point, no further experiments were undertaken to analyse whether any of the smaller ORFs in *NkxUS* or *NkxDS* encode a micropeptide.
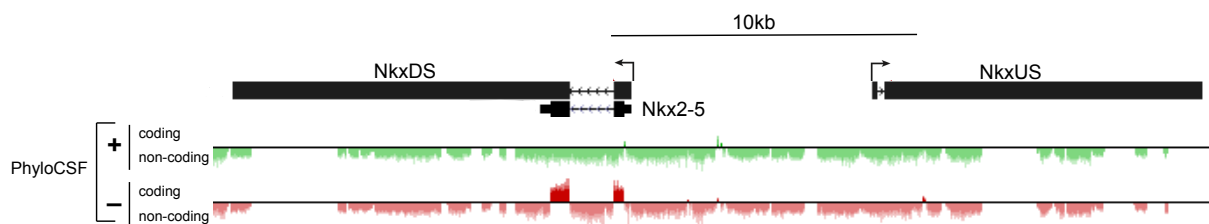
Figure 14 Coding potential of mouse *NkxUS* and *NkxDS*. PhyloCSF score plot for the region surrounding *Nkx2-5* as seen in the UCSC genome browser, depicting in green the plus (+) strand and in red the minus (-) strand. PhyloCSF showed a positive score for the *Nkx2-5* exons and negative scores for the exons of *NkxUS* and *NkxDS*. *NkxUS* and *NkxDS* also contained regions with no score, indicating that the PhyloCSF had not sufficient statistical power to calculate a meaningful score.

2.5   Conservation of *NkxUS* and *NkxDS* in the human genome

2.5.1   Identification of similar transcripts in humans

To determine if similar transcripts are present in humans, we investigated RNA-sequencing from the Roadmap Epigenomics Project and Genotype-Tissue Expression (GTEx) data on human heart [204, 205]. A wealth of unannotated transcripts surrounding *NKX2-5* was observed in these datasets (Figure 15A). Whilst a large number of reads map to the highly expressed *Nkx2-5* mRNA regions, lower abundant reads clearly demarcate transcripts both upstream and downstream of *NKX2-5* in human heart (Figure 15A). A 5.1kb transcript (BC033632) has been annotated downstream of *NKX2-5,* suggesting conservation of functionality with *NkxDS* [206]. Furthermore, a scan of the long intergenic ncRNA (lincRNA) Illumina human body map revealed a heart specific and spliced 4.4kb long transcript (TCONS_00010558) located 1.4kb upstream of *NKX2-5*, transcribed from the same strand (Figure 15A) [207].
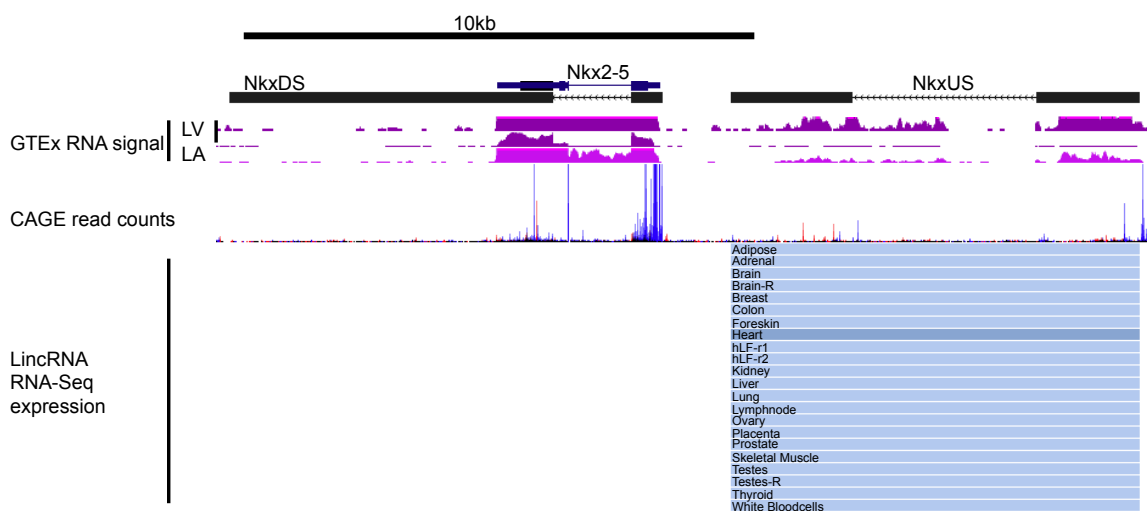
Figure 15 Similar transcripts are expressed in humans. UCSC browser view of the *NKX2-5* locus showing the upstream and downstream identified transcripts in black. Human left ventricle (LV) RNA sequencing data, showing two tracks, in which thresholds have been adjusted to visualise *NkxUS* in the upper track and *Nkx2-5* in the lower track, and left atrial (LA) RNA sequencing data from the GTEx project is shown in purple. CAGE read counts are shown underneath the GTEx data, with blue indicating CAGE tags for the (-) strand. Further, Illumina Body map RNA sequencing data is illustrated below the CAGE read counts track and identified a heart specific lincRNA upstream of *NKX2-5.*

2.5.2 Human and mouse *NkxDS* and *NkxDS* are not conserved on sequence level

Unlike *NKX2-5*, the sequence conservation of the *NKXDS* alternative spliced 3'UTR and *NKXUS* across mammalian species is very low, consistent with previous findings that thousands of lncRNAs have a faster evolutionary turnover and are therefore often species specific and less well conserved than protein-coding genes [208]. Pairwise sequence alignment of human and mouse *NkxUS* and *NkxDS* using the freely available bioinformatic tool EMBOSS identified a sequence identity of 30.3% and 42% for *NkxUS* and *NkxDS*, respectively [209]. Additionally, mouse *NkxUS* is expressed in the plus direction whereas human *NkxUS* is expressed in the minus direction and sits much closer to *NKX2*-5 (Figure 15A). *NkxDS* did not change position, however, seems to be much shorter in humans (Figure 15A).

2.5.3 *NkxUS* is conserved between mouse and human on structural level

In contrast to sequence conservation, many lncRNAs were shown to contain highly conserved secondary structures, which in turn might relate to their function, highlighting the importance of lncRNA structural analysis [143, 210]. Therefore, structural alignment analysis was performed by Dr. Stefan Seemann (University of Copenhagen, Denmark) to explore if common signals exists between the human and mouse transcripts. Foldalign was used to make structural alignments of the human and mouse RNA sequences, which were compared and scanned for common structural domains to determine conservation [211, 212].

Three highly conserved structural alignments with a p-value <0.01 were identified to be conserved between mouse and human *NkxUS*, shown in red in Figure 16A. The three conserved alignments in human *NkxUS* map to the same region in mouse *NkxUS*, which is the start of the second exon (Figure 16A and B). Listed in Table 3 are the positions of the alignments in human and mouse *NkxUS* as well as the corresponding p-values.

No significantly conserved structural domains were identified for *NkxDS*, when excluding the *Nkx2-5* mRNA sequence (Figure 16C).
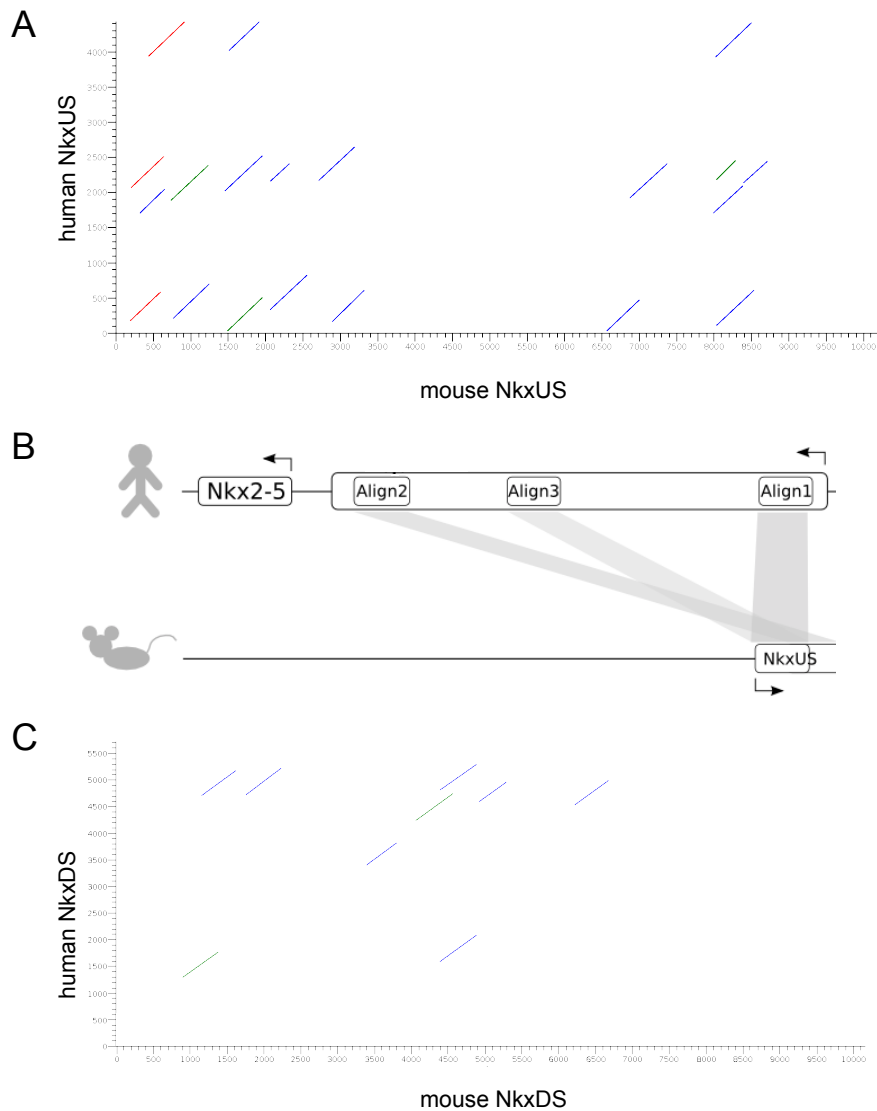
Figure 16 *NkxUS* **contains an RNA structure conserved between mice and humans.** A) Location of structural alignments for *NkxUS*, showing mouse *NkxUS* on the X-axis and human *NkxUS* on the Y-axis. The different color scheme indicates different p-values: red – p-value <0.01; green – p-value <0.1; blue p-value <1. B) Schematic showing human *NKXUS* in the upper and mouse *NkxUS* in the lower panel. Human *NKXUS* contains three conserved structural alignments (Align1-3), which all map to the start of the second exon of mouse *NkxUS*. C) No conserved structural alignments were identified for *NkxDS*.

Table 3 Structural alignment positions

	Nucleotide position		
Alignment number	**Mouse *NkxUS***	**Human *NKXUS***	**P-value**
1	190-597	178-588	0.001
2	433-915	3934-4417	0.004
3	201-640	2067-2506	0.009

2.5.4 *NkxUS* and *NkxDS* are expressed in the adult human heart

Using commercially available total RNA from human heart (Clonetech), we validated expression of human *NkxUS* and *NkxDS*. RT-PCR results showed that both transcripts are expressed, with *NkxUS* being more abundant than *NkxDS* (Figure 17). Relative to *Nkx2-5*, *NkxUS* and *NkxDS* are much more lowly expressed, as already seen from the GTEx RNA sequencing data (Figure 15) [205]. This is also similar to what we observed in the mouse data (Figure 9C). However, unlike in mouse whole heart (Figure 9C), *NkxUS* in human heart seems to be more highly expressed than *NkxDS* (Figure 17).
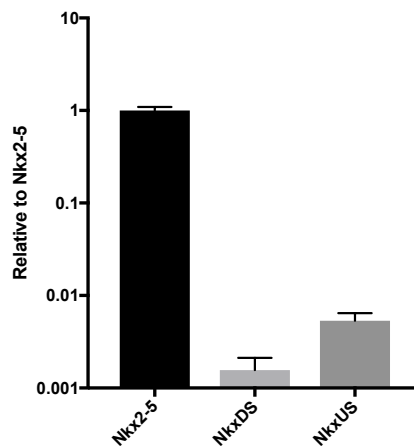
Figure 17 *NkxUS* and *NkxDS* are expressed in total RNA from human heart. RT-PCR results displayed in log10 scale showed that *NkxUS* and *NkxDS* are both expressed in the human heart, although much lower than *Nkx2-5*. *NkxUS* in human heart is more abundant than *NkxDS*. Data presented as mean ± SD from technical replicates, n=1.

2.5.5 *NkxUS* and *NkxDS* show similar expression dynamics to *Nkx2-5* in differentiating human cardiomyocytes

We further confirmed expression of both human transcripts in a differentiation time course of human induced pluripotent stem cells (iPSC) into cardiomyocytes. The differentiation was performed by Hananeh Fonoudi following a published protocol [213]. RNA was extracted every fifth day from day 0 to day 20 of differentiation and expression levels of *Nkx2-5, NkxDS* and *NkxUS* were analysed using RT-PCR.

All three transcripts were expressed in the time course (Figure 18). Similar to the mEB time course (Figure 10A), human *NkxUS* and *NkxDS* transcript levels follow *Nkx2-5* expression in the iPSC time course (Figure 18). Expression levels peak when the cells start beating at around day 10,

followed by a steep decrease from day 15, and levels continue to stay low until day 20 (Figure 18).
Since expression of *NkxUS, NkxDS* and *Nkx2-5* were observed at the same timepoints, it is
reasonable to propose that the transcripts could function at similar stages and perhaps perform
related functions.

In summary, it was confirmed that *NkxUS* and *NkxDS,* despite being only minimally
conserved at the sequence level, are expressed in human cardiac tissue and both transcripts
showed similar expression dynamics in both human and mouse pluripotent stem cell-derived
cardiac tissue. Furthermore, *NkxUS* contains an RNA structure conserved between mice and
humans.
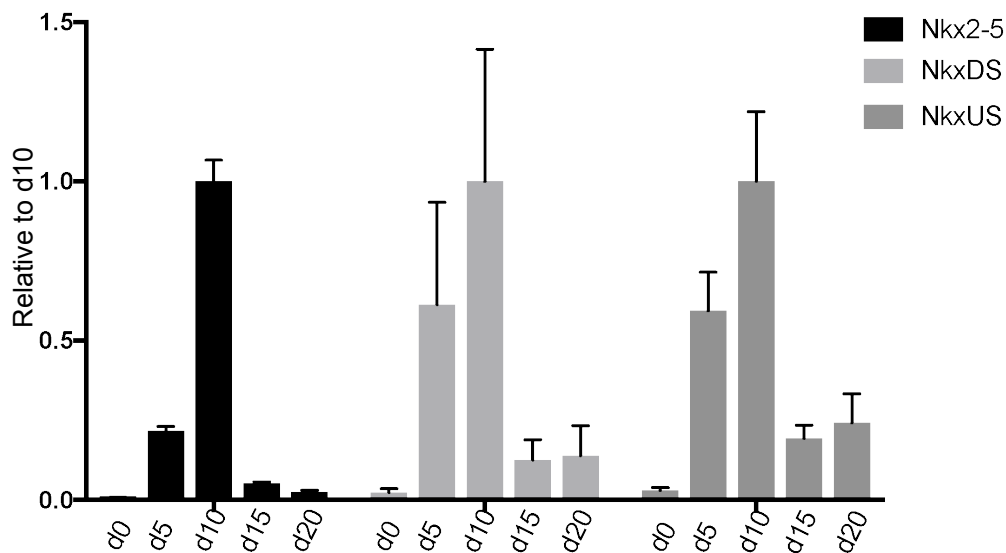
**Figure 18 *NkxUS* and *NkxDS* are expressed in human cardiomyocytes and follow *Nkx2-5*
expression in a differentiation time course of iPSCs into cardiomyocytes.** RT-PCR for *Nkx2-5,
NkxUS* and *NkxDS* using RNA taken from different days during the differentiation. The results are
presented relative to day 10. Data presented as mean ± SD, n=3.

2.5.6   Human *NkxUS* consists of two exons

As seen from the Illumina body map lincRNA data, human *NkxUS* seems to have two exons, similar
to mouse *NkxUS*. To confirm the exon-exon junction, we performed a PCR on human heart cDNA
using primers on either site of the exon junction (Figure 19). PCR products were run on a 1%
agarose gel, bands were excised from the gel and after DNA extraction cloned into pGEMTeasy.

DNA sequencing of several clones confirmed the junction with the exact sequence as annotated from the Illumina body map data (Figure 19).
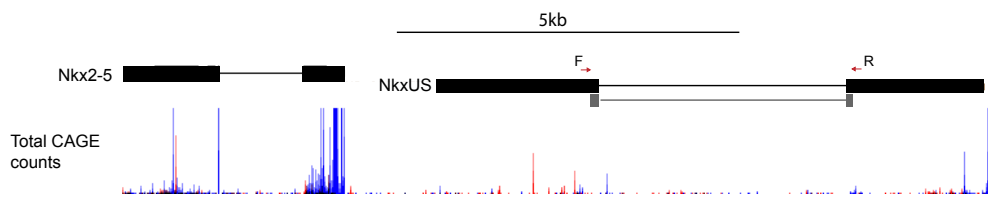
Figure 19. Human *NkxUS* consists of two exons. PCR across the proposed *NkxUS* junction showing in gray the results from the DNA-sequencing. Red arrows indicate positions for the forward (F) and reverse (R) primers. Prominent CAGE tags for the (+) strand (blue) line up with the approximate start site of *Nkx2-5* and *NkxUS*.

2.5.7 Evidence of human *NkxUS* in the literature

A recent study generated an atlas of around 28.000 human lncRNAs with their accurate start sites [214]. The study integrated CAGE data to obtain high-confidence 5'ends of thousands of lncRNAs across major human tissues and cell-types. Based on the accurately mapped 5'ends, the authors were able to analyse the regulatory regions and found that lncRNAs are more conserved than previously thought. The study identified that lncRNAs implicated in a specific GWAS traits are often expressed in the specific cell-type or tissue relevant to the trait and that these GWAS lncRNAs are often conserved. The authors generated an interactive browser, the FANTOM CAT browser (http://fantom.gsc.riken.jp/cat/v1/#/) allowing viewing of the genomic location of identified lncRNA genes. Additionally, the study integrated data of gene expression, conservational and genetic studies. In total, the results suggest that 19,175 lncRNA may be functional, almost as many as human protein coding genes.

Searching for *NkxUS* and *NkxDS* only revealed annotation for *NkxUS*. In the atlas, *NkxUS* is described as an intergenic lncRNA, termed CATG00000082047.1. The accurate 5'end of *NkxUS* as annotated by FANTOM CAT differs by 58bp from that found in the Illumina body map with FANTOM CAT mapping the start site of *NkxUS* to the exact position of the strongest and most proximal CAGE signal. The study also searched for dynamically expressed lncRNAs. 25 sets of FANTOM5 experiments, which were either differentiation time courses or paired control and treatment experiments, were analysed for differential expression [215, 216]. When a lncRNA was significantly differentially expressed (FDR < 0.05) in at least one comparison, it was defined as dynamically

regulated. *NkxUS* was found to be dynamically regulated during differentiation of human embryonic stem cells (ESC) into cardiomyocytes. Compared to day 0, *NkxUS* was significant upregulated from day 5 onwards (FDR= 9.68E-23), similar to what we observed in the iPSC time course (Figure 18). The study also analysed coding potential based on several resources (CPAT, RNACode, phyloCSF, RiboSeq) and found a maximal ORF length of 258, but altogether no coding potential.

Taken together, this study showed similar results to our study for *NkxUS*, in that it is dynamically expressed during cardiomyocyte differentiation. In addition, the study also predicted human *NkxUS* to be non-coding.

2.6 Relevance in disease

2.6.1 Human *NKXUS* contains a GWAS SNP associated with an increased heart rate

As mentioned in the introduction, many trait- and disease-associated SNPs were found to map to intergenic regions [161]. Interestingly, we identified a GWAS SNP (*rs6882776*), associated with an increased heart rate and increased risk of atrial fibrillation, that lies within one of the conserved RNA structures of human *NkxUS* (Figure 20) [217]. The SNP overlaps with alignment number 2 of the structural conservation analysis (Figure 16A, B and Table 3), which lies at the end of the second exon (Figure 20). An elevated resting heart rate is associated with an increased risk for cardiovascular diseases including heart failure, hypertension and arrhythmia as well as all-cause mortality [218-220].

As per the dbSNP database, *rs6882776* is a common SNP in the total population with 58.2% of individuals having a G allele that elevates resting heart rate, compared to individuals with an A at this position. In the GWAS study, ~95% of the participants were of European ancestry with *rs6882776* showing an effect allele frequency (EAF) of 68%, meaning that 68% of participants have the G allele (Table 4). The SNP was associated with an increase in resting heart rate by 0.301 beats per minute (bpm). Whist this overall effect might be small, the same study showed that the heart rate of individuals harboring a combination of heart rate increasing alleles increased by up to 4.1bpm, which was previously shown to be clinically relevant [217, 221].

The heart rate increasing allele at this locus was further associated with prolonged PR duration (+1.29ms) and reduced QT duration (-0.49ms), independent of heart rate, as well as a reduced QRS duration (-0.27ms) (Table 4) [217].

In addition, two independent GWAS studies on mostly individuals with European ancestry associated this SNP with an increased risk of atrial fibrillation with an odds ratio (OR) of 1.11 and 1.06 (P-value = 4.41×10^{-8} and 3×10^{-14}) (Table 4) [217, 222].
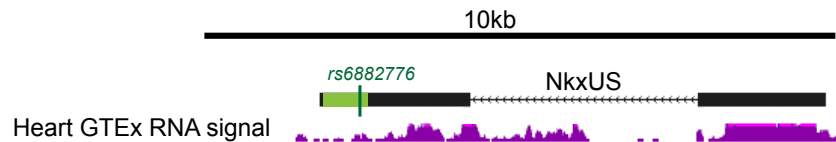
Figure 20 A GWAS SNP lies within the conserved structural alignment of human *NkxUS*. GTEx data of human LV RNA (purple) showing human *NkxUS* and illustrating the position of the GWAS SNP, *rs6882776*, in dark green and the conserved structure in light green.

Table 4 GWAS study details for *rs6882776*. EAF=effect allele frequency based on meta-analysis of stages 1 and 2 combined, β= effect size, SE= standard error, OR=odds ratio, CI=confidence interval.

	Alelles			Per-allele change in heart rate		Stage 1 and 2	
SNP	Effect	Other	EAF	β (bpm)	SE	# people	P-value
rs6882776	G	A	0.68	0.301	0.051	158.807	2.29×10^{-12}

QRS duration				QT duration			
# people	β (ms)	SE (ms)	P-value	# people	β (ms)	SE (ms)	P-value
30.877	-0.27	0.09	1.87×10^{-2}	60.768	-0.49	0.12	2.27×10^{-5}

PR duration				Atrial fibrillation		
# people	β (ms)	SE (ms)	P-value	OR (95% CI)	P-value	
18.484	1.29	0.27	1.46×10^{-6}	1.110 (1.07– 1.15) 1.06 (1.05-1.08)	4.41×10^{-6} 3×10^{-14}	

2.6.2 The GWAS SNP is predicted to totally disrupt the conserved secondary structure of *NkxUS*

To examined if the SNP has an impact the conserved RNA structure, an *in silico* analysis using RNAsnp was performed by Dr. Stefan Seemann (University of Copenhagen, Denmark) [223]. RNAsnp predicts local RNA secondary structural changes due to SNPs based on base pairing probabilities [223]. The results generated by RNAsnp can be visualised in a dot plot, in which a dot indicates base pairing at this position [223]. The larger the dot, the higher the base pairing probability [223].

Analysis of the second exon of human *NkxUS*, which is 2393bp in length, with and without the SNP using RNAsnp predicted a local structure change in the region 1904 to 1964, highlighted in gray background (Figure 21A). The predicted secondary structural change maps to the region surrounding the SNP, which lies at position 1934 (Figure 21B) and thus also affects the conserved RNA structure, which spans the region from 1902 to 2385. Looking at the dot plot of the magnified view of the highlighted gray region of Figure 21A, showed a clear difference in base pairing positions for wildtype and mutant *NkxUS* (Figure 21B). Regions with high base pairing probability for wildtype *NkxUS* showed in some regions none at all for mutant *NkxUS,* indicating a total disruption of the conserved RNA structure. This disruption becomes obvious when displaying the minimum free energy structures in a planar graphic for wildtype (left) and mutant *NkxUS* (right) (Figure 21C). The colored region in Figure 14C indicates the local structure change and the position of the SNP is highlighted with arrows at position 101.

Taken together, the disruptive effect of the GWAS SNP on the predicted conserved RNA structure of *NkxUS* may potentially associate this structure to the SNP associated phenotype.
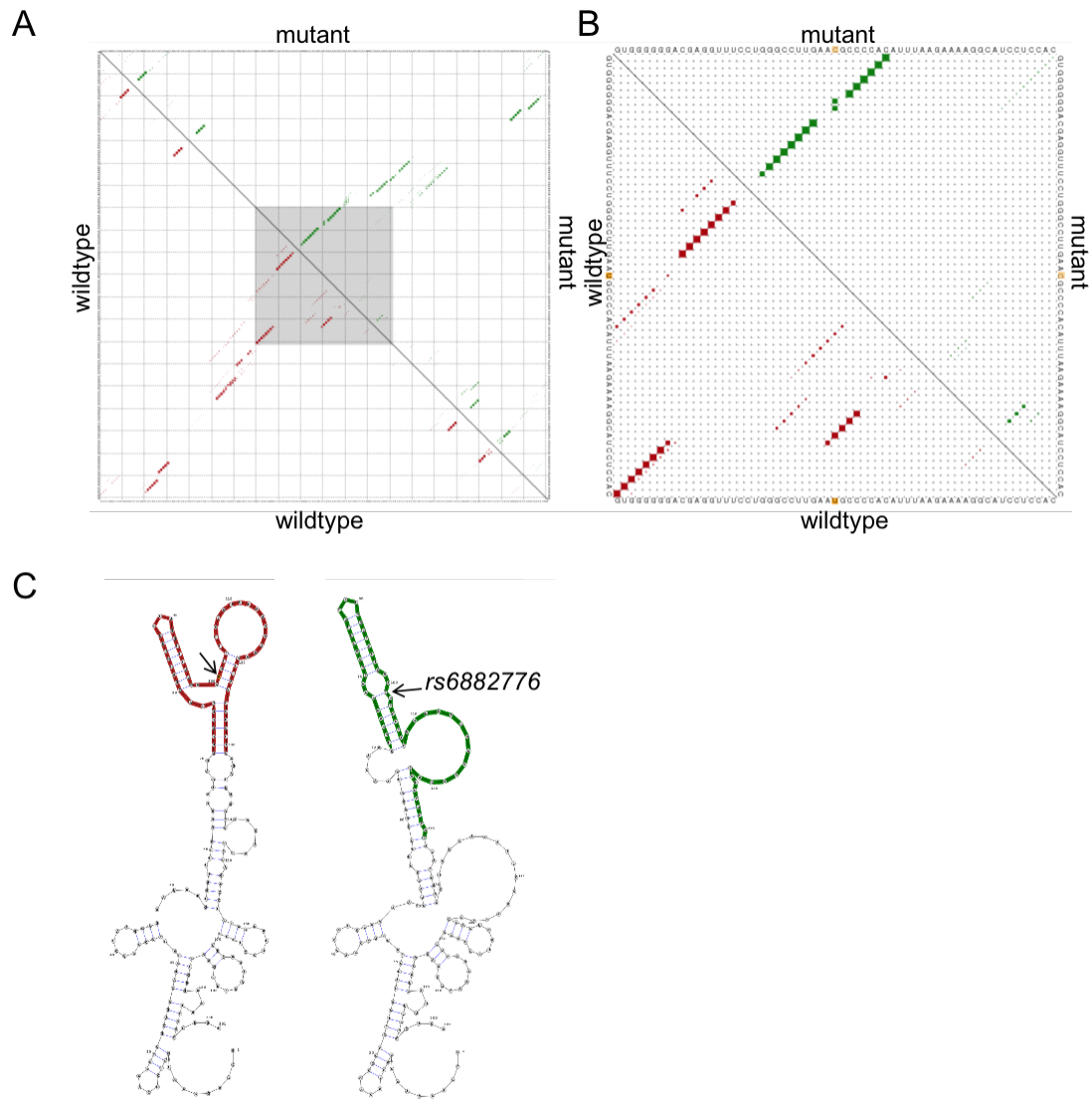
Figure 21 The SNP totally disrupts the conserved RNA structure in human *NkxUS*. A) RNAsnp analysis comparing the secondary structure of the second exon of human wildtype (red) and mutant (*rs6882776*, green) *NkxUS* revealed a structural change in the region 1904 to 1964, highlighted in dark gray. B) Magnified view of the highlighted area in A) showing the local structural change around the SNP position (orange). C) Displayed is the secondary structure for the second exon of human *NkxUS* in a planar graphic with wildtype on the left and mutant on the right. The colored region highlights the local structural change around the SNP and the arrows indicate the position of the SNP.

2.6.3 Haplotype analysis of *rs6882776* in Europeans identified five linked SNPs

GWAS studies are usually run on commercial SNParrays, which use only one SNP per haplotype (region in linkage disequilibrium (LD)), termed lead SNP. Therefore, they may not identify the causal variant, as any SNP in high LD with the lead SNP could be the disease-causing SNP, and additional

studies are generally required to identify the causal variant. Since the *NKXUS* SNP was identified in a study of individuals with mostly European ancestry, we performed a haplotype analysis in Europeans using LDlink (https://ldlink.nci.nih.gov/) [224]. SNPs with a squared coefficient of correlation (r^2) threshold of 0.8 qualify as linked, meaning that during meiotic recombination, these SNPs are almost never separated and most likely inherited together. Five SNPs, all within a ~16kb radius of *NKXUS*, were identified to be in high LD with *rs6882776* and therefore linked (Figure 22A and B). Three of the five SNP's lie within *NkxUS*'s first exon, one lies in an intergenic region and one in the *NKX2-5* coding region (Figure 22A). Looking at heart RNA Sequencing data as well as histone marks, no expression or methylation marks were found at the position of the intergenic SNP, suggesting that the chance of this being the causative SNP is very low. The *NKX2-5* SNP, *rs2277923,* sits at amino acid position 21 of NKX2-5, the last amino acid of the tinman domain, which is one of the six domains conserved between members of the NK2 family of proteins [225]. However, the SNP does not cause a change in the amino acid (GAA -> GAG), thus is a synonymous SNP. Further, the codon usage changed only slightly from a value of 0.4 to 0.58, suggesting that this SNP most likely does not affect NKX2-5 function. However, previous studies showed synonymous mutations can still affect protein function by affecting miRNA or transcription factor binding sites, changing the secondary structure of mRNAs or disrupting splicing, which we did not further follow up on [226]. The SNPs within the first exon of *NkxUS* do not overlap any of the structural conserved alignments thus questioning their relevance.

Interestingly, another recent GWAS study of > 1.000.000 Europeans associated one of these linked SNPs (*rs6891790-G*) with atrial fibrillation (P-value = 1 x 10^{-14}, OR= 1.07) [227]; thus, in total, three independent GWAS studies associated *NkxUS* with atrial fibrillation.

In Europeans, there are three possible haplotypes an individual can carry in the LD region of the *NKXUS* SNP with the most common having a frequency of about 70% (Figure 22C), which closely matches the EAF of the initial GWAS study.

In summary, the likelihood of *rs6882776* being the causal variant is quite high since we identified that it lies within and is predicted to totally disrupt the conserved RNA structure of *NkxUS*, whereas the other linked SNPs are not within conserved alignments, are synonymous mutations or lie in regions with seemingly no cardiac expression.
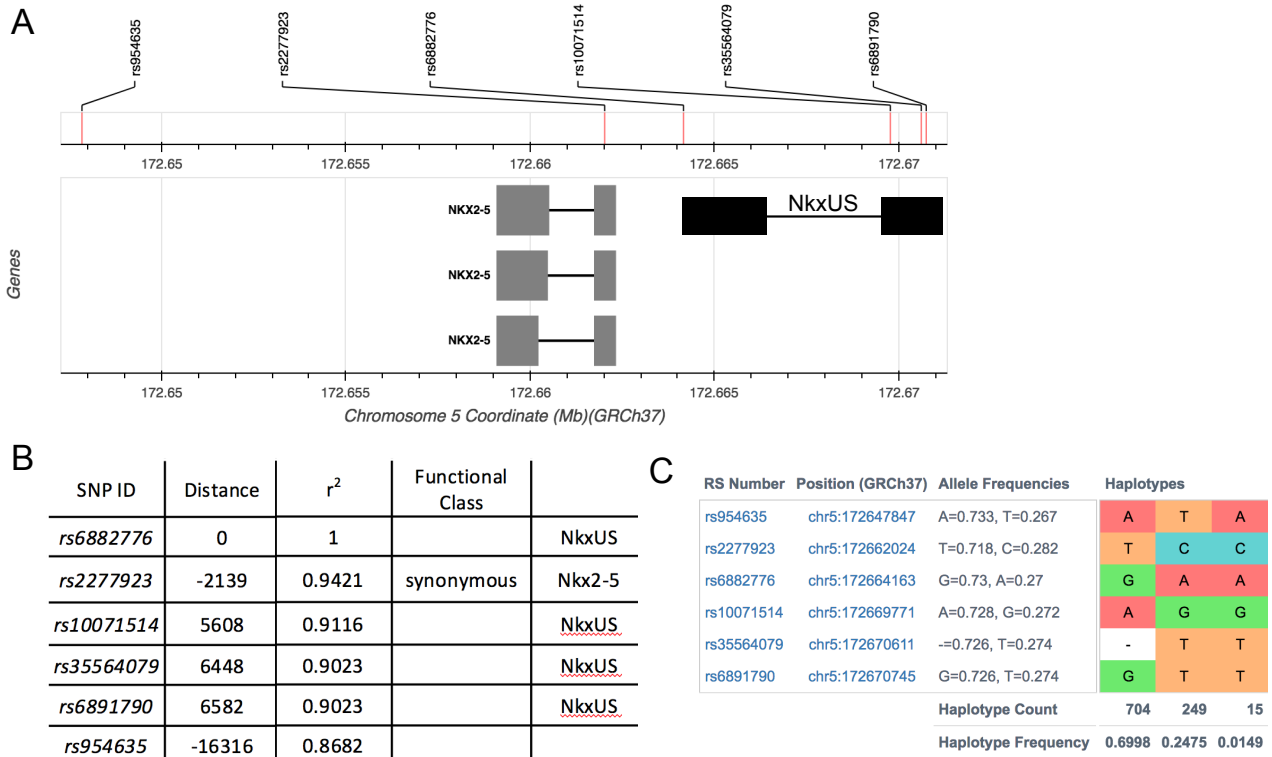
Figure 22 Haplotype analysis of *rs6882776* in Europeans. A) Genomic location of the SNPs in LD with *rs6882776*. B) Table of linked SNPs showing their distance in bp to *rs6882776*, the r^2 value, functional class and gene location. C) European specific haplotype frequencies of all haplotypes observed for the linked SNPs.

2.7 Functional analysis of *NkxUS in vitro*

Based on the association with the GWAS SNP, as well as the conserved structure, we henceforth concentrated on analysing *NkxUS*. To identify its function, a series of functional assessments were performed to determine the effect of partial loss of *NkxUS* at the level of the transcriptome. Further to this, *NkxUS* DNA-binding targets were investigated, as well as protein interacting partners. All three experiments were performed using HL-1 cells.

2.7.1 Knockdown of *NkxUS* in HL-1 cells results in mostly upregulated genes

To generate a general hypothesis about *NkxUS* function, we conducted a knockdown experiment in HL-1 cells followed by RNA sequencing to analyse changes upon downregulation of *NkxUS*. We targeted the transcript using locked nucleotide acid (LNA) GapmeR antisense oligonucleotides

(ASO) (Qiagen). LNA GapmeRs are single stranded ASOs consisting of LNA oligonucleotides in the flanking region and DNA oligonucleotides, free from LNAs, in the middle [228]. LNA oligonucleotides are "locked" by a methylene bridge in the ideal conformation for binding resulting in a higher binding affinity, increased target specificity and increased stability of the formed RNA:DNA complex [228]. Upon binding to the target, the LNA-free DNA center activates RNAseH, which in turn cleaves the RNA resulting in degradation of the transcript [228]. LNA GapmeR ASOs were designed using Exiqon's GapmeR Design Algorithm to identify the most potent and target specific ASOs with minimal off-target effects. Initially, three different ASOs (ASO1-3) targeting *NkxUS* at different positions and one control ASO (Control, Qiagen), which is a commercially available and validated negative control, were purchased in order to perform the experiment once with the three different ASOs resembling biological replicates. However, ASO3 was excluded during the optimisation process due to insufficient *NkxUS* knockdown. To increase specificity of the experiment, we chose to knockdown *NkxUS* with the two remaining ASOs (ASO1 and ASO2) (Figure 23A), which then allowed us to analyse for targets which are deregulated with both ASOs to diminish potential off-target effects specific to one set of ASO.

HL-1 cells were transfected with either ASO1, ASO2 or Control. One day after transfection, the cells were harvested and tested for knockdown efficiency via RT-PCR. The most successful knockdown achieved was a ~80% reduced *NkxUS* expression with ASO1 and ASO2 when analysed relative to cells grown in transfection medium (Lipofectamine) only (Figure 23A). Therefore, biological replicates showing each ~70-80% knockdown of *NkxUS* were sequenced using the Illumina HiSeq 2500 platform and analysed for differentially expressed genes using DESeq2 (Figure 23A) [229]. In order to identify dysregulated genes in common between ASO1 and ASO2, we tested for differentially expressed genes by evaluating ASO1 and ASO2 combined against Control.

Clustering of the three conditions and replicates was visualised on a principal component analysis (PCA) plot (Figure 23B). The technical replicates clustered together and ASO and Control samples were separated on the first component axis (Figure 23B), indicating that the highest variance is observed between the two conditions (ASOs versus Control) rather than between the biological replicates. ASO1 and ASO2 are separated on the second component axis, indicating variance, potentially from off-target effects of the individual ASOs (Figure 23B). However, since we analyse the combined effect of both ASOs against Control, noise from off-targets should be filtered out in the differentially expressed gene list, since only targets showing a similar fold change direction (up- or downregulated) in both ASOs are expected to reach significance.

Differential gene expression analysis using DESeq2 with a false discovery rate (FDR) <0.05 cut-off resulted in 1226 upregulated and 570 downregulated genes in ASO1 and ASO2 samples when computed against Control (Figure 23C). A heatmap further illustrates that comparatively more genes are upregulated for both ASO1 and ASO2 (red) (Figure 23D). In the heatmap, genes are grouped together based on their expression pattern. A dendrogram is added to the left side and the top according to cluster analysis, highlighting that both ASOs behave similarly and are distinct from the Control (Figure 23D). Furthermore, several clusters are observed in the heatmap, as seen from the dendrogram on the left side, suggesting that several different pathways or processes are affected by the knockdown of *NkxUS* (Figure 23D).

The gene list generated using DESeq2 with a FDR <0.05 and log2-fold change (logFC) cut-off of ±0.5 included 538 genes of which 432 are upregulated and 106 are downregulated. This list is given in Appendix II and was used for further analysis. The relative changes in transcript expression levels were relatively low, ranging from a logFC of -1.974 for *NkxUS* to a logFC of +2.452 for *Rnf31* (Ring Finger Protein 31).
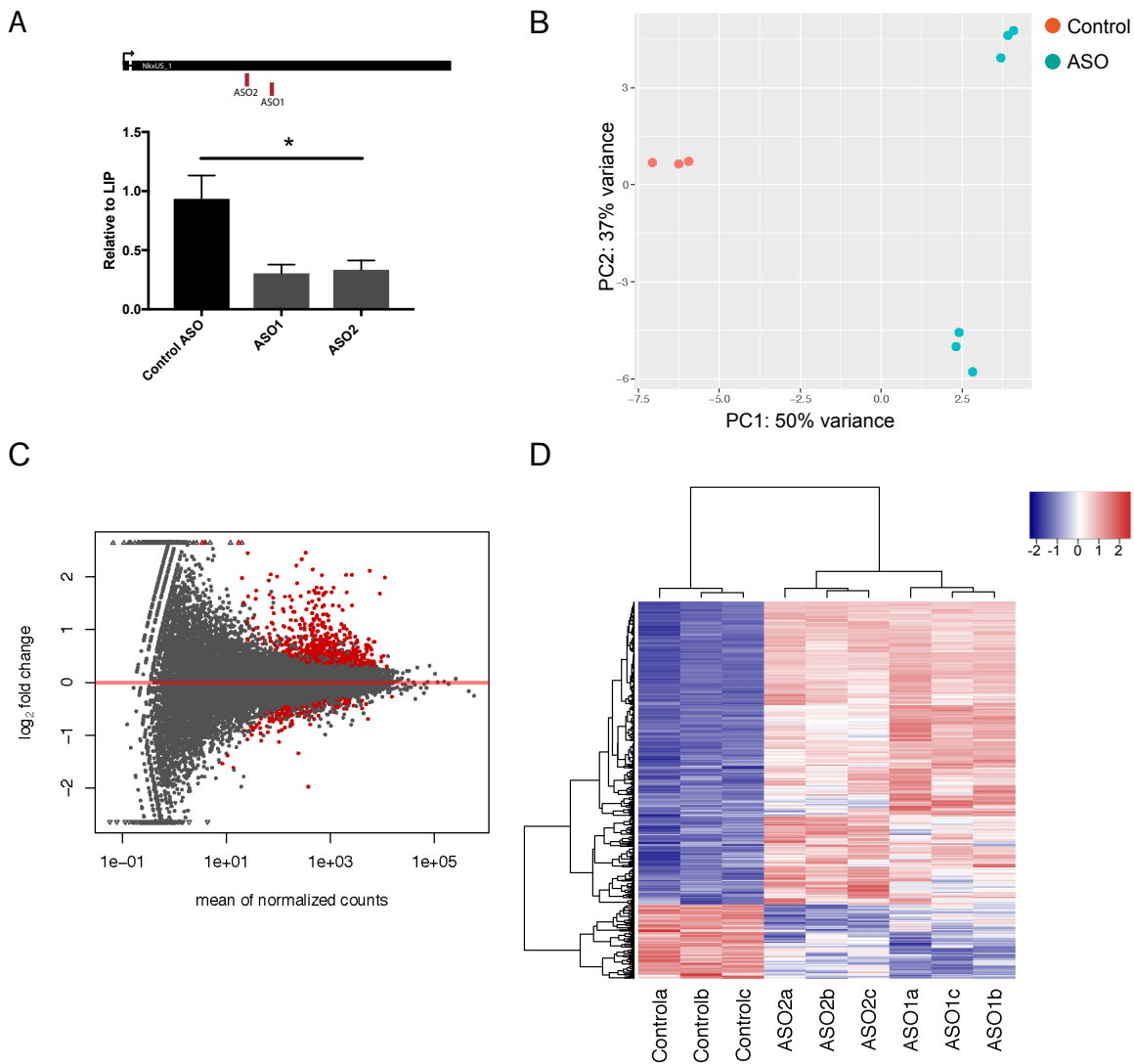
Figure 23 Knockdown of *NkxUS* followed by RNA sequencing results in mostly upregulated genes.
A) ASO1 and ASO2 target the second exon of *NkxUS*. RT-PCR showing ~80% knockdown of *NkxUS* using ASO1 and ASO2 relative to cells with Lipofectamine only. ASO Control did not have an effect on *NkxUS* expression. Data presented as mean ± SD, n=3, *P<0.05, unpaired two-tailed t-test. B) PCA plot showing that ASO1 and ASO2 separate from ASO Control on the first principal component axis. C) Differential expression analysis resulted in mostly upregulated transcripts. Here showing a scatter plot of log2 fold changes (on the y-axis) versus the mean of normalised counts (on the x-axis). Significant differentially expressed transcripts are indicated with red dots. D) Heatmap showing the global differences between ASOs and Control identified using DeSeq2. The scaled expression of each transcript, denoted as the row Z-score, is plotted with red representing higher expression and blue representing lower expression. a, b and c denote biological replicates.

2.7.1.1 Transcripts involved in cardiac contraction are deregulated upon reduced expression of *NkxUS*

In order to interpret the gene expression data and determine functional association of the differentially expressed genes, a gene set enrichment analysis was performed using both network analysis and gene ontology (GO) tools [230]. This enrichment analysis is based on the function of annotated genes and is a useful tool to investigate if the differentially expressed genes are associated with a common biological process or pathway. In our analysis, a background list consisting of all the genes expressed in the Control dataset was used. Using the GO consortium tool to search for enriched biological processes (BP) did not reveal any significantly enriched terms for the whole dataset, nor when split into down- and upregulated genes [231]. Therefore, we searched for networks within the differentially expressed genes using protein-protein association connections obtained from the STRING data-base (STRINGdb) [232], after filtering for high confidence connections. We used the STRINGdb R package to retrieve the network connections, which were used to search for enriched GO terms [232]. Running STRINGdb with the 538 genes significantly differentially expressed (logFC > ±0.5, FDR < 0.05) between both ASOs and Control resulted in a total of 125 connected genes, of which 102 genes were upregulated and 23 downregulated. The complete list of all differentially expressed genes included in the STRINGdb network is given in Appendix III.

The upregulated STRING network genes were enriched for several GO BP terms involved in cardiac contraction and conduction (Figure 24A). The most significant terms (FDR <0.05) included *regulation of transport, signaling, cell communication* and *regulation of localisation*, followed by more specific terms like *action potential*, *regulation of vesicle-mediated transport* and *neuron differentiation* and, interestingly, *regulation of cardiac action potential* and *cardiac muscle contraction* (Figure 24A).

The downregulated STRING network genes were significantly (FDR < 0.05) enriched for terms involved in *cell communication* and *signaling*, as well as *regulation of developmental processes*, more specifically *blood vessel* and *glomerulus development* as well as *cardiac cell fate commitment* (Figure 24B). Genes included in the cardiac cell fate term are *Tbx3* (T-box transcription factor TBX3), which is involved in valve and sinoatrial node development, and *Wt1* (Wilms tumor protein homolog), which is involved in epicardial development [17, 233, 234].

Based on the GWAS SNP being linked to increased heart rate and higher risk of atrial fibrillation, we focused on the cardiac conduction and contraction related terms. Genes included in these terms were mostly ion channels including potassium (*Kcnj3, Kcnj5, Hcn4*) and calcium (*Cacna1c, Cacna1d, Ryr2*) channels as well as regulators of signal transduction (Table 5), all of which also contributed to the most significant term, *regulation of transport*.

Analysis for enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways did not result in any enriched terms for either down- or up-regulated STRING network gene list.

To validate the RNA sequencing results, we used ASO1, ASO2 and a new Control ASO (Control-2). Control-2 was used to rule out possible off-target effects caused by a single Control ASO. We performed independent knockdown experiments and analysed a subset of deregulated genes involved in heart development and cardiac conduction using RT-PCR. Analysis of *Nkx2-5, NkxDS, NkxUS, Cacna1c, Cacna1d, Hcn4, Ryr2, Pde3a, Pde5b, Kcnj3 and Notch1* resulted in an upregulation of all transcripts, verifying the transcriptome data (Figure 24C).
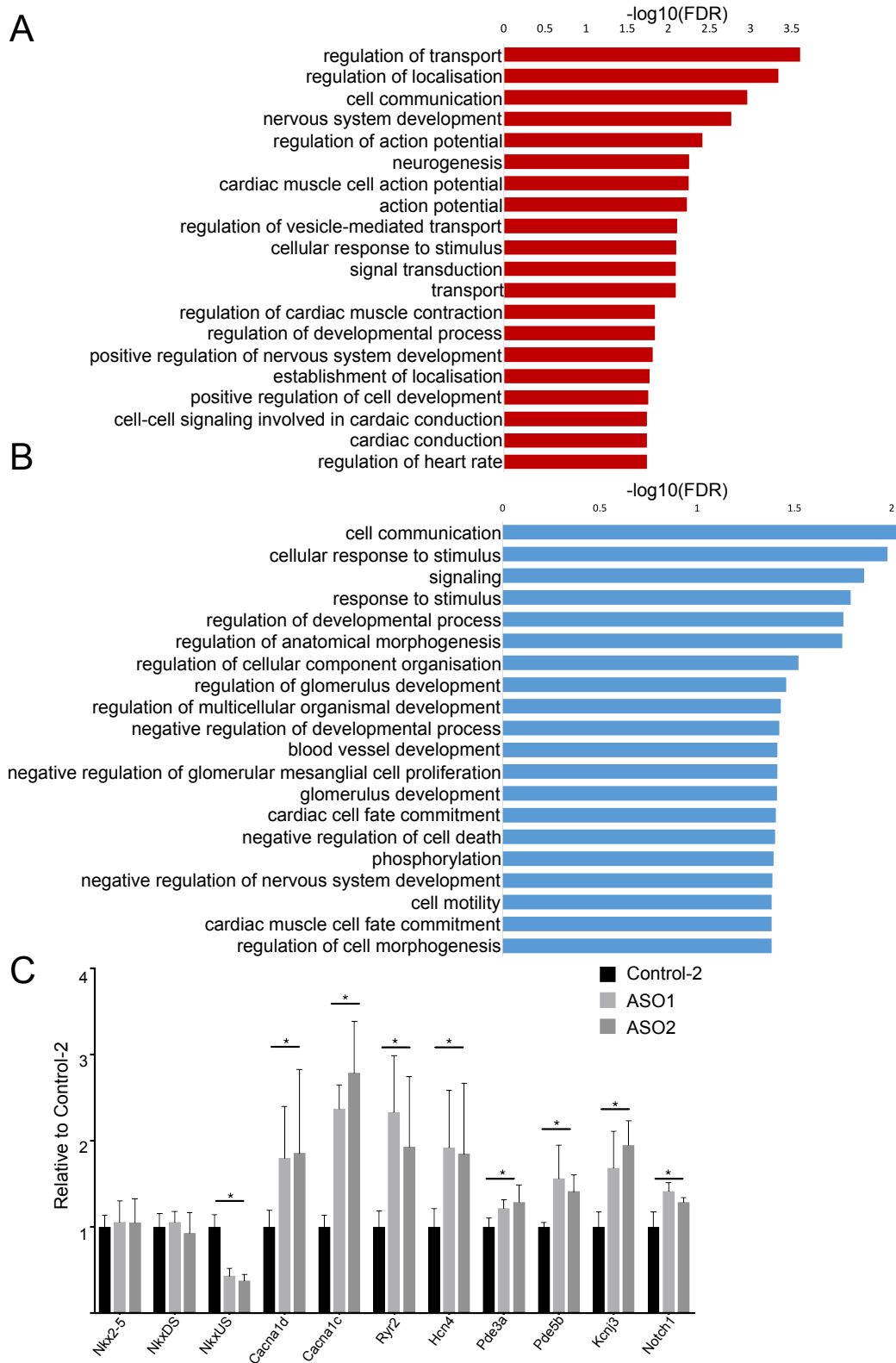
Figure 24 Knockdown of *NkxUS* results in upregulation of genes involved in cardiac contraction. A-B) Summary of the top 20 most significant GO terms for BP for the upregulated (A) and downregulated (B) STRINGdb identified genes showing associated FDRs in a −log10 scale. C) RT-PCR on selected differentially expressed transcripts with ASO1 and ASO2 and a new ASO Control (Control-2) confirmed the RNA sequencing results. Data presented as mean ± SD, *P<0.05, unpaired two-tailed t-test, n=4.

Table 5 Genes included in cardiac conduction and contraction terms (logFC > ±0.5, FDR < 0.05)

Symbol	logFC	FDR	Description
Pde5a	1.62	5.67E-19	Phosphodiesterase 5A
Kcnj3	1.17	5.14E-15	G protein-activated inward rectifier potassium channel
Dmd	0.54	4.55E-07	Dystrophin
Cacna1c	0.79	2.39E-06	Voltage-dependent L-type calcium channel subunit alpha-1C
Cav1	0.57	6.19E-06	Caveolin 1
Ryr2	0.76	5.55E-05	Ryanodine receptor 2
Cacna1d	0.59	1.07E-04	Voltage-dependent L-type calcium channel subunit alpha-1D
Kcnj5	0.31	1.11E-02	G protein-activated inward rectifier potassium channel 4
Hcn4	0.39	4.98E-02	Hyperpolarization Activated Cyclic Nucleotide Gated Potassium Channel 4

2.7.1.2 Discovery of a novel alternative first exon of mouse *NkxUS*

Looking at the RNA sequencing data with the Integrative genome browser (IGV), we observed a possible splice junction to a further upstream exon from the second exon of mouse *NkxUS* (Figure 25A). Subsequently, we performed a PCR on mouse heart cDNA across the junction using a forward primer located within the approximate location of the potential exon and the reverse primer within the start of the second exon (Figure 25B). The results revealed that there is indeed an alternative first exon spliced to the second exon of *NkxUS* (Figure 25B). Further, 5'RACE using mouse heart cDNA identified the same splice junction and the potential start of this exon, indicating that *NkxUS* has two alternative first exons (Figure 25C). Therefore, we termed *NkxUS* containing the earlier identified shorter exon *NkxUS-1* and *NkxUS* containing the further upstream first exon *NkxUS-2*. From the ENCODE heart RNA sequencing data, the first exon of *NkxUS-1* was observed to be more lowly expressed compared to the one of *NkxUS-2,* suggesting that *NkxUS-2* is the predominant isoform (Figure 25C).
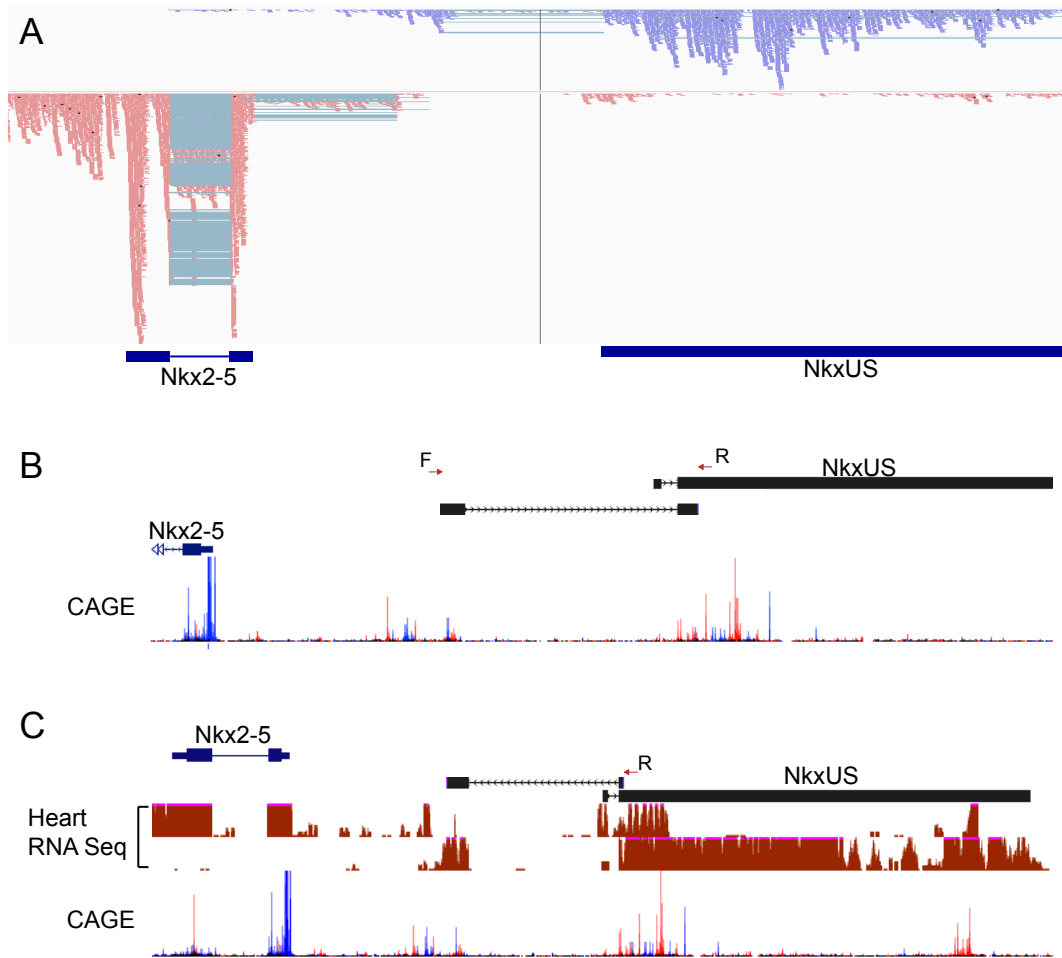
Figure 25 Identification of an additional further upstream first exon of mouse *NkxUS*. A) IGV image of the ASO Control RNA sequencing results showing the mapped reads to the *NkxUS* and *Nkx2-5* region showing in purple reads belonging to the plus strand and in red reads mapping to the minus strand. B) PCR using mouse heart cDNA across the potential junction with the reverse primer (R, red arrow) close to the start of the second exon and the forward primer (F, red arrow) ~4kb upstream showed that *NkxUS* has an additionally first exon which is spliced directly to the second exon; n=4.C) 5'RACE on mouse heart RNA with a gene specific reverse primer (R, red arrow) close to the start of the second exon also identified the further upstream first exon of *NkxUS.* Here showing the results of DNA sequencing of the 5'RACE product; n=4.

2.7.2 Identification of DNA and protein interacting partner of *NkxUS*

RNA can interact with DNA, RNA or proteins. To analyse if *NkxUS* binds to certain chromatin regions genome-wide or interacts with a specific set of proteins, which would in turn give additional insight into its related function, we performed chromatin isolation by RNA purification (ChIRP) [235]. A method established by the Howard Chang group utilising tiling antisense oligonucleotides (oligos) to capture the target RNA along with its bound DNA, RNA or protein partners [235, 236]. The

approach uses 48 biotinylated complementary oligos, which are approximately 20 nucleotides long and cover the entire RNA, excluding repetitive regions and regions which are complementary to other parts of the genome (Figure 26A). The tiling approach makes the method feasible for any RNA without prior knowledge of the secondary structure. To improve signal accuracy, the oligos were ranked according to their location along *NkxUS* and then divided into two pools. All the even numbered oligos were grouped together as the "EVEN" probe set and the odd numbered oligos as the "ODD" probe set (Figure 26A). As a control for unspecific binding to oligos and beads, oligos against LacZ mRNA, which is not expressed in mammalian cells, were used. The key to specificity is the split of the oligo probe sets. If a signal is *NkxUS* specific, it should appear in EVEN and ODD, but not in the LacZ samples.

The efficiency and specificity of *NkxUS* capture can be measured after pull-down by RT-PCR, where EVEN and ODD samples should be enriched specifically for *NkxUS* and no other transcripts. In Figure 26B, a gel of the RT-PCR products is illustrated showing in the ODD and EVEN samples enrichment for *NkxUS* and no enrichment for other highly expressed genes such as *beta-Actin* and *Nkx2-5*, suggesting that the probes are specific and captured *NkxUS* efficiently. In contrast, the negative controls, LacZ and NTC (non-template control), showed no strong expression for any of the analysed transcripts (Figure 26B). Additionally, the positive controls, input and mouse heart cDNA, showed high levels of expression of all tested transcripts (*NkxUS, beta-Actin, Nkx2-5*) (Figure 26B). Some minor unspecific bands were observed in the gel, which most likely are due to unspecific amplification seen after 40 cycles of RT-PCR.

In order to identify possible chromatin interactions, DNA sequencing is performed after successful enrichment of the RNA. RNA sequencing is performed to identify RNA targets and mass spectrometry to identify protein interacting partner. In our study, we were interested in discovering potential DNA-binding sites and protein interacting partners of *NkxUS*.
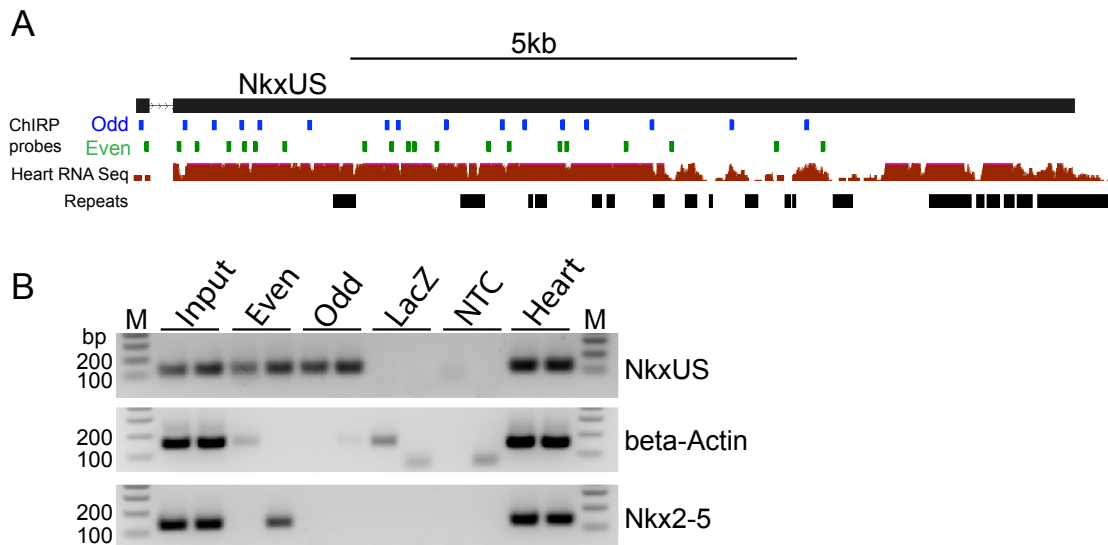
Figure 26 ChIRP successfully captures *NkxUS*. A) Location of the EVEN (green) and ODD (blue) probes tiling along *NkxUS* not overlapping repeat regions. B) Gel following RT-PCR after *NkxUS* pull-down of one replicate showing specific enrichment for *NkxUS* with EVEN and ODD probes but not *Nkx2-5* or *beta-Actin*. LacZ and NTC (non template control) do not show specific retrieval of any of the tested transcripts, whereas input and heart cDNA showed expression of all three transcripts.

2.7.3 *NkxUS* shows limited association with genomic DNA

Several nuclear lncRNAs were shown to associate with chromatin and influence chromatin state and thus transcriptional activity through association with chromatin modulating factors [112, 177, 237-239].

Initially, the DNA ChIRP was performed with five biological replicates, which all showed enrichment for *NkxUS* after capture. *NkxUS* ChIRP enriched DNA samples were sequenced on the Illumina Hi-Seq 2500 and peak enrichment analysis using the software MACS was performed by Thomas Kavanagh (Garvan medical institute, Australia) [240]. Three replicates had to be excluded either due to low DNA library yield or low read/peak counts after DNA sequencing. Altogether, we analysed two DNA ChIRP replicates (EVEN2/ODD2 and EVEN3/ODD3) which showed 10-20% yield for *NkxUS* following capture (Figure 27A). Yield is defined as the amount of *NkxUS* retrieved using ChIRP compared to an input sample, which was taken before probe hybridisation. The input sample was further used as a background to compute peak enrichment over background. When looking at the shared peaks between EVEN and ODD, excluding peaks detected by LacZ probes, of individual replicates, 59 and 23 high confidence DNA-binding targets for EVEN2/ODD2 and EVEN3/ODD3,

respectively, were identified (Figure 27B). Overlapping these two replicates, resulted in 11 shared peaks (Figure 27B). Nine out of those 11 peaks mapped the *NkxUS* locus (Figure 27C), one to *Ptprk* (protein tyrosine phosphatase, receptor type K) (Figure 27D) and one to *Grm8 (*glutamate metabotropic receptor 8) (Figure 27E).

Ptprk is involved in regulating cell adhesion as well as cell contact and is a negative regulator of the epidermal growth factor receptor (EGFR) signaling pathway. GO terms for biological process associated with Ptprk include *cell adhesion, cell migration, negative regulation of cell cycle* and *transcription, neuron projection development, protein dephosphorylation* and *cellular response to reactive oxygen species*.

Grm8 is a G protein-coupled receptor for glutamate and its signaling is known to inhibit adenylate cyclase activity. Associated GO terms for biological process include *regulation of neuron transmitter secretion, synaptic transmission (glutamatergic)* and *adenylate cyclase-inhibiting G protein coupled receptor signaling pathway.*

Therefore, using the EVEN and ODD strategy, we did not see compelling enrichment across the two biological replicates and probe sets and we further investigated the EVEN and ODD overlap of individual replicates.
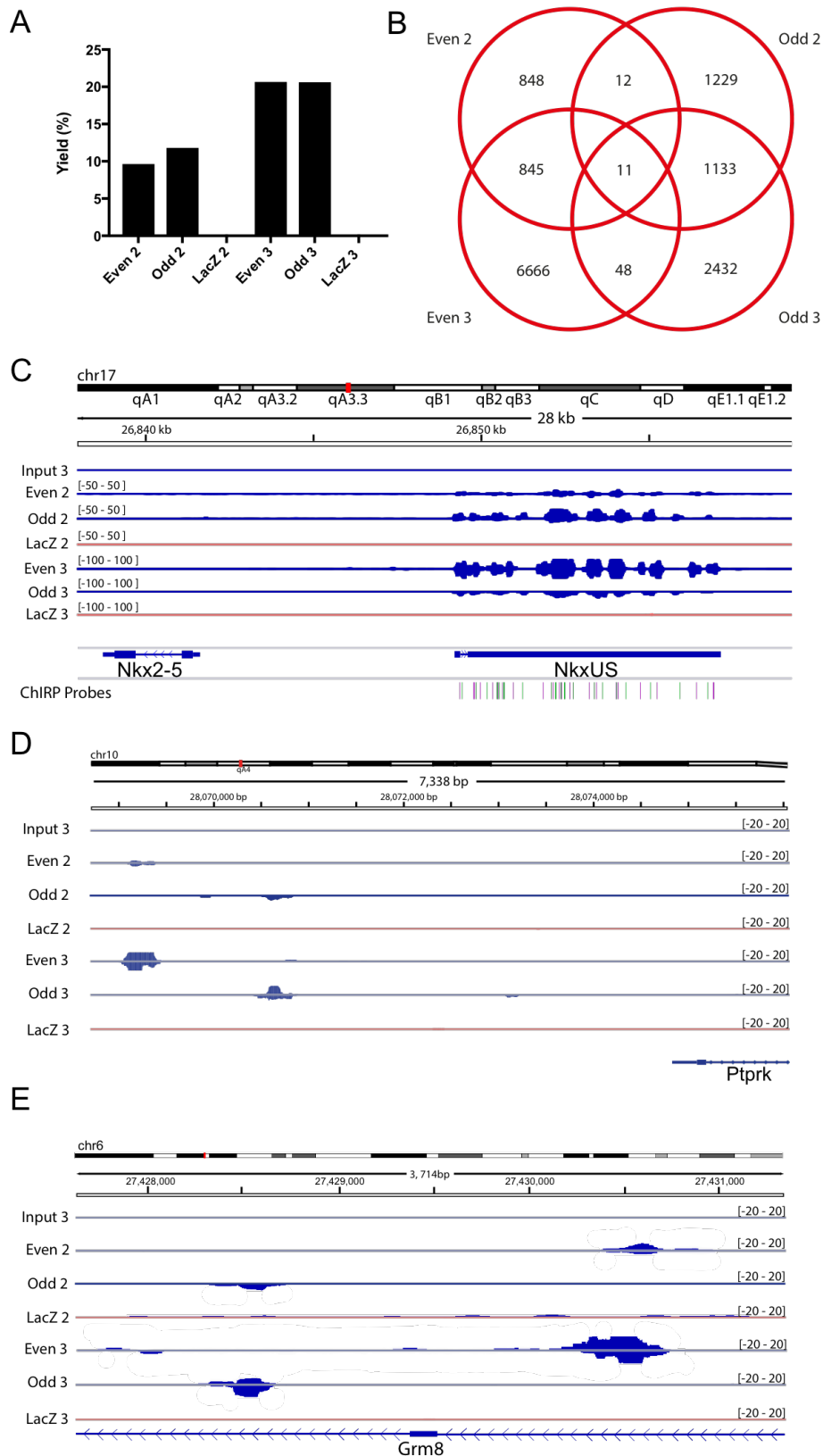
Figure 27 *NkxUS* pull-down followed by DNA sequencing identified two DNA binding regions. A) Results for *NkxUS* yield calculated from the RT-PCR results. B) Venn Diagram of overlapping peaks identified using the MACS package. C-E) Peak enrichment at the *NkxUS* locus (C), *Ptprk* locus (D) and *Grm8* locus (E).

2.7.3.1 Analysis of individual experiments

Due to a lack of enrichment of possible DNA targets when using the overlap of biological replicates, we also investigated the chromatin targets within the EVEN and ODD overlap of individual replicates. Looking at the overlap of experiment 2 revealed, among several predicted genes, one additional candidate, *Osr1* (Protein odd-skipped-related 1). *Osr1* is a transcription factor that is involved in the regulation of embryonic heart and urogenital development [241]. *Osr1* was shown to regulate atrial septum formation in the heart [241]. Studies in mice have shown that loss of *Osr1* expression is embryonic lethal due to deformed atrioventricular junctions and hypoplastic venous valves [242]. Associated GO terms for biological function include *regulation of transcription, heart* and *kidney development* and *cell differentiation*.

Experiment 3 revealed, again, many predicted genes with no known function, as well as 16 known genes listed in Table 6. Analysis of the biological function of these genes, revealed four genes that are involved in *G protein-coupled receptor signaling pathway*, while others are involved in *cell adhesion, cell cycle, transcription* and *circadian rhythm* (Table 6). However, further studies have to be conducted to validate if these genes are specific targets of *NkxUS*.

Table 6 DNA target results from replicate 3

Gene	Description	BP
Aida	Axin interactor, dorsalization-associated protein	Determination of ventral identity
Ank1	Ankyrin-1	Cytoskeleton organisation
Cadm2	Cell adhesion molecule 2	Cell adhesion
Cdc73	Parafibromin	Cell cycle, transcription
Ercc6	DNA excision repair protein	Transcription
Erdr1	Erdr1 protein	Cell proliferation
Ghr	Growth hormone receptor	Endocytosis
Kctd16	BTB/POZ domain-containing protein KCTD16	G protein-coupled receptor signaling pathway
Nampt	Nicotinamide phosphoribosyltransferase	Circadian rhythm
Npr3	Atrial natriuretic peptide receptor 3	G protein-coupled receptor signaling pathway

Olfr211	Olfactory receptor	G protein-coupled receptor signaling pathway
Opcml	Opioid-binding protein/cell adhesion molecule	Cell adhesion
Rgs6	Regulator of G protein signaling 6	G protein-coupled receptor signaling pathway
Sfi1	Protein SFI1 homolog	Cell cycle
Tnik	TRAF2 and NCK-interacting protein kinase	Wnt signaling pathway
Tspan9	Tetraspanin-9	Cell surface receptor signaling

2.7.3.2 Four DNA targets are deregulated upon *NkxUS* knockdown

Binding of *NkxUS* to these DNA targets may lead to the activation or inhibition of their transcription for example by recruiting chromatin modifying complexes to these specific genomic loci as shown for many well-characterised lncRNAs [101, 112]. The lncRNA *HOTAIR* was shown to associate with PRC2 to silence distant loci, whereas the lncRNA *HOTTIP* binds to TrxG/MLL to activate transcription of target genes [141, 243].

Therefore, the list of DNA targets of each replicate was overlapped with the list of significant (FDR <0.05) deregulated transcripts following *NkxUS* knockdown. The intersection revealed four genes (Figure 28), which are listed in Table 7 with their corresponding logFC after *NkxUS* knockdown and their associated function. All four genes are upregulated indicating that *NkxUS* may, upon binding, repress transcription. *Aida* and *Tnik* are both involved in embryonic development, *Sfi1* is involved in the regulation of cell cycle as well as organelle biogenesis and maintenance and *Rgs6* is known to modulate cardiovascular activity. *Rgs6* is a regulator of parasympathetic activation in the heart, functioning as a negative regulator of G protein activation of β-adrenergic receptors [244, 245].

Altogether, the DNA ChIRP with two biological replicates identified only a limited amount of targets, *Ptprk* and *Grm8*, when looking for the overlap between EVEN and ODD, excluding LacZ, and between the biological replicates. Investigating the overlap between individual replicates resulted in 17 additional candidates. Some of which share a common biological function with *Ptprk* and *Grm8* (Table 6). However, further follow up experiments have to be performed to evaluate specific *NkxUS* interaction.

Table 7 Annotation of DNA ChIRP targets that are deregulated upon *NkxUS* knockdown showing the gene name, the logFC as seen from the *NkxUS* knockdown RNA sequencing results and associated gene function

Gene	Description	logFC	Function
Aida	Axin interactor, dorsalization-associated protein	0.22	Acts as a ventralizing factor during embryogenesis.
Rgs6	Regulator of G protein signaling 6	0.52	Regulates G protein-coupled receptor signaling cascades. Inhibits signal transduction by increasing the GTPase activity.
Sfi1	Protein SFI1 homolog	0.67	Plays a role in the dynamic structure of centrosome-associated contractile fibers.
Tnik	TRAF2 and NCK-interacting protein kinase	0.18	Serine/threonine kinase that acts as an essential activator of the Wnt signaling pathway. It is recruited to promoters of Wnt target genes and required to activate their expression.
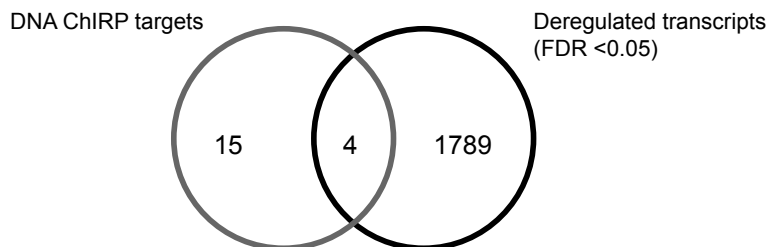
Figure 28 Overlap of *NkxUS* knockdown and DNA ChIRP results. Venn diagram intersecting the deregulated transcripts with a FDR < 0.05 following *NkxUS* knockdown and the DNA ChIRP identified targets.

2.7.4 Identification of protein interacting partners of *NkxUS*

To identify the protein interacting partners of *NkxUS*, we performed ChIRP followed by mass spectrometry (MS). After *NkxUS* capture, proteins were precipitated and separated on a gel, which was then cut into at least four fragments per lane. Proteins were eluted from individual gel slices

and then analysed using a QExactive Plus Orbitrap mass spectrometer by Dr. Ben Crosset (University of Sydney, Australia). For ChIRP-MS, several parameters had to be optimised, including input HL-1 cell number, cell harvesting method, crosslinking condition, fragmentation method and reverse crosslinking time (Table 8). Efficiency of reverse crosslinking can be determined by intersecting the proteins identified in the individual gel pieces. Ideally, after breaking the crosslink, proteins should separate in the gel according to its molecular weight and individual gel pieces should not show many shared proteins. Therefore, if after protein digestion and peptide extraction, peptides are still crosslinked to other proteins or RNA, they will not match any protein in the database as the tolerances for detection are very tight (less than 0.01Da). Thus, in our analysis we rely on the non-crosslinked peptides and might loose a large amount of peptides if the reverse crosslinking step was not efficient.

As for the DNA-ChIRP, we determined the number of proteins shared between EVEN and ODD, which were not present in LacZ, for each replicate (Table 8). Unfortunately, no replicates were done for any one method and therefore difficulties were experienced in the data analysis.

Table 8 Optimisation parameters for ChIRP-MS. Listed are the number of cells used, whether cells were harvested by scraping or trypsin digestion, whether whole or nuclear enriched cell lysates were used, the formaldehyde concentration used for crosslinking, if the samples were fragmented using a probe or water bath sonicator as well as the time for reverse crosslinking and resulting protein number after MS identification.

#	~ Million cells	Cell harvest	Cell lysate	Formaldehyde concentration	Sonication method	Reverse crosslinking	# proteins
1	100	scrape	whole	3%	probe sonicator	30min	1
2	100	trypsin	nuclear enriched	3%	water bath	30min	6
3	100	trypsin	whole	0.5%	water bath	30min	46
4	400	trypsin	whole	0.5%	water bath	90min	168
5	400	scrape	whole	0.5%	water bath	90min	20

2.7.4.1 Analysis of proteins shared between experiments

Since no biological replicates were performed, an initial analysis was focused on looking at the most enriched proteins from all experiments with the idea that true interacting partners should be consistently represented above noise. The unique EVEN/ODD specific proteins of each replicate were intersected to search for common proteins from several independent experiments. The first experiment was excluded as it only identified one protein, which was not shared with any other experiments. Intersection of experiments 2-5 did not reveal any proteins shared between all experiments (Figure 29). An overlap was only observed between two experiments at a time, with in total 17 shared proteins (Figure 29). As a different approach we analysed the overlap of all the EVEN and ODD experiments taken individually. Both approaches revealed the same proteins with only one additional one, HNRNPAB, using the second approach. Listed in Table 9 are the proteins shared between the experiments and ranked based on the number of times observed in any EVEN or ODD experiment.
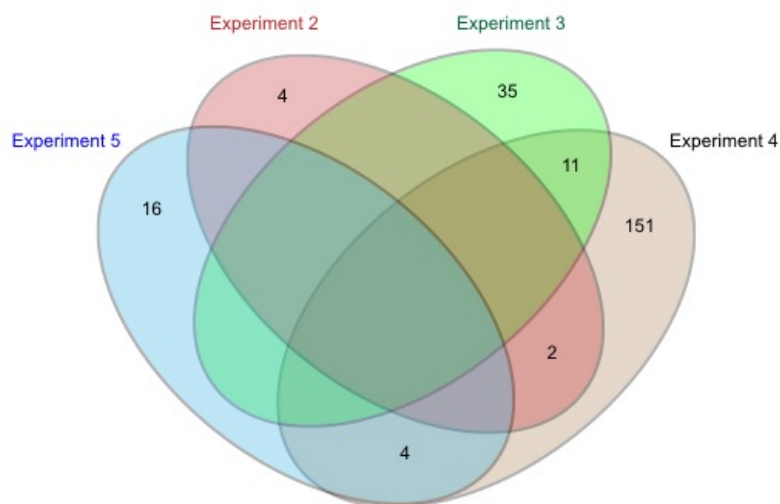
Figure 29 Venn Diagram showing the intersection of proteins identified from the individual replicates. Shown are the proteins shared between EVEN and ODD, excluding proteins found in LacZ. Number of proteins shared between individual replicates are indicated in the overlap.

Table 9 Proteins ranked based on the number of times observed in any EVEN or ODD sample

Protein	Description	# observed
DLAT	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex	5
PRDX3	Thioredoxin-dependent peroxide reductase	5
ACOT13	Acyl-coenzyme A thioesterase 13	4
ATP5H	ATP synthase subunit e	4
CYCS	Cytochrome c	4
FH	Fumarate hydratase	4
GOT2	Aspartate aminotransferase	4
GSTP1	Glutathione S-transferase P 1	4
HINT2	Histidine triad nucleotide-binding protein 2	4
HNRNPAB	Heterogeneous nuclear ribonucleoprotein A/B	4
HSPD1	60 kDa heat shock protein	4
LDHA	L-lactate dehydrogenase A chain	4
OXCT1	Succinyl-CoA:3-ketoacid coenzyme A transferase 1	4
PCBP1	Poly(rC)-binding protein 1	4
PDHA1	Pyruvate dehydrogenase E1 component subunit alpha	4
UQCRC1	Cytochrome b-c1 complex subunit 1	4
NDUFV1	NADH dehydrogenase [ubiquinone] flavoprotein 1	4
SOD2	Superoxide dismutase	4

2.7.4.2 Highest *NkxUS* enrichment was achieved with Experiment 4

Given that only a limited number of proteins were identified to overlap between experiments, we performed a more detailed analysis of Experiment 4. Experiment 4 worked best in terms of visual inspection of the gel, reverse crosslinking efficiency and subsequent protein identification as well as highest *NkxUS* retrieval (~40%) (Figure 30A). Two sharp bands were detected in EVEN and ODD and not LacZ when running the gel after reverse crosslinking (Figure 30B). The lanes were sliced into six pieces in order to identify the specific proteins in these bands via MS (Figure 30C).
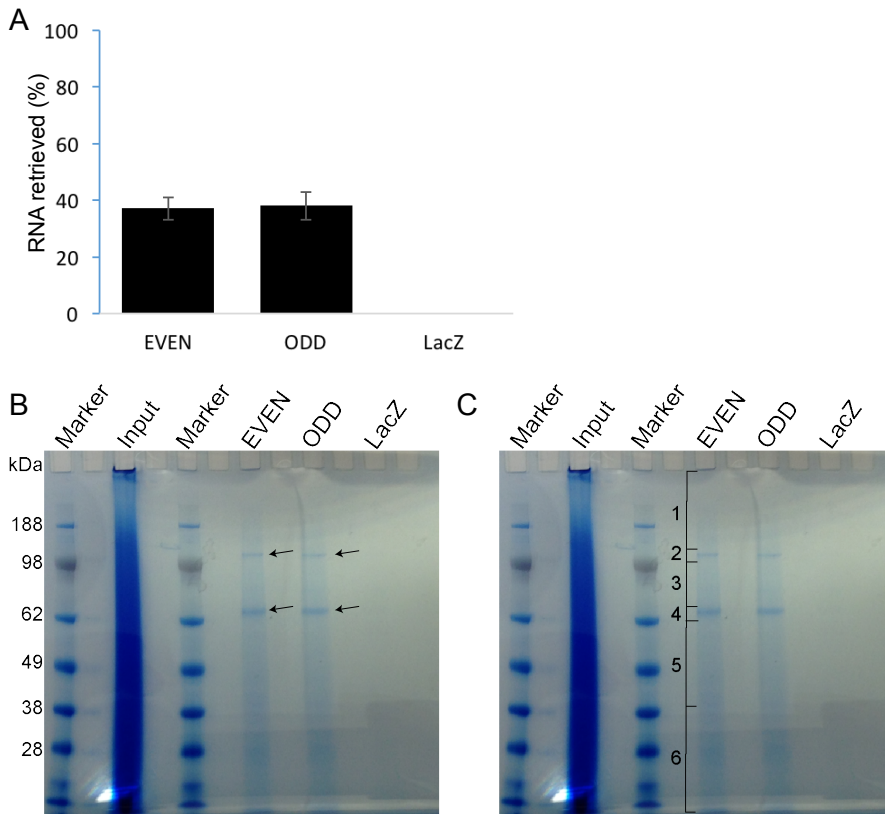
Figure 30 ChIRP successfully enriched for *NkxUS*. A) ~40% of *NkxUS* was retrieved with EVEN and ODD probes, whereas no enrichment was observed with LacZ probes. B) Two clear bands were detected (indicated by arrows) for EVEN and ODD, which were absent in LacZ. C) Each lane was cut into 6 fragments, which were analysed individually using MS.

2.7.4.3 Analysis of the *NkxUS* specific bands of the protein gel

Experiment 4 showed clear visualisation of two sharp bands in EVEN and ODD lanes (Figure 30B), suggesting that these might be specific *NkxUS* interacting proteins as no bands of that size were observed in the LacZ lane. MS of these gel fragments, however, did not identify any high confidence targets for the upper band. The most abundant proteins in this slice were also present in LacZ and no protein matched the expected size, potentially due to incomplete reverse crosslinking and thus the band could represent a protein complex or a still crosslinked protein, which we were unable to identify.

For the lower band, the trifunctional enzyme subunit alpha (HADHA) was the only protein identified within the particular gel slice that was not shared with LacZ. Interestingly, HADHA was also identified in the EVEN probe set of Experiment 3. The absence of HADHA in LacZ and its substantial enrichment in the gel suggests that it may be specific.

2.7.4.4 *NkxUS* interacting proteins are involved in metabolic processes

When overlapping the proteins identified per lane from each sample in Experiment 4, 168 proteins were shared between EVEN and ODD and absent in LacZ (Figure 31A). Looking at the most significant GO terms for biological process revealed largely metabolism related terms but also some non-metabolic terms, including *ion transport, cell communication, transcription* and *muscle contraction* (Figure 31B). The vast majority and most enriched captured proteins have a mitochondrial associated function (Table 10).

Taken together, the majority of proteins shared between experiments, detected in individual experiments and even in the prominent gel band have annotated mitochondrial functions. It remains to be evaluated if these interact specifically with *NkxUS* or represent experimental noise.
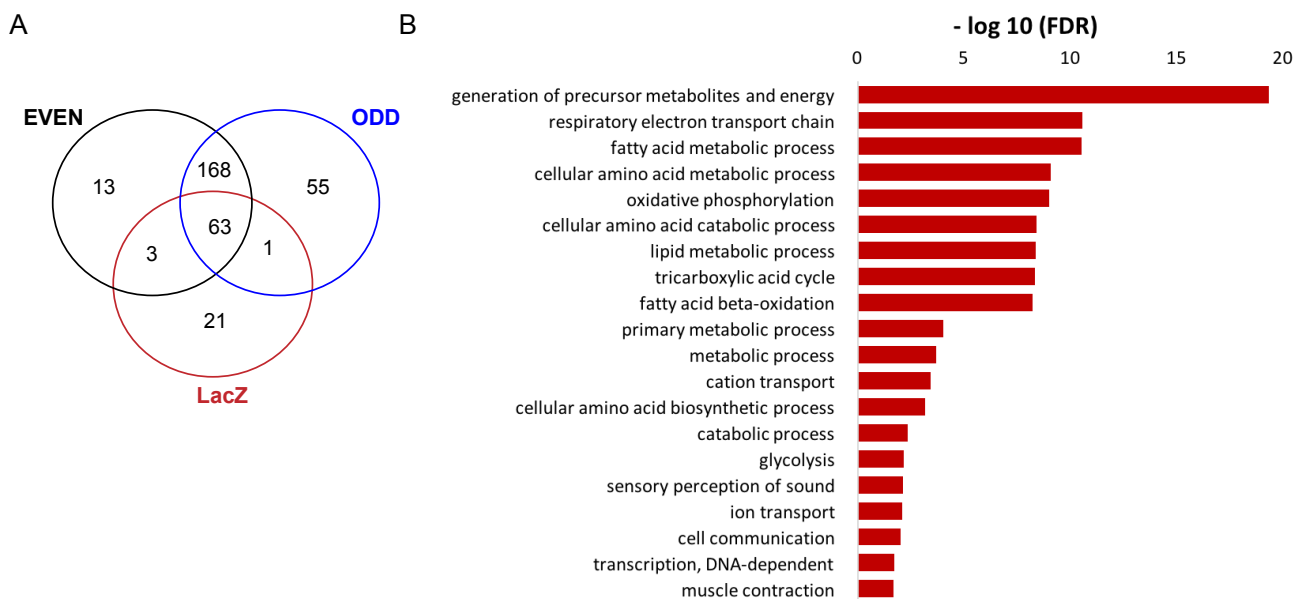
Figure 31 Results for *NkxUS* interacting proteins from Experiment 4. A) Overlap of all three samples resulted in 168 proteins shared between EVEN and ODD but not LacZ. B) Most significant GO term for biological process for the shared proteins. Scale in –log10(FDR).

Table 10 Top 20 most enriched ChIRP-MS identified proteins of Experiment 4 showing the unique peptide count for the EVEN and ODD probes.

Protein	Description	EVEN	ODD
ACO2	Aconitate hydratase	29	28
HSPD1	60 kDa heat shock protein	24	22
PCCB	Propionyl-CoA carboxylase beta chain	22	22
TUFM	Elongation factor Tu	20	18
ACAA2	3-ketoacyl-CoA thiolase	18	18
LRPPRC	Leucine-rich PPR motif-containing protein	19	16
HADHA	Trifunctional enzyme subunit alpha	15	17
ALDH2	Aldehyde dehydrogenase	16	15
LONP1	Lon protease homolog	14	15
ACADM	Medium-chain specific acyl-CoA dehydrogenase	14	14
PDHA1	Pyruvate dehydrogenase E1 component subunit alpha	13	14
ALDH18A1	Delta-1-pyrroline-5-carboxylate synthase	11	16
OAT	Ornithine aminotransferase	13	13
TRAP1	Heat shock protein 75 kDa	12	14
SUCLG2	Succinate--CoA ligase [GDP-forming] subunit beta	14	11
ETFB	Electron transfer flavoprotein subunit beta	12	12
FH	Fumarate hydratase	12	12
TUBB5	Tubulin beta-5 chain	12	12
ACAT1	Acetyl-CoA acetyltransferase	12	11
ACADL	Long-chain specific acyl-CoA dehydrogenase	11	12

2.8   The mouse model to study the function of *NkxUS in vivo*

Functional experiments have so far been performed *in vitro* using an immortalised cell line. However, as these cells can have substantial abnormal functions, we expanded our analysis of *NkxUS* function *in vivo* using a mouse model.

2.8.1 Generation of a mouse model to study the conserved RNA structure

Since we identified an RNA structure conserved between mouse and human *NkxUS* and since conservation often implies function, we intended to generate a mouse model by altering this domain. In our approach, we wanted to target part of the conserved structure of *NkxUS*. Initially, we intended to insert three polyadenylation signals using clustered regularly interspaced short palindromic repeats/CRISPR-associated protein 9 (CRISPR-Cas9) in order to terminate *NkxUS* transcription early, which in turn would lead to reduced or no *NkxUS* expression. However, at that time insertion of the polyadenylation signals was unsuccessful and we decided to study the effect of a partial deletion of the conserved structure.

For *NkxUS*, a single guide RNA (sgRNA) targeting the conserved structure was used. In mouse, the conserved structure lies at the start of the second exon not overlapping any of the known *Nkx2-5* enhancers (Figure 32A) [84-88]. Histone marks also do not show any evidence of enhancer marks, as seen from the absence of H3K4me1 marks (Figure 32A). One limitation is that the entire conserved structural region overlaps a transcript antisense to *NkxUS*, as seen from the ENCODE heart RNA sequencing data (Figure 32A). Thus, by deleted this region we may also alter the antisense transcript and potentially abort its transcription (Figure 33A). However, as we believe that this RNA structure has functional relevance we decided to pursue this approach.

A region of 286-289bp was successfully deleted and two homozygous female founders were generated (Figure 32). The line was termed NkxUSΔ and the two founder females were named NkxUSΔ-F5 and NkxUSΔ-F10. NkxUSΔ-F5 harbors a 286bp deletion (Figure 32B) and NkxUSΔ-F10 a 289bp deletion and 1bp insertion (Figure 32C).
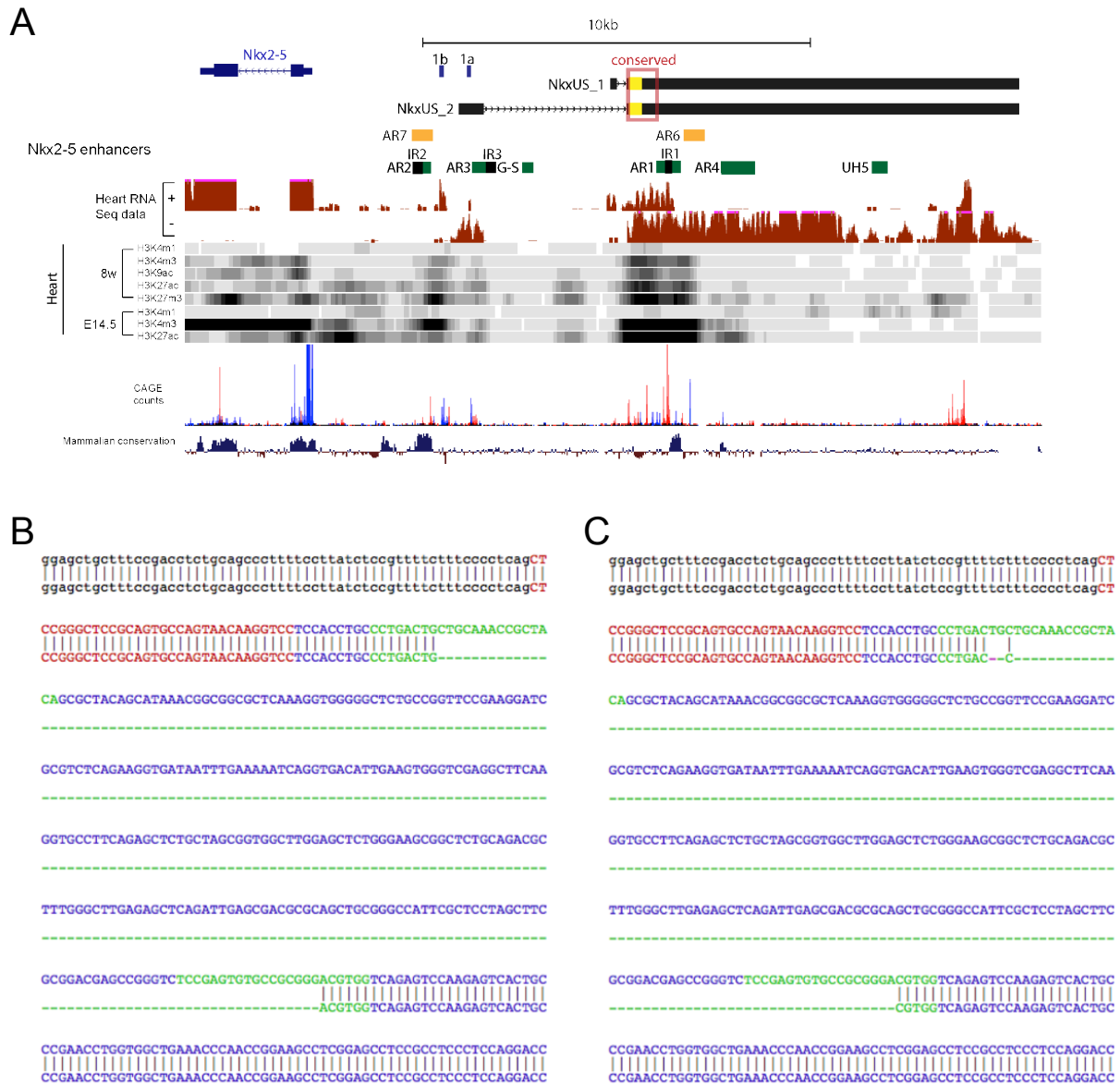
Figure 32 Deletion of part of the conserved structure using CRISPR. A) Depiction of the *Nkx2-5* and *NkxUS* locus highlighting the deleted (yellow) and the conserved (red rectangle) region in *NkxUS*. *Nkx2-5* enhancers are shown in green (cardiac) and yellow (non-cardiac) and inhibitory regions are shown in black. Mouse heart RNA sequencing data from ENCODE is shown in red, also showing the short *NkxUS* antisense transcript on the plus strand. Histone marks from adult and embryonic hearts are shown in the grey tracks. Along with CAGE counts and mammalian sequence conservation in the lower two tracks. B) NkxUSΔ-F5 harbors a 286bp deletion C) NkxUSΔ-F10 harbors a 289bp deletion and an 1bp insert. Legend: CAPITALS = Exon 2 of *NkxUS*, Red = start of the exon prior to conserved region; Blue = conserved region; Green = sgRNA sites; Wild-type sequence is illustrated on the top.

2.8.2 No abnormal phenotype was observed in NKXUSΔ$^{-/-}$ mice

NKXUSΔ$^{-/-}$ mice appeared normal. Crossing of NKXUSΔ$^{+/-}$ mice resulted in average litter sizes of 4-6 mice with normal Mendelian ratios. NKXUSΔ$^{-/-}$ mice are viable, fertile and show no obvious abnormal phenotype or change in body weights (Figure 33).
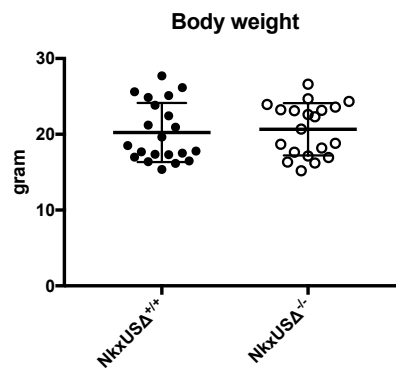
Figure 33 Body weights are unchanged in adult NKXUSΔ$^{-/-}$ mice. Body weight measurements of NKXUSΔ$^{-/-}$ and NKXUSΔ$^{+/+}$ mice did not reveal any differences. Data presented as mean ± SD, n=20.

2.8.3 *NkxUS* RNA levels are unchanged in NkxUSΔ$^{-/-}$ hearts

To analyse if deletion of part of the conserved region affected the RNA levels of *NkxUS* we checked expression levels in whole hearts from adult NkxUSΔ$^{-/-}$ and NkxUSΔ$^{+/+}$ mice. RT-PCR showed that the RNA levels of *NkxUS*, *NkxDS*, *Nkx2-5* and the *NkxUS antisense transcript* (*NkxUSas*) were not changed, indicating that the deletion did not affect the transcript levels of *NkxUS*, nor had any affect on *NkxDS*, *Nkx2-5* or *NkxUSas* expression (Figure 34).
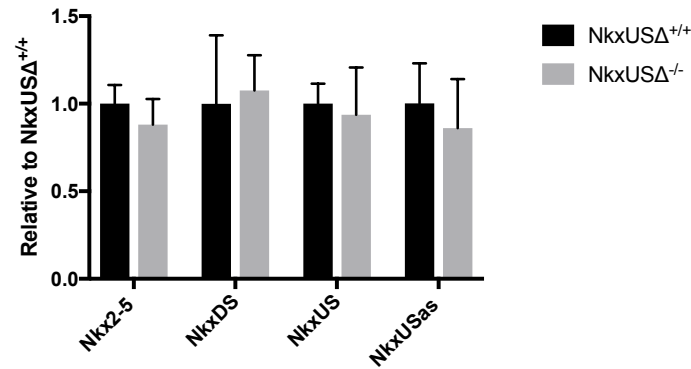
Figure 34 *NkxUS* RNA levels are unchanged in NkxUSΔ[-/-] hearts. RT-PCR showed no change in the expression levels of *Nkx2-5, NkxDS, NkxUS* and *NkxUSas* in NKXUSΔ[-/-] hearts when normalised to four housekeeping genes and relative to NKXUSΔ[+/+]. Data presented as mean ± SD, n=4.

2.8.4 NKX2-5 protein levels are not altered in NkxUSΔ[-/-] hearts

Nkx2-5 RNA levels were unchanged in NKXUSΔ[-/-] mice, however, since mRNA levels may not necessarily reflect protein levels, western blot analysis was performed for NKX2-5 on NKXUSΔ[-/-] and NKXUSΔ[+/+] adult hearts. Western blot analysis with an antibody against NKX2-5 (NKX2-5 N19, Santa Cruz) and Vinculin (H300, Santa Cruz) as a control revealed no significant changes between NKXUSΔ[+/+] and NKXUSΔ[-/-] mice (Figure 35A and B), suggesting that the conserved RNA structure affects neither *Nkx2-5* protein nor RNA levels. However, whether potential *Nkx2-5* and *NkxUS* interactions are affected by this deletion remains to be elucidated.
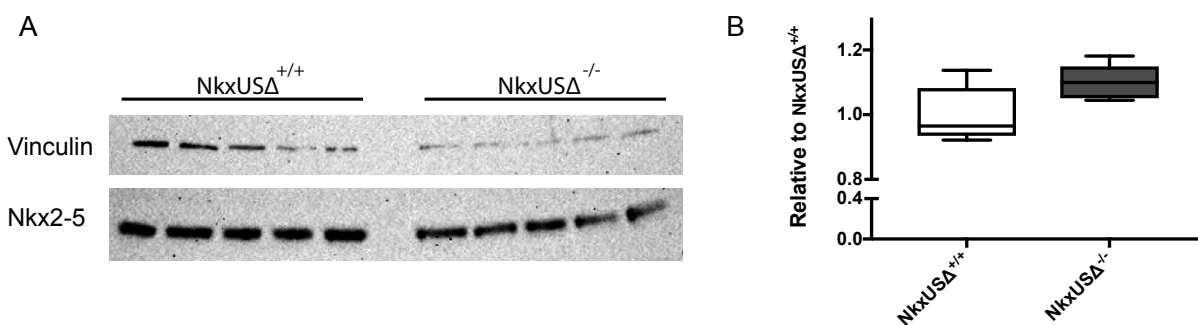
Figure 35 NKX2-5 protein levels are unchanged in NkxUSΔ adult hearts. A) Protein extracts from NKXUSΔ[+/+] and NKXUSΔ[-/-] hearts were analysed by western blot using Vinculin as a control and NKX2-5; n=5. B) Analysis of western blot band intensity, normalised to Vinculin and relative to NKXUSΔ[+/+], revealed no significant changes on protein level for NKX2-5. Data presented as mean ± SD, n=5. The complete western blot can be found in Appendix IV.

2.8.5 No heart defects were detected in NKXUSΔ$^{-/-}$ embryos

Hearts of E17.5 embryos were collected to analyse for cardiac morphological defects using optical projection tomography (OPT). The OPT creates high resolution three-dimensional (3D) images by taking images every few seconds and rotating the samples 360 degrees. Biological samples up to 15mm thick can be imaged [246]. Following reconstruction and compilation of images, the software Amira allows researches to scroll through the heart, rotate and flip it in any direction to analyse for heart defects. The samples were analysed for defects in the size and morphology of the ventricles and atria, as well for any septal and valve defects. Four NkxUSΔ$^{-/-}$ and NKXUSΔ$^{+/+}$ embryos were analysed and no defects were identified.

2.8.6 NkxUSΔ$^{-/-}$ adult mice exhibit normal atrial septal development

In humans, NKX2-5 is commonly associated with several congenital heart abnormalities including atrial septal defects (ASD). ASD is a relatively common congenital heart defect and occurs during heart development when the atrial septum fails to close completely resulting in a permanent opening between both atria. The atrial septum is therefore arguably the most vulnerable structure to developmental perturbations affecting the heart. The main pathological consequence is cardiac shunt, meaning a deviation of normal blood flow [247]. After birth, the pressure in the left atrium is higher than in the right atrium. An ASD allows blood flow from the left into the right atrium, which will ultimately lead to an increase in right atrial pressure and hence progressive enlargement and myocardial thickening [247].

In *Nkx2-5* heterozygous mice, ASD was rarely observed [75]. Instead, these mice were shown to have an increased prevalence of patent foramen ovale (PFO) [75]. PFO is generally a benign condition. It is evident after birth, when the foramen ovale fails to close. The foramen ovale is a small opening in the septum secundum (Figure 37A), one of the two distinct walls (septum primum and secundum) which form the interatrial septum [248]. Before birth, the lungs are not yet functional and oxygen rich blood is received from the mother's placenta. The foramen ovale therefore allows blood to bypass the lungs and facilitates the passage of oxygenated blood to the left atrium. After birth, pressure in the left atrium rises, forcing the septum secundum against the septum primum, permanently closing the opening [248]. In individuals with PFO, a small amount of

blood may leak from the right atrium to the left, which, for the vast majority, is asymptomatic. Clinically, an ASD can be described as a large PFO. Studies suggested that the occurrence of PFO in mice might be linked genetically to the causation of ASD in humans [75].

Hearts of adult (6-8 weeks old) NKXUSΔ$^{-/-}$ and NKXUSΔ$^{+/+}$ mice were dissected and anatomically analysed by Dr. Edwin Kirk (University of New South Wales, Australia) for ASD and PFO as well as differences in atrial septal morphology, more specifically crescent width (CRW), foramen ovale width (FOW) and flap valve length (FVL) (Figure 36A). Previously, mean flap valve length was measured and shown to be negatively correlated with the incidence of PFO, whereas the foramen ovale and crescent width were positively correlated [75]. These quantitative parameters have been previously demonstrated to be highly correlated with prevalence for PFO [75, 249].

20 animals of each genotype were analysed but no significant differences were found in any of the septal quantitative measurements (Figure 36B). In total, one ASD as well as five PFOs were identified in NKXUSΔ$^{-/-}$ mice, whereas, no ASD, but 6 PFOs were found in NKXUSΔ$^{+/+}$ mice, indicating that deletion of part of the conserved region did not affect atrial septal development, as was seen for *Nkx2-5* heterozygous mice. Based on gross morphology and quantitative assessment of atrial septal parameters, we conclude that the region deleted in NKXUSΔ has no impact on heart development.
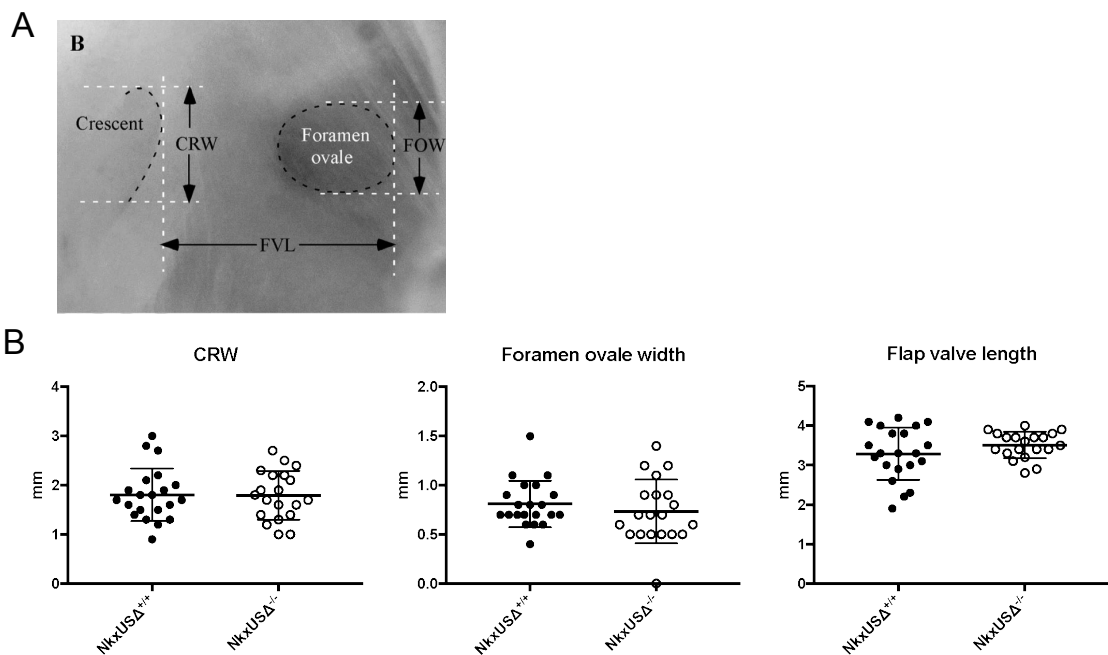
Figure 36 Atrial morphology analysis of NKXUSΔ$^{-/-}$ and NKXUSΔ$^{+/+}$ mice. A) Light micrograph of atrial septal details as seen from the left aspect after removal of the left atrial appendage. The

annulus of the mitral valve is toward the lower left of this panel. Septal landmarks and quantitative septal measurements used in this study are highlighted (taken from [249]). B) No significant differences between genotypes were detected for any of the quantitative measurements. CRW= crescent width. Data presented as mean ± SD, n=20.

2.8.7 Functional analysis of NkxUSΔ$^{-/-}$ mice in a model of left ventricle hypertrophy

Often phenotypes develop under stress. Several models to stress the heart are well established and are important tools to gain further insight into the potential function of genes in the heart. Based on the possible association of human NKXUS with an increased heart rate, we decided to use a model of pressure overload to study the consequences of overload induced hypertrophy on cardiac function and contractility. Pressure overload of the left ventricle (LV) leads to an increased stretching force on the myocardium, which in turn results in hypertrophy and an altered contractile function of the myocardium [250].

We performed transverse aortic constriction (TAC) surgery on male NKXUSΔ$^{-/-}$ and NKXUSΔ$^{+/+}$ mice. TAC is a commonly used method to induce LV hypertrophy and heart failure in mice [251]. It is designed to increase the resistance of pumping for the LV, which leads to an elevated energy demand and eventually to disturbed cardiac contraction and heart failure. As the name suggests, the inside diameter of the aorta close to the right ventricle will be constricted. Banding of the aorta leads to left ventricular overload and hypertrophy as it has to work harder to eject the normal amount of blood. Over time, the heart cannot adjust adequately anymore to the chronic pressure overload, resulting in cardiac dilatation and heart failure [251].

We performed TAC on 10-11 male adult mice of each genotype and sham surgery (placebo surgery) on 9 mice of each genotype. In sham surgery, the chest was opened but no banding was performed. At 21 days post-surgery, echocardiography as well as left ventricle micromanometry were performed. Echocardiography is a non-invasive ultrasound method used to study cardiac structure and function in anaesthetised mice. LV micromanometry was used to determine the development of LV pressure over time in each contractile cycle to analyse ventricular systolic and diastolic function.

After completion of left ventricle micromanometry, mice were sacrificed. The hearts and tibia were collected to measure heart weight, left ventricular weight and left atrial weight relative to tibia length. The surgery and LV micromanometry were performed by Jianxin Wu and the

echocardiography by Scott Kesteven (both of the Victor Chang Cardiac Research Institute, Australia).

We decided to complete the study at day 21 to achieve a model of left ventricle hypertrophy. Heart failure develops at later timepoints as the chronic overload worsens cardiac function.

2.8.7.1 Tissue collection data showed that LV hypertrophy was successfully induced

Tissues were collected and weighted. To account for variations in mouse size, the body and heart weight parameters were corrected for tibia length. The results showed no significant difference in body weight between the genotypes nor TAC and sham animals (Figure 37). LV hypertrophy was successfully induced as seen from the increased whole heart and LV weight in TAC compared to sham mice (Figure 37). However, no significant changes between NKXUSΔ$^{-/-}$ and NKXUSΔ$^{+/+}$ TAC mice regarding the hypertrophic response were detected. The left atrial weight did not increase significantly, indicating that no atrial dilation occurred, which would be expected at later time points when the mice proceed into heart failure (Figure 37).
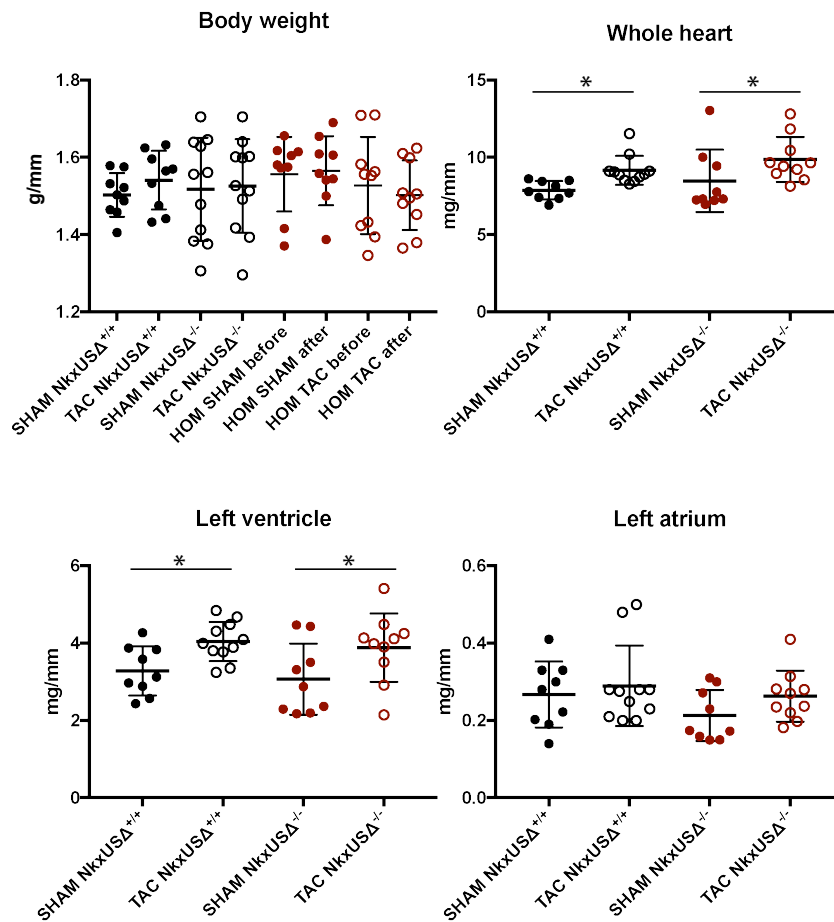
Figure 37 Body weight and tissue collection data. Body weights of NKXUSΔ$^{+/+}$ and NKXUSΔ$^{-/-}$ mice did not change during the experiment and are not different between the groups. The tissue collection data showed a significant increase in whole heart and LV weight in TAC animals in both groups. No differences between NKXUSΔ$^{-/-}$ TAC and NKXUSΔ$^{+/+}$ TAC mice were observed. All parameters were corrected for tibia length. Data presented as mean ± SD, n=9-11, *P<0.05, unpaired two-tailed *t*-test.

2.8.7.2 Echocardiography results revealed an impairment in LV relaxation

Echocardiography was performed on anaesthetised mice to analyse cardiac function. No difference in heart rate under anaesthesia was identified between the groups or genotypes (Figure 38). TAC animals in both groups had a heart rate of ~470bpm and sham animals one of ~505bpm (Figure 38). In line with the left atrial weight, the diameter did not change significantly (Figure 38). As expected, the wall thickness was increased in TAC mice compared to sham animals as the muscle undergoes hypertrophy in response to the pressure overload, further confirming the tissue collection data

(Figure 38). However, again, no difference in the response to hypertrophy was observed between the genotypes (Figure 38).

While all parameters for ejection fraction (measurement of the fraction of blood ejected from the ventricle per heartbeat), stroke volume (volume of blood pumped from the left ventricle per beat) and cardiac output (volume of blood pumped from the ventricles per minute) showed a trend for worsening cardiac function in NKXUSΔ$^{-/-}$ TAC mice compared to NKXUSΔ$^{+/+}$ TAC mice, none of these parameters reached significance due to the increased variance in the response in NKXUSΔ$^{-/-}$ TAC animals (Figure 38).

However, when looking at the blood flow velocity through the mitral valve (E/A), measured as the ratio of the early (E) to late (A) filling phase of the LV, NKXUSΔ$^{-/-}$ sham animals had a significant lower value than NKXUSΔ$^{+/+}$ sham animals (Figure 38). This difference was attenuated in TAC animals. The mitral valve lies between the left atrium and LV and the mitral flow signal is used to assess diastolic function. The LV has two filling phases, the early (passive) phase, and the late (active) phase. During diastole (ventricular relaxation), the mitral valve opens as soon as the pressure in the left atrium exceeds the pressure of the LV. Once opened, blood flows rapidly into the LV due to the high-pressure gradient. This early, passive filling accounts for about 70% of LV filling. In the late phase, blood is actively pumped into the LV during atrial contraction, accounting for 30% of total filling. Therefore, under normal conditions, the E-wave is higher than the A-wave. The difference detected in our study is driven by a significant difference in the early phase of filling, the E-wave, whereas no significant change was observed for the A-wave (Figure 38). The E-wave in NKXUSΔ$^{+/+}$ sham animals had an averaged value of 900cm/s and in NKXUSΔ$^{-/-}$ sham mice one of 750cm/s. The A-wave had a value of about 555-565cm/s; as such, the E-wave is still higher than the A-wave in NKXUSΔ$^{-/-}$ sham mice, suggesting a mild diastolic dysfunction. Diastolic dysfunction could either be due to disturbed LV relaxation or differences in left atrial pressure. However, the E/e' value (e' measured at mitral annulus, a fibrous ring at the mitral valve), which is a measure of atrial pressure, was not different between the groups, suggesting that NKXUSΔ$^{-/-}$ mice have a mild impairment in LV relaxation when unchallenged (Figure 38).
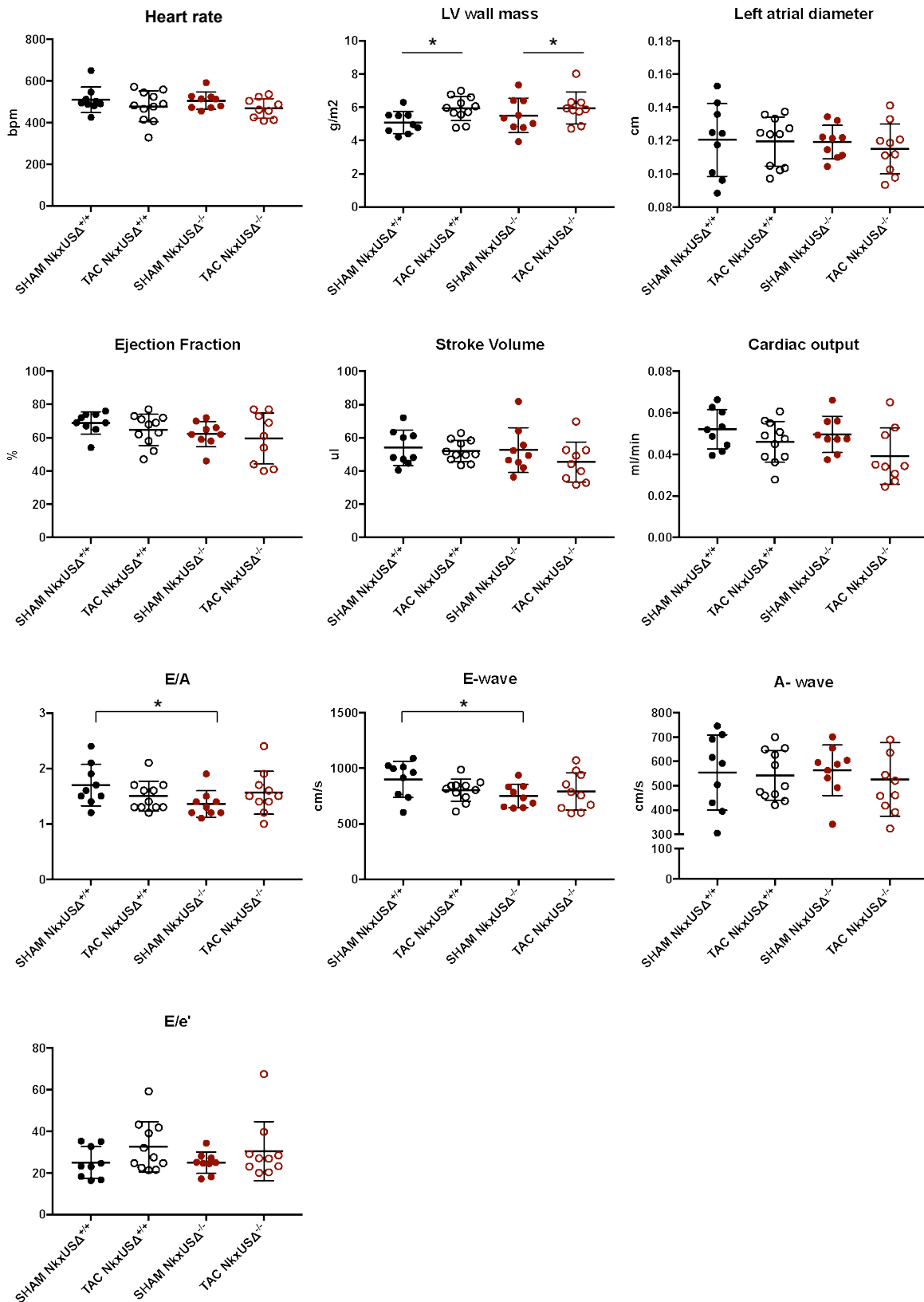
Figure 38 Echocardiography data revealing a mild diastolic dysfunction in NKXUSΔ$^{-/-}$ sham animals. Heart rates were unchanged between the groups. As expected LV wall mass was significantly increased in TAC animals compared to sham animals, but similar between TAC animals. Left atrial diameter, ejection fraction, cardiac output and stroke volume were not significantly

changed. Measuring the blood flow velocity at the mitral valve (E/A) identified a significant difference between NKXUSΔ$^{+/+}$ and NKXUSΔ$^{-/-}$ sham animals, which is attenuate in TAC animals. This difference is driven by the E-wave as the A-wave was not changed. The atrial pressure (E/e') did not show any significant differences, indicating that NKXUSΔ$^{-/-}$ animals have a dysfunction in LV relaxation. Data presented as mean ± SD, n=9-11, *P<0.05, unpaired two-tailed *t*-test.

2.8.7.3 Left ventricular pressure results showed no difference in LV relaxation

To determine if TAC resulted in differences in pressure between NKXUSΔ$^{-/-}$ and NKXUSΔ$^{+/+}$ mice, a catheter was inserted into the LV to measure the pressure continuously for a few minutes.

The end-systolic pressure was significantly increased in TAC animals as expected due to the increased resistance to pumping as we constricted the aorta (Figure 39). However, again no difference between NKXUSΔ$^{+/+}$ and NKXUSΔ$^{-/-}$ TAC mice was detected (Figure 39). The end –diastolic pressure was not changed indicating that the left ventricular filling pressure is unaffected (Figure 39). dp/dt max, which is the derivative of pressure over time and measurement of contractility, is not changed, indicating that cardiac contraction is unaffected by the induced hypertrophy and similar between genotypes. dp/dt min, which is the minimal rate of change of ventricular pressure and a measure for relaxation, was also not changed, indicating that the dysfunction in LV relaxation seen from the echocardiography data is mild.

In summary, the pressure data confirmed the successful induction of LV hypertrophy. Despite this, as in the tissue collection and echocardiography, no differences between the groups were observed regarding the response to hypertrophy. However, a mild but significant reduction of the E-wave velocity with normal atrial (E/e') and LV (dp/dt min) pressures was detected in NKXUSΔ$^{-/-}$ sham compared to NKXUSΔ$^{+/+}$ sham animals, indicating that NKXUSΔ$^{-/-}$ mice have a mild diastolic dysfunction, potentially due to a stiffer ventricle, which is attenuated in hypertrophy, when the ventricles become stiffer in both groups due to thickening of the LV wall.
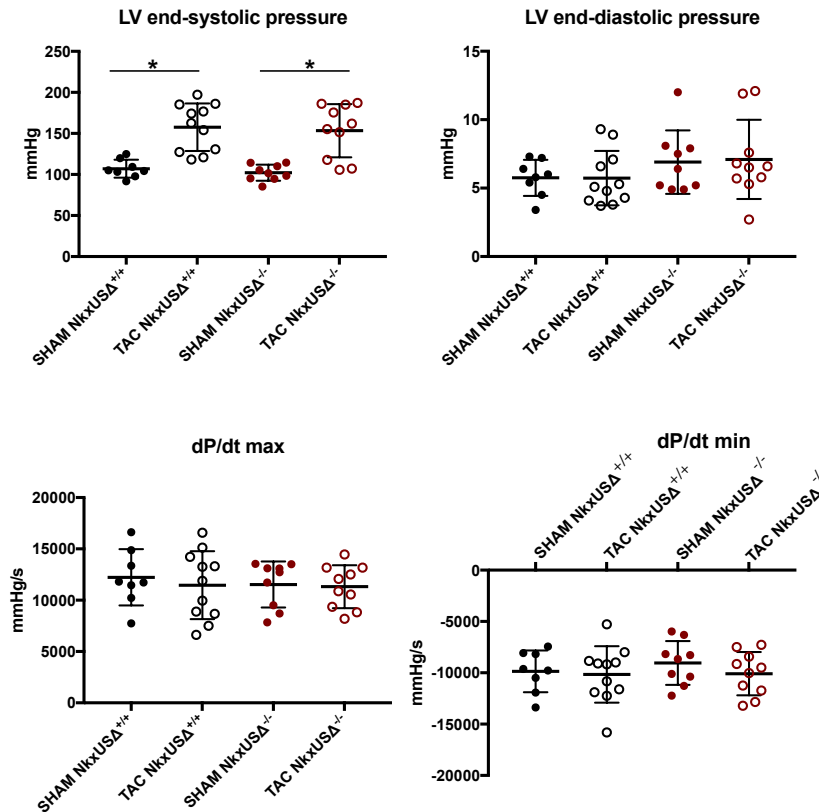
Figure 39 LV pressure data did not show differences in LV relaxation. TAC induced as expected a rise in LV end-systolic pressure as the constriction of the aorta hindered the ejection of blood from the LV, indicating that the model worked. No significant change was observed for the end-diastolic pressure nor in the maximal (dp/dt max) and minimal (dp/dt min) rate of pressure change. Data presented as mean ± SD, n=9-11, *P<0.05, unpaired two-tailed *t*-test.

2.8.8 Heart rate analysis in conscious mice

Since the determination of the precise heart rate of NKXUSΔ$^{-/-}$ and NKXUSΔ$^{+/+}$ mice is crucial for this study and since the TAC study only analysed anaesthetised animals, we next performed a telemetry study to monitor the heart rate in conscious, freely moving mice.

The electrocardiogram (ECG) is the gold standard when it comes to heart rate and cardiac rhythm measurements under normal conditions. Therefore, we recorded the ECG in conscious animals using a telemetry device. Telemetry studies involve the implantation of a small transmitter into the subcutaneous space of a mouse for the long-term recording of physiological parameters.

9-11 adult male animals of each genotype, aged 10-12 week old, were implanted with an ECG telemetry device by Jianxin Wu (Victor Chang Cardiac Research Institute). Four days after implantation, ECG parameters were recorded for 24hrs, while the animals were undisturbed in

their normal cage environment. The data was analysed using the ECG analysis tool ECG Pro (Data Sciences International).

2.8.8.1 NKXUSΔ$^{-/-}$ mice exhibit an increased resting heart rate

Mice rest during the day and are active during the night, therefore, the data was divided into the light (7am-7pm) and dark (7pm-7am) period of the day. The results showed a significant increased resting heart rate for NKXUSΔ$^{-/-}$ mice, whereas no difference was observed in the active period of the day (Figure 40). The heart rates at rest differed by 19.07 ± 7.1bpm with NKXUSΔ$^{-/-}$ mice having on average a heart rate of ~570bpm. When active, NKXUSΔ$^{+/+}$ mice had on average a heart rate of 617.5 ± 8.5bpm and NKXUSΔ$^{-/-}$ mice one of 624 ± 5.5bpm. Furthermore, when looking at the 24h dataset and searching for the lowest heart rate of each animal, NKXUSΔ$^{-/-}$ animals had on average a significantly higher heart rate (444.3 ± 3.4bpm versus 428.1 ± 5.5bpm) (Figure 40).
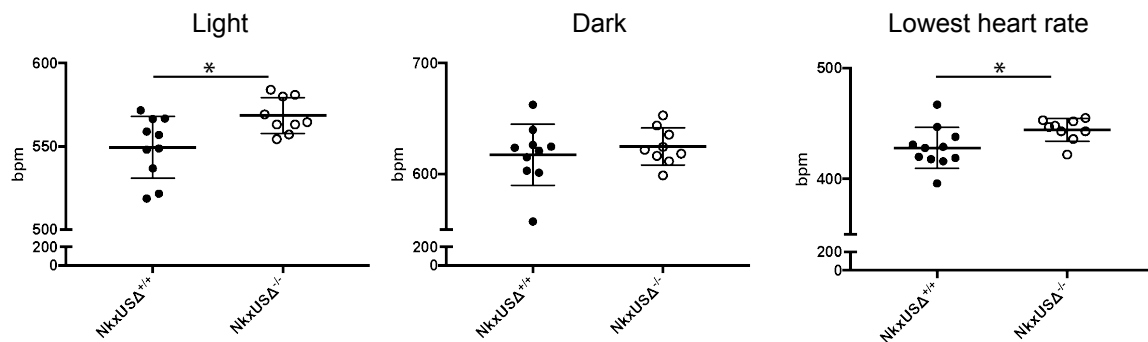
Figure 40 Telemetry data identified an increased resting heart rate in NKXUSΔ$^{-/-}$ mice. A telemetry study monitoring the heart rate in conscious, freely moving mice revealed that during the light (7am - 7pm) period of the day when the mice are at rest, NKXUSΔ$^{-/-}$ mice have a significantly increased heart rate compared to NKXUSΔ$^{+/+}$ mice. This difference is attenuated in the dark (7pm – 7am), active phase of the day. The overall lowest heart rates of the whole 24h dataset per mouse were also significantly higher in NKXUSΔ$^{-/-}$ mice. Data presented as mean ± SD, n=9-10, *P<0.05, unpaired two-tailed *t*-test.

2.8.8.2 Impulse transmission is not altered in NKXUSΔ$^{-/-}$ mice

Changes in heart rate may reflect disturbed electrophysiological properties in other compartments of the heart, which can be analysed using ECG. A normal mouse cardiac rhythm in an ECG shows

five electrical events, the P-, Q-, R-, S- and T-wave. The P-wave reflects electrical activity triggering atrial contraction, the QRS-complex represents ventricular depolarisation and the T-wave ventricular repolarisation. The duration between these complexes gives insight into electrophysiological properties, with the PR-interval indicating atrial and atrioventricular nodal conduction, the QRS-duration showing activation of the ventricles, QT-interval displaying the duration for myocardial repolarisation and ST-interval illustrating the period between depolarisation and repolarisation of the ventricles. However, no difference was observed in any of the intervals, indicating that impulse propagation throughout the heart is not altered (Figure 41).
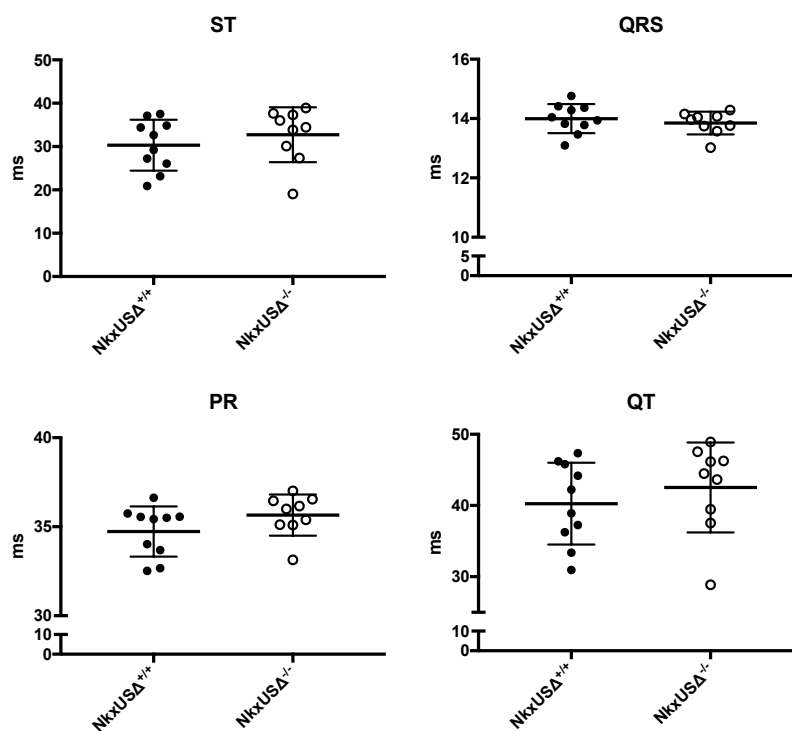
Figure 41 Propagation of action potentials throughout the heart is not affected. ST-, PR-, QRS- or QT-interval were not changed between animal groups. Data presented as mean ± SD, n=9-10.

2.8.8.3 Sinus pauses occurred at a similar rate in both genotypes

The data was additionally analysed for arrhythmias, more specifically the occurrences of sinus pauses. A sinus pause is described as a period where no heart beat is initiated from the sinoatrial node and was defined as a section on the ECG with an R-wave to R-wave interval greater or equal to 400ms. A few sinus pauses were observed in the 24-hour dataset, but with no significant differences between NKXUSΔ[-/-] and NKXUSΔ[+/+] mice (Figure 42). Dividing the dataset into light and

dark phase of the day also did not reveal any differences (Figure 42), indicating that no spontaneous arrhythmias occurred.
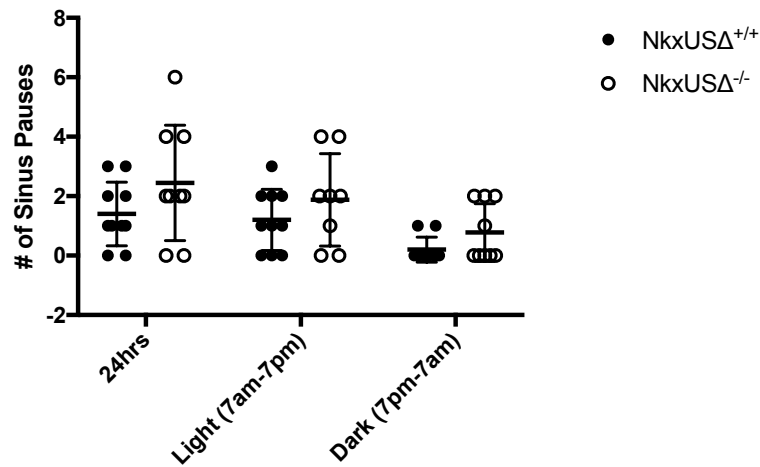
Figure 42 Analysing the telemetry data for the occurrences of sinus pauses revealed no significant differences between the groups when looking at the whole dataset nor when dividing it into day and night time. Data presented as mean ± SD, n=9-10.

2.8.8.4 Circadian rhythm and heart rate variability are not changed in NKXUSΔ[-/-] mice

An increased heart rate could be due to a disturbed circadian rhythm or/and heart rate fluctuation, alternating between tachycardia and bradycardia. Therefore, the telemetry data was divided into one-hour sections and plotted on a graph. On average, mice of both genotypes did not show differences in heart rate variability and showed a similar 24-hour cycle pattern (Figure 43), indicating that the increased resting heart rate is not due to a disturbed circadian rhythm nor heart rate fluctuations.
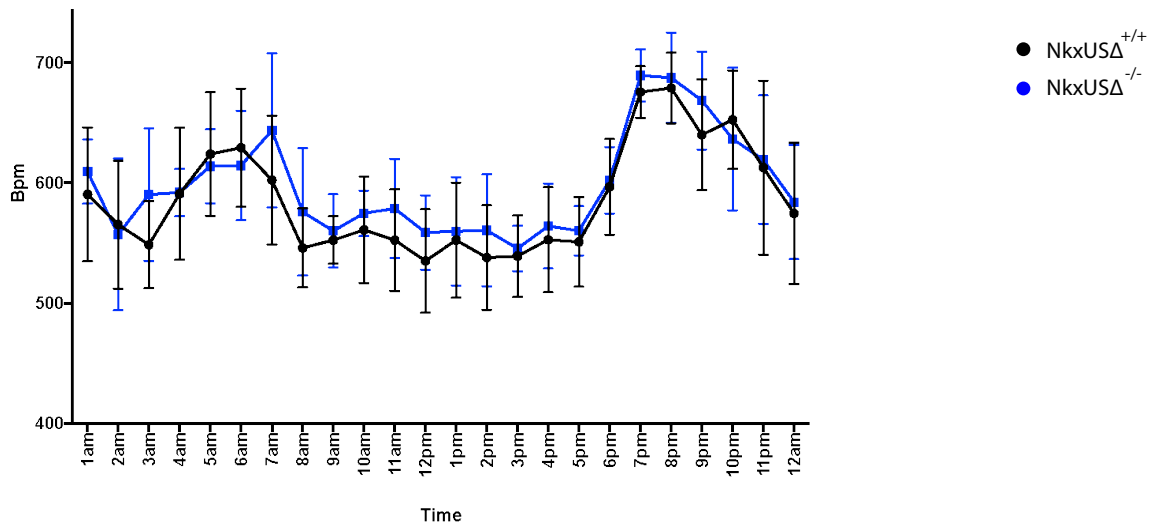
Figure 43 Heart rates of NKXUSΔ$^{-/-}$ and NKXUSΔ$^{+/+}$ animals showed a similar pattern over 24hrs. Plotting of the average heart rate per hour did not show any signs of heart rate variability in NKXUSΔ$^{-/-}$ mice compared to NKXUSΔ$^{+/+}$ mice. NKXUSΔ$^{-/-}$ mice showed the same heart rate pattern as NKXUSΔ$^{+/+}$ mice. Data presented as mean ± SD, n=9-10.

2.8.8.5 Increased resting heart is not due to higher activity or elevated body temperature

Higher heart rates in the resting period of the day could result from NKXUSΔ$^{-/-}$ mice being more active. Looking at the activity data recorded by the telemetry device, did not, however, show any significant differences in mouse activity, nor body temperature, during day or night time (Figure 44).
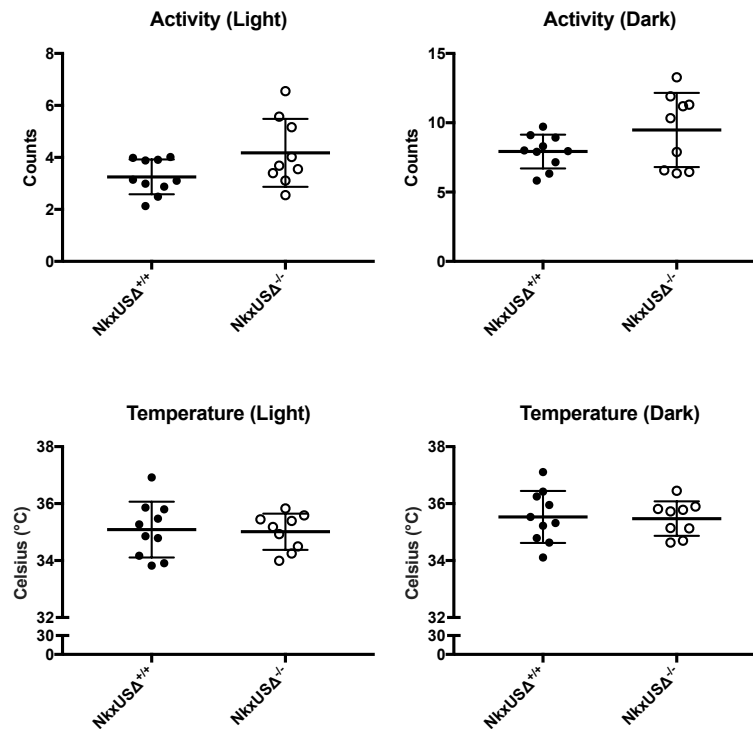
Figure 44 Activity and body temperature are not changed. Mouse activity and body temperature were not significantly changed during the day or night time between NKXUSΔ[-/-] and NKXUSΔ[+/+] mice. Data presented as mean ± SD, n=9-10.

In summary, no difference in any of the ECG parameter intervals was detected, indicating that the propagation of the electrical signal from the sinoatrial node throughout the heart is not altered. Heart rates of NKXUSΔ[-/-] mice did not seem to vary considerably more than NKXUSΔ[+/+] mice over 24hrs, indicating that the averaged increased heart rates are not due to alternating high and low heart rates. Furthermore, circadian rhythm did not seem to be affected, as heart rates were lower during the day and higher during the night time, as was expected. Neither higher activity nor higher body temperature could explain the increased resting heart rate. Therefore, there are two possible causes. An altered heart rate can reflect a disturbed function of the sinoatrial node (SAN) or the autonomic nervous system. Within the autonomic nervous system, the sympathetic nervous system activates the "fight or flight" response, thus increases the heart rate. However, since the sympathetic is less active while the mice are at rest, we suspect that the heart rate difference is due to a dysfunction in the SAN.

2.8.9 Pacemaker activity analysis

The telemetry study identified that NKXUSΔ$^{-/-}$ mice have an increased resting heart rate compared to NKXUSΔ$^{+/+}$ mice, which might be due to an increased firing rate of the SAN. The SAN is the pacemaker of the heart, where the heart beat is initiated and coordinated. To analyse for differences in SAN beating rate, we dissected the SAN regions from NKXUSΔ$^{+/+}$ and NKXUSΔ$^{-/-}$ mice for calcium imaging (Figure 46A).

2.8.9.1 SAN's of NKXUSΔ$^{-/-}$ mice exhibit a higher beating frequency

Calcium imaging was performed on dissected SAN preparations of male NKXUSΔ$^{-/-}$ and NKXUSΔ$^{+/+}$ mice, aged 10-12 weeks old. The SANs were dissected with the right atrium and part of the interatrial septum as well as inferior and superior vena cava still attached (Figure 45A). To detect intracellular calcium mobilisation, a calcium sensitive dye, Cal-520 (AAT Biosystem), was added to the tissue preparation. Cal-520 can cross the cell membrane and, upon binding to calcium inside the cell, its fluorescence is greatly increased (Figure 45A). The tissue samples were imaged while perfused with Tyrode solution at 37°C.

Calcium imaging of SAN preparation at 37°C revealed a significant increased beating rate in NKXUSΔ$^{-/-}$ (388 ± 10.4bpm) compared to NKXUSΔ$^{+/+}$ SANs (340 ± 14.8bpm) (Figure 45B). To analyse for any differences in the response to heart rate lowering and increasing drugs, two drugs were administered. First, we applied Ivabradine (IVA), which specifically blocks the pacemaker HCN channels, thereby decreasing the heart rate. Second, we applied Isoproterenol (ISO), which is a β-adrenoreceptor agonists that functions by binding to β-adrenoceptors, thereby increasing the heart rate by mimicking the actions of sympathetic adrenergic stimulation. Perfusion with 3µM IVA showed the expected decrease in heart rate; however, no significant differences were observed in the response to the drug (Figure 45B and C). The heart rate decreased in NKXUSΔ$^{+/+}$ SANs by ~15% and NKXUSΔ$^{-/-}$ SANs by ~20% (Figure 45C). Upon β-adrenergic stimulation (1µM ISO) the heart rate increased to ~490bpm in both, NKXUSΔ$^{+/+}$ and NKXUSΔ$^{-/-}$, preparations (Figure 45B) with a significant difference in the mean change before and after ISO treatment (Figure 45D). NKXUSΔ$^{+/+}$ SANs showed a greater response to ISO with a heart rate increase by ~45% whereas NKXUSΔ$^{-/-}$ SANs were less responsive (~25% increase) (Figure 45D).

Furthermore, to identify any disturbances in impulse formation, the tissues were paced for 20s at 50V to measure the sinoatrial node recovery time (SANRT). The SANRT is defined as the interval between the last paced beat and the first spontaneous one. The SANRT was corrected (cSANRT) for the cycle length measured immediately prior to pacing. The cSANRT was slightly prolonged in NKXUSΔ$^{+/+}$ preparations, however did not reach significance (Figure 45E), indicating that the recovery of normal beat pattern is not altered.

In summary, we identified an increased beating rate in NKXUSΔ$^{-/-}$ SAN preparations, which showed a normal response to IVA and normal impulse formation, however are less responsive to ISO treatment.
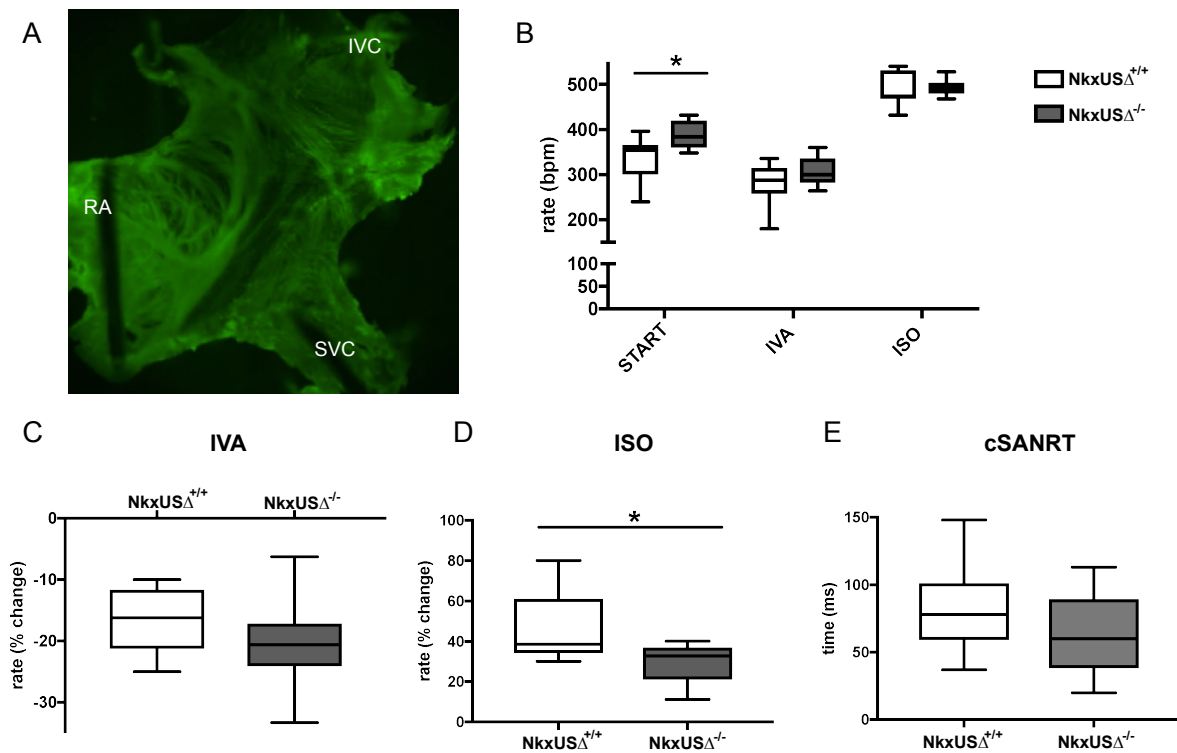
Figure 45 Increased beating rate in isolated SAN preparation of NKXUSΔ$^{-/-}$ mice. A) SAN preparation during calcium imaging. RA= right atrium; IVC= inferior vena cava; SVC= superior vena cava. B) Beating rate of NKXUSΔ$^{-/-}$ SAN preparations was significantly increased under control conditions (start) but not after perfusion with IVA (3μM) and ISO (1μM). C-D) The change in heart rate before and after drug treatment showed no difference for IVA (C), but a significant change in the response to ISO (D). E) The cSANRT is slightly changed but not significant. Data presented as mean ± SD, n=9-10, *P<0.05, unpaired two-tailed *t*-test.

3 Discussion

Cardiovascular disease is a global problem and especially affects the elderly population (<65 years) [252, 253]. Worldwide the average life span is increasing leading to an increased incidence of CVD, which will place a major burden on health care systems [252]. Despite decades of research to better understand the mechanisms driving CVD, the molecular and genetic basis of CVD development remains largely unknown. Heart development and function are regulated by a network of transcription factors and disruption of this network has catastrophic outcomes for the cardiovascular system. Though many of these transcriptional regulators are known and well studied, the complete picture remains unresolved.

The recent discovery of lncRNAs as a new class or regulators reveals an additional layer of regulation of gene regulatory networks. LncRNAs play major roles in many biological processes, including development, cell cycle regulation, response to stress and regulation of transcription, and are associated with numerous human diseases [101, 254, 255]. To date, thousands of lncRNAs have been identified, but their precise function *in vivo* has only been determined for a small fraction, leaving the *in vivo* function for most lncRNAs less understood [255]. So far, most studies have focused on global expression or potential function of lncRNAs, by analysing healthy versus disease tissue or comparison of different tissues or developmental stages, and only a limited amount of lncRNAs have been investigated in detail. While a number of lncRNAs have been identified to play central roles in cardiac development and disease, still thousands remain uncharacterised [101, 151, 152]. Studies have shown that at least 100 are specifically expressed in the heart, highlighting the abundant opportunities for novel discoveries of yet uncharacterised cardiac lncRNAs [156-159]. Interestingly, GWAS studies have identified that a large proportion (>90%) of trait- and disease associated SNPs lie within non-coding regions of the genome, of which ~85% lie within haploblocks expressing non-coding transcripts, raising the possibility that the GWAS phenotype could be due to disruption of non-coding transcripts [161, 163, 256, 257]. Indeed, many lncRNAs have been identified to contain genetic risk factors and have already contributed to a better understanding of the genetics of human diseases [108, 160]. Thus, functional characterisation of novel non-coding transcripts harboring disease associated SNPs remains a crucial goal to understand how these transcripts contribute to disease and will thus lead to advancements in precision medicine and human health.

Regarding the heart, NKX2-5 is one of the key regulators of heart development and thus essential for survival of mouse embryos [258, 259]. Numerous mutations within *NKX2-5* have been associated with congenital heart disease [258, 260]. However, even though *Nkx2-5* has been the

focus of many studies in the last two decades, the exact molecular mechanism of its function and regulation remains still largely unknown. By scanning the transcriptional landscape of the *Nkx2-5* locus, we detected previously unrecognised cardiac RNA expression downstream and upstream of *Nkx2-5*. For this reason, the potential involvement of these transcripts (*NkxDS* and *NkxUS*) in the regulation of *Nkx2-5* or other genes important for heart development and/or function was investigated.

This discussion is divided into two parts. The first part will describe the identification and characterisation of *NkxDS* (downstream transcript) and its relationship to *Nkx2-5.* The second part will first describe the localisation and expression of *NkxUS* (upstream transcript), followed by the functional characterisation of *NkxUS in vitro* as well as *in vivo*.

3.1 Discovery of a novel isoform of *Nkx2-5*

Detailed structural analysis showed that mouse *NkxDS* is a long, polyadenlyated transcript that can be continuous with and contains the coding region of *Nkx2-5*, suggesting that *NkxDS* represents a novel transcript isoform of *Nkx2-5* generated by differential termination and polyadenylation (Figure 8). A similar transcript was identified in humans (Figure 15), although continuity has not been confirmed yet. Interestingly, *Nkx2-5* and *NkxDS* shared a similar expression pattern in mouse tissues being cardiac enriched, more precisely in the myocardium, but also showing expression, although at much lower levels, in adult spleen and tongue, which has been previously described for *Nkx2-5* (Figure 9) [187]. Moreover, *NkxDS* followed the expression dynamics of *Nkx2-5* in mouse and human differentiating cardiomyocytes (Figure 10 and 18), although *NkxDS* was more lowly expressed, suggesting that both transcripts are co-regulated and may function at similar timepoints. As many lncRNAs were identified to be cell-type specific, it would be of interest to investigate whether *NkxDS*, as for *Nkx2-5,* is also expressed in other cardiac cell-types, or if it is cardiomyocyte-specific [261]. In contrast, even though *NkxDS* contains the *Nkx2-5* protein-coding region, its cellular localisation and polysome association was quite different. *NkxDS* was detected almost exclusively in the nucleus whereas *Nkx2-5* was enriched in the cytoplasm (Figure 12), as expected for an mRNA. Consistent with its subcellular localisation, *NkxDS* showed no or minimal association with polysomes (Figure 13), whereas *Nkx2-5* was clearly polysome-associated as expected for a well translated mRNA. This suggests that, when the extended 3'UTR *NkxDS* tail is

attached, the transcript is retained in the nucleus and thus may not be accessible for translation. In fact, several recent studies have identified genes producing protein-coding and non-coding isoforms [262-265]. These bifunctional RNAs were found to be generated by various post-transcriptional mechanisms including alternative splicing, intron retention, RNA modifications as well as, similar to our case, differential polyadenylation [262, 263]. Their coding and non-coding isoforms were shown to have distinct functions and can be separated either temporally with expression and function at different developmental stages, or spatially in different cellular or subcellular compartments [263]. Since *NkxDS* and *Nkx2-5* showed a similar expression pattern, but different cellular localisations, they could have distinct functions within these different compartments. However, since NKX2-5 is as a transcription factor and thus expressed in the nucleus, NKX2-5 and *NkxDS* may be involved in similar nuclear processes. The mechanism of how differential polyadenylation is regulated and how *NkxDS* is retained in the nucleus remains unknown and would be of great interest for future studies. Elements residing within the long *NkxDS* tail could determine non-coding versus coding or nuclear versus cytoplasmic localisation as was identified for several bifunctional RNAs [264, 266]. In several cases, differential polyadenylation was shown to be regulated by proteins. In the case of *NEAT1*, HNRNPK was identified to inhibit polyadenylation of the shorter transcript, *NEAT1_1*, which then resulted in the production of the long isoform, *NEAT1_2* [267]. Using online prediction tools, a few HNRNPK sites have been found around the *Nkx2-5* polyadenylation site (data not shown); however, association of HNRNPK to any of these sites has yet to be investigated.

Analysis of the known bifunctional transcripts revealed diverse functions as observed for lncRNAs [262, 263, 265, 268]. *NkxDS* and NKX2-5 could act antagonistically to regulate each other, crosstalk, or the long non-coding *NkxDS* tail could harbor miRNA or protein binding sites and thereby act as a miRNA sponge or competing RNA. It is noteworthy that previous *Nkx2-5* knockout mouse models targeted the coding region of *Nkx2-5*, thus would have also disrupted *NkxDS* [74, 75, 269]. To determine the specific functions of *Nkx2-5* and *NkxDS*, systematic analyses of mutants in which specifically *NkxDS* is disrupted versus those that affect the protein-coding capacity of *Nkx2-5* have to be performed. One *Nkx2-5* mouse model exists, which expresses nuclear-localising Cre Recombinase protein from the 3'UTR of *Nkx2-5*, thus does not affect the *Nkx2-5* coding region [71]. In contrast to *Nkx2-5* null mutations, these mice are viable as homozygous, but interestingly express only ~50% of the normal NKX2-5 protein levels [71]. However, since the insertion might

have altered miRNA or protein binding sites within the 3'UTR of *Nkx2-5*, this mouse model is not suitable to determine a specific *NkxDS* function [71].

In conclusion, the essential cardiac transcription factor *Nkx2-5* may have a coding and non-coding function and analysis of the specific function of *NkxDS* has the potential to reveal a new regulatory mechanism of *Nkx2-5* and/or of heart development.

3.2 *NkxUS* is a cardiac specific lncRNA involved in the regulation of cardiac conduction

3.2.1 *NkxUS* is retained in the nucleus

This work identified *NkxUS* as a novel, polyadenylated lncRNA (Figure 8). *NkxUS* is expressed in cardiac tissues throughout mouse development with nuclear-enriched expression but no specific sub-nuclear localisation (Figure 49D[4] and 12), suggesting that it functions within the nucleus but not in a particular sub-nuclear organelle. Nuclear expression was observed for the majority of lncRNAs, which were shown to be involved in chromatin organisation and regulation of transcriptional and post-transcriptional gene expression [101, 112, 270]. Consistent with its exclusive nuclear expression, no or minimal association with polysomes and no coding potential was identified (Figure 13 and 14), suggesting that *NkxUS* is truly non-coding. However, there is still the possibility that *NkxUS* encodes a micropeptide, although analysis using PhyloCSF makes this possibility unlikely (Figure 14) and was not further investigated.

Taken together, nuclear with no sub-nuclear expression of *NkxUS* narrows down the list of the potential functions and suggests that *NkxUS* may play a role in epigenetic, transcriptional and/or posttranscriptional control or chromatin remodeling.

3.2.2 Mouse *NkxUS* is tissue specific in adult mice

Expression analysis revealed that *NkxUS* is in adults specifically expressed in cardiac tissues (Figure 9A). *In situ* hybridisation showed that early in development *NkxUS* is also expressed in lung and neural tube (Figure 9G), however no expression in these tissues was observed in later stages, and the function of *NkxUS* in these tissues was not further explored. During differentiation of mouse

embryonic stem cells into embryoid bodies enriched for cardiomyocytes, *NkxUS* was lowly expressed at day 0 to 4 (Figure 10A), suggesting that it does not play a role in maintaining pluripotency or early differentiation of germ layers. In the time course analysis, *Nkx2-5* and *NkxUS* showed a similar dynamic expression pattern (Figure 10A), suggesting that they may function at similar timepoints and/or through related pathways. The expression was highest at day 7 during cardiomyocyte enriched differentiation (Figure 10A), which approximately resembles the cardiac crescent stage when cardiac mesodermal tissues start to develop, suggesting that *NkxUS* may be important for cardiomyocyte differentiation, similar to what was shown for *Nkx2-5* [180, 271, 272]. However, *NkxUS* was much more lowly expressed compared to *Nkx2-5* (Figure 10B).

Interestingly, the *in situ* hybridisation showed some distinct differences between both transcripts. *NkxUS* showed a predominately endocardial expression whereas *Nkx2-5* was expressed mostly in the myocardium (Figure 9D-H), although *Nkx2-5* is also known to be expressed in the endocardium of the forming heart tube [71, 261]. Nevertheless, low-level *NkxUS* expression was also observed in the myocardium early in development (Figure 9D) and also in isolated and differentiating cardiomyocytes (Figure 9B, 10 and 11). However, *NkxUS* expression in endocardial cells may be higher, which remains to be analysed, since no endocardial cell line was available. Moreover, *NkxUS* showed a strong *in situ* expression in the valve leaflets before and after birth, whereas no expression was observed for *Nkx2-5* nor *NkxDS* in this tissue before birth (Figure 9F, 49E[3] and F[3]). The endocardium is crucial for valve development and since *NkxUS* is predominately expressed in endocardial cells, *NkxUS* could be important for the function or development of the valves, which could be analysed using a full *NkxUS* knockout mouse [273]. Furthermore, although the mESC differentiation time course was enriched for cardiomyocytes, there was also differentiation of other cell-types such as endothelial cells. Therefore, the much lower expression of *NkxUS* relative to *Nkx2-5* could also mean that *NkxUS* is not critical for cardiomyocyte differentiation but rather for the differentiation of the minor cell types, most likely endothelial cells, since *NkxUS* is at later stages predominately expressed in this cell type.

Taken together, *NkxUS* and *Nkx2-5* share many characteristics, both are predominately cardiac expressed, expressed in cardiomyocytes and at the same timepoints during mESC differentiation and mouse heart development, suggesting co-regulation and a potential role for *NkxUS* in pathways and time points related to those involving *Nkx2-5.*

3.2.3 Human *NkxUS* is associated with an increased heart rate

Evolutionary conservation is considered a powerful resource to identify functionally important genomic regions. Despite the lack of sequence conservation of most lncRNAs, many to date have been identified or predicted to contain conserved RNA secondary structures, which are currently used to propose functional relevance and interrogated to understand their broader roles [128-130]. Structural studies of the lncRNA *Braveheart* identified a short G-rich secondary structure, which upon genetic perturbation prevented cardiomyocyte differentiation [274].

For *NkxUS*, a similar heart specific transcript was detected in humans (Figure 15), although the relative proximity to the transcriptional start site of *Nkx2-5* and orientation was different. Expression of *NkxUS* was confirmed in human cardiac tissues (Figure 17) and showed similar expression dynamics to *Nkx2-5* in pluripotent stem cell-derived cardiac tissue (Figure 18). This supports functional conservation of *NkxUS* between mouse and human and further suggests that, as in mouse, human *NkxUS* is not involved in maintaining pluripotency and may function at time points and processes similar to those of *Nkx2-5*. Sequence conservation between species is low, consistent with the idea that lncRNAs evolved only recently. However, three conserved RNA structures were identified within human *NkxUS* which map to the start of the second exon of mouse *NkxUS* (Figure 16), implying that these might be important for the biological function of *NkxUS*. Intriguingly, a GWAS SNP associated with an increased resting heart rate and increased risk for atrial fibrillation lies within and is predicted to totally disrupt one of these conserved structures of human *NkxUS* (Figure 20 and 21), supporting functional relevance [217]. The magnitude of the SNP effect was shown to be small (0.301bpm), however, in combination with other heart rate increasing SNPs the effect size reached clinical relevance with an increase in resting heart rate of up to 4bpm in adults [217]. Moreover, the GWAS SNP was part of a common haplotype and it is possible that the *NKXUS* SNP is not the causal variant. Haplotype analysis revealed five SNPs in high LD with the *NKXUS* SNP (Figure 22); three of them map to the first exon of *NKXUS* not overlapping any of the conserved structures, one maps to an intergenic region and one is a synonymous mutation within *NKX2-5*. While NKX2-5 protein is likely not affected by this SNP, the SNP could still affect binding sites for interacting partners or splicing of the mRNA, which we did not test [226]. However, since the *NKXUS* SNP lies within the conserved structure of a cardiac specific transcript and is predicted to alter this structure, we proposed that the *NKXUS* SNP is the causal variant and further explored the potential function of *NkxUS*.

3.2.4 *NkxUS* knockdown in HL-1 cells affects genes involved in cardiac contraction

LncRNAs employ a variety of gene regulatory mechanisms [112, 139, 142]. They are able to directly bind to target genes, act as decoys, guides, signals or as scaffolds for transcription factors and histone modifiers to activate or inhibit the expression of target genes [112, 142]. To explore the function of *NkxUS*, we intended to first identify the biological processes and pathways *NkxUS* is influencing

Downregulation of *NkxUS* in HL-1 cells resulted in predominately an upregulation of transcripts (Figure 23), suggesting that *NkxUS* has an inhibitory role either directly for example by associating with histone modifying proteins or indirectly for example through the act of its transcription or genome organisation, as shown for many lncRNAs [104, 144, 145]. Of note, *Nkx2-5* mRNA and *NkxDS* were not altered upon *NkxUS* knockdown, indicating that *NkxUS* does not affect their gene expression. However, whether *Nkx2-5* is altered at the protein level following knockdown has not been investigated.

GO terms associated with the upregulated transcripts were enriched for *regulation of transport, signaling* and *cell communicatio*n, as well as *cardiac conduction* and *contraction* associated terms (Figure 24). Since the GWAS SNP in *NKXUS* is associated with an altered heart rate and since NKXUSΔ$^{-/-}$ mice exhibit an increased resting heart rate (Figure 40), we focused on the cardiac contraction and conduction associated terms (Figure 24). The transcripts within these terms were also included in the most significant terms, *regulation of transport*, *signaling* and *cell communication*. Most of these transcripts were ion channels, regulators of ion channels as well as genes involved in cAMP production (Figure 24 and 46).

cAMP is an important secondary messenger, which in the heart triggers the activation of ion channels and is therefore critical for cardiac contraction. Stimulation of cardiomyocyte G protein-coupled β-adrenergic receptors (β-AR) by epinephrine or norepinephrine, leads to the activation of an associated stimulating G protein (G$_s$), which in turn activates adenylate cyclases to form cAMP from ATP (Figure 46). The resulting cAMP interacts with and regulates cyclic nucleotide-gated ion channels, such as HCN channels, as well as protein kinase A (PKA) (Figure 46). PKA enhances contractility and calcium transients by phosphorylating a range of proteins, including the L- and T-Type calcium channels, HCN channels, G protein-coupled inwardly-rectifying potassium channels (GIRKs), the ryanodine receptor (RYR) and phospholamban (PLB) (Figure 46). Upon phosphorylation, PLB detaches from SERCA, which is then activated. SERCA is an ATPase which, at the cost of ATP,

transfers calcium back into the SR during diastole. Additionally, parasympathetic triggered release of acetylcholine activates cholinergic receptors, which are also GPCRs and can either activate an inhibitory G protein (G_i) subunit, which inhibits adenylate cyclases, or an activating β-G protein ($G_β$) subunit which activates GIRKs (Figure 46) [26]. Parasympathetic activation can be modulated by RGS6, which inhibits $G_β$ as well as G_i (Figure 46) [245, 275]. Interestingly, GIRKS are also activated upon sympathetic induced β-adrenergic stimulation through PKA (Figure 46) [276].

All of these proteins mentioned and illustrated in Figure 46 (except β-adrenergic receptors, PKA and T-Type calcium channels) are upregulated at the mRNA level in our dataset, indicating that *NkxUS* may play a role in regulating gene expression of cardiac cAMP signaling components. Interestingly, *Rgs6* was also identified in the DNA ChIRP (Table 7), indicating that *NkxUS* may under normal circumstances bind to *Rgs6* to negatively influence its transcription. However, *Rgs6* was only identified in one ChIRP replicate and chromatin association has not been verified yet.
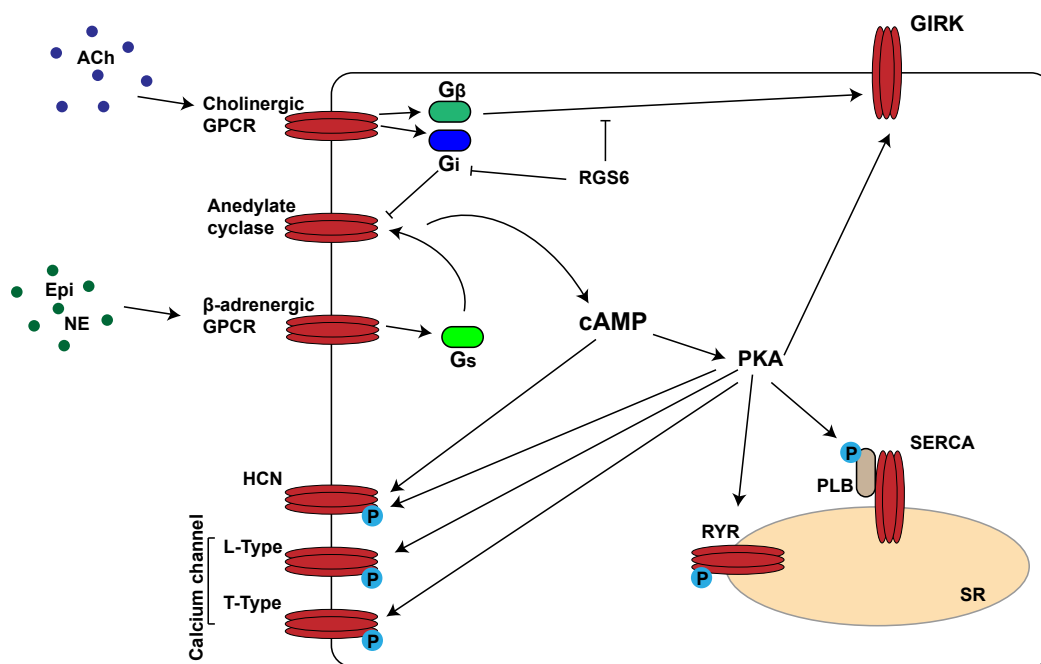
Figure 46. Several upregulated genes are involved in cardiac cAMP signaling. Schematic of components involved in regulation of cardiac contraction via cAMP signaling. Ach= Acetylcholine; Epi= Epinephrine; NE= Norepinephrine; GPCR= G protein-coupled receptor; G_i= inhibitory G protein; G_a= activating G protein; $G_β$= activating β-G protein; PKA= Protein kinase A; Girk= G protein-coupled inwardly-rectifying potassium channels; PLB= Phospholamban; RYR= ryanodine receptor; SR= Sarcoplasmic reticulum; SERCA= Sarcoplasmic/endoplasmic reticulum calcium ATPase 2a; HCN= Hyperpolarisation-activated cyclic nucleotide–gated channel; RGS6= Regulator of G protein signaling 6.

3.2.5 *NkxUS* may bind to genes involved in the regulation of transcription and/or G protein-coupled receptor signaling

Many lncRNAs reside in the nucleus and target specific genomic loci directly to regulate gene expression [112, 139]. Analysis of DNA binding region of *NkxUS* using DNA ChIRP revealed a limited number of candidates (Figure 27). Using the EVEN and ODD strategy and comparing biological replicates, we did not see compelling enrichment across the two replicates and probe sets and only identified two targets (*Grm8* and *Ptprk*), which could have been due to several factors. It could either be that 1) *NkxUS* does not bind DNA, 2) we did not enrich for *NkxUS* efficiently or 3) *NkxUS* only binds at these few sites. However, 2) seems less likely since we observed peak enrichment at the *NkxUS* locus (9 out of the 11 shared peaks) (Figure 27), which is a positive control for the technical part of the experiment as it represents nascent transcription of *NkxUS*. In our experiment we retrieved 10-20% of *NkxUS* (Figure 27). However, higher *NkxUS* yields might improve the signal-to-noise ratio and thus potentially identify additional targets and/or result in an improved overlap between replicates if *NkxUS* indeed binds to chromatin. In comparison, the original ChIRP publication retrieved 88-95% of their lncRNAs of interest [239]. However, these were quite abundant lncRNAs and hence lead to more efficient pull-down. Nevertheless, with the peak enrichment detected at the *NkxUS* locus and the observed specific *NkxUS* yield (Figure 27), we can be confident that *NkxUS* was successfully captured. Therefore, it could either be that *NkxUS* does not bind DNA or only to these few locations.

Looking at the overlap between biological replicates revealed two genes, *Grm8* and *Ptprk* (Figure 27). We could have lost a lot of shared peaks due to differences in *NkxUS* yield between the two biological replicates, EVEN3/ODD3 retrieved more *NkxUS* RNA than EVEN2/ODD2 (Figure 27A) and four times more shared peaks, which could explain the limited number of overlap between the two biological replicates. Therefore, we thought it is reasonable to also investigate the 12 and 48 peaks shared between EVEN2/ODD2 and EVEN3/ODD3, respectively, and analyse whether these have a related function to *Grm8* and *Ptprk* (Figure 27B). When intersecting the individual experiments, several additional potential DNA targets with similar function to *Grm8* and *Ptprk* were identified (Table 6).

Grm8 is a G protein-coupled receptor for glutamate and its signaling is known to inhibit adenylate cyclase activity. Interestingly, altogether five genes detected by DNA ChIRP (*Grm8, Npr3, Rgs6, Kctd16* and *Olfr211*) were identified to be involved in G protein-coupled receptor signaling

and, all except *Rgs6*, were shown to inhibit adenylate cyclase activity, thus preventing the formation of cAMP [277, 278]. The observation that several of the ChIRP identified genes inhibit adenylate cyclases is intriguing and consistent with the RNA sequencing results, in which adenylate cyclases and its downstream targets such as calcium and HCN channels were upregulated. Potentially, *NkxUS* binding to these genomic loci under normal circumstances enhances gene expression and therefore inhibits adenylate cyclases, which are important to regulate cardiac contraction. Although, apart from *Rgs6*, for none of the other genes a cardiac role has been described to date, even though they are expressed in cardiac tissue. *Grm8* is known to function in the neurotransmission of glutamate and it is associated with several neuropathological conditions including schizophrenia and depression [279, 280]. Its function in the heart is unexplored. However, a GWAS study associated *Grm8* with PR interval, supporting a role in cardiac conduction, although the study did not indicated whether the PR interval increased or decreased [281]. Another recent study analysed brain tissues from patients with sudden unexpected death in epilepsy, which proposed that among others, cardiac dysfunction has a causal link to epilepsy [282]. The study identified several variants in genes associated with cardiac arrhythmia as well as *Grm8*, further supporting a cardiac role [282]. This is consistent with several ion channels, among other for example *Kcnh2*, which are associated with long QT syndrome and epilepsy as most ion channels are expressed in excitable cells of the heart and brain [283]. Additionally, as previously mentioned, *Rgs6* was also upregulated upon *NkxUS* knockdown in cardiac cells, making it a valuable candidate for follow up studies as its function in regulating cardiac contraction is well described (Figure 46) [245]. *Rgs6* is a negative regulator of parasympathetic regulation of heart rate and mice lacking RGS6 exhibit bradycardia, therefore an upregulation of *Rgs6* would result in an increased heart rate (Figure 46) [245, 275, 284].

The second gene identified through ChIRP, *Ptprk,* is involved in the *regulation of cell adhesion, cell cycle* and *transcription* as are *Cadm2, Cdc73, Ercc6, Opcml, Tnik* and *Sfi1*. The role of *Ptprk* in tumor suppression and in the nervous system is well studied, but although expressed in the heart, a specific cardiac function has not yet been determined [285, 286]. Mechanisms of tumor suppression involve regulation of β-catenin, which is a multifunctional protein playing major roles in cell-cell adhesion and Wnt signal transduction [285, 287]. Wnt/β-catenin signaling plays an important role during heart development and is activated in response to cardiac injury [288, 289]. Interestingly, a recent study identified active Wnt/β-catenin signaling in developing SAN cells and functional analysis revealed that this signaling regulates the parasympathetic control of pacemaker

activity and thereby heart rate [290]. However, it remains unclear if *Ptprk* plays a similar role in the heart to regulate β-catenin. Intriguingly, three additional ChIRP identified genes were shown to interact with β-catenin. *Cdc73*, which is a component of the RNA polymerase II associated protein complex (PAF), was shown to positively regulate Wnt signaling by directly binding to β-catenin upon dephosphorylation and to be important for cardiomyocyte specification and elongation of the heart tube [291-294]. The PAF complex has been shown to regulate a number of transcription-related processes including 3'end formation of polyadenylated and nonpolyadenylated RNA Polymerase II transcripts, histone modification and gene transcription [295, 296]. *Opcml* is another cell adhesion protein and tumor suppressor. Knockdown in mice resulted in cell cycle arrest and an upregulation of β-catenin [297]. Moreover, a GWAS study associated a mutation in *Opcml* with increased heart rate variability [298]. *Tnik* is a serine/threonine kinase that acts as an essential activator of the Wnt signaling pathway and was shown to directly interact with β-catenin [299]. It is recruited to target genes in a β-catenin dependent manner to activate gene expression [299]. Moreover, it was also slightly upregulated upon *NkxUS* knockdown (Table 7), indicating that *NkxUS* may negatively regulate *Tnik* gene expression. Although all of these genes are expressed in the heart, a cardiac function, except for *Cdc73*, has not been determined to date and studies mainly focused on their role in cancer and the nervous system.

Moreover, the downregulated transcripts, *Tbx3* and *Wt1*, which are involved in cardiac cell fate commitment are also associated with Wnt/β-catenin signaling (Figure 24). *Wt1* is a positive upstream regulator of the Wnt/β-catenin signaling pathway. Loss of *Wt1* resulted in a downregulation of *Ctnnb1* (β-catenin) and other components of the signaling pathway [300]. *Tbx3* is a transcription factor and plays a major role in cardiac development and the formation of the sinoatrial node [301]. *Tbx3* lies downstream of the β-catenin signaling pathway and was in cancer studies shown to be directly and positively transcriptionally regulated by β-catenin [302, 303].

In summary, since the DNA ChIRP identified several genes with similar functions to *Grm8* and *Ptprk* and to transcripts deregulated in the *NkxUS* knockdown experiment, *NkxUS* may indeed bind to those genomic locations and function in the negative regulation of adenylate cyclases or in Wnt/β-catenin signaling. However, similar to ChIP-sequencing results, not all binding sites are often functional and additional follow up experiments using independent ChIRP experiments followed by RT-PCRs are required to validate *NkxUS* binding to these regions.

3.2.6 *NkxUS* potentially associates with proteins that can localise to the nucleus and mitochondrion

Cardiac contraction and relaxation is a highly energy demanding process, relying on the production and consumption of an immense amount of ATP every day, which matches roughly our body weight [304]. ~90% of ATP is produced by mitochondrial oxidative phosphorylation in the inner mitochondrial membrane, thus mitochondria play an essential role in controlling cardiac contraction [305]. To meet the hearts energy requirement for contraction and relaxation, cardiomyocytes have a high mitochondrial content and rely heavily upon its energy supply. In cardiomyocytes, ATP is needed to release actin from myosin during cardiac contraction, to restore SR calcium load via the sarcoplasmic/endoplasmic reticulum calcium ATPase (SERCA) during relaxation and for the synthesis of cAMP via adenylate cyclases [305]. Most mitochondrial proteins are encoded in the nucleus, thus communication between these compartments is critical in order for the nucleus to respond to metabolic changes within the mitochondria [306].

ChIRP followed by MS to identify protein interacting partners of *NkxUS* resulted in enrichment of numerous mitochondrial proteins (Figure 31). Based on the nuclear localisation of *NkxUS* and higher mitochondrial content in cardiomyocytes, we were unsure about the accuracy and specificity of the results. However, recent studies discovered that the nucleus harbors several mitochondrial proteins, with some involved in the direct mitochondria-to-nucleus communication, indicating that these might directly link mitochondrial metabolism to gene expression [307-310]. These mitochondrial proteins were found to exert their canonical functions within the nucleus as well as nuclear specific functions including regulation of transcription, DNA repair and kinase activity [309]. Based on the nuclear localisation of *NkxUS* we first focused on proteins with an annotated nuclear localisation. Several mitochondrial proteins identified in our work are indeed also described to localise to the nucleus. Among the most enriched ChIRP detected mitochondrial and nuclear proteins is the pyruvate dehydrogenase complex (PDC) (Table 9), which converts pyruvate into acetyl-CoA that contributes to oxidative phosphorylation and consists of three enzymes [311]. Two subunits of the first enzyme, Pyruvate dehydrogenase (PDHA1 and PDHB), as well as the second enzyme, Dihydrolipoyl transacetylase (DLAT) (Table 9), were captured following *NkxUS* pull-down. Moreover, DLAT was one of the two most enriched proteins identified. PDC was shown to translocate from the mitochondria to the nucleus in response to growth signals and disrupted oxidative phosphorylation [311]. Within the nucleus it exerts its normal function by

producing acetyl-CoA, which can normally not cross the mitochondrial membrane and is necessary as a cofactor for histone acylation [311]. Histone acylation alters chromatin accessibility and subsequently regulates gene transcription, DNA replication and DNA repair [312]. Another protein identified in our study, fumarate hydratase (FH) (Table 9 and 10), has emerged as a mitochondrial and nuclear localised protein [313, 314]. Within the mitochondrion it participates in the tricarboxylic acid (TCA) cycle and within the nucleus in DNA repair by inhibiting histone demethylation [313, 315]. Interestingly, it was shown that two groups of *Fh* mRNAs are generated by the same gene, one group containing the mitochondrial targeting sequence, while the other lacks it and localises to the cytoplasm [313]. Upon DNA damage, FH translocates to the nucleus, produces fumarate, which in turn is required for the DNA damage response [313]. Additional examples of nuclear localised proteins found in this work are listed in Table 11.

Altogether, it becomes evident that numerous mitochondrial enzymes also reside in the nucleus and potential nuclear roles for mitochondrial proteins have to be considered in the future until shown that they exclusively localise to mitochondria [307]. How many more of the proteins identified in this work localise to the nucleus or may represent noise or indirect interacting partners is difficult to determine and requires follow up studies. No other ChIRP study to date has used a muscle cell line, which would have been helpful to compare whether they also identified many mitochondrial proteins.

Of the identified proteins with to date no annotated nuclear localisation, many are involved in oxidative phosphorylation, which fuels the synthesis of ATP (Figure 32). The protein identified within one of the visible bands in the gel, HADHA, is part of the fatty acid metabolism pathway which leads to the generation of acetyl-CoA, which is then oxidised for energy production via oxidative phosphorylation. ATP is essential for cardiac contraction and needed for the basal beating of the SAN in the absence of β-adrenergic receptor stimulation. The more ATP available, the more can be used for actin-myosin contraction, by adenylate cyclases to produce cAMP and by SERCA to pump calcium back into the sarcoplasmic reticulum to prime the cell for the next depolarisation. ATP consumption within the SAN is tightly linked to ATP production within mitochondria. It was shown that reducing ATP consumption through inhibition of adenylate cyclases in the SAN resulted in a reduced mitochondrial ATP synthesis [316].

Furthermore, ChIRP-MS would benefit from additional optimisation especially in the crosslink/reverse-crosslink and washing conditions. We discovered that the reverse-crosslinking was still incomplete after several optimisation attempts. We might achieve a better overlap

between replicates once this step is optimised. Additionally, ChIRP might benefit from more stringent washing conditions to enable the efficient removal of unspecific bound proteins. A recent paper established a new method to identify RNA-Protein interactions by targeting the RNA using antisense probes consisting of a mix of LNA and DNA nucleotides [317]. The incorporation of LNA nucleotides increases the melting temperature of the oligonucleotides allowing more stringent washing condition. If the mitochondrial proteins non-specifically bind to the probes or beads, they might be removed using these more stringent washing steps.

In summary, even though we did not fully optimise the ChIRP method and did not perform biological replicates, we identified several proteins over several experiments which exert a nuclear function. Based on the nuclear function of these proteins *NkxUS* might play a role in regulating transcription and/or histone modification to control cell cycle progression, cell proliferation or gene expression of nuclear genes and potentially mitochondrial genes in response to mitochondrial stress or higher metabolic demand (Table 11). Additionally, if the DNA and protein ChIRP targets are indeed specific, *NkxUS* could recruit some these proteins to the DNA targets to regulate transcription. However, validation of interaction has not been performed and could be verified using immunoprecipitation of the protein of interest followed by RT-PCR to check for *NkxUS* expression.

Table 11 Nuclear function of proteins enriched in *NkxUS* ChIRP

Protein	Description	Nuclear function	Reference
PDC	Pyruvate dehydrogenase complex	acetyl-CoA for histone acetylation	[307, 311]
OXCT1	Succinyl-CoA:3-ketoacid coenzyme A transferase 1	acetyl-CoA for histone acetylation	[318]
LDH	L-lactate dehydrogenase	lacatate and NAD+ for histone acetylation; DNA replication/repair; Transcription	[309, 319, 320]
PCBP1	Poly(rC)-binding protein 1	regulation of transcription and splicing	[321]

FH	Fumarate hydratase	DNA replication/repair; fumarate for inhibition of histone/DNA demethylation	[313-315]
HNRNPAB	Heterogeneous nuclear ribonucleoprotein A/B	mRNA processing and transport	[322]
MYH6	Myosin 6	enhanced RNA Polymerase II dependent transcription	[323]
ACO2	Aconitase 2	unknown	[324]
CRAT	Carnitine O-acetyltransferase	acetyl-CoA for histone acetylation	[325]
PGK	Phosphoglycerate kinase	DNA replication/repair	[326]
PGM	Phosphoglycerate mutase	unknown	[327]
SDH	Succinate dehydrogenase	fumarate for inhibition of Histone/DNA demethylation	[309, 328]

3.2.7 The conserved RNA structure is not essential for development

As conservation is assumed to indicate function, we interrogated using a mouse model the potential function of the conserved structure. Deletion of part of the conserved structure including the SNP homolog did not affect the transcript levels of *NkxUS* nor *NkxUSas* in isolated hearts (Figure 34), indicating that it is not essential for the stability or production of these RNAs. Moreover, *Nkx2-5* is not altered on RNA or protein level in NKXUSΔ$^{-/-}$ hearts (Figure 34 and 35), indicating that this structure is not involved in the regulation of *Nkx2-5* expression. This is consistent with the knockdown experiment, in which *Nkx2-5* was also not deregulated. However, it is still possible that *NkxUS* acts downstream of *Nkx2-5*, which has not been tested yet.

Since heart morphology was normal in NKXUSΔ$^{-/-}$ embryos and NKXUSΔ$^{-/-}$ adult mice did not show any defects in atrial septal development (Figure 36), which is the most prone structure for

developmental defects, we conclude that the conserved RNA structure is not crucial for heart development.

3.2.8 Left ventricular relaxation is impaired in NKXUSΔ$^{-/-}$ mice

To gain insights into the potential function of the conserved region in cardiac contraction, we challenged NKXUSΔ$^{-/-}$ and NKXUSΔ$^{+/+}$ mice using a model of LV hypertrophy. Under anaesthesia the heart rates showed no difference between the groups (Figure 38). Although, LV hypertrophy was successfully induced (Figure 37), no difference was observed between the genotypes regarding the response (Figure 37, 38 and 39). Both groups behaved similarly, showing an increase in left ventricular wall mass and systolic pressure and no significant difference in cardiac output, ejection fraction, stroke volume nor contractility (Figure 37 and 38). Although, NKXUSΔ$^{-/-}$ mice showed a slight worse (non-significant) cardiac output, ejection fraction and stroke volume, it would be interesting to study if these terms reach significance at later timepoints of the study when the mice proceed into heart failure.

However, a significant difference between the genotypes was observed in the flow velocity at the mitral valve during the passive filling phase of the LV in sham mice (Figure 37), indicating that NKXUSΔ$^{-/-}$ mice have a disturbed left ventricular relaxation, potentially due to a stiffer ventricle. This difference was not observed in TAC animals (Figure 37), as induced hypertrophy and thus growth of the LV is known to decrease ventricular compliance and LV diastolic filling, likely masking the effect seen in sham NKXUSΔ$^{-/-}$ mice [329].

In summary, under control condition NKXUSΔ$^{-/-}$ mice have a mild diastolic dysfunction due to altered LV relaxation, which could be due to a stiffer ventricle and has not been further investigated.

3.2.9 NKXUSΔ$^{-/-}$ mice exhibit an increased resting heart rate

As a GWAS SNP in human *NKXUS* was associated with an increased resting heart rate and the LV hypertrophy study only measured the heart rate in anaesthetised mice (Figure 38), we analysed the heart rates in conscious, freely moving mice.

Monitoring of the heart rate over 24hrs showed that NKXUSΔ$^{-/-}$ mice have an increased resting heart rate (Figure 40), as observed for the GWAS SNP in humans. Conduction of excitation throughout the heart was not affected as no difference in the PR-, QT- or QRS-intervals were observed (Figure 41). This is consistent with the RNA sequencing results of the *NkxUS* knockdown in HL-1 cells where gap junction transcripts were not deregulated. Gap junctions are essential for signal transmission as they connect the cytoplasm of neighbouring cells to provide cell-cell communication [11]. Thus, changes in gap junction composition lead to an altered transmission of the electrical excitation between neighbouring cells, which would be detected while interpreting the ECG data. Furthermore, the increased heart rate was not due to increased mouse activity (Figure 44), altered circadian rhythm or heart rate fluctuations (Figure 43), which are partly regulated by the autonomic nervous system. Since the heart rate difference was detected during the day when mice are at rest and since *NkxUS* expression was also detected in the SAN (Figure 9G), we investigated if the increase is due to an elevated pacemaker activity.

The analysis of spontaneous beating frequency and the corresponding calcium transients of dissected SAN preparation of NKXUSΔ$^{-/-}$ and NKXUSΔ$^{+/+}$ mice revealed a significant faster beating in NKXUSΔ$^{-/-}$ SANs (Figure 45), indicating that the phenotype observed *in vivo* is due to an increased intrinsic SAN firing rate. NKXUSΔ$^{+/+}$ SANs showed a similar beating frequency as reported by other studies using the same buffer temperature (Figure 45) [49, 330]. Beating rates were lower in dissected tissues compared to the *in vivo* results for both genotypes (Figure 40 and 45), consistent with previous studies and is most likely attributable to the absence of sympathetic tone as well as other signaling molecules such as peptides and hormones, which also modulate cardiac and neuronal activity [331]. The observed statistical difference between genotypes was higher in dissected SAN tissues compared to *in vivo* (Figure 40 and 45), potentially due to experimental factors, such as the physiological buffer system, perfusion rate and/or temperature.

Collectively, we identified a dysfunction in the SAN network. However, this impairment could be caused by enhanced intrinsic properties of single pacemaker cells. The next step would be to dissociate the SAN tissue into single cells and then analyse the firing rate using calcium imaging experiments. If single SAN cells of NKXUSΔ$^{-/-}$ mice still beat faster than NKXUSΔ$^{+/+}$ SAN cells, then the increased heart rate is due to a cell autonomous disturbance in single SAN cells, which can arise from heightened basal ion channel activity in the absence of β-stimulation. If the results show no difference, then the faster beating is due to enhanced properties of the SAN cellular network itself,

which also contains fibroblasts, fat and nervous cells, and which could result from an altered structure or cell composition of the network or enhanced conduction [15, 332].

To investigate the mechanism initiating spontaneous pacemaker action potentials, we challenged the SAN tissues with two drugs. Blocking of the HCN channels using IVA showed the expected decrease in beating rate (Figure 45). For both genotypes, the firing rate decreased by the same relative amount (Figure 45), indicating that the HCN channels are most likely not the source of the increased beating frequency in NKXUSΔ$^{-/-}$ mice. Changes in the overall rate of calcium transients after IVA perfusion was consistent with results of other studies (Figure 45) [49]. A concentration of 3µM IVA was chosen as it equals the half maximal inhibitory concentration (IC$_{50}$), hence the concentration which leads to about 50% inhibition of the pacemaker current. However, although we specifically blocked the HCN channels, pacemaker automaticity is complex and multiple pathways evolved to compensate for the loss of one component [29]. Here, we specifically inhibited HCN channels, which are part of the membrane clock, thus the calcium clock is not affected and remains sufficient to depolarise the cells. Therefore, we did not see a 50% decrease in beating rate. Inhibition using IVA was shown to be dose-dependent, however in this study we have not investigated whether higher IVA doses might lead to an altered response between the genotypes.

In contrast, the second drug, ISO, was used to activate the cAMP signaling pathway via activation of β-adrenoreceptors, which are essential regulators of SAN pacemaker activity. β-adrenergic stimulation with ISO lead to the expected strong increase in beating frequency (Figure 45). However, NKXUSΔ$^{-/-}$ SANs were less responsive to ISO treatment than NKXUSΔ$^{+/+}$ SANs (Figure 45). ISO treatment resulted in a ~25% increase in beating for NKXUSΔ$^{-/-}$ SANs and a ~45% increase for NKXUSΔ$^{+/+}$ SANs (Figure 45). The reduced responsiveness of NKXUSΔ$^{-/-}$ SANs to ISO may indicate that downstream targets of β-adrenoreceptor signaling are potentially already more active. Consistent with the results of the telemetry study for the active period of the day when sympathetic stimulation is elevated (Figure 40), NKXUSΔ$^{-/-}$ and NKXUSΔ$^{+/+}$ SANs reached similar heart rates after ISO treatment (Figure 45).

As mentioned previously, the mechanism of SAN automaticity is facilitated by two oscillators. Therefore, it would be of interest to investigate whether the increased heart rate is due to a dysfunction in the membrane or calcium clock, or potentially both. Transcripts of both clocks are upregulated in the knockdown experiment, suggesting that both clocks may be altered. However, whether the conserved structure affects both clocks is unknown. As previously mentioned *NkxUS* transcript levels were unchanged in NKXUSΔ$^{-/-}$ hearts, however since the

phenotype of NKXUSΔ$^{-/-}$ mice is consistent with the results of the knockdown of the entire *NkxUS* transcript, the conserved structure potentially may affect the main function of *NkxUS.*

To distinguish between the two clocks, an SAN experiment analysing the effect of ISO in presence of IVA could give further insight. Previously, we analysed ISO and IVA separately (Figure 45), therefore the effect we observed with ISO was solely from this drug. If the heart rate difference is due to the membrane clock, a similar result to the one we measured with ISO alone is expected (Figure 45). The idea is that if the membrane clock is already dysfunctional, blocking it using IVA should not alter the outcome much more and therefore show a similar outcome as with ISO alone, whereas if the membrane clock was functional then blocking it in the presence of ISO should results in a much smaller increase in beating rate.

Taken together, this work discovered a role for the conserved RNA structure in the regulation of cardiac contraction. The presence of an increased heart rate in NKXUSΔ$^{-/-}$ mice at the *in vivo* and *in vitro* level, indicates a dysfunction at the level of the SAN network rather than the autonomic nervous system. It is unclear if this is caused by an intrinsic problem of individual cells or within the SAN network. An increased beating rate may have arisen as a consequence of higher protein abundance or activity of β-adrenoreceptors or downstream components, which requires further analysis. As a next step, studies to dissect which specific components of the network are altered and responsible for the increased heart rate will be of great interest and help to decipher a more detailed function of the conserved *NkxUS* RNA structure in the regulation of heart rate.

3.2.9.1 Insights into mechanisms of *NkxUS* function

Analysis of the *NkxUS* transcript in HL-1 cells identified a connection of *NkxUS* to the negative regulation of gene expression of cardiac contraction associated genes. Moreover, *NkxUS* may interact with proteins involved in transcription, histone modification and possibly nuclear-to-mitochondrial communication and ATP synthesis. On the DNA level *NkxUS* may bind to genes involved in G protein coupled receptor signaling to inhibit adenylate cyclase activity and Wnt signaling. Analysis of the conserved RNA structure resulted in the association with an altered heart rate due to enhanced intrinsic pacemaker activity, which aligns with the results of the knockdown

experiment in HL-1 cells, indicating that the conserved structure may be an essential domain for *NkxUS* function.

Considering the association of the GWAS SNP with heart rate, the phenotype of the mouse and the results of *NkxUS* knockdown, we suspect that an impairment in the cAMP signaling pathway may cause the increased resting heart rate. cAMP is an important secondary messenger and plays an essential role in the regulation of several ion channels important for pacemaker activity. Numerous genes involved in this pathway are upregulated upon *NkxUS* knockdown (Figure 46), suggesting that perhaps a subset of these genes are also altered in NKXUSΔ[-/-] mice. Upregulated genes included downstream targets of cAMP signaling, such as adenylate cyclases, calcium channels, including L-type calcium channels and the ryanoide receptor, ATPases and the pacemaker specific sodium channel *Hcn4* (Figure 46). β-adrenoreceptors which activate adenylate cyclases were not altered. An increased influx of sodium and calcium ions depolarises the cell more rapidly and together with increased uptake of calcium back into the SR via SERCA as well as potassium efflux, which facilitate repolarisation, results in increased contractility and thus heart rate.

However, it has not been tested yet whether these genes are also deregulated in NKXUSΔ[-/-] SANs and RT-PCRs as well as western blots on SANs have to be performed. If the results show no difference in expression, the channels may be more active but not more abundant, which could be caused by elevated cAMP levels which then activate PKA, which in turn activates the previously mentioned channels (Figure 46). This can be investigated by using drugs to specifically inhibit individual channels as we have done for the HCN channels and then analyse for the differences in calcium transients between both genotypes.

Knockout mouse models of several components required for SAN automaticity exist. Loss of function was demonstrated to results as expected in SAN dysfunction, including bradycardia (HCN1, HCN4, RYR2, SERCA2a) and tachycardia-bradycardia syndrome (NCX, CACNA1D) [22, 49, 331, 333-335]. Accordingly, overexpression of, for example, HCN4 and SERCA2a resulted in enhanced contractility [336, 337]. Elevated heart rate induced by SERCA2a overexpression resulted from an increased calcium uptake into the SR, thus increased calcium load within the SR, which can then be used in the next cycle to depolarise the cells [337]. On the other hand, gain-of-function mutations within RYR2 lead to higher intracellular calcium concentrations, but did not accelerate the heart rate. On the contrary, it resulted in spontaneous calcium release without being in synchrony with the membrane or calcium clock, which in turn lead to cardiac arrhythmia [338]. This suggests that in

our model both clocks still work in concert and that potentially several clock components are deregulated. As we do not see sinus dysrhythmia, components facilitating depolarisation as well as repolarisation are functional and likely more active, which is consistent with the RNA sequencing results as we also discovered upregulated GIRK potassium channels (*Kcnj3, Kcnj5*) and *Serca2a*, which facilitate repolarisation. Furthermore, this is consistent with the observation that action potential duration is rate dependent and shortens during higher heart rates to ensure regular pacing [339]. This is facilitated through heightened sympathetic signaling at higher heart rates, which activates β-adrenoreceptors and thereby also its downstream targets including potassium channels (Figure 46).

Taken together, the mechanism underlying sinoatrial node automaticity is clearly complicated and does not rely on a single component. Given that we identified an increased but regular resting heart rate, we suggest that several components facilitating pacemaker depolarisation as well as repolarisation are altered. Either gene expression of cAMP signaling components is upregulated as in the HL-1 knockdown experiment, or intracellular cAMP or ATP levels are higher in NKXUSΔ$^{-/-}$ SANs, which in turn activate important contraction and relaxation mediating proteins, or both. Additionally, the protein ChIRP identified several proteins with both mitochondrial and nuclear function (Table 11). Mitochondria are the main ATP source and many of the identified proteins are involved in ATP synthesis, thus perhaps the crosstalk between *NkxUS* and the dual-localised proteins could play a role in controlling ATP production within mitochondria via nuclear-to-mitochondrial signaling and disruption of *NkxUS* could lead to enhanced ATP production. Most mitochondrial proteins are encoded in the nucleus and mitochondrial communication with the nucleus is well described and essential for homeostasis [340]. However, we still have to verify these interactions. Similarly, the DNA ChIRP results support a function of *NkxUS* in G protein-coupled receptor signaling which regulates cAMP signaling. The functions of these candidates were mostly studied in excitable neurons, in which they played a role in the negative regulation of adenylate cyclases. This is consistent with our phenotype as a disrupted *NkxUS* function leads to an increased heart rate, potentially due to reduced inhibition of genes involved in the cAMP signaling pathway. However, initial results require validation.

Altogether, follow up experiments to validate the ChIRP results and explore whether ion channels of the calcium and/or membrane clock are altered and whether the cAMP/ATP levels are elevated in NKXUSΔ$^{-/-}$ SANs would provide a detailed understanding of the underlying mechanism of the phenotype. Moreover, it would be of interest to analyse whether NKXUSΔ$^{-/-}$ mice are more

prone to develop atrial fibrillation, since three independent GWAS studies associated *rs6882776* with atrial fibrillation. Preliminary data of pacing induced atrial fibrillation on NKXUSΔ$^{-/-}$ and NKXUSΔ$^{+/+}$ mice suggests a sensitivity of NKXUSΔ$^{-/-}$ mice to develop atrial fibrillation, however so far this has not been statistically significant.

In conclusion, we identified a novel nuclear and cardiac specific lncRNA, just upstream of one of the most important genes for heart development. Our hypothesis that *NkxUS* might regulate *Nkx2-5* and/or heart development was not confirmed. However, *NkxUS* is important for proper heart function as it seems to regulate the expression of genes involved in cardiac contraction. Most significantly, we identified a GWAS SNP associated with an increased resting heart rate within the conserved RNA structure of *NkxUS* and by deleting this region in the mouse, even though the genomic location and orientation of the mouse transcript is different, we recapitulated the same phenotype as the GWAS SNP. We pinpointed the phenotype to be originating from the SAN, most likely due to altered gene expression, increased activity of cAMP signaling components or elevated ATP/cAMP levels.

Importantly, from our understanding this is the first study achieving such a connection *in vivo.* Since the majority of GWAS SNPs map to non-coding regions, many more are likely to affect lncRNAs and potentially lie within conserved RNA structures. This study could lead the way to motivate other researchers to investigate GWAS SNPs within lncRNAs and/or conserved RNA structures using animal models, which ultimately will lead to a better understanding of underlying causes of diseases and advancements in precision medicine.

4 Material & Methods

4.1 Molecular Biology and Microbiology

4.1.1 Polymerase Chain reaction (PCR)

KAPA HiFi DNA Polymerase (KAPABiosystems) was used for PCR reactions as follows:

5X KAPA HiFi Buffer (Fidelity)	5µl
10mM KAPA dNTP Mix	0.75µl
10 µM Forward Primer	0.75µl
10 µM Reverse Primer	0.75µl
Template DNA	1µg
KAPA HiFi DNA Polymerase	1µl
PCR-grade water	Up to 25µl

PCRs was performed with the following cycling protocol:

Step	Temperature	Duration	Cycles
Initial denaturation	95°C	3min	1
Denaturation	98°C	20 s	
Annealing	60°C	15 s	30
Extension	72°C	1min	
Final extension	72°C	3min	1

A complete list of used primers can be found in Appendix V.

4.1.2 Agarose gel electrophorese

DNA fragments were separated based on size using agarose gel electrophoresis. The gels were prepared using 1% agarose (Sigma-Aldrich) dissolved in 1x TAE buffer (40mM Tris, 1mM EDTA, pH 8) containing 1ug/ml ethidiumbromide (Sigma-Aldrich). A 1 kilobase (kb) or 100 basepair (bp) DNA marker (Promega) was used for size determination. Prior to loading the gels, the DNA samples were mixed with 6x loading buffer (Promega). The samples were run at 120V for 30-60min and the DNA was visualized using an UV-transilluminator (BioRad).

4.1.3 Isolation of DNA fragments from agarose gels

Bands of interest were excised using scalpels under UV light (UV transilluminator, Thermo Fisher Scientific) and the DNA was extracted using the Wizard SV Gel and PCR Clean-Up System (Promega) following the manufacturer's protocol.

4.1.4 Bacterial transformation

PCR fragments were A-tailed using Taq-Polymerase (Roche) for 15min at 75°C as follows:

T4 DNA Ligase Buffer	5µl
pGEM-T Easy Vector	1µl
PCR Fragment	3µl
T4 DNA Ligase	1µl
Total	10µl

Afterwards, the fragments were cloned into the pGEM-T Easy Vector Systems (Promega). The ligation was set up as follows:

T4 DNA Ligase Buffer	5µl
pGEM-T Easy Vector	1µl
PCR Fragment	3µl
T4 DNA Ligase	1µl
Total	10µl

The reaction was incubated for one hour at room temperature (RT). Afterwards, the plasmids, which carry an ampicillin resistance cassette, were transformed into chemically competent 10-beta *Escherichia coli cells (*NEB). The competent cells were first thawed on ice. 2µl of plasmid was added to 50µl of competent cells, carefully mixed and then incubated on ice for 20min. This was followed by a heat shock at 42°C for 40 sec and an incubation on ice for 5min. 200µl of stable outgrowth medium (NEB) was added to the mixture and incubated for one hour at 37°C. Afterwards, cells were transferred on Luria-Bertani (LB) agar plates with ampicillin (Sigma-Aldrich, final concentration: 100µg/µl) and incubated overnight at 37°C. LB agar plates were examined for colony formation on the following day.

4.1.5   Preparation of plasmid DNA from bacterial cultures

Single colonies were picked and inoculated in 5ml LB medium supplemented with ampicillin (final concentration: 100µg/µl) overnight at 37°C with shaking at 180 RPM. Plasmid DNA was isolated using the Wizard Plus SV Minipreps Start-Up Kit (Promega) following the manufacturer's protocol. Afterwards, DNA concentrations were determined using a NanoDrop2000 (Thermo Fisher Scientific).

4.1.6   DNA sequencing

The plasmid inserts were sequenced at an external facility (Garvan Molecular Genetics, Garvan Institute of Medical Research, Sydney, Australia). A sequencing reaction mix of 40ng plasmid and 2 pmol of primer was prepared for Sanger sequencing. A list of all used primers can be found in Appendix V.

4.1.7   Digestion of DNA

Plasmids were digested using restriction digestion enzymes from NEB. Reaction were set up as follows:

NEB Buffer	1.5µl
Plasmid DNA	1µg
Enzyme	1µl
PCR-grade water	Up to 15µl

After an one hour incubation at 37°C , the linearised plasmids were separated on a 1% agarose gel in 1x TAE Buffer and extracted from the gel using Wizard SV Gel and PCR Clean-Up System (Promega) following the manufacturer's protocol.

The following enzymes were used: *Spe*I, *Nco*I, *Afe*I.

4.1.8 Rapid amplification of cDNA ends (RACE)

RACE was performed on heart RNA from C57BL/6J mice. The experiment was performed using the FirstChoice RLM-RACE Kit (Thermo Fisher Scientific) following the manufacturer's protocol. All reagents detailed in this section were provided with the kit. 10µg of total RNA was used for 3' and 5'RACE. Gene specific primers used for the nested PCRs are listed in Table 12. The PCRs was run as follows:

Step	Temperature	Duration	Cycles
Initial denaturation	94°C	3min	1
Amplification	94°C	30 s	35
	60°C	30 s	
	72°C	30 s	
Final extension	72°C	7min	1

The PCR products were run on a 2% agarose gel containing 1µg/µl of ethidiumbromide. Bands were isolated from the gel, purified, cloned into the pGEM-T Easy vector and the inserts were sequenced using m13 forward and reverse primer (Appendix V). DNA sequences were visualized using the University of California Santa Cruz (UCSC) genome browser (https://genome.ucsc.edu/)[175].

Table 12 List of gene specific primers used for the nested RACE PCR in 5'-3' direction

Gene specific primer	Sequence (5'-3' direction)
NkxUS 5'Outer	CGACCACTTACAACCTGATTCATGC
NkxUS 5'Inner -1	GCAGTGACTCTTGGACTCTGACCAC
NkxUS 5'Inner-2	CTTTAATCCCTGAGCCATCTCTCCA
NkxUS 5'Inner-3	CTGTAGCGCTGTAGCGGTTT
NkxUS 3'Outer	GGAATCTGTTAGTAAAATCGCAGCA
NkxUS 3'Inner	GCAGCCTGTATATTCTCCAGCTCTT
NkxDS 3'Outer	GTGAGGGTGAAGGAGAA
NkxDS 3'Inner	GGTCTCAGAATGGACAAGCA

4.2 Gene expression analysis

4.2.1 RNA extraction

RNA was extracted using Trizol (Life Technologies) following the manufacturer's protocol. Briefly, 1mL Trizol was added to the samples, vortexed for 10s and incubated for 5min at RT. Afterwards, 200µl chloroform (Sigma-Aldrich) was added, the tube was shaken for 15 s, incubated for 2-3min at RT and then centrifuged for 15min at 4°C at 12,000 g. The aqueous, upper phase was transferred into a new tube and 500µl isopropanol (Sigma-Aldrich) was added, the tubes were mixed by shaking for 15s and incubated at RT for 10min. Following centrifugation at 12,000g for 10min at 4°C, the supernatant was removed and the pellet was washed with 1ml of 75% ethanol. The tubes were inverted several times to mix and then centrifuged for 5min at 7,500g at 4°C. The supernatant was removed, the pellet was left to air-dry and then resuspended in 30µl of nuclease-free water (Life Technologies). The samples were stored at -80°C or used directly for DNaseI treatment.

4.2.2 DNase treatment

RNA was either treated with DNaseI from NEB or Promega. The reactions were set up as follows using either DNaseI:

RNA	1µg
10x DNase buffer	1µl
DNaseI (1U/µl)	1µl
water	up to 10µl

The samples mixed carefully by pipetting and incubated for 10min at 37°C. Afterwards, the samples were cleaned up using the RNA Clean & Concentrator kit from Zymo Research.

4.2.3 RNA clean up and concentration

RNA was cleaned up and concentrated using the RNA Clean & Concentrator- 5 kit (Zymo Research) following the manufacturer's protocol. Briefly, nuclease-free water was added to the samples up to 50µl. 100µl of RNA binding buffer was added, the samples were vortexed and centrifuged briefly. Then an equal amount of 100% ethanol (150µl) was added to the samples. After vortexing and a

brief spin. The samples were spin through the columns for 30s at 16,000 g. The flow-through was discarded and 400µl of RNA Prep buffer was added to the columns. The samples were centrifuged again for 30 s, the flow-through was discarded and 700µl of RNA Wash buffer was added to the columns. After another centrifugation for 30s the samples were washed again with 400µl RNA Wash buffer and centrifuged for 2min. The columns were transferred into new collection tubes, 15µl of nuclease free water was added and the columns were centrifuged for 30 s. Eluted RNA was quantified using a Nanodrop2000 (Thermo Fisher Scientific) and then stored at -80°C or used directly for downstream experiments.

4.2.4   cDNA synthesis and real-time PCR (RT-PCR)

cDNA was synthesised using the SuperScript III First-Strand Synthesis SuperMix (Thermo Fisher Scientific) for real-time PCR (RT-PCR). The reaction was set up as follows:

RT Reaction Mix	10µl
RNA	up to 1µg
RT Enzyme Mix	2µl
DEPC-treated water	up to 20µl

The reaction mix was incubated at 25°C for 10min, then 50°C for 30min and then terminated at 85°C at 5min. The mix was chilled on ice and then used for RT-PCR. RT-PCR was performed using the SYBR green Master Mix (Roche). The reaction was set up in 15µl and in 384 well plates (BioRad) as follows:

cDNA	1µl
SYBR green Master Mix	7.5µl
DEPC-treated water	5.3µl
10 µM Forward primer (FW)	0.6µl
10 µM Reverse primer (RV)	0.6µl
Total	15µl

 The samples were run on the BioRad CFX384 with the following cycle conditions:

Step	Temperature	Duration	Cycles
Pre-incubation	95°C	7min	1
Amplification	95°C	20 s	40
	60°C	15 s	
	72°C	1min	
Melting curve	95°C	5 s	1
	65°C	1min	
	97°C	continuous	
Cooling	40°C	10 s	1

Primers used for gene expression analysis in mouse and human samples are listed in Table 13 and 14. Levels of gene expression were normalised against a combination of two to three housekeeping genes listed in Table 15. Calculations were performed by comparative method ($2^{-\Delta\Delta CT}$).

Table 13 List of RT-PCR primers for mouse transcripts in 5'-3' direction

Symbol	Transcript name	Sequence
Cacna1c	Calcium Voltage-Gated Channel Subunit Alpha1 C	FW: TCCCGAGCACATCCCTACTC
		RV: ACTGACGGTAGAGATGGTTGC
Hcn1	Potassium/Sodium Hyperpolarization-Activated Cyclic Nucleotide-Gated Channel 1	FW: CAAATTCTCCCTCCGCATGTT
		RV: TGAAGAACGTGATTCCAACTGG
Kcnj3	Potassium Voltage-Gated Channel Subfamily J Member 3	FW: GGGGACGATTACCAGGTAGTG
		RV: CGCTGCCGTTTCTTCTTGG
Kcnj5	Potassium Voltage-Gated Channel Subfamily J Member 5	FW: AAAACCTTAGCGGCTTTGTATCT
		RV: AAGGCATTAACAATCGAGCCC
Cacna1d	Calcium Voltage-Gated Channel Subunit Alpha1 D	FW: GCTTACGTTAGGAATGGATGGAA
		RV: GAAGTGGTCTTAACACTCGGAAG
Hcn4	Potassium/sodium hyperpolarization-activated cyclic nucleotide-gated channel 4	FW: GCATGATGCTTCTGCTGTGT
		RV: CCAGCTTTCGGCAGTTAAAG
Nt5m	5',3'-Nucleotidase	FW: GAGCAGATTGTGTTGACCAGA
		RV: CAGGTGGTAGTTGTGGCAGG
Ptprs	Protein Tyrosine Phosphatase, Receptor Type S	FW: GGTGAACAACATACCCCCGAC
		RV: TCCCACCTCTGTGTAAGCCA
Ryr2	Ryanodine receptor 2	FW: AAAAGTGCGTGTTGGAGATGA
		RV: CACCGCCAATGAGATAGCCT
Nkx2-5	NK2 Homeobox 5	FW: CCCAAGTGCTCTCCTGCTTTC

		RV: TCCAGCTCCACTGCCTTCTG
NkxDS	Nkx2-5 Downstream	FW: GGAATGTGTGAGGCTCTGCT
		RV: TGAATTGACACCCAGAACACA
NkxUS	Nkx2-5 Upstream	FW: CACTCTGACCCAGCACCTTT
		RV: GGCATAGCGAACCTTGTCTC
NkxUSas	NkxUS antisense	FW: CCCGCACATTCTGCTATTTGAG
		RV: GTCCCCGTTAGTCAGTCCAG
Pde3b	Phosphodiesterase 3B	FW: AGTATCAGTAGCTTGATGGGTGC
		RV: CCCTTGTGAAGTTTTCGATCTCC
Pde5a	Phosphodiesterase 5a	FW: CGGCCTACCTGGCATTCTG
		RV: GCAAGGTCAAGTAACACCTGATT

Table 14 List of RT-PCR primers for human transcripts in 5'-3' direction

Symbol	Transcript name	Sequence
Nkx2-5	NK2 Homeobox 5	FW: CTATCCACGTGCCTACAGCGAC
		RV: GCACAGCTCTTTCTTTTCGGC
NkxDS	Nkx2-5 Downstream	FW: CTGGCACAAAGTAAGCACCA
		RV: CCAGTGGTGAGCAGTGAAGA
NkxUS	Nkx2-5 Upstream	FW: CTCCTCTCTCCTGCAACACC
		RV: GTCAGTCCTGTGGGTCAGGT

Table 15 List of primer sequences for house keeping genes in 5'-3' direction

Symbol	Transcript Name	Sequence
Eef1e1	Eukaryotic Translation Elongation Factor 1 Epsilon 1	FW: TCCAGTAAAGAAGACACCCAGA
		RV: GACAAAACCAGCGAGACACA
Hprt1	Hypoxanthine Phosphoribosyltransferase 1	FW: GCTTGCTGGTGAAAAGGACCTCTCGAAG
		RV: CCCTGAAGTACTCATTATAGTCAAGGGCAT
Rpl4	Ribosomal Protein L4	FW: GCCGCTGGTGGTTGAAGATAA
		RV: CGTCGGTTTCTCATTTTGCCC
Tbp	TATA-Box Binding Protein	FW: TATGACCCCTATCACTCCTG
		RV: TTCTTCACTCTTGGCTCCTGT
Actb	Actin Beta	FW: GGCTGTATTCCCCTCCATCG
		RV: CCAGTTGGTAACAATGCCATGT
Gapdh	Glycerinaldehyd-3-phosphat-Dehydrogenase	FW: GGTCCTCAGTGTAGCCCAAG
		RV: AATGTGTCCGTCGTGGATCT

4.3 Northern blot analysis

4.3.1 Probe synthesis

For DIG-labeled RNA probes, specific regions of *Nkx2-5, NkxUS* and *NkxDS* were amplified from mouse cDNA using primers listed in Table 16 and cloned into the pBlueScriptII SK vector. 1µg of plasmid, linearised using SpeI (NEB), was used for *in vitro* transcription reactions with digoxigenin-11-UTP labelling mixture (Roche) as follows:

Lineraised plasmid	1µg
10x transcription buffer (NEB)	1.5µl
10x DIG reaction mix (Roche)	1.5µl
0.1M DTT (Sigma)	1.5µl
rRNasin (Promega)	0.75µl
T7 Polymerase (NEB)	0.5µl
water	up to 15µl

The samples were incubated at 37°C for 2hrs and then treated with 1µl of DNaseI (NEB). Afterwards, the samples were purified through G-50 columns (GE Healthcare) according to manufactures' instructions. Aliquots of probes were stored at -80°C.

Table 16 List of primers used to generate northern probes in 5'-3' direction

Transcript	Sequence	Probe length
NkxDS	FW: CAGCTTTCTCCCAGGTCAAG	1025 bp
	RV: ATTGTCCCAGAAGCCAACAC	
Nkx2-5	FW: CTACGGCGTGGGTCTCAAT	601 bp
	RV: CTCTTCCCATTAAAGTGAGTGCG	
NkxUS	FW: CACTCTGACCCAGCACCTTT	719 bp
	RV: GACGTGGCTAGGTGGACTGT	

4.3.2 Northern blot run

Northern Blots were performed as outlined in the NorthernMax Procedure (Ambion) and the DIG Northern Starter Kit (Roche) with a few modifications. Briefly, total RNA was extracted from wild-type adult mouse heart (C57BL/6J) using Trizol, DNase treated and quantified on the Nanodrop 2000 (Thermo Fisher Scientific). 10µg of RNA in a volume of 6µl was prepared in 20µl of loading

buffer containing 2µl 10x MOPS (200mM MOPS (Sigma-Aldrich), 50mM Sodium Acetate, 10mM EDTA), 10µl formamide and 2µl 37% formaldehyde, heated at 65°C for 15min and chilled on ice. 3µl 6x loading dye (Promega) and 0.5µl EtBr (10µg/ul) were added and the samples were loaded onto a 1.2% agarose gel made with 1x MOPS buffer and containing 2% formaldehyde. 5µl of denatured ssRNA ladder (NEB) was also loaded. The gel was run in 1x MOPS buffer at 105V (5 volts/cm) for 3hrs and visualized for RNA integrity and size on a UV transilluminator. The gel was blotted overnight by capillary transfer using 20x SSC onto a positively charged Nylon membrane (Amersham). The membrane was cross-linked for 15s and pre-hybridised in 7ml pre-warmed NorthernMax ULTRAhyb Buffer (Life Technologies) at 68°C for 1 h. 100ng/ml Dig-labeled probe was added to the ULTRAhyb buffer and incubated overnight at 68°C with rolling. The membrane was washed twice for 5min each with Low Stringency Wash Buffer (2x SSC, 0.1% SDS) at RT and then twice for 15min washes in High Stringency Wash Buffer (0.1% SSC, 0.1% SDS) at 68°C. The membrane was rinsed briefly in MAB Buffer pH7.5 (0.1M Maleic Acid, 0.15M NaCl) and blocked for 30min at RT in Blocking Solution (Roche). DIG anti-AP (1:10.000) in blocking solution was added to the membrane, incubated for 30min at RT and then washed three times for 10min in MABT (MAB + 0.3% Tween 20). The membrane was equilibrated for 3min in Detection Buffer (0.1M Tris pH 9.5, 0.1M NaCl) and probes were detected with Amersham Hyperfilm ECL (GE Healthcare Life Sciences) using CSPD (1:100 in detection buffer).

4.4 Western Blot

For western-blotting, mouse heart tissues were homogenized using a homogenizer in using RIPA buffer (20mM Tris-HCl pH 8.0, 150mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 1 protease inhibitor (cOmplete™ Mini EDTA-free Protease Inhibitor Cocktail, Sigma Aldrich)). Proteins were measured using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific) using the manufacturer's protocol. 30µg of protein was prepared for SDS- polyacrylamide gel electrophoresis (SDS-PAGE) by adding 4x LDS sample buffer (Thermo Fisher Scientific) with 100mM DTT to the samples, followed by an incubation at 70°C for 10min. Proteins were separated on a 4-12% Bis-Tris gel (Thermo Fisher Scientific) in 1x MOPS buffer (Thermo Fisher Scientific), which was run at 200V for 40min. Afterwards, the gel was blotted onto a PVDF membrane (BioRad). Proteins were transferred onto the PVDF membrane at 30V for 1 h. After transfer, the membrane was cut into at

~60kDa. The two membrane pieces were incubated in 5% milk powder in TBST (0.1% Tween20, Sigma Aldrich) for at least 1h at RT. Incubation with primary antibodies (Table 17) was performed over night at 4°C in 5% milk powder in TBST. The next day, the membranes were washed three times with TBST before incubating with the respective fluorescently labelled secondary antibody (Table 17) in 5% milk powder in TBST for 1h at RT. The membranes were washed three times with TBST and once with TBS, before signals were detected using the Pierce ECL Western Blotting Substrate (Thermo Fisher Scientific) and documented using the ChemiDoc (BioRad). Signal intensities were quantified using the Image Lab software (BioRad).

Table 17 List of antibodies used for western blotting

Antibody	Concentration	Species	Company
Primary			
Nkx2-5 N19	1:1000	Goat	Santa Cruz
Vinculin H300	1:1000	Rabbit	Santa Cruz
Secondary			
anti goat	1:8000	Donkey	Thermo Fisher Scientific
anti rabbit	1:8000	Rat	Sigma Aldrich

4.5 Cell culture techniques

4.5.1 Cell lines

The HL-1 cell line was donated by Prof W C Claycomb (Department of Biochemistry and Molecular Biology, Louisiana State University, New Orleans, LA, USA) [179].
HL-1 cells were cultured in Claycomb Medium (Sigma-Aldrich) supplemented with 10% heat inactivated fetal bovine serum (FBS; SAFC), 2mM L-glutamine (Life Technologies) and 0.1mM Norepinephrine (Sigma-Aldrich) at 37°C in 5% CO_2. Cells were seeded in 60-180cm^2 dishes (Corning) and split when confluent.

Mouse muscle C2C12 cells (ATCC) were cultured in high Glucose DMEM (Sigma-Aldrich) supplemented with 10% FBS (SAFC) and 1% sodium pyruvate (Sigma-Aldrich) at 37°C in 5% CO_2. Cells were seeded in 60-180cm^2 dishes and split when reached about 80% confluence.

Mouse embryonic fibroblasts (MEF, ATCC) were cultured in high glucose DMEM, 10% FBS and 5ml

Glutamax (life Technologies) at 37°C in 5% CO_2. Cells were cultured in 6-well culture plates (Corning) and split when about 80% confluent.

4.5.2   Mouse embryonic stem cell differentiation into embryoid bodies

Embryonic stem cells (ESC) were maintained on irradiated MEFs in ESC medium (Knockout-DMEM (Life Technologies) supplemented with 20% Knockout Serum Replacement (Life Technologies), 0.1mM MEM minimum essential amino acids solution (Life Technologies), 1mM sodium-pyruvate (Sigma-Aldrich), 0.1mM β-mercaptoethanol (Sigma-Aldrich), 100U/ml Penicillin/Streptomycin (Sigma-Aldrich), and 1000U/ml Leukemia Inhibitory Factor (LIF, Sigma-Aldrich)).

For direct cardiomyocyte differentiation, 800 ESCs were cultured for 2 days in 20µl of ESC medium that contained 10^{-4}M ascorbic acid (Sigma-Aldrich) in hanging drops to produce embryoid bodies (EBs). Subsequently, EBs were cultured as suspensions in bacterial dishes for 5 additional days. On Day 7, EBs were plated separately in 1% gelatin-coated wells of a 24-well tissue culture plate for an additional 23 days to allow adherence and development of beating cardiomyocytes.

4.5.3   Differentiation of hiPSCs into Cardiomyocytes

The experiment was started with hiPSCs cultured on irradiated MEF cells in a 1:3 ratio which are approximately 70-80% confluent. Any differentiated colonies were removed using the stereomicroscope. 0.5ml Dissociation Solution (DS; 10ml 0.05% Trypsin (Life Technologies), 4ml Knockout Serum Replacement (Life Technologies), 1ml Collagenase type IV (1mg/ml) (ScimaR), 5ml Knockout-DMEM (Life Technologies) and 20µl 1M $CaCl_2$) was added to the cells and incubate for 0.5–1min until MEF cells were dissociated and the edges of hiPSCs colonies became clear. DS was removed and 0.75ml collagenase type IV (1mg/ml) was added. The cells were incubated for 5-15min at 37°C and checked under the microscope during the incubation period. When the colonies started to lift off and detach, the solution was removed and 1.5ml hES medium (390ml Knockout-DMEM, 100ml Knockout Serum Replacement, 5ml Glutamax (Life Technologies), 5ml MEM

minimum essential amino acids solution (Life Technologies) and 0.5ml β-Mercaptoethanol (Sigma-Aldrich)) was added. Using a p1000 pipette, the colonies were gently detached as a whole, taking care not to separate the cells. The cells were transferred into a 15ml tube using a 5ml pipette. The well was washed with 1ml hES medium, which was afterwards added to the same tube. The tube was left to stand for 5min until all colonies sediment. Medium was removed and replaced with 2ml hES medium supplemented with 100ng/ml basic fibroblast growth factor (bFGF; Life Technologies). The cells were transferred into an ultra-low attachment plate using a 5ml pipette (1ml into each well of a 24-well plate and 3ml into each well of a 6-well plate) and incubated at 37°C/5% CO_2 for at least 6 hr. During this time, the required number of plates were coated with laminin (Life Technologies) and incubated for 1h at 37°C. After 6 hrs, the colony aggregates were transferred into a 15ml tube using 5ml pipette, the plate was washed with RB medium (390ml RPMI 1640 (Life Technologies), 10ml B27 minus insulin (Life Technologies), 5ml Glutamax, 5ml MEM minimum essential amino acids solution, 5ml Penicillin/Streptomycin (Sigma-Aldrich) and 0.5ml β-Mercaptoethanol) to collect any remaining aggregates and the medium was added to the same 15ml tube (0.5ml per well for 24-well and 1ml per well for 6-well plates). The aggregates were left standing for 5min and then the medium was carefully aspirated. 2ml RB medium supplemented with 12μM CHIR99021 (WNT activator) (Miltenyi Biotec Australia) was added to the aggregates and carefully transferred with a 5ml pipette to laminin coated wells. After 24 hrs, the medium was carefully removed in each well and 1ml RB medium was added.

After another 24 hrs, the medium was carefully removed and exchanged with RB medium supplemented with IWP2 (inhibitor of Wnt ligand production), Purmorphamine (activator of the Hedgehog pathway) and SB431542 (inhibitor of the TGF-beta/Activin/Nodal pathway) (5μM each, all from Miltenyi Biotec Australia). The medium was changed every second day for four days and then every 3-4 days with RB medium.

4.5.4 Cell fractionation

HL-1 cell fractionation was performed according to the protocol described by Weil *et al*. [341]. HL-1 cells were grown as monolayers in 10 cm dishes until 80% confluent. Cells were trypsinised and transferred into 1ml of 1x PBS pH 7.4. 100μl was removed as "total cell extract" and added to 1ml of Trizol. Remaining cells were spun at 3000rpm for 3min and carefully resuspended in 200μl

Dautry Buffer (10mM Tris pH7.8, 140 nM NaCl, 1.5mM MgCl$_2$, 10mM EDTA, 0.5% NP-40 (Sigma-Aldrich), 100U/ml RNaseOUT (Life Technologies)), incubated on ice for 5min and then centrifuged at 4°C for 5min at 3000 RPM. The supernatant was transferred to a new tube and the pellet was kept. The supernatant was centrifuged again at full speed for one minute to eliminate any potential nuclei and transferred to a new tube together with 1ml of Trizol for the "Cytoplasmic Fraction". The first pellet, was washed once with 500μl Dautry Buffer, mixed by pipetting and centrifuged at 4°C, 3000rpm for 5min. The supernatant was eliminated and the pellet kept and resuspended in 1ml of Trizol for the "Nuclear Fraction". RNA was extracted, DNase digested and 500ng RNA per fraction was used in cDNA reactions. Cellular localisation was checked via RT- PCR.

4.5.5 Polysome gradient profiling

HL-1 cells were grown as described in [179]. 10μl of cycloheximide (CHX, 100mg/ml, Sigma-Aldrich) or puromycin (10mg/ml, Thermo Fisher Scientific) was added to the cells. After an incubation of 10min at 37°C, the media was removed, the cells washed and harvested. Cytoplasmic lysates were prepared and polysomes were extracted by adding 1ml of polysome extraction buffer (20mM Tris, pH 7.5, 100mM KCl, 5mM MgCL$_2$, 0.5% NP-40, supplemented with Protease-Inhibitor (PI, 50x stock, Roche), 100U/ml RNaseOUT and 100μg/ml CHX or puromycin) to the cell pellet. The samples were incubated on ice for 10min and then centrifuged for 12,000g for 10min at 4°C to pellet nuclei and debris. The supernatant was transferred to a new tube and RNA concentrations were measured using the Nanodrop2000. Equal amounts of cytoplasmic lysate were loaded onto 10%–50% linear sucrose gradients and centrifuged at 36,000rpm for 1h and 45min at 8°C in a SW41 rotor (Beckman Coulter). Twelve fractions were collected from the top of the gradient using a piston gradient fractionator (BioComp Instruments). The absorbance at 254 nm was measured with a UV-M II monitor (Bio-Rad).

Following fractionation, 110μl of 10% SDS and 12μl of proteinase K (20mg/mL; Invitrogen) were added to each fraction and incubated with shaking at 1000rpm for 30min at 42°C to digest protein debris and to inactivate nucleases. Polysomal RNA was extracted using Trizol (Thermo Fisher Scientific) following the manufacturer's protocol. RNA quality was assessed on the Bioanalyzer 2100 (Agilent), cDNA was synthesised and used for RT-PCR.

4.5.6 RNA Fluorescence in situ hybridisation (RNA FISH)

Custom Stellaris FISH Probes were designed towards the non-repetitive regions of mouse *NkxUS and NkxDS*, using the Stellaris RNA FISH Probe Designer (Biosearch Technologies) available online at www.biosearchtech.com/ stellaris designer. 48x 20mer oligonucleotides were designed per target transcript and conjugated to Quasar670 flurophores (Table 18). A *NEAT1* probe set (Cat #: SMF-3010-1), conjugated to Quasar 570 fluorophore, was used as a predesigned control. Staining was carried out following the manufacturer's protocol for adherent mammalian cells. *NkxDS*, *NkxUS* and *NEAT1* probe sets were used at 200 nM with hybridisation overnight at 37°C. Briefly, HL-1 and C2C12 cells were grown on coated 13mM glass coverslips (Menzel Glaeser) in 24 well plates and grown for 2-3 days prior to staining. Cells were washed once in 1x PBS and fixed for 10min at RT in fixation solution (10 mL: 1ml 37% formaldehyde solution (Sigma-Aldrich), 1ml 10X RNase-free PBS (Life Technologies), and 8ml RNase-free water). The fixation solution was removed and cells were washed twice with 1x PBS before permeabilization with 70% ethanol at 4°C for at least one hour. Ethanol was then aspirated and cells were incubated for 5min in wash buffer (50 mL: 5ml 20x SSC, 5ml deionized formamide (Sigma Aldrich), 40ml RNase-free water). Coverslips were removed from the wells and inverted onto a 40μl drop of probe (200 nM) in hybridisation buffer (5 mL: 0.5g dextran sulfate (Fisher BioReagents), 0.5ml 20x SSC, 0.5ml deionized formamide, 4ml RNase-free water) on a clean parafilm surface, and placed at 37°C in a dark humidified chamber overnight. The next day, coverslips were carefully transferred to a new 24 well plate, cell side up, containing 0.5ml wash buffer, and incubated in the dark at 37°C for 30min with shaking. The wash buffer was replaced with wash buffer containing Hoechst (Life Technologies, 1:20.000), and cells were incubated again in the dark at 37°C for 30min with shaking. Hoechst solution was replaced with 2x SSC and incubated for 5min at RT. The coverslips were mounted on glass microscope slides in one drop of Prolong gold antifade mountant (Life Technologies) and sealed with nail polish. Slides were imaged using a Zeiss LSM 700 Upright confocal microscope (Zeiss) with the manufacturer provided ZEN imaging software.

Table 18 RNA-FISH probe sequences for *NkxUS* and *NkxDS*

Final probe name	Probe (5'-> 3')	Final probe name	Probe (5'-> 3')
NkxUS_1	ggttcctatatttcaccgaa	NkxDS_1	cttttatgagtgcaggttct
NkxUS_2	gcggaagatgtacttcttca	NkxDS_2	cccagcaacagtgaatcaag
NkxUS_3	cggagcccggagaaataaaa	NkxDS_3	cccagcaacagtgaatcaag

NkxUS_4	tttatgctgtagcgctgtag	NkxDS_4	gtcaaattactgccataccg
NkxUS_5	attatcaccttctgagacgc	NkxDS_5	cagagcgattatgttggtca
NkxUS_6	ctcaatctgagctctcaagc	NkxDS_6	aaacttgggacacccagtaa
NkxUS_7	accacttacaacctgattca	NkxDS_7	agcagctgagctgaatacat
NkxUS_8	taccggttcacaatcctctc	NkxDS_8	cattagtttagtcccaggaa
NkxUS_9	cgcttttcaccaattcgtgc	NkxDS_9	gattggtagaggagtgctac
NkxUS_10	gtgacattgtgttcctctag	NkxDS_10	acgttgagaatgtggtggtt
NkxUS_11	atgtgcaccttgaaagcttg	NkxDS_11	cctgaaagagaggagttgca
NkxUS_12	gttcacactaattggtgtgc	NkxDS_12	gctctttcttcagtaggtaa
NkxUS_13	tgcactccggaattgtgaac	NkxDS_13	gactactttgactctgggac
NkxUS_14	ataccagagcagatttggtg	NkxDS_14	ttgcctttgtcattgttaga
NkxUS_15	ttggctgttccttgtgtttg	NkxDS_15	cttccactaccactattttt
NkxUS_16	ctctgatcaggttgtcttag	NkxDS_16	gaagtgttccaatctggcaa
NkxUS_17	gtattgcagccaagaagtga	NkxDS_17	ttccatagtgtatggtgaca
NkxUS_18	ctcagcagtttgagtaatcc	NkxDS_18	gctgtttggctaaattctgg
NkxUS_19	ccctgataaatgacaaggga	NkxDS_19	aagcagtcagttcactgttt
NkxUS_20	ggagaaggctgatctaaaga	NkxDS_20	ctctttcagagggctcaaag
NkxUS_21	gaatcagcttcagttagggc	NkxDS_21	caccaactgggagctaattt
NkxUS_22	atcaaccttttgctattgcc	NkxDS_22	ggggatgggggaataaggaa
NkxUS_23	cactgagttgttcggtttgc	NkxDS_23	tcttgtgttactaccttcac
NkxUS_24	tgttagagggaaagacaccc	NkxDS_24	tctaaatggcgctggatcta
NkxUS_25	aaccaggcactatagcttcc	NkxDS_25	gattcaaacctgagtgacct
NkxUS_26	ctagttttcgggtcaagagt	NkxDS_26	ttgacttctgggtcaaagca
NkxUS_27	gtcaagaatgctcaacctgg	NkxDS_27	ggaaatgggttgctcagatc
NkxUS_28	ccatttgataggcagacaga	NkxDS_28	gtgtagaagcctgtaacagt
NkxUS_29	tatctgaagcatcacctacc	NkxDS_29	taggtgcttgtactagtctg
NkxUS_30	tctctaggtgtccagataac	NkxDS_30	gcacttcaggattacatgga
NkxUS_31	taagccattggagatagctc	NkxDS_31	ggttaactctgggatgtgac
NkxUS_32	agagatagacaggcaccttg	NkxDS_32	ggctctgtgtcttatggaaa
NkxUS_33	tcagagtgggagacattgtg	NkxDS_33	tttccaacattatccctaga
NkxUS_34	gacgagaagattgctcagct	NkxDS_34	caccaggactgtgtatagat
NkxUS_35	ttaggattgagagacgtggc	NkxDS_35	actgagcaaagtgatcctca
NkxUS_36	agaagcatgaagacagctga	NkxDS_36	cacagtgatggctacattca
NkxUS_37	ttgtcctgtggacttaacac	NkxDS_37	aagccctgaggattactctg
NkxUS_38	ttccaatctgtgcagaagtc	NkxDS_38	gcagttcattgacagtgctg
NkxUS_39	aagcaaggatcttgcatgct	NkxDS_39	actggagtttctggagacag
NkxUS_40	ctgtaccccagaaaaacagt	NkxDS_40	acatcattttgtacctgact
NkxUS_41	ctttccatttgataggcaga	NkxDS_41	tctgagggtggttaatgtga
NkxUS_42	tcatcctggagttagggaaa	NkxDS_42	ggtcctttgtaactgcagat
NkxUS_43	tgagtgggcagtagagagaa	NkxDS_43	tgactgtgagtgctcactgc
NkxUS_44	gttccattcgatggatcatt	NkxDS_44	gtgacagtgtctatgtgctg
NkxUS_45	gatgtccaggatggaaatgc	NkxDS_45	atgggaccatgtgcattaag

NkxUS_46	aggcccaggtaaaagaactc	NkxDS_46	ctgcatgtgttccatattta
NkxUS_47	ctcttttcccactcaattac	NkxDS_47	aggaaacctataggtcgcat
NkxUS_48	cctaaagcaagggtcaagga	NkxDS_48	tttggctgatgcgtttagag

4.5.7 Cell transfection

Two locked nucleotide acid (LNA) GapmeR antisense oligonucleotides (ASOs) (Exiqon) were custom designed for *NkxUS*, named ASO1 and ASO2 as well as a ready to use control LNA GapmeR. The target sequences are listed in Table 19. Lipofectamine 2000 (Life Technologies) was used as a transfection reagent and the transfection was set up in a 24-well format. 40 nM of ASO was mixed with 50µl of OPTI-MEM (Thermo Fisher Scientific) directly in the well and incubated for 5min at RT. 4µl of Lipofectamine 2000 was used per well and mixed with 50µl of OPTI-MEM. After 5min, the Lipofectamine mixture was added to the ASO solution, mixed well and incubated for 20min at RT. Meanwhile, the cells were harvested using Trypsin (Life Technologies) as previously described. 2×10^5 cells in 1ml of HL-1 medium were used per well and directly added to the transfection mix and incubated at 37°C for 24 hrs. The following day the cells were washed twice with 1x PBS before harvesting in 700µl of Trizol. The samples were stored at –80°C until further use.

RNA was extracted following the Trizol protocol. The RNA was treated with 1µl DNAseI for 10min at 37°C. RNA samples were quantified using the NanoDrop2000 (Thermo Fisher Scientific) and knockdown efficiency was determined using RT-PCR.

Table 19 ASO target sequences

	Target sequence
ASO1	CACGTCTCTCAATC
ASO2	GTCTATCTCTTTTGTC

4.5.7.1 RNA sequencing after cell transfection

High quality RNA with an RNA integrity number (RIN) of > 9, extracted from three biological replicates for ASO1, ASO2 and control ASO, were used to prepare in total twelve libraries using the TruSeq Stranded mRNA Library Prep kit from illumina following the standard protocol. The library preparation was performed by Dominik Kaczorowski (Garvan Institute of Medical Research, Sydney,

Australia). The RNA-sequencing libraries were sequenced on the Illumina HiSeq2500 v4 platform in high output mode at the Kinghorn Centre for Clinical Genomics (Garvan Institute of Medical Research, Sydney, Australia). For all samples, over 20 million 125 nucleotide paired-end reads were generated.

4.6 Chromatin isolation by RNA immunoprecipitation (ChIRP)

4.6.1 Design of ChIRP probes

Antisense DNA tiling probes for selective retrieval of *NkxUS* were designed using the online probe designer at singlemoleculefish.com as recommended by the original ChIRP paper [235].
40 biotintylated probes were designed, labelled according to their position along *NkxUS* and then divided into two pools, "EVEN" (all even numbered probes) and "ODD" (all odd numbered probes) (Table 20). Additionally, LacZ probes were used as a negative control as LacZ is not present in mammalian cells (Table 20). The oligos with standard desalting purification were produced by Sigma-Aldrich. The probes were diluted to 100 μM concentration and stored at -20°C. All ChIRP experiments were done with the three probe pools.

Table 20 NkxUS and LacZ ChIRP probes

Final probe name	Probe (5'-> 3')	Final probe name	Probe (5'-> 3')
NkxUS_Probe 1	ggttcctatatttcaccgaa	LacZ_Probe 1	ccagtgaatccgtaatcatg
NkxUS_Probe 2	gcggaagatgtacttcttca	LacZ_Probe 3	attaagttgggtaacgccag
NkxUS_Probe 3	cggagcccggagaaataaaa	LacZ_Probe 5	aatgtgagcgagtaacaacc
NkxUS_Probe 4	tttatgctgtagcgctgtag	LacZ_Probe 7	aataattcgcgtctggcctt
NkxUS_Probe 5	attatcaccttctgagacgc	LacZ_Probe 9	aattcagacggcaaacgact
NkxUS_Probe 6	ctcaatctgagctctcaagc	LacZ_Probe 11	atcttccagataactgccgt
NkxUS_Probe 7	accacttacaacctgattca	LacZ_Probe 13	gctgatttgtgtagtcggtt
NkxUS_Probe 8	taccggttcacaatcctctc	LacZ_Probe 15	aactgttacccgtaggtagt
NkxUS_Probe 9	cgcttttcaccaattcgtgc	LacZ_Probe 17	tttcgacgttcagacgtagt
NkxUS_Probe 10	gtgacattgtgttcctctag	LacZ_Probe 19	accattttcaatccgcacct
NkxUS_Probe 11	atgtgcaccttgaaagcttg	LacZ_Probe 21	ttcatcagcaggatatcctg
NkxUS_Probe 12	gttcacactaattggtgtgc	LacZ_Probe 23	tggttcggataatgcgaaca
NkxUS_Probe 13	tgcactccggaattgtgaac	LacZ_Probe 25	agacgattcattggcaccat
NkxUS_Probe 14	ataccagagcagatttggtg	LacZ_Probe 27	atttgatccagcgatacagc
NkxUS_Probe 15	ttggctgttccttgtgtttg	LacZ_Probe 29	tttgatggaccatttcggca

NkxUS_Probe 16	ctctgatcaggttgtcttag	LacZ_Probe 31	aaacggggatactgacgaaa
NkxUS_Probe 17	gtattgcagccaagaagtga	LacZ_Probe 33	atacagaactggcgatcgtt
NkxUS_Probe 18	ctcagcagtttgagtaatcc	LacZ_Probe 35	tattcgctggtcacttcgat
NkxUS_Probe 19	ccctgataaatgacaaggga	LacZ_Probe 37	tttaccttgtggagcgacat
NkxUS_Probe 20	ggagaaggctgatctaaaga	LacZ_Probe 39	aaatccatttcgctggtggt
NkxUS_Probe 21	gaatcagcttcagttagggc		
NkxUS_Probe 22	atcaaccttttgctattgcc		
NkxUS_Probe 23	cactgagttgttcggtttgc		
NkxUS_Probe 24	tgttagagggaaagacaccc		
NkxUS_Probe 25	aaccaggcactatagcttcc		
NkxUS_Probe 26	ctagttttcgggtcaagagt		
NkxUS_Probe 27	gtcaagaatgctcaacctgg		
NkxUS_Probe 28	ccatttgataggcagacaga		
NkxUS_Probe 29	tatctgaagcatcacctacc		
NkxUS_Probe 30	tctctaggtgtccagataac		
NkxUS_Probe 31	taagccattggagatagctc		
NkxUS_Probe 32	agagatagacaggcaccttg		
NkxUS_Probe 33	tcagagtgggagacattgtg		
NkxUS_Probe 34	gacgagaagattgctcagct		
NkxUS_Probe 35	ttaggattgagagacgtggc		
NkxUS_Probe 36	agaagcatgaagacagctga		
NkxUS_Probe 37	ttgtcctgtggacttaacac		
NkxUS_Probe 38	ttccaatctgtgcagaagtc		
NkxUS_Probe 39	aagcaaggatcttgcatgct		
NkxUS_Probe 40	ctgtaccccagaaaaacagt		

4.6.2 ChIRP followed by mass spectrometry

The protein ChIRP was performed following a published protocol [236]. 100-500 million HL-1 cells were harvested and crosslinked with 0.5% or 3% Formaldehyde for 10min followed by quenching with 1.25M Glycine (Sigma-Aldrich) for 5min. After spinning, the pellet was washed with cold 1x phosphate buffered saline (PBS) and spin again. The pellet was resuspended in 1ml of cold 1x PBS, transferred to a 1.5ml tube and spin again at 4°C. Afterwards, the pellet was flash frozen and stored at -80°C.

Pellets were thawed at RT and weighed. Per 100mg of pellet, 1ml of lysis buffer (50mM Tris-Cl pH7.0, 10mM EDTA, 1% SDS) supplemented with Protease-Inhibitor (PI, 50x stock, Roche), Superase-IN (200x stock, Ambion) and Phenylmethylsulfonyl fluoride (PMSF, 100x stock, Sigma-

Aldrich) was added and resuspended. The chromatin was placed on ice and sonicated at 4°C using the Bioruptor Pico (Diagenode) until the fragments were under 500bp long. After sonication the chromatin was centrifuged at 16,000 RCF for 10min, the supernatant was transferred to a new tube, flash frozen and stored at -80°C.

Chromatin of 100-500 million cells were thawed at RT and pooled. Per ml of lysate, 30µl of C-1 magnetic beads (Invitrogen) were washed three times with unsupplemented lysis buffer and resuspended in original volume of lysis buffer supplemented with PI, Superase-IN and PMSF. The beads were added to the lysate and incubated for 30min, shaking at 37°C. Afterwards, the beads were removed twice from the lysate using a magnetic stand. Per 1ml of lysate, 10µl was removed to serve as RNA input and 50µl to serve as protein input. The input samples were kept on ice until further use. In the meanwhile, 2ml of hybridisation buffer (750mM NaCl, 1% SDS, 50mM Trsi-Cl pH7.0, 1mM EDTA, 15% formamide (Sigma-Aldrich)) supplemented with fresh PI, Superase-IN and PMSF was prepared per 1ml of lysate. The lysate was divided equally into three 15ml Falcon tubes, one for EVEN, one for ODD and one for LacZ. 2ml of Hybridisation buffer and 1µl of 100µmol of probe pool was added per 1ml of lysate and incubated overnight at 37°C shaking.

The next day, 100µl of beads per 100 pmol of probe were prepared as the day before and added to the samples. After 30min at 37°C, the beads were collected using a magnetic stand and washed five times with wash buffer (2x NaCL and socium citrate, 0.5% SDS) freshly supplemented before each wash with PMSF. The samples were washed each time for 5min at 37°C shaking. At the last wash, 100µl per 1ml of lysate was removed for RNA isolation and the rest was used for protein isolation.

For the RNA isolation, 85µl and 95µl of Proteinase K buffer (100mM NaCl, 10mM Tris-Cl pH8, 1mM EDTA, 0.5% SDS) was added to the RNA input and bead sample, respectively, as well as 5µl Proteinase K (20mg/ml, Ambion) each. The samples were incubated at 50°C for 45min, briefly spin down and incubated for 10min at 95°C. The samples were chilled on ice and 500µl Trizol (Thermo Fisher Scientific) was added. After vortexing the samples were incubated for 10min at RT and then either stored at -80°C or further processed directly. 100µl of chloroform (Sigma-Aldrich) was added to each sample, the samples were vortexed and then spin at 16,000 RCF for 15min at 4°C. The aqueous phase was transferred to a new tube and 1.5 volume of 100% ethanol was added. The samples were vortexed and the spin through miRNeasy mini columns (Qiagen) following the manufacturer's protocol. An on-column DNAseI (Qiagen) digest was performed and the RNA was eluted in 30µl of nuclease-free water. 8µl of each sample was used for cDNA synthesis using Superscript III Supermix. The cDNA was diluted 1:3 and used for RT-PCR to check for *NkxUS*

enrichment. Primers for *Nkx2-5, Gapdh and Actb* (Table 13 and 15) were used to check for specific enrichment.

For protein elution, the beads of the protein fraction were collected and resuspended in 50µl biotin elution buffer (12.5mM biotin (Invitrogen), 7.5mM HEPES pH 7.5, 75mM NaCl, 1.5mM EDTA, 0.15% SDS, 0.075% sarkosyl (Sigma-Aldrich), and 0.02% Na-Deoxycholate (Sigma-Aldrich)). 50µl of biotin elution buffer was also added to the protein input. The samples were mixed at RT for 20min followed by 10min at 65°C. The elution was repeated once more with fresh biotin elution buffer. The two eluents were pooled, leading to in total 100µl of sample. Residual beads were removed again and 25% total volume of Trichlorotic acid (Sigma-Aldrich) was added to the samples. After vortexing, the samples were rotated overnight at 4°C for protein precipitation. The following day, the samples were centrifuged at 16,000 RCF for 60min at 4°C. The supernatant was removed and the pellets were washed with cold acetone (Sigma-Aldrich). After another spin for 5min at 4°C the acetone was removed completely and the pellet was air-dried. The pellet was the resuspended in 7.5µl of nuclease-free water and 2.5µl of 4x LDS sample buffer (Thermo Fisher Scientific). The samples were boiled for 30min at 95°C for reverse crosslinking and then loaded onto a 6% Bis-Tris gel (Thermo Fisher Scientific). The gel was run for 50min at 200V and afterwards stained with Coomassie Brilliant Blue G250 (Thermo Fisher Scientific) as follows. A dye stock solution was prepared consisting of 0.1% Coomassie Brilliant Blue G250, 2% ortho-phopshotic acid (Sigma-Aldrich) and 10% ammonium sulphate (Sigma-Aldrich). The gel was fixed with 40% ethanol, 10% acetic acid for one hour. Afterwards, the gel was washed twice in water for 10min. The gel was stained with Coomassie Brilliant Blue G250 working solution (80% stock solution, 20% methanol) overnight at RT. The next day the gel was destained with 1% acetic acid for three hours, changing the solution several times.

The lanes of the stained gel were cut into at least 4 pieces each. The samples were prepared and analysed using mass spectrometry by Dr. Ben Crosset (University of Sydney) as follows.

The excised gel fragments were destained in a 60:40 solution of 40mM NH_4HCO_3 (pH 7.8)/100% acetonitrile (MeCN) for 1 h. Gel pieces were vacuum-dried and the rehydrated with a 12ng/µL trypsin (Promega) solution at 4°C for one hour. Excess trypsin was removed and gel pieces covered with 40mM NH_4HCO_3 and incubated overnight at 37°C. Peptides were concentrated and desalted using C18 Zip- Tips (Millipore) as per the manufactures instructions. Peptides were resuspended in 25µl 3% (v/v) acetonitrile/0.1% (v/v) formic acid, and briefly sonicated. Samples were separated by nano-LC using an Ultimate 3000 HPLC and autosampler system (Thermo Fisher Scientific) coupled

to an in-house fritless nano 75µm × 40cm column packed with ReproSil Pur 120 C18 stationary phase (1.9µm, Dr, Maisch GmbH, Germany). LC mobile phase buffers were comprised of A: 0.1% (v/v) formic acid and B: 80% (v/v) acetonitrile/0.1% (v/v) formic acid. Peptides were eluted using a linear gradient of 5% B to 40% B over 60min and then 95% B wash over 1min at a flow rate of 250nL/min. The LC was coupled to a QExactive Plus Orbitrap mass spectrometer (Thermo Fisher Scientific). Column voltage was 2300V and the heated capillary set to 275°C. Positive ions were generated by electrospray and the Orbitrap operated in data-dependent acquisition mode. A survey scan of 350-1550m/z was acquired (resolution = 70,000, with an accumulation target value of 1,000,000 ions) with lockmass enabled. Up to 10 of the most abundant ions (>1.7e5 ions), with charge states ≥ +2 were sequentially isolated and fragmented and target value of 100,000 ions collected. Ions selected for MS/MS were dynamically excluded for 20s. The data were analysed using Proteome Discoverer vr 2.1 (Thermo Fisher Scientific) and Mascot vr 2.4 (Matrix Science). The search parameters included the following variable modifications: oxidized methionine and carbamidomethyl cysteine. The enzyme was set to trypsin, precursor mass tolerance was 10ppm while the fragment tolerance was 0.1Da. The database was the SwissProt database restricted to Mus musculus taxonomy.

4.6.3   ChIRP followed by DNA Sequencing

The DNA ChIRP was performed following the methods paper [235]. The method is similar to the one used for the protein ChIRP except that 1% glutaraldehyde (Sigma-Aldrich) was used for crosslinking and about 100 million cells per experiment were used. Otherwise, the method was followed as described for the protein ChIRP until the protein elution step. Instead of protein elution, DNA was isolated as follows. Each bead sample was resuspended in 150µl DNA elution buffer (50mM sodium bicarbonate (Sigma-Aldrich), 1% SDS), 10µl RNaseA (10mg/ml, Sigma-Aldrich) and 10µl RNaseH (10U/µl, Life Technologies). The samples were incubated at 37°C for 30min with shaking. Afterwards, beads and supernatant were separated using a magnetic stand. The supernatant was transferred to a new 1.5ml tube. This step of DNA elution was repeated once more. The supernatants were pooled, yielding in total 300µl of sample. 15µl of Proteinase K (Sigma-Aldrich) was added to each sample, mixed and incubated at 45°C for 50min with shaking. The DNA samples were transferred into yellow phase lock gel tubes. 300µl of Phenol:Chloroform:Isoamyl (Sigma-

Aldrich) was added per sample and incubated for 10min with vigorous shaking. The samples were spin down for 5min at full speed at 4°C and the aqueous phase was transferred into a new tube. 3µl GlycoBlue (Ambion), 30µl sodium acetate (Sigma-Aldrich) and 900µl 100% ethanol were added, the samples were vortexed and then stored overnight at -20°C. The next day the samples were spin down for 30min at 4°C at full speed. The supernatant was removed and 1ml 70% ethanol was added to the samples. After vortexing, the samples were centrifuged at full speed for 5min. The supernatant was removed and the pellet was left to air dry for 1min. The pellet was resuspended in 30µl elution buffer (Qiagen) and the samples were stored at – 20°C until library preparation. Efficient enrichment of *NkxUS* was determined by RT-PCR as described for the protein ChIRP. DNA from five biological replicates for EVEN, ODD, LacZ and DNA input were assessed for quantity and quality using a NanoDrop2000 (Thermo Fisher Scientific) and the Bioanalyser2100 (Agilent) using the high sensitivity DNA chip as per the manufacturer's protocol. Samples were end-repaired using Klenow DNA polymerase (NEB) and T4 DNA polymerase (NEB). Briefly, Klenow DNA polymerase was diluted 1:5 to 1U/µl and added to the reaction mix (Table 21). The sample mix was incubated at 20°C for 30min and then the DNA was purified using 90µl of AMPure XP beads and resuspended in 8.75µl of resuspension buffer. End-repaired samples were then used in the Illumina TruSeq stranded mRNA kits from the 3' adenylation step using half-reactions. The library preparation was performed by Dominik Kaczorowski (Garvan Institute of Medical Research, Sydney, Australia). The samples were sequenced over two lanes using 125bp paired-end and the Illumina HiSeq2500 v4 platform in high output mode at the Kinghorn Centre for Clinical Genomics (Garvan Institute of Medical Research, Sydney, Australia)

Table 21 Reaction mixture for DNA end repair

ChIRP DNA	30µl
Nuclease-free water	10µl
10x T4 DNA Polymerase buffer	5µl
dNTP mix	2µl
T4 DNA Polymerase (10U/µl)	1µl
Klenow DNA Polymerase (5U/µl)	1µl
T4 PNK (10U/µl)	1µl
Total	50µl

4.7 Animal experimentation

Animal experimentation was performed with approval of the Garvan Institute/St Vincent's Hospital Animal Ethics Committee (Project numbers 16/03 and 16/10). All mice were kept on a full C57BL/6J background and housed at the BioCore facilities in the Victor Chang Cardiac Research Institute. Mice were sacrificed by CO_2 inhalation.

4.7.1 Experimental mouse model

For the investigation of the conserved region in *NkxUS*, mice lacking part of this region were generated using CRISPR/Cas9 (Figure 33). This mouse line was termed NkxUSΔ and is on a C57BL/6J background. Genome editing using CRISPR may have off-target mutation sites [342, 343]. Typically, mice are backcrossed for several generations to remove off-targets. However, in the interest of time we crossed the founder females with C57BL/6J mice to expand the colonies and then crossed heterozygous offspring of NkxUSΔ_F5 to the ones of NkxUSΔ_F10 to limit the chance of off-target effects. In the meantime, NkxUSΔ-F5 and NkxUSΔ-F10 lines were continued to backcross until the 10th generation to C57BL/6J mice to ensure removal of off-target sites. For all experiments described in this thesis, animals generated from intercrossing NkxUSΔ-F5 and NkxUSΔ-F10 were used and are referred to as NkxUSΔ. NkxUSΔ homozygous (NKXUSΔ$^{-/-}$) and NkxUSΔ wildtype (NKXUSΔ$^{+/+}$) mice were analysed in the following experiments.

4.7.2 Mouse genotyping

For genotyping, genomic DNA was extracted from tail clips using Extract-N-AmpTM Tissue PCR Kit (Sigma-Aldrich) following manufacturer's protocol. PCR was performed on tail clip DNA using two different primer pairs (USD1 and USD2) listed in Table 22, which span the deletion. The REDExtract-N-Amp PCR ReadyMix (Sigma-Aldrich) was used for the PCR and set up as follows:

REDExtract-N-Amp PCR ReadyMix	5µl
DNA	1µl
10µM Forward primer	1µl
10µM Reverse primer	1µl
nuclease-free water	2µl
Total	10 µl

The samples were run with the following PCR conditions:

Temperature	Duration	Cycles
94°C	3min	1
95°C	15 s	
63°C	1min	34x
72°C	1min	
72°C	10min	1

Table 22 Genotyping primers (FW=forward; RV= reverse)

USD1	FW: AGATTCCCAGCAGTCCGTAG
	RV: GCAGTGCCAGTAACAAGGTC
USD2	FW: TAAGGGCTGTAAAGCACGG
	RV: CCAGTAACAAGGTCCTCCAC

4.7.3 *In situ* hybridisation

In situ hybridisation was performed on C57BL/6J mice using the same RNA probes as for the Northern blot (Table 16) and based on a protocol described previously [344]. Tissues were collected and fixed overnight in 4% paraformaldehyde (PFA) in 1x PBS by rocking 4°C. Tissues were dehydrated in a graded ethanol series, embedded in paraffin and sectioned. Sectioned tissues were deparaffinised and rehydrated, followed by proteinase K treatment (20µg/mL, Roche) in 1x PBS for 10min at 37°C. Slides were washed two times 5min with 1x PBS, post-fixed in 4% PFA for 5min at RT and washed once with 1x PBS fro 5min. The tissues were permeabilised with 0.07M HCl for 15min at RT. Afterwards, the slides were washed washed two times 5min with 1x PBS and incubated with 0.25% acetic acid in 0.1M triethanolamine (pH8, Sigma Aldrich) for 10min at RT to block non-specific probe binding. The slides were washed once for 5min with 1x PBS and once with nuclease-free water. The slides were incubated in hybridisation buffer (50% formamide, 5 × SSC, 1% block solution (Roche), 5mM EDTA, 0.1% Tween-20, 0.1% Chaps (Sigma-Aldrich), 0.1mg/ml heparin

(Becton-Dickinson), and 1mg/ml yeast total RNA (Roche)) without probe for 2hrs at 70°C in a humidified chamber and then with probe (1ng/µl) overnight at 70°C. The next day, the slides were briefly rinsed in 2x SSC and washed three times for 30min in 50% formamide in 2x SSC at 65°C, followed by three times 5min washed with PBST. Probe bound to the section was immunologically detected using sheep anti-digoxigenin Fab fragment covalently coupled to alkaline phosphatase (Sigma-Aldrich) overnight at 4°C, according to the manufacturer's protocol.

The next day, slides were washed twice for 10min with MABT at RT, then three times for 1h at RT. BM purple (Sigma-Aldrich) was to the slides and incubated until staining developed. The slides were washed two times for 5min in 1x PBS, then incubated in 4% PFA at RT for 20min and washed again in 1x PBS for 5min. The slides were dehydrated using a graded ethanol series and mounted using DPX (Sigma-Aldrich). Images were taken using a Leica microscope.

4.7.4   RNA extraction from mouse tissues

Tissues were collected from mice and flash frozen. Tissues were homogenized in 1ml Trizol for 30-50s until the samples were completely homogenized. Total RNA was extracted as described in 2.2.1. The extracted RNA was treated with DNAseI and subsequently cleaned up using the RNA Clean and Concentrator kit as described in 2.2.2 and 2.2.3. RNA was eluted in 10-15µl. After measuring the RNA concentration using the NanoDrop2000, the RNA was stored at -80°C.

4.7.5   Optical projection tomography (OPT)

Embryonic day (E) 17.5 old NkxUSΔ$^{-/-}$ and NkxUSΔ$^{+/+}$ mouse embryos were analysed by OPT for heart defects. Heart were dissected in 1x PBS containing 40mM potassium chloride (KCl) and 1mM Heparin (Clifford Hallam Healthcare). Following dissection, the hearts were incubated in 3% glutaraldehyde (Sigma-Aldrich) for at least one hour at RT. Afterwards, the hearts were transferred to 70% methanol (MeOH), followed by 85% and two times 100%, each for at least one hour, before placing in fresh 100% MeOH overnight. The next day the hearts were optically cleared using a 1:2 mixture of benzyl alcohol and benzyl benzoate (BA:BB) for at least one hour. The BA:BB was replaced with fresh one and the hearts were left to clear overnight. The next day, the samples were

analysed using the OPT scanner by placing a single heart in the tip of a cut-off and sealed glass pipette, which was then fixed to the mounting post of the OPT. OPT scanning was performed using the FITC channel and a custom microscope controlled by the software OPTimum (James Springfield, Institute for Molecular Biosciences, University of Queensland, Australia). About 800 images were obtained per heart. The images were reconstructed using the software Nrecon (Bruker microCT) and heart morphology analysed using the software AMIRA (version 5.5.0, FEI Visualization Services Group).

4.7.6 Atrial septal morphology analysis

Twenty 6-8 week old NkxUS$\Delta^{-/-}$ and NkxUS$\Delta^{+/+}$ mice were anatomically dissected by Dr. Edwin Kirk (University of New South Wales, Sydney) to analyse the atrial septal morphology following his published article [249]. Briefly, the mouse body weight was measured before the mice were euthanized using CO_2 chambers. The mouse chest was opened and the heart was removed and placed into a petridish containing 1x PBS. The left atrium (LA) was opened to analyse for atrial septal defects (ASD), patent foramen ovale (PFO), crescent width (CRW), flap valve length (FLV), foramen ovale width (FOW). CRW, FLV and FOW were measured using an eyepiece graticule. To determine whether a PFO was present, the right atrium (RA) was pressurised. If remaining blood contained within the RA would pass across the septum into the LA, then a PFO was present. To measure the atrial septal anatomy, the cut edges of the LA were hold by dissection forceps to expose the atrial septum and quantitative parameters were measured using an eyepiece graticule as in Figure 47. Care was taken to maintain a consistent amount of stretch for all the animals. Dr. Edwin Kirk was kept blinded to the genotype throughout the whole experiment.
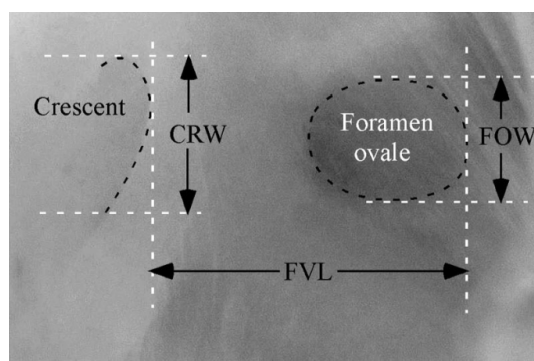
Figure 47 Atrial septal morphology analysis. Measurements of quantitative parameters are shown for CRW, FVL ad FOW.

4.7.7 Thoracic Aortic Constriction surgery (TAC)

TAC is a surgery in which one of the major arteries leading from the heart, the aorta, is constricted. Constriction causes an increase in the resistance to pumping for the left ventricle (LV) leading to LV hypertrophy (growth). In this model, the inside diameter of the aorta close to the LV was reduced by tying a suture around it.

10-12 week old male NkxUSΔ-/- and NkxUSΔ+/+ mice were used for this study. The surgery was performed by Jianxin Wu (Victor Chang Cardiac Research Institute), an experienced small animal surgeon. Anaesthesia was induced via an intraperitoneal injection of xylazine and ketamine mixed in the same syringe (mice: ketamine 80-100mg/kg, xylazine 10-20mg/kg and atropine 0.02-0.05mg/kg; Troy Laboratories). Xylazine is a muscle relaxant and was used to ensure safe intubation. Atropine is an anticholinergic agent, not an anaesthetic agent, and is used in mice to counteract ketamine-xylazine-induced bradycardia (slow heart rates). Following induction, the mice were shaved on the belly side and intubated using a 20G cannula (Insyte BD). After intubation, the animal was transferred to the warming pad and a ventilator is connected to the intubation cannula. Anaesthesia was maintained throughout the surgical procedure by 1-2% of isoflurane. Once an adequate plane of anaesthesia was reached a 1-2cm horizontal skin incision was made over the suprasternal notch. The thyroid was retracted and a small incision was made into the muscles overlying the sternum. A longitudinal cut (approximately 2mm) was then made through the proximal sternum. The ribs were retracted to allow visualisation of the thoracic organs under low power magnification. The aorta was separated between the right and left carotid arteries and

isolated from the surrounding tissues. Once the vessel was isolated, forceps are passed underneath to capture a ligature, which is then pulled back. A 25-gauge shortened and blunted needle was placed on top of the artery and the silk suture was then tied firmly around the needle and the artery. The needle was subsequently removed as quickly as possible to leave the internal diameter of the vessel the size of the needle. The thoracic muscles was closed in one layer by horizontal mattress pattern using Prolene suture material as appropriate to the size of the animal. Air was evacuated from the chest by briefly over-inflating the lungs by occluding the gas-out arm of the ventilation circuit. The skin was closed with one cruciate suture in Prolene as appropriate to the size of the animal. At this point all surgical interventions were complete. A sham procedure is exactly the same except that no ligature was tied around the aorta. Following surgery, the mice were placed in a recovery cage overnight. The next day, the mice were transferred back into their home cage. The mice were housed singly for the whole experiment and were monitored until completion of the experiment. 21 days after surgery cardiac function, heart size and blood flow were analysed in real time using echocardiography. Following echocardiography, chamber pressure was measured using LV micromanometry. After completion, the mice were sacrificed and the hearts and tibias were collected. After the whole hearts were weighted, the LV and LA were dissected and weighted. The tibia length was determined using a Vernier caliper (RS pro). Jianxin Wu was kept blinded to the genotype for the whole experiment.

4.7.8 Echocardiography

Echocardiography is the non-invasive method for imaging the heart in real time using ultrasound. We used a high frequency probe operating at a central frequency of 30MHz (MX 400, FujiFilm Visualsonics). To induce anaesthesia, the mouse was placed in an induction chamber with 3-4% isoflurane and when unconscious was transferred to a warmed platform with integrated ECG amplifier. Anaesthesia was maintained by 1-2% isoflurane inhalation via a nose cone. The animal was positioned on its back with each foot held by tape above each ECG electrode which were then connected electrically individually with ultrasound gel, which is conductive. The chest was shaved and washed with water to minimise hair and ultrasound gel applied to enable transmission of the ultrasound beam. The echo probe is brought into contact with the gel and echo images were obtained and recorded.

The warming platform is on a moveable table in the horizontal plane (X and Y axes) controlled by micrometre mechanisms (rack and pinion). The probe is held above the animal in a circular clamp on an armature that is independent of the table and can be adjusted vertically. The circular clamp allows the probe to rotate about its central vertical axis.

The first image acquired was the 2D B-mode image of the parasternal long axis (LAX) and was achieved with the probe beam aligned with the Y axis of the table and the mouses heart aligned with the beam, with the aortic valve and the apical dimple in view and orthogonal to the beam. Next the short axis B-mode image was acquired by rotating the probe in its circular clamp 90° and adjusting the table in the Y axis to observe the mid-papillary plane of the LV. By moving the table so the probe moves toward the head of the mouse the pulmonary artery (PA) bifurcation was observed just beyond the aortic annulus and the pulse wave Doppler (PW) voxel is located proximal to the bifurcation to limit turbulence. The LA was imaged in B-mode from a modified LAX orientation with the beam set orthogonal to the posterior wall of the LA where it aligns with the aortic valve. Aortic PW data was obtained from a supra-sternal view with the voxel located above the aortic valve in line with the crossing of the PA. Mitral valve (MV) PW and MV annulus tissue Doppler (TD) was observed from a sub-costal 4-chamber view.

LV mass and volumes were derived using a modified bullet formula (V = 5 x LAX length x SAX area / 6) where SAX area was obtained by planimetry of the border and LAX length taken from the inferior point of the valve leaflet co-optation to the apical dimple. The conservation of mass rule (LV mass at diastole = LV mass at systole) was applied when determining LV mass and was determined first providing an internal control to the chamber volume estimates. Papillary muscles were excluded from the internal chamber border planimetry estimates. Echocardiography was performed by Scott Kesteven (Victor Chang Cardiac Research Institute) who was kept blinded for the whole study.

4.7.9 LV Micromanometry

Micromanometry uses a transducer tipped catheter which is inserted into the LV to measure the chamber pressure in anaesthetized mice allowing the determination of ventricular systolic and diastolic function. LV micromanometry was performed by Jianxin Wu. Anaesthesia was induced and maintained as for Echocardiography. Once the mouse was anaesthetized, a midline incision was

made in the ventral neck and the submandibular glands are gently separated. The right carotid artery was isolated and retracted using two sutures placed proximally and distal around the artery. A second proximal ligature was looped around the vessel between the other two sutures. The loop and two free ends were tight together with a loose knot. The artery was elevated and ligated with a distal suture. Gentle tension was applied to the first proximal suture to occlude blood flow. Using a bent-tipped 25-gauge syringe needle as a catheter introducer, the carotid artery was punctured close to the distal suture and the catheter was inserted to a point beyond the second proximal suture. A knot was firmed tightly around the vessel and catheter to minimize blood flow. The first proximal ligature was released and the catheter tip advanced into the vessel and progressed into the LV. LV pressure was recorded for several minutes. After completion, mice were euthanized while still under deep anaesthesia.

4.7.10 Telemetry

A small transmitter was implanted by Jianxin Wu into the subcutaneous space of male NkxUSΔ$^{-/-}$ and NkxUSΔ$^{+/+}$ mice (10-12 weeks of age) for the long term recording of physiological parameters. For the implantation, anesthesia was induced and maintained by isoflurane inhalation as per LV micromanometry.

We used the PhysioTel ETA-F10 transmitter (Data Sciences International, DSI), which is currently the smallest of the single channel bipotential transmitter series available. The device was placed subcutaneously on the left flank of the mice between the fore and hind limb. The ETA-F10 is fitted with flexible leads that extend from the transmitter body. Two incisions were made for the leads. The positive one was be located on the lower left of chest wall and made big enough to enable insertion of the transmitter. The negative lead incision was made on the upper right side of the chest. After inserting the transmitter, the leads were tunneled subcutaneously to their incision site and secured to the skeletal muscle to serve as sensing electrodes across the heart on the chest wall. All incisions were closed and the animal was placed overnight in a recovery cage.

After surgery, the mice were housed singly for 3 days to recover from the surgery. At day 4 post-surgery the ECG was recorded for 24 hrs. While recording, the mice were in their normal home cage. A receiving plate was placed under the cage of each implanted mouse, the transmitter was activated using a magnet and recording was started using the Ponemah software (DSI).

After completion of recording, mice were sacrificed using CO_2 chambers. Transmitters were recovered from euthanised animals and prepared for reuse following the manufacturer's recommendation. Briefly, the transmitters were cleaned from as much tissue as possible and were soaked overnight in an enzymatic cleaner (Edwards group holdings). The next day, the implants were gently cleaned using a soft brush, rinsed in water and then soaked again if any remaining tissue was attached. After removal of all tissue, the transmitters were sterilized in Actril (OnBoard Solutions) overnight, then rinsed in water and stored in sterile packaging until further use. Telemetry data was analysed using the Ponemah ECG pro and Data Insights softwares (DSI).

4.7.11   Dissection of the sinoatrial node (SAN)

Mice were sacrificed using CO_2 chambers. The hearts along with lungs and thymus were dissected from the mouse's chest and washed in warmed Tyrode solution (140mM NaCl, 5.4mM KCl, 5mM Hepes (Life Technologies), 5.5mM glucose, 1mM MgCl2, 1.8mM CaCl2, pH7.4). The tissue was transferred to a with silicon (Dow Corning) coated dish containing warmed Tyrode solution. The heart was oriented so that the posterior vessels were facing up, with the animals' RA on the right and immobilised by pinning through the ventricles, lungs and thymus (Figure 48A). The vessels were separated so that the inferior vena cava (IVC) and superior vena cava (SVC) are clearly visible (Figure 48B). Both vessels were cleaned from fat tissue as best as possible (Figure 48B). Then the ventricles, lungs, thymus and left atrium were removed. The tissue was re-pinned through the IVC, SVC and RA. Any remaining ventricular tissue was removed. The anterior wall of the RA and vena cava was opened using dissection scissors. The pins were repositioned to stretch the tissue gently revealing the sinoatrial node region, which lies in between the interatrial septum, RA, IVC and SVC (Figure 48C).
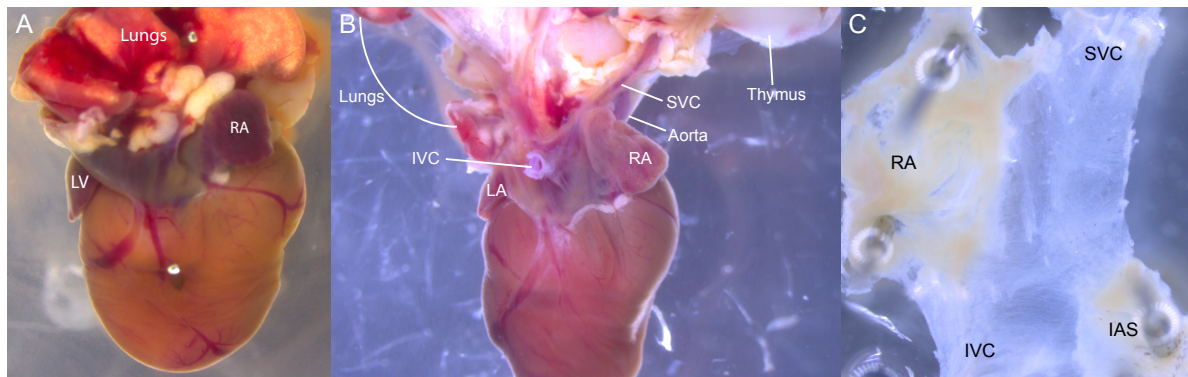
Figure 48 Dissection of the mouse SAN. A) After dissecting the heart from the mouse's chest, it was pinned to a siliconised dish, through the lungs, ventricle and thymus. B) Vessels were separated, cleaned from fat and stretched apart to free the SVC and IVC. C) The RA along with SVC, IVC and part of the IAS were dissected from the heart, cut open and spread open to reveal the SAN region. RA= right atrium, LA= left atrium, SVC= superior vena cava, IVC= inferior vena cava, IAS= interatrial septum.

4.7.11.1   Calcium imaging of the SAN

The SAN's were dissected as described in 4.8.11. Dissected tissues were transferred to a with ~2mM silicon (Dow Corning) coated petri dish and warmed Tyrode solution. The tissue was pinned to the new dish with the endocardial surface facing down. To stain the tissue, 10µM Cal-520 dye (AAT Bioquest) in 2ml Tyrode solution was added and the tissue was incubated for one hour shaking gently at RT in the dark. Afterwards, the tissue was washed three times with dye free warmed Tyrode and then incubated for 30min in 2ml of warm Tyrode with 5µM Blebbistatin (Sigma-Aldrich). The tissue was imaged on the Eclipse Ti2 inverted microscope (Nikon), while perfused at a rate of 8 ml/min with warmed Tyrode containing 5µM Blebbisstatin. The Tyrode solution was either kept at 37°C. During the whole experiment, perfusion was never stopped. Before capturing the first video to measure the heart rate, the tissue was allowed to adjust to the perfusion for 15min. Videos were taken for 5s, unless otherwise stated, with 8x8 binning, 5ms exposure time and 100-120 frames per second (fps). First, a video was taken at a 2x magnification visualizing the entire tissue preparation. After, identifying the SAN region, an additional video was taken at 10x magnification. The heart rate was determined by counting the calcium spikes within the 5s video and multiplying by 12 to determine the heart rate in beats per minute (bpm).

4.7.11.2   SAN tissue pacing

Following identification of the SAN region, the SAN tissue was paced while imaging to determine the SAN recovery time (SANRT). Two silver wires attached to two electrodes were placed on opposite sides of the SAN tissue, one close to the inferior and one close to the superior vena cava. The tissue was paced for 20s at 50V with a width of 0.2ms and at an interval of 0.1ms. A video was taken for 30s with the pacing starting after 5 s. The SANRT was calculated as the time interval between the last stimulation spike and the first spontaneous one. Corrected SNRT (cSNRT) was calculated as the difference between the SNRT and the cycle length immediately before pacing.

4.7.11.3   SAN drug administration

The effects of two drugs were tested, Ivabradine (IVA, Sigma-Aldrich) and Isoproterenol (ISO, Sigma-Aldrich). IVA was used at a concentration of 3µM and ISO at a concentration of 1µM. IVA is a selective inhibitor of the pacemaker specific HCN channels and is thus a heart rate reducing agent. ISO is a ß-adrenergic receptor agonist and thus a heart rate increasing agent. The drugs were added to the perfusion and after 10min for IVA and 2min for ISO a video was taken to measure the heart rate changes due to the drugs. The tissue was washed with fresh Tyrode solution for at least 1h following the first drug treatment.

4.8   Bioinformatics analyses

4.8.1 Structure alignment with *Foldalign*

FoldAlign was run with the parameters `-plot_score -max_length 500 -no_backtrack` to align human and mouse *NkxUS* [212]. Local non-overlapping hits were identified with *locateHits*. Afterwards, *FoldAlign* was run with parameters `-plot_score -max_length 500 -no_backtrack` to align human *TCONS_00010558* and 1,000 di-nucleotide mouse *NkxUS* sequences, and to estimate kappa and lambda of an extreme value distribution describing the alignment score distribution. Based on this distribution, p-values were calculated for the real alignments.

RNAfold with the parameters *–p –d2* was run for the identified structural human alignments to analyse for common domains between the conserved structures and thermodynamic structures.

4.8.2   Impact of SNP prediction

RNAsnp version 1.1 was run in global mode (parameters `-m 1 -M 1 -w 100`) on the second exon of human *NkxUS* containing the *rs6882776* SNP [223].

4.8.3   RNA sequencing analysis

500ng of RNA was used per sample. The libraries were prepared using the TruSeq Library Prep kit (Illumina) and were sequenced on the Illumina HiSeq on two lanes.

After sequencing, reads from the two lanes were merged. FastQC (version 0.11.3) was used to assess the initial quality of the collated read files. Adapters sequences, over-represented reads, 5'GC bias and low-quality reads were trimmed using trimmomatic (version 0.32) and quality was checked again using fastqc [345]. Reads were then aligned to the mouse genome (mm10)using STAR (version 2.5.1a) [346]. Bam files were indexed and sorted using samtools (version 1.2.5) [347]. Read counts from the indexed bam files were quantified using RSEM (version 1.3.0) [348]. Post alignment mapped files were quality checked with RSeQC (version 2.6.1), which collects multiple statistics including gene body coverage, read distributions, quality and error rates, duplication rates, deletion rates, insertion rates, junction coverage and saturation. [349]. The read counts results from RSEM were collated and were analysed using DESeq2 for differential expression in R-Studio (version 1.1.419) [229].

4.8.4   DNA ChIRP analysis

The analysis of the DNA sequencing of the ChIRP samples was performed by Thomas Kavanagh (Garvan Institute of Medical Research, Sydney, Australia). Short reads were checked for quality and trimmed as described in section 2.4.1. The reads were then mapped to the mouse genome (mm10)

using STAR as detailed in section 2.4.1. Read distribution was assessed with RSeQC v2.6.1 using the *read_distribution.py* script. Peaks were called using MACS v2.0.10. [350]. BAM files were used as input with an EVEN or ODD pool as the test dataset and the corresponding replicates LacZ pool as the control data. Separately all sample peaks were called with MACS without using LacZ as the background. Peak files were intersected between the EVEN and ODD pool and then to the LacZ for each replicate using bedtools v2.22.0 (bedtools window function). Any peaks within 2kb of each other, in both the EVEN and ODD pool, were considered as true peaks if they didn't intersect a peak in the LacZ controls.

4.9 Statistics

GraphPad Prism software (Version 7) was used for all statistical analyses. For all comparisons of means, unpaired student *t*-test was performed and standard deviation was used as the measure of spread.

References

1. Buckingham, M., S. Meilhac, and S. Zaffran, *Building the mammalian heart from two sources of myocardial cells.* Nat Rev Genet, 2005. **6**(11): p. 826-35.

2. Gavaghan, M., *Cardiac anatomy and physiology: a review.* AORN J, 1998. **67**(4): p. 802-22; quiz 824-8.

3. Flanigan, M. and S.M. Gaskell, *A review of cardiac anatomy and physiology.* Home Healthc Nurse, 2004. **22**(1): p. 45-51.

4. Sternberg, C.N., et al., *Abiraterone acetate for patients with metastatic castration-resistant prostate cancer progressing after chemotherapy: final analysis of a multicentre, open-label, early-access protocol trial.* Lancet Oncol, 2014. **15**(11): p. 1263-8.

5. Anderson, R.H., R. Razavi, and A.M. Taylor, *Cardiac anatomy revisited.* J Anat, 2004. **205**(3): p. 159-77.

6. Fukuta, H. and W.C. Little, *The cardiac cycle and the physiologic basis of left ventricular contraction, ejection, relaxation, and filling.* Heart Fail Clin, 2008. **4**(1): p. 1-11.

7. Pinto, A.R., et al., *Revisiting Cardiac Cellular Composition.* Circ Res, 2016. **118**(3): p. 400-9.

8. Brutsaert, D.L., *Cardiac endothelial-myocardial signaling: its role in cardiac growth, contractile performance, and rhythmicity.* Physiol Rev, 2003. **83**(1): p. 59-115.

9. Drawnel, F.M., C.R. Archer, and H.L. Roderick, *The role of the paracrine/autocrine mediator endothelin-1 in regulation of cardiac contractility and growth.* Br J Pharmacol, 2013. **168**(2): p. 296-317.

10. Anderson, R.H., et al., *The anatomy of the cardiac conduction system.* Clin Anat, 2009. **22**(1): p. 99-113.

11. Lo, C.W., *Role of gap junctions in cardiac conduction and development: insights from the connexin knockout mice.* Circ Res, 2000. **87**(5): p. 346-8.

12. Rohr, S., *Role of gap junctions in the propagation of the cardiac action potential.* Cardiovasc Res, 2004. **62**(2): p. 309-22.

13. van Weerd, J.H. and V.M. Christoffels, *The formation and function of the cardiac conduction system.* Development, 2016. **143**(2): p. 197-210.

14. Opthof, T., *The mammalian sinoatrial node.* Cardiovasc Drugs Ther, 1988. **1**(6): p. 573-97.

15. Boyett, M.R., H. Honjo, and I. Kodama, *The sinoatrial node, a heterogeneous pacemaker structure.* Cardiovasc Res, 2000. **47**(4): p. 658-87.

16. Moorman, A.F. and V.M. Christoffels, *Cardiac chamber formation: development, genes, and evolution.* Physiol Rev, 2003. **83**(4): p. 1223-67.

17. Christoffels, V.M., et al., *Development of the pacemaker tissues of the heart.* Circ Res, 2010. **106**(2): p. 240-54.

18. Tse, G., *Mechanisms of cardiac arrhythmias.* J Arrhythm, 2016. **32**(2): p. 75-81.

19. DiFrancesco, D., *Characterization of single pacemaker channels in cardiac sino-atrial node cells.* Nature, 1986. **324**(6096): p. 470-3.

20. DiFrancesco, D., *The cardiac hyperpolarizing-activated current, if. Origins and developments.* Prog Biophys Mol Biol, 1985. **46**(3): p. 163-83.

21. Baruscotti, M., A. Bucchi, and D. Difrancesco, *Physiology and pharmacology of the cardiac pacemaker ("funny") current.* Pharmacol Ther, 2005. **107**(1): p. 59-79.

22. Mangoni, M.E., et al., *Functional role of L-type Cav1.3 Ca2+ channels in cardiac pacemaker activity.* Proc Natl Acad Sci U S A, 2003. **100**(9): p. 5543-8.

23. Bohn, G., et al., *Expression of T- and L-type calcium channel mRNA in murine sinoatrial node.* FEBS Lett, 2000. **481**(1): p. 73-6.

24. Mesirca, P., A.G. Torrente, and M.E. Mangoni, *Functional role of voltage gated Ca(2+) channels in heart automaticity.* Front Physiol, 2015. **6**: p. 19.

25. Lakatta, E.G., V.A. Maltsev, and T.M. Vinogradova, *A coupled SYSTEM of intracellular Ca2+ clocks and surface membrane voltage clocks controls the timekeeping mechanism of the heart's pacemaker.* Circ Res, 2010. **106**(4): p. 659-73.

26. Aziz, Q., Y. Li, and A. Tinker, *Potassium channels in the sinoatrial node and their role in heart rate control.* Channels (Austin), 2018. **12**(1): p. 356-366.

27. Unudurthi, S.D., R.M. Wolf, and T.J. Hund, *Role of sinoatrial node architecture in maintaining a balanced source-sink relationship and synchronous cardiac pacemaking.* Front Physiol, 2014. **5**: p. 446.

28. Joung, B., et al., *The calcium and voltage clocks in sinoatrial node automaticity.* Korean Circ J, 2009. **39**(6): p. 217-22.

29. Cingolani, E., J.I. Goldhaber, and E. Marban, *Next-generation pacemakers: from small devices to biological pacemakers.* Nat Rev Cardiol, 2018. **15**(3): p. 139-150.

30. Bogdanov, K.Y., T.M. Vinogradova, and E.G. Lakatta, *Sinoatrial nodal cell ryanodine receptor and Na(+)-Ca(2+) exchanger: molecular partners in pacemaker regulation.* Circ Res, 2001. **88**(12): p. 1254-8.

31. Rossi, A.E. and R.T. Dirksen, *Sarcoplasmic reticulum: the dynamic calcium governor of muscle.* Muscle Nerve, 2006. **33**(6): p. 715-31.

32. Philipson, K.D. and A.Y. Nishimoto, *Na+-Ca2+ exchange is affected by membrane potential in cardiac sarcolemmal vesicles.* J Biol Chem, 1980. **255**(14): p. 6880-2.

33. DiFrancesco, D. and C. Tromba, *Acetylcholine inhibits activation of the cardiac hyperpolarizing-activated current, if.* Pflugers Arch, 1987. **410**(1-2): p. 139-42.

34. Mangoni, M.E. and J. Nargeot, *Genesis and regulation of the heart automaticity.* Physiol Rev, 2008. **88**(3): p. 919-82.

35. Behar, J., et al., *The Autonomic Nervous System Regulates the Heart Rate through cAMP-PKA Dependent and Independent Coupled-Clock Pacemaker Cell Mechanisms.* Front Physiol, 2016. **7**: p. 419.

36. Kemp, C.D. and J.V. Conte, *The pathophysiology of heart failure.* Cardiovasc Pathol, 2012. **21**(5): p. 365-71.

37. Pasotti, M., F. Prati, and E. Arbustini, *The pathology of myocardial infarction in the pre- and post-interventional era.* Heart, 2006. **92**(11): p. 1552-6.

38. Sutton, M.G. and N. Sharpe, *Left ventricular remodeling after myocardial infarction: pathophysiology and therapy.* Circulation, 2000. **101**(25): p. 2981-8.

39. Towbin, J.A. and N.E. Bowles, *The failing heart.* Nature, 2002. **415**(6868): p. 227-33.

40. Sisakian, H., *Cardiomyopathies: Evolution of pathogenesis concepts and potential for new therapies.* World J Cardiol, 2014. **6**(6): p. 478-94.

41. Burke, M.A., et al., *Clinical and Mechanistic Insights Into the Genetics of Cardiomyopathy.* J Am Coll Cardiol, 2016. **68**(25): p. 2871-2886.

42. Hershberger, R.E. and J.D. Siegfried, *Update 2011: clinical and genetic issues in familial dilated cardiomyopathy.* J Am Coll Cardiol, 2011. **57**(16): p. 1641-9.

43. Muchtar, E., L.A. Blauwet, and M.A. Gertz, *Restrictive Cardiomyopathy: Genetics, Pathogenesis, Clinical Manifestations, Diagnosis, and Therapy.* Circ Res, 2017. **121**(7): p. 819-837.

44. Burkett, E.L. and R.E. Hershberger, *Clinical and genetic issues in familial dilated cardiomyopathy.* J Am Coll Cardiol, 2005. **45**(7): p. 969-81.

45. Li, Y.G., et al., *Cardiac Arrhythmias: Update on Mechanisms and Clinical Managements.* Cardiol Res Pract, 2016. **2016**: p. 8023723.

46. Antzelevitch, C. and A. Burashnikov, *Overview of Basic Mechanisms of Cardiac Arrhythmia.* Card Electrophysiol Clin, 2011. **3**(1): p. 23-45.

47. Cooney, M.T., et al., *Elevated resting heart rate is an independent risk factor for cardiovascular disease in healthy men and women.* Am Heart J, 2010. **159**(4): p. 612-619 e3.

48. Tverdal, A., V. Hjellvik, and R. Selmer, *Heart rate and mortality from cardiovascular causes: a 12 year follow-up study of 379,843 men and women aged 40-45 years.* Eur Heart J, 2008. **29**(22): p. 2772-81.

49. Torrente, A.G., et al., *Burst pacemaker activity of the sinoatrial node in sodium-calcium exchanger knockout mice.* Proc Natl Acad Sci U S A, 2015. **112**(31): p. 9769-74.

50. Harvey, R.P., *Patterning the vertebrate heart.* Nat Rev Genet, 2002. **3**(7): p. 544-56.

51. Moorman, A., et al., *Development of the heart: (1) formation of the cardiac chambers and arterial trunks.* Heart, 2003. **89**(7): p. 806-14.

52. Kelly, R.G., M.E. Buckingham, and A.F. Moorman, *Heart fields and cardiac morphogenesis.* Cold Spring Harb Perspect Med, 2014. **4**(10).

53. Vincent, S.D. and M.E. Buckingham, *How to make a heart: the origin and regulation of cardiac progenitor cells.* Curr Top Dev Biol, 2010. **90**: p. 1-41.

54. Epstein, J.A., *Franklin H. Epstein Lecture. Cardiac development and implications for heart disease.* N Engl J Med, 2010. **363**(17): p. 1638-47.

55. Kelly, R.G., *The second heart field.* Curr Top Dev Biol, 2012. **100**: p. 33-65.

56. Anderson, R.H., et al., *Development of the heart: (2) Septation of the atriums and ventricles.* Heart, 2003. **89**(8): p. 949-58.

57. Mitchell, S.C., S.B. Korones, and H.W. Berendes, *Congenital heart disease in 56,109 births. Incidence and natural history.* Circulation, 1971. **43**(3): p. 323-32.

58. Dunwoodie, S.L., *Combinatorial signaling in the heart orchestrates cardiac induction, lineage specification and chamber formation.* Semin Cell Dev Biol, 2007. **18**(1): p. 54-66.

59. Hoffman, J.I. and S. Kaplan, *The incidence of congenital heart disease.* J Am Coll Cardiol, 2002. **39**(12): p. 1890-900.

60. Tchervenkov, C.I., et al., *The nomenclature, definition and classification of hypoplastic left heart syndrome.* Cardiol Young, 2006. **16**(4): p. 339-68.

61. Apitz, C., G.D. Webb, and A.N. Redington, *Tetralogy of Fallot.* Lancet, 2009. **374**(9699): p. 1462-71.

62. Olson, E.N., *Gene regulatory networks in the evolution and development of the heart.* Science, 2006. **313**(5795): p. 1922-7.

63. Cripps, R.M. and E.N. Olson, *Control of cardiac development by an evolutionarily conserved transcriptional network.* Dev Biol, 2002. **246**(1): p. 14-28.

64. Pashmforoush, M., et al., *Nkx2-5 pathways and congenital heart disease; loss of ventricular myocyte lineage specification leads to progressive cardiomyopathy and complete heart block.* Cell, 2004. **117**(3): p. 373-86.

65. Davidson, E.H. and D.H. Erwin, *Gene regulatory networks and the evolution of animal body plans.* Science, 2006. **311**(5762): p. 796-800.

66. Benson, D.W., et al., *Mutations in the cardiac transcription factor NKX2.5 affect diverse cardiac developmental pathways.* J Clin Invest, 1999. **104**(11): p. 1567-73.

67. Harvey, R.P., *NK-2 homeobox genes and heart development.* Dev Biol, 1996. **178**(2): p. 203-16.

68. Lints, T.J., et al., *Nkx-2.5: a novel murine homeobox gene expressed in early heart progenitor cells and their myogenic descendants.* Development, 1993. **119**(3): p. 969.

69. Kim, Y. and M. Nirenberg, *Drosophila NK-homeobox genes.* Proc Natl Acad Sci U S A, 1989. **86**(20): p. 7716-20.

70. Bodmer, R., *The gene tinman is required for specification of the heart and visceral muscles in Drosophila.* Development, 1993. **118**(3): p. 719-29.

71. Stanley, E.G., et al., *Efficient Cre-mediated deletion in cardiac progenitor cells conferred by a 3'UTR-ires-Cre allele of the homeobox gene Nkx2-5.* Int J Dev Biol, 2002. **46**(4): p. 431-9.

72. Lyons, I., et al., *Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5.* Genes Dev, 1995. **9**(13): p. 1654-66.

73. Tanaka, M., et al., *The cardiac homeobox gene Csx/Nkx2.5 lies genetically upstream of multiple genes essential for heart development.* Development, 1999. **126**(6): p. 1269-80.

74. Furtado, M.B., et al., *A novel conditional mouse model for Nkx2-5 reveals transcriptional regulation of cardiac ion channels.* Differentiation, 2016. **91**(1-3): p. 29-41.

75. Biben, C., et al., *Cardiac septal and valvular dysmorphogenesis in mice heterozygous for mutations in the homeobox gene Nkx2-5.* Circ Res, 2000. **87**(10): p. 888-95.

76. Elliott, D.A., et al., *Cardiac homeobox gene NKX2-5 mutations and congenital heart disease: associations with atrial septal defect and hypoplastic left heart syndrome.* J Am Coll Cardiol, 2003. **41**(11): p. 2072-6.

77. Briggs, L.E., et al., *Perinatal loss of Nkx2-5 results in rapid conduction and contraction defects.* Circ Res, 2008. **103**(6): p. 580-90.

78. Jay, P.Y., et al., *Nkx2-5 mutation causes anatomic hypoplasia of the cardiac conduction system.* J Clin Invest, 2004. **113**(8): p. 1130-7.

79. George, V., S. Colombo, and K.L. Targoff, *An early requirement for nkx2.5 ensures the first and second heart field ventricular identity and cardiac function into adulthood.* Dev Biol, 2015. **400**(1): p. 10-22.

80. Prall, O.W., et al., *An Nkx2-5/Bmp2/Smad1 negative feedback loop controls heart progenitor specification and proliferation.* Cell, 2007. **128**(5): p. 947-59.

81. Schott, J.J., et al., *Congenital heart disease caused by mutations in the transcription factor NKX2-5.* Science, 1998. **281**(5373): p. 108-11.

82. Gutierrez-Roelens, I., et al., *A novel CSX/NKX2-5 mutation causes autosomal-dominant AV block: are atrial fibrillation and syncopes part of the phenotype?* Eur J Hum Genet, 2006. **14**(12): p. 1313-6.

83. Stallmeyer, B., et al., *Mutational spectrum in the cardiac transcription factor gene NKX2.5 (CSX) associated with congenital heart disease.* Clin Genet, 2010. **78**(6): p. 533-40.

84. Reecy, J.M., et al., *Identification of upstream regulatory regions in the heart-expressed homeobox gene Nkx2-5.* Development, 1999. **126**(4): p. 839-49.

85. Schwartz, R.J. and E.N. Olson, *Building the heart piece by piece: modularity of cis-elements regulating Nkx2-5 transcription.* Development, 1999. **126**(19): p. 4187-92.

86. Searcy, R.D., et al., *A GATA-dependent nkx-2.5 regulatory element activates early cardiac gene expression in transgenic mice.* Development, 1998. **125**(22): p. 4461-70.

87. Lien, C.L., et al., *Control of early cardiac-specific transcription of Nkx2-5 by a GATA-dependent enhancer.* Development, 1999. **126**(1): p. 75-84.

88. Tanaka, M., et al., *Complex modular cis-acting elements regulate expression of the cardiac specifying homeobox gene Csx/Nkx2.5.* Development, 1999. **126**(7): p. 1439-50.

89. Chi, X., et al., *Complex cardiac Nkx2-5 gene expression activated by noggin-sensitive enhancers followed by chamber-specific modules.* Proc Natl Acad Sci U S A, 2005. **102**(38): p. 13490-5.

90. Brown, C.O., 3rd, et al., *The cardiac determination factor, Nkx2-5, is activated by mutual cofactors GATA-4 and Smad1/4 via a novel upstream enhancer.* J Biol Chem, 2004. **279**(11): p. 10659-69.

91. Gong, W., et al., *Inferring dynamic gene regulatory networks in cardiac differentiation through the integration of multi-dimensional data.* BMC Bioinformatics, 2015. **16**: p. 74.

92. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.

93. International Human Genome Sequencing, C., *Finishing the euchromatic sequence of the human genome.* Nature, 2004. **431**(7011): p. 931-45.

94. Taft, R.J., M. Pheasant, and J.S. Mattick, *The relationship between non-protein-coding DNA and eukaryotic complexity.* Bioessays, 2007. **29**(3): p. 288-99.

95. Liu, G., J.S. Mattick, and R.J. Taft, *A meta-analysis of the genomic and transcriptomic composition of complex life.* Cell Cycle, 2013. **12**(13): p. 2061-72.

96. Consortium, E.P., *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.

97. Ransohoff, J.D., Y. Wei, and P.A. Khavari, *The functions and unique features of long intergenic non-coding RNA.* Nat Rev Mol Cell Biol, 2018. **19**(3): p. 143-157.

98. Mercer, T.R., et al., *Specific expression of long noncoding RNAs in the mouse brain.* Proc Natl Acad Sci U S A, 2008. **105**(2): p. 716-21.

99. Quinn, J.J. and H.Y. Chang, *Unique features of long non-coding RNA biogenesis and function.* Nat Rev Genet, 2016. **17**(1): p. 47-62.

100. Di Mauro, V., M. Barandalla-Sobrados, and D. Catalucci, *The noncoding-RNA landscape in cardiovascular health and disease.* Noncoding RNA Res, 2018. **3**(1): p. 12-19.

101. Schonrock, N., R.P. Harvey, and J.S. Mattick, *Long noncoding RNAs in cardiac development and pathophysiology.* Circ Res, 2012. **111**(10): p. 1349-62.

102. Klattenhoff, C.A., et al., *Braveheart, a long noncoding RNA required for cardiovascular lineage commitment.* Cell, 2013. **152**(3): p. 570-83.

103. Grote, P., et al., *The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse.* Dev Cell, 2013. **24**(2): p. 206-14.

104. Anderson, K.M., et al., *Transcription of the non-coding RNA upperhand controls Hand2 expression and heart development.* Nature, 2016. **539**(7629): p. 433-436.

105. Aguilo, F., S. Di Cecilia, and M.J. Walsh, *Long Non-coding RNA ANRIL and Polycomb in Human Cancers and Cardiovascular Disease.* Curr Top Microbiol Immunol, 2016. **394**: p. 29-39.

106. Kumarswamy, R., et al., *Circulating long noncoding RNA, LIPCAR, predicts survival in patients with heart failure.* Circ Res, 2014. **114**(10): p. 1569-75.

107. Sun, C., et al., *Long non-coding RNA MIAT in development and disease: a new player in an old game.* J Biomed Sci, 2018. **25**(1): p. 23.

108. Ishii, N., et al., *Identification of a novel non-coding RNA, MIAT, that confers risk of myocardial infarction.* J Hum Genet, 2006. **51**(12): p. 1087-99.

109. Morris, K.V. and J.S. Mattick, *The rise of regulatory RNA.* Nat Rev Genet, 2014. **15**(6): p. 423-37.

110. Mattick, J.S. and J.L. Rinn, *Discovery and annotation of long noncoding RNAs.* Nat Struct Mol Biol, 2015. **22**(1): p. 5-7.

111. Harrow, J., et al., *GENCODE: the reference human genome annotation for The ENCODE Project.* Genome Res, 2012. **22**(9): p. 1760-74.

112. Sun, Q., Q. Hao, and K.V. Prasanth, *Nuclear Long Noncoding RNAs: Key Regulators of Gene Expression.* Trends Genet, 2018. **34**(2): p. 142-157.

113. Cabili, M.N., et al., *Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution.* Genome Biol, 2015. **16**: p. 20.

114. Clemson, C.M., et al., *An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles.* Mol Cell, 2009. **33**(6): p. 717-26.

115. Tripathi, V., et al., *The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation.* Mol Cell, 2010. **39**(6): p. 925-38.

116. Bond, C.S. and A.H. Fox, *Paraspeckles: nuclear bodies built on long noncoding RNA.* J Cell Biol, 2009. **186**(5): p. 637-44.

117. Kugel, J.F. and J.A. Goodrich, *Non-coding RNAs: key regulators of mammalian transcription.* Trends Biochem Sci, 2012. **37**(4): p. 144-51.

118. Fatica, A. and I. Bozzoni, *Long non-coding RNAs: new players in cell differentiation and development.* Nat Rev Genet, 2014. **15**(1): p. 7-21.

119. Gloss, B.S. and M.E. Dinger, *The specificity of long noncoding RNA expression.* Biochim Biophys Acta, 2016. **1859**(1): p. 16-22.

120. Djebali, S., et al., *Landscape of transcription in human cells.* Nature, 2012. **489**(7414): p. 101-8.

121. Deveson, I.W., et al., *The Dimensions, Dynamics, and Relevance of the Mammalian Noncoding Transcriptome.* Trends Genet, 2017. **33**(7): p. 464-478.

122. Liu, S.J., et al., *Single-cell analysis of long non-coding RNAs in the developing human neocortex.* Genome Biol, 2016. **17**: p. 67.

123. Yan, L., et al., *Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells.* Nat Struct Mol Biol, 2013. **20**(9): p. 1131-9.

124. Nitsche, A., et al., *Comparison of splice sites reveals that long noncoding RNAs are evolutionarily well conserved.* RNA, 2015. **21**(5): p. 801-12.

125. Guttman, M., et al., *Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals.* Nature, 2009. **458**(7235): p. 223-7.

126. Carninci, P., et al., *The transcriptional landscape of the mammalian genome.* Science, 2005. **309**(5740): p. 1559-63.

127. Derrien, T., et al., *The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression.* Genome Res, 2012. **22**(9): p. 1775-89.

128. Smith, M.A., et al., *Widespread purifying selection on RNA structure in mammals.* Nucleic Acids Res, 2013. **41**(17): p. 8220-36.

129. Seemann, S.E., et al., *The identification and functional annotation of RNA structures conserved in vertebrates.* Genome Res, 2017. **27**(8): p. 1371-1383.

130. Ulitsky, I., et al., *Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution.* Cell, 2011. **147**(7): p. 1537-50.

131. Gupta, R.A., et al., *Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis.* Nature, 2010. **464**(7291): p. 1071-6.

132. Zhang, Q., et al., *NEAT1 long noncoding RNA and paraspeckle bodies modulate HIV-1 posttranscriptional expression.* MBio, 2013. **4**(1): p. e00596-12.

133. Dykes, I.M. and C. Emanueli, *Transcriptional and Post-transcriptional Gene Regulation by Long Non-coding RNA.* Genomics Proteomics Bioinformatics, 2017. **15**(3): p. 177-186.

134. Rinn, J.L., *lncRNAs: linking RNA to chromatin.* Cold Spring Harb Perspect Biol, 2014. **6**(8).

135. Holoch, D. and D. Moazed, *RNA-mediated epigenetic regulation of gene expression.* Nat Rev Genet, 2015. **16**(2): p. 71-84.

136. Kretz, M., et al., *Control of somatic tissue differentiation by the long non-coding RNA TINCR.* Nature, 2013. **493**(7431): p. 231-5.

137. Kretz, M., *TINCR, staufen1, and cellular differentiation.* RNA Biol, 2013. **10**(10): p. 1597-601.

138. Faghihi, M.A., et al., *Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase.* Nat Med, 2008. **14**(7): p. 723-30.

139. Schmitz, S.U., P. Grote, and B.G. Herrmann, *Mechanisms of long noncoding RNA function in development and disease.* Cell Mol Life Sci, 2016. **73**(13): p. 2491-509.

140. Kornfeld, J.W. and J.C. Bruning, *Regulation of metabolism by long, non-coding RNAs.* Front Genet, 2014. **5**: p. 57.

141. Rinn, J.L., et al., *Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs.* Cell, 2007. **129**(7): p. 1311-23.

142. Wang, K.C. and H.Y. Chang, *Molecular mechanisms of long noncoding RNAs.* Mol Cell, 2011. **43**(6): p. 904-14.

143. Mercer, T.R. and J.S. Mattick, *Structure and function of long noncoding RNAs in epigenetic regulation.* Nat Struct Mol Biol, 2013. **20**(3): p. 300-7.

144. Brockdorff, N., *Noncoding RNA and Polycomb recruitment.* RNA, 2013. **19**(4): p. 429-42.

145. Davidovich, C. and T.R. Cech, *The recruitment of chromatin modifiers by long noncoding RNAs: lessons from PRC2.* RNA, 2015. **21**(12): p. 2007-22.

146. Ebert, M.S. and P.A. Sharp, *Emerging roles for natural microRNA sponges.* Curr Biol, 2010. **20**(19): p. R858-61.

147. Geisler, S. and J. Coller, *RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts.* Nat Rev Mol Cell Biol, 2013. **14**(11): p. 699-712.

148. Wang, K., et al., *CARL lncRNA inhibits anoxia-induced mitochondrial fission and apoptosis in cardiomyocytes by impairing miR-539-dependent PHB2 downregulation.* Nat Commun, 2014. **5**: p. 3596.

149. Wang, K., et al., *The long noncoding RNA CHRF regulates cardiac hypertrophy by targeting miR-489.* Circ Res, 2014. **114**(9): p. 1377-88.

150. Kopp, F. and J.T. Mendell, *Functional Classification and Experimental Dissection of Long Noncoding RNAs.* Cell, 2018. **172**(3): p. 393-407.

151. Devaux, Y., et al., *Long noncoding RNAs in cardiac development and ageing.* Nat Rev Cardiol, 2015. **12**(7): p. 415-25.

152. Kurian, L., et al., *Identification of novel long noncoding RNAs underlying vertebrate cardiovascular development.* Circulation, 2015. **131**(14): p. 1278-1290.

153. Han, P., et al., *A long noncoding RNA protects the heart from pathological hypertrophy.* Nature, 2014. **514**(7520): p. 102-106.

154. Ounzain, S., et al., *CARMEN, a human super enhancer-associated long noncoding RNA controlling cardiac specification, differentiation and homeostasis.* J Mol Cell Cardiol, 2015. **89**(Pt A): p. 98-112.

155. Liu, Y., et al., *Expression profiling and ontology analysis of long noncoding RNAs in post-ischemic heart and their implied roles in ischemia/reperfusion injury.* Gene, 2014. **543**(1): p. 15-21.

156. Yang, K.C., et al., *Deep RNA sequencing reveals dynamic regulation of myocardial noncoding RNAs in failing human heart and remodeling with mechanical circulatory support.* Circulation, 2014. **129**(9): p. 1009-21.

157. Matkovich, S.J., et al., *Epigenetic coordination of embryonic heart transcription by dynamically regulated long noncoding RNAs.* Proc Natl Acad Sci U S A, 2014. **111**(33): p. 12264-9.

158. Werber, M., et al., *The tissue-specific transcriptomic landscape of the mid-gestational mouse embryo.* Development, 2014. **141**(11): p. 2325-30.

159. Ounzain, S., et al., *Genome-wide profiling of the cardiac transcriptome after myocardial infarction identifies novel heart-specific long non-coding RNAs.* Eur Heart J, 2015. **36**(6): p. 353-68a.

160. Consortium, C.A.D., et al., *Large-scale association analysis identifies new risk loci for coronary artery disease.* Nat Genet, 2013. **45**(1): p. 25-33.

161. Bartonicek, N., et al., *Intergenic disease-associated regions are abundant in novel transcripts.* Genome Biol, 2017. **18**(1): p. 241.

162. Zhang, F. and J.R. Lupski, *Non-coding genetic variants in human disease.* Hum Mol Genet, 2015. **24**(R1): p. R102-10.

163. Maurano, M.T., et al., *Systematic localization of common disease-associated variation in regulatory DNA.* Science, 2012. **337**(6099): p. 1190-5.

164. Kumar, V., et al., *Human disease-associated genetic variation impacts large intergenic non-coding RNA expression.* PLoS Genet, 2013. **9**(1): p. e1003201.

165. Gloss, B.S. and M.E. Dinger, *Realizing the significance of noncoding functionality in clinical genomics.* Exp Mol Med, 2018. **50**(8): p. 97.

166. Li, M., et al., *Circulating Long Noncoding RNA LIPCAR Acts as a Novel Biomarker in Patients with ST-Segment Elevation Myocardial Infarction.* Med Sci Monit, 2018. **24**: p. 5064-5070.

167. Lucas, T., A. Bonauer, and S. Dimmeler, *RNA Therapeutics in Cardiovascular Disease.* Circ Res, 2018. **123**(2): p. 205-220.

168. Wang, K., et al., *MDRL lncRNA regulates the processing of miR-484 primary transcript by targeting miR-361.* PLoS Genet, 2014. **10**(7): p. e1004467.

169. Zhu, X.H., et al., *LncRNA MIAT enhances cardiac hypertrophy partly through sponging miR-150.* Eur Rev Med Pharmacol Sci, 2016. **20**(17): p. 3653-60.

170. Shao, M., et al., *LncRNA TINCR attenuates cardiac hypertrophy by epigenetically silencing CaMKII.* Oncotarget, 2017. **8**(29): p. 47565-47573.

171. Viereck, J., et al., *Long noncoding RNA Chast promotes cardiac remodeling.* Sci Transl Med, 2016. **8**(326): p. 326ra22.

172. Bell, R.D., et al., *Identification and initial functional characterization of a human vascular cell-enriched long noncoding RNA.* Arterioscler Thromb Vasc Biol, 2014. **34**(6): p. 1249-59.

173. Kodzius, R., et al., *CAGE: cap analysis of gene expression.* Nat Methods, 2006. **3**(3): p. 211-22.

174. Kawai, J., et al., *Functional annotation of a full-length mouse cDNA collection.* Nature, 2001. **409**(6821): p. 685-90.

175. Karolchik, D., A.S. Hinrichs, and W.J. Kent, *The UCSC Genome Browser.* Curr Protoc Bioinformatics, 2007. **Chapter 1**: p. Unit 1 4.

176. Ustyantsev, I.G., et al., *[Canonical and noncanonical RNA polyadenylation].* Mol Biol (Mosk), 2017. **51**(2): p. 262-273.

177. Mercer, T.R., M.E. Dinger, and J.S. Mattick, *Long non-coding RNAs: insights into functions.* Nat Rev Genet, 2009. **10**(3): p. 155-9.

178. Karolchik, D., et al., *The UCSC Genome Browser Database.* Nucleic Acids Res, 2003. **31**(1): p. 51-4.

179. Claycomb, W.C., et al., *HL-1 cells: a cardiac muscle cell line that contracts and retains phenotypic characteristics of the adult cardiomyocyte.* Proc Natl Acad Sci U S A, 1998. **95**(6): p. 2979-84.

180. Dinger, M.E., et al., *Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation.* Genome Res, 2008. **18**(9): p. 1433-45.

181. Bradley, A., et al., *Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines.* Nature, 1984. **309**(5965): p. 255-6.

182. Donovan, P.J. and J. Gearhart, *The end of the beginning for pluripotent stem cells.* Nature, 2001. **414**(6859): p. 92-7.

183. Evans, M.J. and M.H. Kaufman, *Establishment in culture of pluripotential cells from mouse embryos.* Nature, 1981. **292**(5819): p. 154-6.

184. Takahashi, T., et al., *Ascorbic acid enhances differentiation of embryonic stem cells into cardiac myocytes.* Circulation, 2003. **107**(14): p. 1912-6.

185. Moraveji, S.F., et al., *Inhibition of glycogen synthase kinase-3 promotes efficient derivation of pluripotent stem cells from neonatal mouse testis.* Hum Reprod, 2012. **27**(8): p. 2312-24.

186. Wamstad, J.A., et al., *Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage.* Cell, 2012. **151**(1): p. 206-20.

187. Kasahara, H., et al., *Cardiac and extracardiac expression of Csx/Nkx2.5 homeodomain protein.* Circ Res, 1998. **82**(9): p. 936-46.

188. Yaffe, D. and O. Saxel, *Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle.* Nature, 1977. **270**(5639): p. 725-7.

189. Fox, A.H., et al., *Paraspeckles: a novel nuclear domain.* Curr Biol, 2002. **12**(1): p. 13-25.

190. Beilharz, T.H. and T. Preiss, *Translational profiling: the genome-wide measure of the nascent proteome.* Brief Funct Genomic Proteomic, 2004. **3**(2): p. 103-11.

191. Molotski, N. and Y. Soen, *Differential association of microRNAs with polysomes reflects distinct strengths of interactions with their mRNA targets.* RNA, 2012. **18**(9): p. 1612-23.

192. Kim, J., et al., *Identification of many microRNAs that copurify with polyribosomes in mammalian neurons.* Proc Natl Acad Sci U S A, 2004. **101**(1): p. 360-5.

193. Wang, L., et al., *CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model.* Nucleic Acids Res, 2013. **41**(6): p. e74.

194. Fickett, J.W., *Recognition of protein coding regions in DNA sequences.* Nucleic Acids Res, 1982. **10**(17): p. 5303-18.

195. Makarewich, C.A. and E.N. Olson, *Mining for Micropeptides.* Trends Cell Biol, 2017. **27**(9): p. 685-696.

196. Anderson, D.M., et al., *A micropeptide encoded by a putative long noncoding RNA regulates muscle performance.* Cell, 2015. **160**(4): p. 595-606.

197. Bazzini, A.A., et al., *Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation.* EMBO J, 2014. **33**(9): p. 981-93.

198. Makarewich, C.A., et al., *MOXI Is a Mitochondrial Micropeptide That Enhances Fatty Acid beta-Oxidation.* Cell Rep, 2018. **23**(13): p. 3701-3709.

199. Nelson, B.R., et al., *A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle.* Science, 2016. **351**(6270): p. 271-5.

200. Bi, P., et al., *Control of muscle formation by the fusogenic micropeptide myomixer.* Science, 2017. **356**(6335): p. 323-327.

201. Andrews, S.J. and J.A. Rothnagel, *Emerging evidence for functional peptides encoded by short open reading frames.* Nat Rev Genet, 2014. **15**(3): p. 193-204.

202. Mackowiak, S.D., et al., *Extensive identification and analysis of conserved small ORFs in animals.* Genome Biol, 2015. **16**: p. 179.

203. Lin, M.F., I. Jungreis, and M. Kellis, *PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions.* Bioinformatics, 2011. **27**(13): p. i275-82.

204. Roadmap Epigenomics, C., et al., *Integrative analysis of 111 reference human epigenomes.* Nature, 2015. **518**(7539): p. 317-30.

205. Consortium, G.T., *The Genotype-Tissue Expression (GTEx) project.* Nat Genet, 2013. **45**(6): p. 580-5.

206. Strausberg, R.L., et al., *Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.* Proc Natl Acad Sci U S A, 2002. **99**(26): p. 16899-903.

207. Cabili, M.N., et al., *Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses.* Genes Dev, 2011. **25**(18): p. 1915-27.

208. Washietl, S., M. Kellis, and M. Garber, *Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals.* Genome Res, 2014. **24**(4): p. 616-28.

209. Chojnacki, S., et al., *Programmatic access to bioinformatics tools from EMBL-EBI update: 2017.* Nucleic Acids Res, 2017. **45**(W1): p. W550-W553.

210. Johnsson, P., et al., *Evolutionary conservation of long non-coding RNAs; sequence, structure, function.* Biochim Biophys Acta, 2014. **1840**(3): p. 1063-71.

211. Havgaard, J., S. Kaur, and J. Gorodkin, *Comparative ncRNA gene and structure prediction using Foldalign and FoldalignM.* Curr Protoc Bioinformatics, 2012. **Chapter 12**: p. Unit12 11.

212. Havgaard, J.H., R.B. Lyngso, and J. Gorodkin, *The FOLDALIGN web server for pairwise structural RNA alignment and mutual motif search.* Nucleic Acids Res, 2005. **33**(Web Server issue): p. W650-3.

213. Fonoudi, H., et al., *A Universal and Robust Integrated Platform for the Scalable Production of Human Cardiomyocytes From Pluripotent Stem Cells.* Stem Cells Transl Med, 2015. **4**(12): p. 1482-94.

214. Hon, C.C., et al., *An atlas of human long non-coding RNAs with accurate 5' ends.* Nature, 2017. **543**(7644): p. 199-204.

215. Arner, E., et al., *Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells.* Science, 2015. **347**(6225): p. 1010-4.

216. Consortium, F., et al., *A promoter-level mammalian expression atlas.* Nature, 2014. **507**(7493): p. 462-70.

217. den Hoed, M., et al., *Identification of heart rate-associated loci and their effects on cardiac conduction and rhythm disorders.* Nat Genet, 2013. **45**(6): p. 621-31.

218. Mensink, G.B. and H. Hoffmeister, *The relationship between resting heart rate and all-cause, cardiovascular and cancer mortality.* Eur Heart J, 1997. **18**(9): p. 1404-10.

219. Jensen, M.T., J.L. Marott, and G.B. Jensen, *Elevated resting heart rate is associated with greater risk of cardiovascular and all-cause mortality in current and former smokers.* Int J Cardiol, 2011. **151**(2): p. 148-54.

220. Jensen, M.T., et al., *Resting heart rate is associated with cardiovascular and all-cause mortality after adjusting for inflammatory markers: the Copenhagen City Heart Study.* Eur J Prev Cardiol, 2012. **19**(1): p. 102-8.

221. Inoue, T., et al., *Higher heart rate predicts the risk of developing hypertension in a normotensive screened cohort.* Circ J, 2007. **71**(11): p. 1755-60.

222. Roselli, C., et al., *Multi-ethnic genome-wide association study for atrial fibrillation.* Nat Genet, 2018. **50**(9): p. 1225-1233.

223. Sabarinathan, R., et al., *The RNAsnp web server: predicting SNP effects on local RNA secondary structure.* Nucleic Acids Res, 2013. **41**(Web Server issue): p. W475-9.

224. Machiela, M.J. and S.J. Chanock, *LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants.* Bioinformatics, 2015. **31**(21): p. 3555-7.

225. Bartlett, H., G.J. Veenstra, and D.L. Weeks, *Examining the cardiac NK-2 genes in early heart development.* Pediatr Cardiol, 2010. **31**(3): p. 335-41.

226. Hunt, R., et al., *Silent (synonymous) SNPs: should we care about them?* Methods Mol Biol, 2009. **578**: p. 23-39.

227. Nielsen, J.B., et al., *Biobank-driven genomic discovery yields new insight into atrial fibrillation biology.* Nat Genet, 2018. **50**(9): p. 1234-1239.

228. Lundin, K.E., et al., *Biological activity and biotechnological aspects of locked nucleic acids.* Adv Genet, 2013. **82**: p. 47-107.

229. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.

230. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.

231. Mi, H., et al., *PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements.* Nucleic Acids Res, 2017. **45**(D1): p. D183-D189.

232. Franceschini, A., et al., *STRING v9.1: protein-protein interaction networks, with increased coverage and integration.* Nucleic Acids Res, 2013. **41**(Database issue): p. D808-15.

233. Singh, R., et al., *Tbx2 and Tbx3 induce atrioventricular myocardial development and endocardial cushion formation.* Cell Mol Life Sci, 2012. **69**(8): p. 1377-89.

234. Duim, S.N., M.J. Goumans, and B.P.T. Kruithof, *WT1 in Cardiac Development and Disease*, in *Wilms Tumor*, M.M. van den Heuvel-Eibrink, Editor. 2016: Brisbane (AU).

235. Chu, C., J. Quinn, and H.Y. Chang, *Chromatin isolation by RNA purification (ChIRP).* J Vis Exp, 2012(61).

236. Chu, C., et al., *Systematic discovery of Xist RNA binding proteins.* Cell, 2015. **161**(2): p. 404-16.

237. Khalil, A.M., et al., *Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression.* Proc Natl Acad Sci U S A, 2009. **106**(28): p. 11667-72.

238. Koziol, M.J. and J.L. Rinn, *RNA traffic control of chromatin complexes.* Curr Opin Genet Dev, 2010. **20**(2): p. 142-8.

239. Chu, C., et al., *Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions.* Mol Cell, 2011. **44**(4): p. 667-78.

240. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS).* Genome Biol, 2008. **9**(9): p. R137.

241. Wang, Q., et al., *Odd-skipped related 1 (Odd 1) is an essential regulator of heart and urogenital development.* Dev Biol, 2005. **288**(2): p. 582-94.

242. Davisson, M.T., et al., *Kidney adysplasia and variable hydronephrosis, a new mutation affecting the odd-skipped related 1 gene in the mouse, causes variable defects in kidney development and hydronephrosis.* Am J Physiol Renal Physiol, 2015. **308**(12): p. F1335-42.

243. Wang, K.C., et al., *A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression.* Nature, 2011. **472**(7341): p. 120-4.

244. Rorabaugh, B.R., et al., *Regulator of G Protein Signaling 6 Protects the Heart from Ischemic Injury.* J Pharmacol Exp Ther, 2017. **360**(3): p. 409-416.

245. Yang, J., et al., *RGS6, a modulator of parasympathetic activation in heart.* Circ Res, 2010. **107**(11): p. 1345-9.

246. Sharpe, J., et al., *Optical projection tomography as a tool for 3D microscopy and gene expression studies.* Science, 2002. **296**(5567): p. 541-5.

247. Krasuski, R.A., *When and how to fix a 'hole in the heart': approach to ASD and PFO.* Cleve Clin J Med, 2007. **74**(2): p. 137-47.

248. Anderson, R.H., N.A. Brown, and S. Webb, *Development and structure of the atrial septum.* Heart, 2002. **88**(1): p. 104-10.

249. Kirk, E.P., et al., *Quantitative trait loci modifying cardiac atrial septal morphology and risk of patent foramen ovale in the mouse.* Circ Res, 2006. **98**(5): p. 651-8.

250. Mihl, C., W.R. Dassen, and H. Kuipers, *Cardiac remodelling: concentric versus eccentric hypertrophy in strength and endurance athletes.* Neth Heart J, 2008. **16**(4): p. 129-33.

251. deAlmeida, A.C., R.J. van Oort, and X.H. Wehrens, *Transverse aortic constriction in mice.* J Vis Exp, 2010(38).

252. North, B.J. and D.A. Sinclair, *The intersection between aging and cardiovascular disease.* Circ Res, 2012. **110**(8): p. 1097-108.

253. Dhingra, R. and R.S. Vasan, *Age as a risk factor.* Med Clin North Am, 2012. **96**(1): p. 87-91.

254. Hu, G., et al., *Molecular mechanisms of long noncoding RNAs and their role in disease pathogenesis.* Oncotarget, 2018. **9**(26): p. 18648-18663.

255. Li, L. and H.Y. Chang, *Physiological roles of long noncoding RNAs: insight from knockout mice.* Trends Cell Biol, 2014. **24**(10): p. 594-602.

256. Schierding, W., W.S. Cutfield, and J.M. O'Sullivan, *The missing story behind Genome Wide Association Studies: single nucleotide polymorphisms in gene deserts have a story to tell.* Front Genet, 2014. **5**: p. 39.

257. Hrdlickova, B., et al., *Genetic variation in the non-coding genome: Involvement of micro-RNAs and long non-coding RNAs in disease.* Biochim Biophys Acta, 2014. **1842**(10): p. 1910-1922.

258. Chung, I.M. and G. Rajakumar, *Genetics of Congenital Heart Defects: The NKX2-5 Gene, a Key Player.* Genes (Basel), 2016. **7**(2).

259. Akazawa, H. and I. Komuro, *Cardiac transcription factor Csx/Nkx2-5: Its role in cardiac development and diseases.* Pharmacol Ther, 2005. **107**(2): p. 252-68.

260. McElhinney, D.B., et al., *NKX2.5 mutations in patients with congenital heart disease.* J Am Coll Cardiol, 2003. **42**(9): p. 1650-5.

261. Zamir, L., et al., *Nkx2.5 marks angioblasts that contribute to hemogenic endothelium of the endocardium and dorsal aorta.* Elife, 2017. **6**.

262. Dinger, M.E., et al., *Differentiating protein-coding and noncoding RNA: challenges and ambiguities.* PLoS Comput Biol, 2008. **4**(11): p. e1000176.

263. Dhamija, S. and M.B. Menon, *Non-coding transcript variants of protein-coding genes - what are they good for?* RNA Biol, 2018. **15**(8): p. 1025-1031.

264. Jenny, A., et al., *A translation-independent role of oskar RNA in early Drosophila oogenesis.* Development, 2006. **133**(15): p. 2827-33.

265. Szempruch, A. and M. Guttman, *Linking Protein and RNA Function within the Same Gene.* Cell, 2017. **168**(5): p. 753-755.

266. Lim, S., et al., *Dorsal activity of maternal squint is mediated by a non-coding function of the RNA.* Development, 2012. **139**(16): p. 2903-15.

267. Naganuma, T., et al., *Alternative 3'-end processing of long noncoding RNA initiates construction of nuclear paraspeckles.* EMBO J, 2012. **31**(20): p. 4020-34.

268. Williamson, L., et al., *UV Irradiation Induces a Non-coding RNA that Functionally Opposes the Protein Encoded by the Same Gene.* Cell, 2017. **168**(5): p. 843-855 e13.

269. Elliott, D.A., et al., *A tyrosine-rich domain within homeodomain transcription factor Nkx2-5 is an essential element in the early cardiac transcriptional regulatory machinery.* Development, 2006. **133**(7): p. 1311-22.

270. Vance, K.W. and C.P. Ponting, *Transcriptional regulatory functions of nuclear long noncoding RNAs.* Trends Genet, 2014. **30**(8): p. 348-55.

271. Behrens, A.N., et al., *Nkx2-5 mediates differential cardiac differentiation through interaction with Hoxa10.* Stem Cells Dev, 2013. **22**(15): p. 2211-20.

272. Bondue, A., et al., *Defining the earliest step of cardiovascular progenitor specification during embryonic stem cell differentiation.* J Cell Biol, 2011. **192**(5): p. 751-65.

273. Harris, I.S. and B.L. Black, *Development of the endocardium.* Pediatr Cardiol, 2010. **31**(3): p. 391-9.

274. Xue, Z., et al., *A G-Rich Motif in the lncRNA Braveheart Interacts with a Zinc-Finger Transcription Factor to Specify the Cardiovascular Lineage.* Mol Cell, 2016. **64**(1): p. 37-50.

275. Stewart, A., J. Huang, and R.A. Fisher, *RGS Proteins in Heart: Brakes on the Vagus.* Front Physiol, 2012. **3**: p. 95.

276. Mullner, C., et al., *Heterologous facilitation of G protein-activated K(+) channels by beta-adrenergic stimulation via cAMP-dependent protein kinase.* J Gen Physiol, 2000. **115**(5): p. 547-58.

277. Pagano, M. and M.B. Anand-Srivastava, *Cytoplasmic domain of natriuretic peptide receptor C constitutes Gi activator sequences that inhibit adenylyl cyclase activity.* J Biol Chem, 2001. **276**(25): p. 22064-70.

278. Schwenk, J., et al., *Native GABA(B) receptors are heteromultimers with a family of auxiliary subunits.* Nature, 2010. **465**(7295): p. 231-5.

279. Terracciano, A., et al., *Genome-wide association scan of trait depression.* Biol Psychiatry, 2010. **68**(9): p. 811-7.

280. Li, W., et al., *Significant association of GRM7 and GRM8 genes with schizophrenia and major depressive disorder in the Han Chinese population.* Eur Neuropsychopharmacol, 2016. **26**(1): p. 136-146.

281. Deng, X., et al., *Genome wide association study (GWAS) of Chagas cardiomyopathy in Trypanosoma cruzi seropositive subjects.* PLoS One, 2013. **8**(11): p. e79629.

282. Friedman, D., et al., *Cardiac arrhythmia and neuroexcitability gene variants in resected brain tissue from patients with sudden unexpected death in epilepsy (SUDEP).* NPJ Genom Med, 2018. **3**: p. 9.

283. Johnson, J.N., et al., *Identification of a possible pathogenic link between congenital long QT syndrome and epilepsy.* Neurology, 2009. **72**(3): p. 224-31.

284. Wydeven, N., et al., *RGS6, but not RGS4, is the dominant regulator of G protein signaling (RGS) modulator of the parasympathetic regulation of mouse heart rate.* J Biol Chem, 2014. **289**(4): p. 2440-9.

285. Novellino, L., et al., *PTPRK negatively regulates transcriptional activity of wild type and mutated oncogenic beta-catenin and affects membrane distribution of beta-catenin/E-cadherin complexes in cancer cells.* Cell Signal, 2008. **20**(5): p. 872-83.

286. Chen, Y., et al., *Analysis of PTPRK polymorphisms in association with risk and age at onset of Alzheimer's disease, cancer risk, and cholesterol.* J Psychiatr Res, 2018. **96**: p. 65-72.

287. Xu, Y., et al., *Receptor-type protein-tyrosine phosphatase-kappa regulates epidermal growth factor receptor function.* J Biol Chem, 2005. **280**(52): p. 42694-700.

288. Ozhan, G. and G. Weidinger, *Wnt/beta-catenin signaling in heart regeneration.* Cell Regen (Lond), 2015. **4**(1): p. 3.

289. Tian, Y., E.D. Cohen, and E.E. Morrisey, *The importance of Wnt signaling in cardiovascular development.* Pediatr Cardiol, 2010. **31**(3): p. 342-8.

290. Burkhard, S.B. and J. Bakkers, *Spatially resolved RNA-sequencing of the embryonic heart identifies a role for Wnt/beta-catenin signaling in autonomic control of heart rate.* Elife, 2018. **7**.

291. Takahashi, A., et al., *SHP2 tyrosine phosphatase converts parafibromin/Cdc73 from a tumor suppressor to an oncogenic driver.* Mol Cell, 2011. **43**(1): p. 45-56.

292. Mosimann, C., G. Hausmann, and K. Basler, *Parafibromin/Hyrax activates Wnt/Wg target gene transcription by direct association with beta-catenin/Armadillo.* Cell, 2006. **125**(2): p. 327-41.

293. Langenbacher, A.D., et al., *The PAF1 complex differentially regulates cardiomyocyte specification.* Dev Biol, 2011. **353**(1): p. 19-28.

294. Kikuchi, I., et al., *Dephosphorylated parafibromin is a transcriptional coactivator of the Wnt/Hedgehog/Notch pathways.* Nat Commun, 2016. **7**: p. 12887.

295. Zhu, B., et al., *The human PAF complex coordinates transcription with events downstream of RNA synthesis.* Genes Dev, 2005. **19**(14): p. 1668-73.

296. Rosonina, E. and J.L. Manley, *From transcription to mRNA: PAF provides a new path.* Mol Cell, 2005. **20**(2): p. 167-8.

297. Li, C., et al., *OPCML is frequently methylated in human colorectal cancer and its restored expression reverses EMT via downregulation of smad signaling.* Am J Cancer Res, 2015. **5**(5): p. 1635-48.

298. Kerr, K.F., et al., *Genome-wide association study of heart rate and its variability in Hispanic/Latino cohorts.* Heart Rhythm, 2017. **14**(11): p. 1675-1684.

299. Mahmoudi, T., et al., *The kinase TNIK is an essential activator of Wnt target genes.* EMBO J, 2009. **28**(21): p. 3329-40.

300. von Gise, A., et al., *WT1 regulates epicardial epithelial to mesenchymal transition through beta-catenin and retinoic acid signaling pathways.* Dev Biol, 2011. **356**(2): p. 421-31.

301. Hoogaars, W.M., et al., *Tbx3 controls the sinoatrial node gene program and imposes pacemaker function on the atria.* Genes Dev, 2007. **21**(9): p. 1098-112.

302. Renard, C.A., et al., *Tbx3 is a downstream target of the Wnt/beta-catenin pathway and a critical mediator of beta-catenin survival functions in liver cancer.* Cancer Res, 2007. **67**(3): p. 901-10.

303. Gillers, B.S., et al., *Canonical wnt signaling regulates atrioventricular junction programming and electrophysiological properties.* Circ Res, 2015. **116**(3): p. 398-406.

304. Tornroth-Horsefield, S. and R. Neutze, *Opening and closing the metabolite gate.* Proc Natl Acad Sci U S A, 2008. **105**(50): p. 19565-6.

305. Brown, D.A., et al., *Expert consensus document: Mitochondrial function as a therapeutic target in heart failure.* Nat Rev Cardiol, 2017. **14**(4): p. 238-250.

306. Kotiadis, V.N., M.R. Duchen, and L.D. Osellame, *Mitochondrial quality control and communications with the nucleus are important in maintaining mitochondrial function and cell health.* Biochim Biophys Acta, 2014. **1840**(4): p. 1254-65.

307. Monaghan, R.M. and A.J. Whitmarsh, *Mitochondrial Proteins Moonlighting in the Nucleus.* Trends Biochem Sci, 2015. **40**(12): p. 728-735.

308. Yogev, O. and O. Pines, *Dual targeting of mitochondrial proteins: mechanism, regulation and function.* Biochim Biophys Acta, 2011. **1808**(3): p. 1012-20.

309. Boukouris, A.E., S.D. Zervopoulos, and E.D. Michelakis, *Metabolic Enzymes Moonlighting in the Nucleus: Metabolic Regulation of Gene Transcription.* Trends Biochem Sci, 2016. **41**(8): p. 712-730.

310. Cardamone, M.D., et al., *Mitochondrial Retrograde Signaling in Mammals Is Mediated by the Transcriptional Cofactor GPS2 via Direct Mitochondria-to-Nucleus Translocation.* Mol Cell, 2018. **69**(5): p. 757-772 e7.

311. Sutendra, G., et al., *A nuclear pyruvate dehydrogenase complex is important for the generation of acetyl-CoA and histone acetylation.* Cell, 2014. **158**(1): p. 84-97.

312. Williamson, E.A., et al., *Overview for the histone codes for DNA repair.* Prog Mol Biol Transl Sci, 2012. **110**: p. 207-27.

313. Leshets, M., et al., *Fumarase: From the TCA Cycle to DNA Damage Response and Tumor Suppression.* Front Mol Biosci, 2018. **5**: p. 68.

314. Yogev, O., et al., *Fumarase: a mitochondrial metabolic enzyme and a cytosolic/nuclear component of the DNA damage response.* PLoS Biol, 2010. **8**(3): p. e1000328.

315. Jiang, Y., et al., *Local generation of fumarate promotes DNA repair through inhibition of histone H3 demethylation.* Nat Cell Biol, 2015. **17**(9): p. 1158-68.

316. Yaniv, Y., et al., *Ca2+-regulated-cAMP/PKA signaling in cardiac pacemaker cells links ATP supply to demand.* J Mol Cell Cardiol, 2011. **51**(5): p. 740-8.

317. Rogell, B., et al., *Specific RNP capture with antisense LNA/DNA mixmers.* RNA, 2017. **23**(8): p. 1290-1302.

318. Newman, J.C. and E. Verdin, *Ketone bodies as signaling metabolites.* Trends Endocrinol Metab, 2014. **25**(1): p. 42-52.

319. Castonguay, Z., et al., *Nuclear lactate dehydrogenase modulates histone modification in human hepatocytes.* Biochem Biophys Res Commun, 2014. **454**(1): p. 172-7.

320. Liu, Y., et al., *Nuclear lactate dehydrogenase A senses ROS to produce alpha-hydroxybutyrate for HPV-induced cervical tumor growth.* Nat Commun, 2018. **9**(1): p. 4429.

321. Ghanem, L.R., et al., *The Poly(C) Binding Protein Pcbp2 and Its Retrotransposed Derivative Pcbp1 Are Independently Essential to Mouse Development.* Mol Cell Biol, 2016. **36**(2): p. 304-19.

322. Geuens, T., D. Bouhy, and V. Timmerman, *The hnRNP family: insights into their role in health and disease.* Hum Genet, 2016. **135**(8): p. 851-67.

323. Vreugde, S., et al., *Nuclear myosin VI enhances RNA polymerase II-dependent transcription.* Mol Cell, 2006. **23**(5): p. 749-55.

324. Jung, S.J., et al., *Essential function of Aco2, a fusion protein of aconitase and mitochondrial ribosomal protein bL21, in mitochondrial translation in fission yeast.* FEBS Lett, 2015. **589**(7): p. 822-8.

325. Madiraju, P., et al., *Mitochondrial acetylcarnitine provides acetyl groups for nuclear histone acetylation.* Epigenetics, 2009. **4**(6): p. 399-403.

326. Popanda, O., G. Fox, and H.W. Thielmann, *Modulation of DNA polymerases alpha, delta and epsilon by lactate dehydrogenase and 3-phosphoglycerate kinase.* Biochim Biophys Acta, 1998. **1397**(1): p. 102-17.

327. Egea, G., et al., *Nuclear location of phosphoglycerate mutase BB isozyme in rat tissues.* Histochemistry, 1992. **97**(3): p. 269-75.

328. Xiao, M., et al., *Inhibition of alpha-KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors.* Genes Dev, 2012. **26**(12): p. 1326-38.

329. Lalande, S. and B.D. Johnson, *Diastolic dysfunction: a link between hypertension and heart failure.* Drugs Today (Barc), 2008. **44**(7): p. 503-13.

330. Glukhov, A.V., et al., *Functional anatomy of the murine sinus node: high-resolution optical mapping of ankyrin-B heterozygous mice.* Am J Physiol Heart Circ Physiol, 2010. **299**(2): p. H482-91.

331. Fenske, S., et al., *Sick sinus syndrome in HCN1-deficient mice.* Circulation, 2013. **128**(24): p. 2585-94.

332. Fenske, S., et al., *Comprehensive multilevel in vivo and in vitro analysis of heart rate fluctuations in mice by ECG telemetry and electrophysiology.* Nat Protoc, 2016. **11**(1): p. 61-86.

333. Baruscotti, M., et al., *Deep bradycardia and heart block caused by inducible cardiac-specific knockout of the pacemaker channel gene Hcn4.* Proc Natl Acad Sci U S A, 2011. **108**(4): p. 1705-10.

334. Bround, M.J., et al., *Cardiac ryanodine receptors control heart rate and rhythmicity in adult mice.* Cardiovasc Res, 2012. **96**(3): p. 372-80.

335. Andersson, K.B., et al., *Moderate heart dysfunction in mice with inducible cardiomyocyte-specific excision of the Serca2 gene.* J Mol Cell Cardiol, 2009. **47**(2): p. 180-7.

336. Saito, Y., et al., *Enhancement of Spontaneous Activity by HCN4 Overexpression in Mouse Embryonic Stem Cell-Derived Cardiomyocytes - A Possible Biological Pacemaker.* PLoS One, 2015. **10**(9): p. e0138193.

337. del Monte, F., et al., *Restoration of contractile function in isolated cardiomyocytes from failing human hearts by gene transfer of SERCA2a.* Circulation, 1999. **100**(23): p. 2308-11.

338. Wang, Y.Y., et al., *RyR2R420Q catecholaminergic polymorphic ventricular tachycardia mutation induces bradycardia by disturbing the coupled clock pacemaker mechanism.* JCI Insight, 2017. **2**(8).

339. Williams, B.A., D.R. Dickenson, and G.N. Beatch, *Kinetics of rate-dependent shortening of action potential duration in guinea-pig ventricle; effects of IK1 and IKr blockade.* Br J Pharmacol, 1999. **126**(6): p. 1426-36.

340. Nunnari, J. and A. Suomalainen, *Mitochondria: in sickness and in health.* Cell, 2012. **148**(6): p. 1145-59.

341. Weil, D., et al., *Mature mRNAs accumulated in the nucleus are neither the molecules in transit to the cytoplasm nor constitute a stockpile for gene expression.* RNA, 2000. **6**(7): p. 962-75.

342. Hsu, P.D., et al., *DNA targeting specificity of RNA-guided Cas9 nucleases.* Nat Biotechnol, 2013. **31**(9): p. 827-32.

343. Mali, P., et al., *RNA-guided human genome engineering via Cas9.* Science, 2013. **339**(6121): p. 823-6.

344. Moorman, A.F., et al., *Sensitive nonradioactive detection of mRNA in tissue sections: novel application of the whole-mount in situ hybridization protocol.* J Histochem Cytochem, 2001. **49**(1): p. 1-8.

345. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data.* Bioinformatics, 2014. **30**(15): p. 2114-20.

346. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.

347. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.

348. Li, B. and C.N. Dewey, *RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.* BMC Bioinformatics, 2011. **12**: p. 323.

349. Wang, L., S. Wang, and W. Li, *RSeQC: quality control of RNA-seq experiments.* Bioinformatics, 2012. **28**(16): p. 2184-5.

350.   Feng, J., et al., *Identifying ChIP-seq enrichment using MACS.* Nat Protoc, 2012. **7**(9): p. 1728-40.

Appendix

I.   Detailed *in situ* hybridisation results
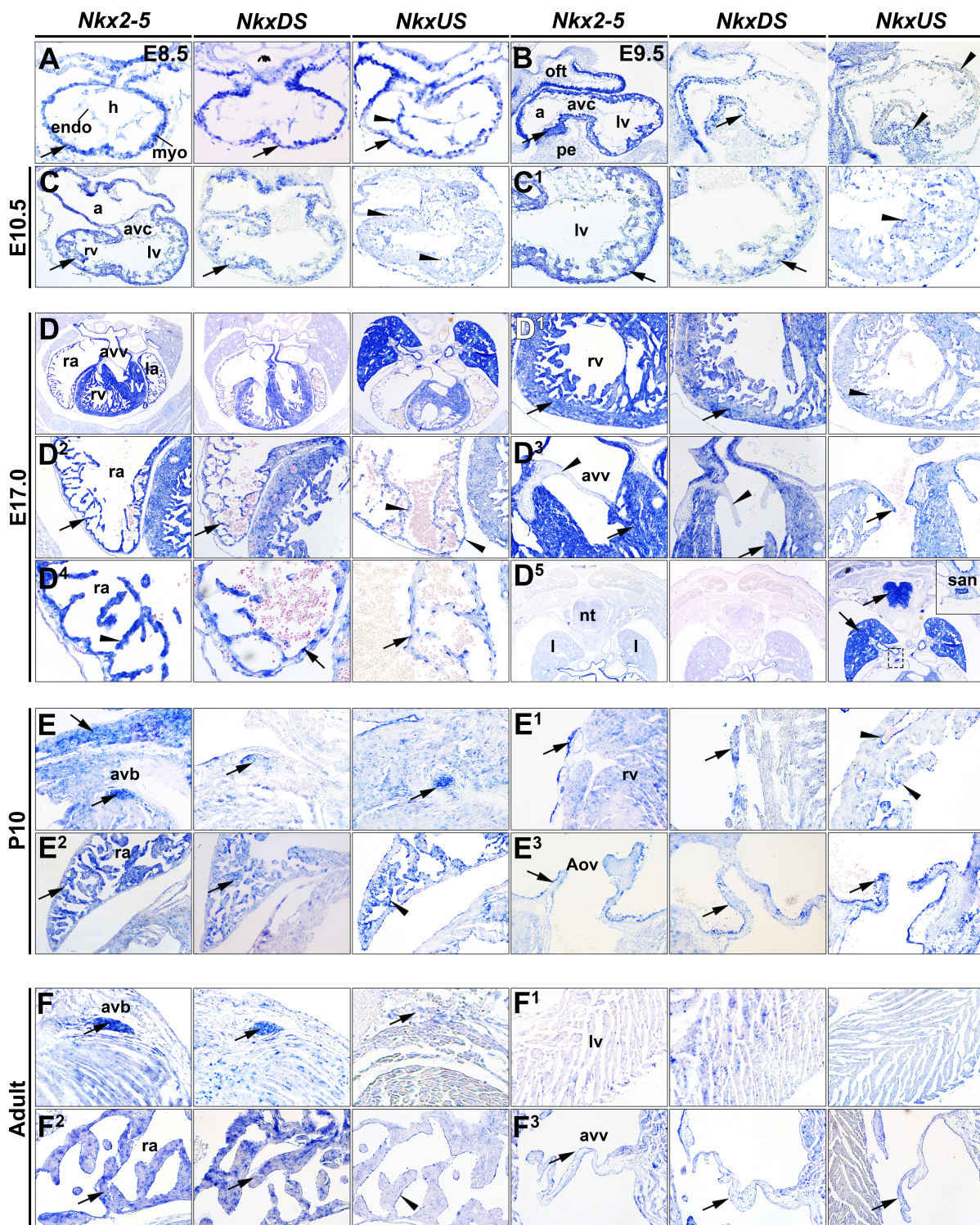
Figure 49 *NkxUS* and *NkxDS* are expressed in the heart throughout mouse development. (A-F³)
Nkx2-5 (left panels), *NkxDS* (middle panels) and *NkxUS* (right panels) expression pattern by *in situ*
hybridisation on sections from embryos at E8.5 (A), E9.5 (B), E10.5 (C-C¹), E17.0 (D-D⁵), postnatal
day 10 (P10, E-E³) and Adult (F-F³). (C-C¹) whole heart view on C and higher magnification of left

ventricle in C[1]. (D-D[5]) whole heart view on D and higher magnification of left ventricle in D[1], right atria in D[2], atrioventricular valves in D[3], an even higher magnification to show the nuclear pattern of NkxDS (middle panels) and *NkxUS* (right panels) compared to *Nkx2-5* (left panels), and extracardiac tissue expression in D[5]. Note the expression of *NkxUS* in the sinoatrial node (D[5] right panel and inset. (E-F[3]) atrioventricular bundle region in E and F, ventricular wall in E[1] and E[1], right atrium in E[2] and F[2], and aortic valve region in E[3] and atrioventricular valve region in F[3]. Heart (h), endocardium (endo), myocardium (myo), atrium (a), atrioventricular canal (avc), outflow tract (oft), left ventricle (lv), right atrium (ra), left atrium (la), right ventricle (rv), left ventricle (lv), atrioventricular valve (avv), neural tube (nt), lungs (l), sinoatrial node (san), atrioventricular bundle (avb), aortic valve (Aov).

II. List of deregulated genes (logFC ± 0.5, FDR <0.05)

Table 23 List of transcripts differentially expressed genes in *NkxUS* knockdown analysis

Symbol	logFC	FDR
NkxUS	-1.974120864	4.83E-76
Orai2	-1.358030924	2.26E-23
Galnt17	-1.142136706	2.86E-05
Hsd3b1	-1.1391927	8.63E-05
Wt1	-1.044732138	0.012887285
Arhgap22	-1.042908738	0.010303458
Zic1	-0.991837961	0.001156969
Garem2	-0.988354033	0.046252393
Tek	-0.961422159	0.028702617
Ttbk1	-0.945494118	0.00280253
Gm17029	-0.918114041	0.047170694
Sh3tc1	-0.918051543	0.005886205
Dpf1	-0.909428954	8.66E-05
Amz1	-0.901187128	0.008631171
Ctgf	-0.89283263	9.25E-06
Duox1	-0.885210452	0.00487689
Bmp4	-0.87870426	0.012146739
Nt5m	-0.870866497	3.82E-09
Hrct1	-0.850166189	0.035334764
Rasgrp3	-0.845104358	1.49E-06
Cdk20	-0.814884167	0.014995941
Fam46b	-0.804781011	0.015568907
Epcam	-0.800247888	0.011516789
Tcaf2	-0.799217157	0.008566907
Mmp17	-0.798339406	0.019364483
Gm29609	-0.795553395	0.013390264
Chd3os	-0.789430301	7.19E-08

Phf19	-0.787422347	1.72E-07
Dntt	-0.786920478	0.017698186
Prkch	-0.779099094	0.004637747
Susd1	-0.771568002	0.011108799
Opn3	-0.763264894	0.032874208
Tmem159	-0.750499567	0.002487276
Dennd3	-0.742928808	0.047431366
Abca12	-0.733934015	0.001664799
Hck	-0.733261244	0.000307319
Cers1	-0.716820235	0.018988133
Sema3g	-0.715471494	6.45E-05
Fbxo40	-0.712666213	0.01757791
Emilin2	-0.710599656	0.008277044
Slco4a1	-0.698823044	0.000235644
Fmnl3	-0.691947318	0.000173049
Baz1b	-0.691402299	1.78E-07
Akna	-0.685616395	0.002673338
Tns4	-0.664274406	0.00035394
Asb12	-0.656043982	0.000369566
Inhba	-0.655015689	0.00755748
9930012K11Rik	-0.652789877	0.027121793
Gss	-0.649898677	1.35E-05
Dmtn	-0.649632325	0.015420245
Cntf	-0.640660575	0.042304777
Naaa	-0.640296594	0.007387519
Abi3	-0.63896704	1.03E-05
Mfsd13a	-0.636936562	5.25E-06
Sh3rf2	-0.632570798	0.00244589
Tnfrsf12a	-0.630222012	2.04E-15
Scn10a	-0.62570447	0.002558532
Dirc2	-0.622655724	0.001704271
Prdm11	-0.620934399	3.72E-06
Dusp27	-0.6178745	0.027183558
Gm42715	-0.617461245	0.024709964
Rhof	-0.609771671	0.030067997
Pus10	-0.608911512	0.001514664
Bdh2	-0.608259038	0.016611478
H2-T10	-0.605524147	0.011889929
Rapsn	-0.604888349	0.001476135
Abca4	-0.60387815	0.006279814
Txnrd3	-0.60199785	0.018739361
Zfp934	-0.599500423	0.028828946

Rab26os	-0.598247136	0.011873459
Fgf13	-0.596593599	0.012511453
Gm10727	-0.595685445	0.004324999
Syt11	-0.591461561	4.45E-09
Osbp2	-0.587857224	0.000147848
Pip5k1b	-0.580387125	0.006026636
Hexim1	-0.576472023	9.25E-08
1190005I06Rik	-0.571983422	0.001337162
9130401M01Rik	-0.57079895	3.50E-06
Gm26795	-0.569022661	0.000373498
Mphosph9	-0.56361739	1.79E-20
Pemt	-0.55882414	0.005967815
AI480526	-0.55224819	0.043304742
Cyp2j6	-0.552037802	0.02145807
Hnf1a	-0.550384482	0.047916817
Lrp8	-0.548957107	0.004380918
Sipa1	-0.547728263	2.71E-07
Gm11627	-0.544808807	0.010240512
Daglb	-0.540520694	4.38E-06
Kcnn1	-0.538779222	0.016952221
Tbx3	-0.538229121	0.04776858
Gpr161	-0.537681086	0.026166133
Kcnv2	-0.529044451	0.015568907
Relt	-0.528812455	0.009994584
Nap1l5	-0.528003196	0.003410031
Mtfp1	-0.527058187	0.006438329
1600002K03Rik	-0.526205958	0.008435839
Ccdc102a	-0.525428873	0.006099038
Dera	-0.521578791	0.000764776
Zfp419	-0.519509066	0.011355676
Sh3rf1	-0.5190187	0.000221983
Nhlrc1	-0.516713669	0.012479401
Palm	-0.513597271	2.41E-06
Mum1l1	-0.512543556	0.033236274
Plppr2	-0.506104652	0.007795673
Gsap	-0.500435447	0.026435263
Hcn4	0.397426257	0.046584237
Sspn	0.501801135	0.000824463
Npdc1	0.502437247	1.88E-05
Chmp6	0.503852053	4.47E-07
Adamts7	0.504625179	0.0342827

Pura	0.505611783	1.05E-13
Dpy19l1	0.505834352	2.00E-06
Ddx6	0.50589409	2.75E-17
Abcc1	0.506004758	0.005537315
Cux1	0.506742156	1.56E-13
Mapk8ip3	0.50792203	4.51E-19
Tcta	0.508283018	0.005679737
AW011738	0.508365956	0.048172239
Tyk2	0.509032269	0.000255444
Usp34	0.509841633	1.12E-11
Ppm1e	0.510621454	0.000157195
Gm45062	0.511197284	0.015063349
Slc12a2	0.511976066	8.89E-07
Nus1	0.512180525	2.67E-12
Phip	0.513149522	7.05E-12
Fgf12	0.513270679	0.003314299
Gm37818	0.515581966	0.034370631
Prune2	0.516341181	0.043304742
Nrbp2	0.51682098	0.012398843
Nav3	0.51754893	3.16E-08
Myzap	0.517977396	0.000540505
Hnrnpc	0.518813664	1.28E-10
march5	0.51949675	7.55E-09
Rpe	0.52084458	8.60E-12
R3hcc1l	0.523083878	1.73E-10
Mapkapk2	0.523295878	2.01E-16
Sema5a	0.524738507	1.09E-05
Nav2	0.525051437	6.37E-07
Fuz	0.526092069	0.047431366
Atf2	0.526998364	4.29E-05
Pakap	0.527452152	1.12E-06
Il6st	0.528484938	3.74E-07
Vopp1	0.528593892	1.37E-10
A730056A06Rik	0.528954433	0.035728039
Tmtc3	0.529249651	1.30E-13
Cep57	0.529400396	8.97E-18
Gm14325	0.529799193	0.000323006
Clip1	0.529800242	2.78E-11
Rnf19b	0.529954652	1.44E-07
Rgs6	0.529955596	0.002044157
Chd1l	0.530222541	1.88E-09
Kazn	0.530763191	0.000428627

Cpeb2	0.531190342	5.00E-10
Utp4	0.53134533	4.43E-07
Slc9a5	0.531470307	0.030928354
Tcf12	0.531911994	6.14E-15
Zfp277	0.532132546	0.002723471
Large1	0.53296516	0.007082902
Kpna4	0.534854635	2.74E-05
Setd1b	0.536568031	9.67E-07
Rexo5	0.537821452	0.000149665
Dhx57	0.538750723	8.62E-06
Gatb	0.539952603	1.27E-07
Commd1	0.541087667	3.85E-05
Tmem87b	0.541937865	1.38E-06
2500002B13Rik	0.542438059	0.003287568
Ccdc112	0.542883757	1.29E-05
Tbc1d1	0.542954547	7.26E-12
Snx25	0.546545874	2.94E-10
Dmd	0.546563094	3.98E-10
Me1	0.546837818	1.50E-14
Senp6	0.547356758	1.10E-17
1110019D14Rik	0.547571272	0.015689074
Arhgef4	0.548109672	0.000127129
Kmt5b	0.549174703	1.15E-15
Zfp704	0.549670662	1.25E-06
Arfip2	0.549850206	5.99E-06
Nrip3	0.55151955	0.002040463
Tmeff2	0.552835601	0.018493203
Bpnt1	0.55308232	5.19E-07
Adcy5	0.555507168	7.41E-06
Lrif1	0.556436868	6.21E-14
Mib1	0.556748903	8.49E-18
Zfp871	0.557767529	1.82E-15
Sec24a	0.557957438	1.06E-07
Rmnd5a	0.559539247	2.92E-15
Brd8	0.560030725	1.49E-10
Mettl8	0.561932514	0.030172976
H19	0.562941168	0.030153482
Setd2	0.563203514	9.89E-30
Gm37084	0.563278478	0.000181552
Stradb	0.56427769	2.90E-05
Slc7a2	0.565127204	0.000176515
Pde3b	0.565342965	9.21E-16

Gramd3	0.565988262	8.90E-05
Fnbp1l	0.566566378	1.36E-16
Vps41	0.567246071	1.34E-09
Prkce	0.570118206	8.18E-07
Cav1	0.570762127	1.26E-06
Rabgap1l	0.571462943	4.56E-09
Tom1l2	0.572772831	2.23E-08
BC017158	0.572836785	2.46E-06
Ide	0.572947421	4.87E-08
Ankrd13d	0.574197447	0.02779753
Kin	0.574554711	0.000163459
Gls	0.577421023	4.48E-16
Dennd4a	0.578185141	2.90E-05
Cand1	0.578643729	1.60E-23
Zfp949	0.581209955	0.000116973
Acbd5	0.581514126	2.61E-14
Alcam	0.581564102	0.00680135
Zrsr2	0.582207815	3.94E-17
Map4k3	0.583444323	1.11E-10
Ptpn23	0.583722172	4.51E-19
Ube3a	0.584849368	6.46E-07
Atxn1	0.58488787	0.006322855
NA	0.58499627	1.37E-10
1700020I14Rik	0.586459405	4.88E-13
Mbd5	0.587169706	5.72E-05
Cnot2	0.587919612	4.93E-13
Tmem127	0.588944877	7.62E-17
Ankrd12	0.590186323	3.92E-17
Dock11	0.594059186	0.007165546
Tbc1d16	0.595842717	1.31E-08
Invs	0.596943788	1.54E-07
Spred2	0.596960772	1.04E-06
Aven	0.597104724	0.000315525
Cacna1d	0.598310831	1.24E-05
Phf8	0.598714955	9.03E-06
Gatad1	0.599555264	3.60E-20
Pitpnc1	0.600109816	3.55E-05
Mgea5	0.600751364	2.23E-27
Ints6	0.601442739	5.62E-15
Katna1	0.601523409	1.61E-07
Tmem209	0.601535858	0.000227056
Wdpcp	0.602165708	0.001007041

Unc13c	0.603094078	2.56E-05
Gm12696	0.60811408	0.020581817
Trim23	0.608181719	1.46E-06
Ccdc28a	0.608319098	0.02655378
Notch3	0.608344593	0.010403402
Zxdb	0.608586333	9.33E-08
Krit1	0.608955524	1.29E-09
Ppm1l	0.611489803	0.011108799
Hells	0.612268477	1.32E-17
Proser1	0.614188373	2.36E-13
Brca2	0.61643117	0.005523516
Gm11400	0.616958306	0.033840457
Gm19721	0.617080025	0.038564556
Coq3	0.617722667	4.39E-08
Scfd2	0.622439704	0.006311148
Atp11a	0.623051758	1.28E-05
Acer3	0.623472077	2.34E-07
Tln1	0.624502243	3.70E-32
Uaca	0.625530149	1.72E-14
Adar	0.625626238	0.004592875
Lrrc28	0.626923263	4.28E-05
Nfxl1	0.630175758	2.98E-06
Appl2	0.630377497	7.90E-06
Twsg1	0.630749626	1.02E-10
Dock7	0.632620114	0.001670481
Gstt1	0.633627435	9.27E-12
Zhx1	0.637172725	6.82E-05
Zcchc6	0.638577895	1.00E-20
Ranbp17	0.641403053	4.42E-06
Exoc4	0.646239528	6.52E-17
Hectd1	0.646463317	5.41E-41
Tcp11l2	0.647596538	0.018807535
Pros1	0.650085667	0.000701967
Sash1	0.653207925	1.05E-13
Gm10138	0.654359701	0.00315873
Gm20186	0.655118433	0.002437542
Zcchc7	0.655858547	3.11E-09
Scaper	0.656149729	7.72E-05
Kdm7a	0.658062996	5.13E-05
Pard3	0.661161082	1.90E-12
Jmy	0.66416667	2.80E-20
Ptpn4	0.665616873	2.13E-18

Ago3	0.66635883	1.23E-08
Nono	0.667361205	6.79E-46
Frem1	0.669151326	5.12E-08
Ahcyl2	0.671758231	1.04E-05
Smarcad1	0.673009413	1.49E-07
Stxbp3	0.67419113	1.72E-11
Tcim	0.674268561	0.019585888
Dgkh	0.674376759	0.002509803
Sfi1	0.675651663	7.26E-12
Dmtf1	0.677162345	7.85E-21
Prim2	0.679078213	2.71E-05
Gm43737	0.679191062	0.005329922
Megf9	0.679555358	7.49E-20
Yaf2	0.680245147	4.70E-09
Plp1	0.682005854	0.033946157
Gfpt2	0.690296785	5.49E-05
Palm2	0.691099186	1.32E-12
Vps8	0.697325358	8.25E-12
Pla2r1	0.700059435	0.021447671
F8	0.701598051	0.003168731
Glcci1	0.701951686	0.013652924
Lonrf3	0.702254498	1.79E-05
Serpinb6a	0.702583827	4.39E-17
2210408I21Rik	0.70286362	0.000710338
Ankrd13b	0.703652105	4.35E-14
Fdx1l	0.704588671	4.50E-08
Ahi1	0.704924864	0.000157429
Lamp2	0.705691487	2.67E-09
Lpp	0.708295246	2.65E-09
Neat1	0.708640702	0.000544882
Ttc19	0.709521297	5.22E-12
Herc3	0.709540579	8.08E-06
Chmp7	0.709879549	6.25E-13
Pdia6	0.711454956	2.52E-14
Agtrap	0.713098632	2.67E-19
Efr3a	0.715691404	5.26E-08
Alg13	0.716046222	1.53E-09
Cdk17	0.716117781	3.73E-11
Atp11c	0.71967315	3.11E-09
Scrn2	0.719962398	5.46E-06
Api5	0.723518498	9.25E-25
Plscr3	0.725890602	5.79E-14

Gab2	0.726193843	3.78E-06
Cutal	0.727554269	0.003820553
Atp5s	0.728124355	0.000492102
Rapgef1	0.728607701	1.38E-12
Gm38300	0.730156805	0.008119
Phka1	0.739070645	4.62E-16
Gm38057	0.739369182	0.01697366
Srgap2	0.740315314	6.77E-18
Celf2	0.740359738	2.70E-07
Ccdc50	0.742062743	1.20E-21
Gm2788	0.748883119	0.031138567
Wdr35	0.749344575	1.37E-22
Tcf20	0.749531184	1.68E-31
Rif1	0.75116264	1.25E-25
Strn	0.755141068	0.001117305
Trim36	0.755323803	3.70E-06
Ccdc171	0.757983232	1.66E-06
Nova1	0.760162334	6.32E-09
Stx7	0.760367877	2.71E-14
Gm17066	0.760434538	0.002865004
Lnpep	0.763115576	2.26E-10
Kat2b	0.763312872	2.59E-36
Gnaq	0.763997667	9.25E-10
Ryr2	0.765445568	0.000138009
Nrp1	0.766682052	6.69E-27
Rmdn1	0.767376298	0.001518245
Sox6	0.769216655	6.64E-07
Nod1	0.770246873	0.002783039
Fbxl17	0.771650782	3.82E-05
Tmem218	0.775838277	0.008566907
Herc1	0.777725148	1.25E-11
Gng12	0.77912364	1.23E-25
Trim34b	0.779344416	0.038887263
Rbms1	0.780735178	2.06E-17
Mcf2l	0.781870641	1.81E-05
Tiam1	0.782070643	1.65E-15
Zfp449	0.783543961	1.05E-05
Tsc1	0.784838295	7.21E-15
Osbpl11	0.784956135	5.12E-27
Gm38009	0.789337822	0.037911201
Cacna1c	0.790837569	8.56E-06
Stag2	0.796719066	2.62E-10

Gm29488	0.80131887	0.02889418
Ago4	0.802879365	5.70E-07
Agfg1	0.802911845	7.54E-40
Ncoa5	0.805478329	6.37E-17
Slc38a3	0.805676915	4.96E-10
Osbpl8	0.808589694	9.05E-09
Cyp20a1	0.811888658	1.26E-12
Slc12a6	0.813125806	0.000450304
Fam227a	0.814961416	0.029768105
Wdfy1	0.819101518	5.68E-07
Lamc1	0.819846211	6.52E-17
2010111I01Rik	0.821211733	3.70E-16
Mocs1	0.823755191	2.80E-05
Pdcd4	0.832948449	0.001088976
Cyld	0.834048947	1.53E-18
Klf7	0.835147476	2.95E-29
Mdm4	0.836036912	3.09E-29
Loxl3	0.836492846	6.28E-07
Mier3	0.83923526	2.05E-17
Ssbp2	0.843219747	4.26E-16
Cldn34c1	0.846288365	0.001580045
Zfp280d	0.84777949	6.12E-06
Sgcb	0.84946443	9.52E-38
Nbeal1	0.850016572	5.95E-24
Cdc14b	0.85342265	4.99E-12
Dtna	0.853707583	5.99E-12
Tfcp2	0.854851192	3.42E-05
Gm21738	0.858337926	0.02269822
Dph6	0.859119008	3.48E-11
Atl1	0.85968985	1.27E-05
Mturn	0.861343771	0.012408788
Fsd1l	0.862961087	2.82E-08
Mdn1	0.864996515	1.76E-31
Efcab11	0.869911374	0.004539079
Polk	0.870864066	1.49E-07
Ranbp3	0.874128478	2.49E-20
Klhl12	0.876060297	7.40E-22
2810030D12Rik	0.881016264	0.044930397
Lrrc1	0.883777332	9.04E-24
Gm6211	0.885905429	0.030144114
Pank2	0.887642874	8.70E-20
Pogz	0.889415948	3.64E-40

Fip1l1	0.890805727	5.89E-48
Galm	0.892466241	1.70E-09
Ttc27	0.893911468	5.78E-24
Usp38	0.895785367	8.23E-32
Gm38257	0.89978993	0.011713787
Spag9	0.902290666	6.20E-15
Otud7b	0.9062282	2.73E-44
Polq	0.908548725	1.06E-10
Dock4	0.908916697	0.016193948
Gbf1	0.924548294	1.34E-34
Nup160	0.933912121	8.40E-17
Exoc2	0.944597562	3.99E-34
D5Ertd579e	0.947063427	3.30E-59
Lama3	0.947761612	0.003001571
Tmem185a	0.970922586	3.96E-20
Dnah7c	0.975654896	0.034247481
Crtc3	0.978736202	4.72E-33
Pikfyve	0.980039817	1.10E-20
Notch1	0.985748153	1.06E-10
Cep83	0.985879816	2.37E-24
Plscr4	0.986962635	3.86E-08
Usp43	0.998638933	1.74E-06
Gm45713	0.999455427	0.00117145
Slc1a4	1.00584915	1.23E-05
Bcas3	1.007973009	1.77E-10
Rexo4	1.008591696	1.26E-21
Tomm70a	1.011800009	1.12E-56
Zmynd11	1.012099687	1.14E-47
A930001C03Rik	1.016514492	7.96E-05
Klf12	1.022369059	0.020156313
Mecp2	1.024677799	7.55E-06
Tanc2	1.02565436	2.93E-44
Rbm6	1.027756735	2.78E-28
Ccdc122	1.038380126	0.003171914
Arap2	1.040471516	1.73E-10
Taf1b	1.044940973	5.29E-13
Asah2	1.047584254	8.37E-18
Hsd17b12	1.049501486	2.71E-07
Mindy2	1.052860391	4.11E-29
Mlip	1.063695444	4.24E-21
Deptor	1.064103883	0.018493203
Arih1	1.067754088	9.30E-51

Prickle2	1.069925711	1.24E-08
Gmds	1.07701884	5.21E-05
Rbms3	1.081159534	4.20E-11
Shprh	1.086341543	3.21E-42
Mtus1	1.100343818	2.17E-09
Slc44a3	1.104543221	2.05E-06
Zfp654	1.104633376	2.81E-41
Poc5	1.111658578	2.49E-24
Zfp984	1.112144528	4.12E-38
Gopc	1.114933363	1.09E-20
Rapgef4	1.128380439	0.008338387
Hint3	1.134919934	3.97E-07
Rptor	1.138106071	4.44E-33
Arl5b	1.140183653	1.14E-34
Nup210l	1.15878975	0.006740568
Rbm27	1.173521588	2.86E-40
Kcnj3	1.176381386	6.14E-15
Slc9a7	1.180037521	3.66E-07
Rbm41	1.183252039	1.54E-22
Rsrc1	1.19079515	1.52E-25
Poc1b	1.197340753	1.63E-11
Nek6	1.207437204	1.60E-15
Lgr4	1.214235525	8.20E-21
Fbln7	1.226007467	8.63E-05
Diaph3	1.227761172	6.98E-28
Lonp2	1.255745691	9.74E-43
Gm21986	1.267190999	0.015155808
Ip6k1	1.267368059	1.04E-32
Fgfr1op	1.27675076	4.95E-36
Hbs1l	1.283156925	4.58E-35
Ammecr1	1.293957005	2.72E-58
Fancc	1.303598428	1.42E-11
Sh3bp4	1.310001821	3.31E-10
Fam219b	1.340268242	6.69E-21
Gnb1l	1.347917518	9.77E-23
Gm43753	1.350331497	0.005329922
Taok3	1.351487243	3.98E-14
Tbc1d4	1.372355551	3.96E-34
Gm37738	1.376424414	1.04E-07
Cwc27	1.388124255	2.30E-15
Fam172a	1.396494174	2.74E-48
Ica1	1.415767746	4.03E-12

Btbd10	1.417721409	7.85E-61
Ankrd55	1.437333604	0.009640606
Acsl4	1.440635066	1.08E-48
Mycbp2	1.444967474	7.01E-57
Ctnnd1	1.457593256	3.35E-33
Vamp7	1.463083131	5.36E-31
Pcx	1.465196651	3.68E-31
Exoc6b	1.467586371	1.47E-55
Palmd	1.495769816	3.96E-34
Gm42892	1.520106767	9.64E-05
Gm26510	1.529337167	8.78E-05
Per3	1.539654047	3.91E-29
Tbck	1.576580385	1.29E-31
Cryl1	1.589409112	3.63E-29
Fam35a	1.607338974	3.42E-10
Pde5a	1.620509551	3.87E-21
Ust	1.622070458	4.20E-09
Ctnnbl1	1.636089019	3.65E-28
Ccdc138	1.644964727	4.57E-13
4833413G10Rik	1.653830449	1.41E-06
Fuca2	1.65942174	2.92E-94
Pyurf	1.678505383	3.41E-80
Cltc	1.682190321	2.81E-86
Scfd1	1.692893302	1.36E-56
Lnpk	1.697797393	9.12E-55
Elavl1	1.705111947	1.01E-102
Alpk1	1.710183771	3.91E-67
Tsc22d2	1.714227811	4.68E-60
Cblb	1.717717725	2.93E-83
Nr6a1	1.721387563	2.63E-06
Reps1	1.742766854	4.61E-77
Gigyf1	1.787123727	6.43E-64
Fam122b	1.796381138	1.54E-57
Corin	1.799794852	1.82E-24
Gm42876	1.901459785	7.51E-05
Lyrm7	1.910861429	1.25E-11
Mthfd1l	1.91562277	3.98E-32
Gm37397	1.974947315	2.62E-21
Canx	1.984053386	7.29E-69
Zfp185	1.988845677	4.62E-08
Zfand3	2.032140633	2.17E-40
Wwc2	2.037375076	1.76E-73

Ptprs	2.038800689	9.57E-118
Trip12	2.116709008	1.01E-65
Slf1	2.14751243	6.29E-58
Arl13b	2.213658877	2.25E-54
Slc4a1ap	2.34550361	1.55E-76
Gm38374	2.42382019	8.92E-08
Rnf31	2.452224203	9.34E-78
Gm9403	2.623884742	1.86E-07

III.   **List of genes identified using StringDB**

Table 24 List of StringDB genes

Symbol	logFC	FDR
Orai2	-1.36	2.26E-23
Wt1	-1.04	1.29E-02
Arhgap22	-1.04	1.03E-02
Zic1	-0.99	1.16E-03
Tek	-0.96	2.87E-02
Ctgf	-0.89	9.25E-06
Bmp4	-0.88	1.21E-02
Nt5m	-0.87	3.82E-09
Rasgrp3	-0.85	1.49E-06
Prkch	-0.78	4.64E-03
Hck	-0.73	3.07E-04
Baz1b	-0.69	1.78E-07
Gss	-0.65	1.35E-05
Cntf	-0.64	4.23E-02
Abi3	-0.64	1.03E-05
Scn10a	-0.63	2.56E-03
Rhof	-0.61	3.01E-02
Rapsn	-0.60	1.48E-03
Fgf13	-0.60	1.25E-02
Pip5k1b	-0.58	6.03E-03
Lrp8	-0.55	4.38E-03
Sipa1	-0.55	2.71E-07
Kcnn1	-0.54	1.70E-02
Tbx3	-0.54	4.78E-02
Gpr161	-0.54	2.62E-02
Kcnv2	-0.53	1.56E-02
Hcn4	0.40	4.66E-02
Sspn	0.50	8.24E-04

Chmp6	0.50	4.47E-07
Ddx6	0.51	2.75E-17
Abcc1	0.51	5.54E-03
Cux1	0.51	1.56E-13
Tyk2	0.51	2.55E-04
Fgf12	0.51	3.31E-03
Prune2	0.52	4.33E-02
Nav3	0.52	3.16E-08
Sema5a	0.52	1.09E-05
Nav2	0.53	6.37E-07
Atf2	0.53	4.29E-05
Il6st	0.53	3.74E-07
Cep57	0.53	8.97E-18
Clip1	0.53	2.78E-11
Rgs6	0.53	2.04E-03
Chd1l	0.53	1.88E-09
Tcf12	0.53	6.14E-15
Kpna4	0.53	2.74E-05
Tmem87b	0.54	1.38E-06
Tbc1d1	0.54	7.26E-12
Dmd	0.55	3.98E-10
Me1	0.55	1.50E-14
Arhgef4	0.55	1.27E-04
Adcy5	0.56	7.41E-06
Mib1	0.56	8.49E-18
Sec24a	0.56	1.06E-07
Mettl8	0.56	3.02E-02
H19	0.56	3.02E-02
Pde3b	0.57	9.21E-16
Vps41	0.57	1.34E-09
Prkce	0.57	8.18E-07
Cav1	0.57	1.26E-06
Gls	0.58	4.48E-16
Cacna1d	0.60	1.24E-05
Notch3	0.61	1.04E-02
Acer3	0.62	2.34E-07
Tln1	0.62	3.70E-32
Twsg1	0.63	1.02E-10
Gstt1	0.63	9.27E-12
Exoc4	0.65	6.52E-17
Pros1	0.65	7.02E-04
Pard3	0.66	1.90E-12

Ago3	0.67	1.23E-08
Nono	0.67	6.79E-46
Stxbp3	0.67	1.72E-11
Sfi1	0.68	7.26E-12
Gfpt2	0.69	5.49E-05
Vps8	0.70	8.25E-12
F8	0.70	3.17E-03
Neat1	0.71	5.45E-04
Chmp7	0.71	6.25E-13
Cdk17	0.72	3.73E-11
Api5	0.72	9.25E-25
Gab2	0.73	3.78E-06
Rapgef1	0.73	1.38E-12
Stx7	0.76	2.71E-14
Lnpep	0.76	2.26E-10
Kat2b	0.76	2.59E-36
Gnaq	0.76	9.25E-10
Ryr2	0.77	1.38E-04
Nrp1	0.77	6.69E-27
Gng12	0.78	1.23E-25
Mcf2l	0.78	1.81E-05
Tiam1	0.78	1.65E-15
Tsc1	0.78	7.21E-15
Cacna1c	0.79	8.56E-06
Stag2	0.80	2.62E-10
Ago4	0.80	5.70E-07
Agfg1	0.80	7.54E-40
Lamc1	0.82	6.52E-17
Mocs1	0.82	2.80E-05
Pdcd4	0.83	1.09E-03
Sgcb	0.85	9.52E-38
Cdc14b	0.85	4.99E-12
Dtna	0.85	5.99E-12
Tfcp2	0.85	3.42E-05
Ranbp3	0.87	2.49E-20
Spag9	0.90	6.20E-15
Nup160	0.93	8.40E-17
Exoc2	0.94	3.99E-34
Lama3	0.95	3.00E-03
Pikfyve	0.98	1.10E-20
Notch1	0.99	1.06E-10
Arap2	1.04	1.73E-10

|		
Taf1b	1.04	5.29E-13
Asah2	1.05	8.37E-18
Deptor	1.06	1.85E-02
Gmds	1.08	5.21E-05
Rapgef4	1.13	8.34E-03
Rptor	1.14	4.44E-33
Kcnj3	1.18	6.14E-15
Nek6	1.21	1.60E-15
Fgfr1op	1.28	4.95E-36
Ammecr1	1.29	2.72E-58
Tbc1d4	1.37	3.96E-34
Acsl4	1.44	1.08E-48
Vamp7	1.46	5.36E-31
Pcx	1.47	3.68E-31
Exoc6b	1.47	1.47E-55
Pde5a	1.62	3.87E-21
Elavl1	1.71	1.01E-102
Cblb	1.72	2.93E-83
Mthfd1l	1.92	3.98E-32
Canx	1.98	7.29E-69
Ptprs	2.04	9.57E-118
Trip12	2.12	1.01E-65
Atp2a2	0.28	4.72E-02

IV. Complete western blot
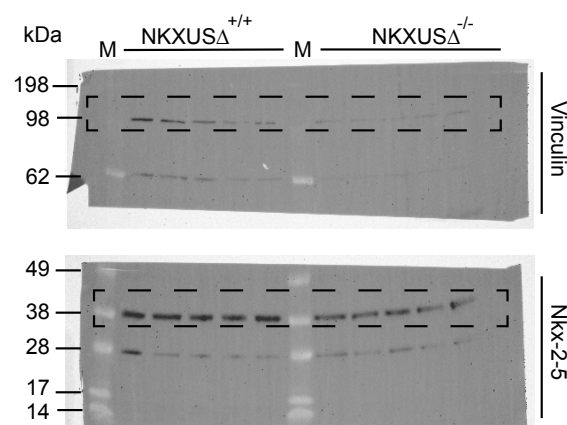
Figure 50 Complete western blot of Figure 35. Complete western blot from five NkxUSΔ$^{-/-}$ and NkxUSΔ$^{+/+}$ mice with antibodies against Vinculin and NKX2-5. The indicated region (dashed rectangle) was used for Figure 35. The protein sizes are indicated on the right in kDa. M= Marker.

V.   List of primers

Table 25 List of PCR primers

Nkx2-5/DS junction 1	FW	CCAATGGCAGGCTGAATCC
Nkx2-5/DS junction 2	FW	GGAACGACTCCCACCTTTAG
Nkx2-5/DS junction 3	FW	ATGTTCCCCAGCCCTGGG
Nkx2-5/DS junction 4	FW	CCCAAGTGCTCTCCTGCTTTC
Nkx2-5/DS junction 5	FW	GAGCTGGACAAAGCCGAGGAC
Nkx2-5/DS junction 6	FW	CTACGGCGTGGGTCTCAAT
Nkx2-5/DS junction 1	RV	CTCTTCCCATTAAA GTGAGTGCG
Nkx2-5/DS junction 2	RV	GGATTTCTGAGCGAGCGCAGAGG
Nkx2-5/DS junction 3	RV	CATCCTCAGGCCAACTTGAT
human NkxUS junction	FW	AGACCCCACACTTCAGGGAG
	RV	GGACTCGCCTCATTCAGGT
mNkxUS upstream first exon	FW	GGGTGTTTGCGC TTTTTGCT
	RV	CTGTAGCGCTGTAGCGGTTT
m13	FW	CGCCAGGGTTTTCCCAGTCACGAC
	RV	TCACACAGGAAACAGCTATGAC

Acknowledgments

I would like to thank Prof. Dr. Richard Harvey for giving me the opportunity to undertake my PhD in his research group, for his continuous support and the possibility to think and act independently while constantly providing with critical feedback where necessary. I really enjoyed working on this project.

I would also like to express my gratitude to Prof. Dr. Thorsten Hoppe for agreeing to be my supervisor at the University of Cologne, for finding the time to meet me when I was in Germany and to evaluate this work.

A big thank you to Dr. Nicole Schonrock, who started this project, co-supervised me during my PhD and has always been there for me. I enjoyed working with her and would like to thank Nicole for her enthusiasm, patience, motivating discussions and encouragement. I greatly appreciate her help and guidance.

I would also like to thank all the previous and present Harvey lab members for the great time inside and outside the lab, their constant support, valuable advice and discussions. I would also like to thank everyone at the Victor Chang Institute who has helped me in many different ways.

I would like to extend my gratitude to Prof. Dr. Niels H. Gehring, Prof. Dr. Werr, and to Dr. Andre Franz for being part of my thesis committee.

Special thanks to Stuart, my family and friends for continuous support, their understanding, patience and encouragement!

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind – in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie – abgesehen von unten angegebenen Teilpublikationen – noch nicht veröffentlicht worden ist, sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Thorsten Hoppe und Prof. Dr. Richard Harvey betreut worden.

_____   _____

Datum, Ort   Unterschrift

Curriculum vitae

Personal Data:

Name:	Ann-Kristin Alteköster
Date of birth:	17.02.1986
Place of birth:	Herford, Germany
Nationality:	German
E-mail:	a.altekoester@victorchang.edu.au

Academic studies:

2015-2019	**Doctoral thesis** at the University of Cologne and the Victor Chang Cardiac Research Institute. Prof. Thorsten Hoppe Prof. Richard Harvey
2014	**Work experience** at the Victor Chang Cardiac Research Institute, Sydney, Australia.
2011-2013	**Master of Science (M.Sc.)** in the field of Biological Sciences, University of Cologne, Germany.
2008-2011	**Bachelor of Sciences (B.Sc.)** in the field of Biological Sciences, University of Cologne, Germany.
2007	**Bachelor of Sciences (1. Term)** in the field of Biological Sciences, Dalhousie University, Halifax, and University of Alberta, Edmonton Canada.
2005-2006	**Diplom (Term 1-2)** in the field of Chemistry, Heinrich-Heine University, Düsseldorf, Germany.

Education:

1996-2005	**Abitur,** Franz-Stock Gymnasium, Arnsberg