



**University of
Nottingham**

UK | CHINA | MALAYSIA

Genetic studies on congenital heart disorders

José Juan Aparicio Sánchez; BSc, MSc

School of Life Sciences

University of Nottingham

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

September 2019

Main supervisor – Prof. J David Brook

DECLARATION

I hereby declare that this thesis is the result of my own work, which has been undertaken during my period of registration for the PhD degree at The University of Nottingham, and that it has not been presented to this or any other university for the award for any degree or diploma other than that for which I am a candidate.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor David Brook for giving me the opportunity to join his lab and supporting me since the very first day of my PhD thesis. His help, guidance and expertise have proven invaluable throughout this journey, and I can say without any doubt that he is the best supervisor any student could ever ask for. I would also like to express my gratitude to all the past and current members of the Brook's lab with whom I have shared so many amazing moments during these past years. Thanks for making of me a better scientist and a better person: Tushar, Sarah, Kerry, Anna, Tasabeeh, Marzena, Marta, Naveed, Philippine, Amy, Zhefeng, Paulina, Belle, Anjani and Owen (sorry if I am forgetting anyone!).

I am also extremely grateful to Siobhan Loughna, which has been like another supervisor to me. Her invaluable insight and advice have had a tremendously positive impact on this thesis. Thanks also to the members of her lab, especially Sophie, Chris and Qazi with whom I have worked closely and have helped me in many occasions. I would also like to thank Tom Killelea (Ed Bolt's lab), Hannah Belsham (Claire Friel's lab) and Thomas Miller (Catarina Gadelha's lab) for their helpful advice and technical help on recombinant protein production and biochemical assays.

It cannot be expressed with words how grateful I am to Laura, my girlfriend. We decided to embark on this journey together more than 3 years ago, and every single day since then, she has proven to be the best partner someone could ever have and an endless source of encouragement, resilience and positivism on all those difficult moments I have had to go through along the years. I would like to thank my father, my mother and my sister for being there for me since I was a child. Without their hard work, extensive education and moral support I could not have achieved as many things as I have throughout the years.

Last but not least, I would like to express my gratitude to The University of Nottingham, which through the EU Vice-Chancellor's Scholarship for Research Excellence have funded my work during these 3 years.

TABLE OF CONTENTS

1. INTRODUCTION.....	16
1.1. Cardiogenesis: a quick overview.....	16
1.1.1. Gastrulation	16
1.1.2. Cardiac crescent formation	17
1.1.3. Linear heart tube looping and chamber formation	17
1.1.4. Endocardial cushions formation	18
1.1.5. Cardiac chambers and outflow tract septation	21
1.2. Congenital heart disease	24
1.2.1. Categories	24
1.2.2. Quality of life	27
1.2.3. Etiology	27
1.2.4. Potential treatments.....	29
1.3. Cardiac transcription factors	29
1.3.1. T-box family and TBX5	30
1.3.2. NK-type homeodomain factor NKX2-5	32
1.3.3. GATA family of transcription factors	33
1.3.4. MEF2 family	34
1.3.5. HAND1 and HAND2.....	35
1.3.6. Other cardiac developmental genes.....	36
1.4. Genome-editing technologies in cardiovascular biology	37
1.4.1. Tools for genome editing.....	38
1.5. Animal models for the study of cardiac development and CHD	43
1.5.1. Mouse	43
1.5.2. Zebrafish	45
1.6. Hypothesis and aims.....	50
2. MATERIALS AND METHODS.....	51
2.1. Basic techniques	51
2.1.1. DNA and RNA manipulation techniques.....	51
2.1.2. Protein manipulation techniques	58
2.1.3. Cell culture.....	61
2.1.4. Bacteriological techniques.....	64
2.2. ACETYLATION OF TBX5 AND ITS ROLE DURING CARDIAC DEVELOPMENT ..	65

2.2.1.	In-vitro studies	65
2.2.2.	In-vivo studies.....	67
2.3.	ANALYSIS OF CHD4 MUTATIONS FOUND IN SYNDROMIC CHD PATIENTS ..	82
2.3.1.	Plasmids and constructs	82
2.3.2.	Expression of CHD4 mutant variants in mammalian cells	84
2.3.3.	Pull-down assays.....	85
2.3.4.	Immunofluorescence and cell imaging.....	85
2.3.5.	Recombinant CHD4 production in E.coli.....	86
2.3.6.	Recombinant CHD4 production using the baculovirus-insect cell expression system	88
2.3.7.	Nucleosome assembly	91
2.3.8.	Nucleosome/DNA binding assay.....	91
2.3.9.	Nucleosome sliding assay	92
2.3.10.	ATPase activity assay	92
2.4.	MOUSE WORK.....	93
2.4.1.	Mouse lines.....	93
2.4.2.	Mice maintenance and welfare	94
2.4.3.	DNA extraction for genotyping.....	94
2.4.4.	Genotyping	95
2.4.5.	Embryonic heart dissection	95
2.4.6.	β -Galactosidase staining of whole embryos/organs	96
2.4.7.	HREM Sample Preparation	97
2.4.8.	High-resolution episcopic microscopy.....	99
2.5.	STATISTICAL ANALYSIS.....	99
3.	ACETYLATION OF TBX5 AND ITS ROLE DURING CARDIAC AND FIN DEVELOPMENT	101
3.1.	Introduction.....	101
3.1.1.	Acetylation as a post-translational modification.....	101
3.1.2.	HDACs and HATs in cardiac development	101
3.1.3.	Hypothesis and aims of this project	106
3.2.	Preliminary results & data	107
3.2.1.	KAT2A and KAT2B interact with and acetylate TBX5.....	107
3.2.2.	TBX5 acetylation potentiates its transcriptional activity.....	108
3.2.3.	Mapping the acetylation residue in TBX5.....	109

3.3.	Results.....	115
3.3.1.	Acetylation of TBX5 and nucleocytoplasmic shuttling	115
3.3.2.	<i>kat2a</i> and <i>kat2b</i> expression in zebrafish	122
3.3.3.	Morpholino-mediated knockdown of <i>kat2a</i> , <i>kat2b</i> and <i>tbx5a</i>	124
3.3.4.	sgRNA design and synthesis.....	129
3.3.5.	Checking target sequences presence in the zebrafish genome	129
3.3.6.	sgRNA activity <i>in vitro</i> assay	130
3.3.7.	CRISPR/Cas9-mediated knockout of <i>kat2a</i> , <i>kat2b</i> and <i>tbx5a</i>	131
3.3.8.	Analysis of mutations in CRISPR/Cas9-mediated KOs	136
3.3.9.	<i>kat2a</i> and <i>kat2b</i> KDs/KOs present abnormal fin development.....	139
3.3.10.	HAT inhibitor treatment of zebrafish embryos during development	141
3.3.11.	Expression levels of <i>Tbx5a</i> target genes in <i>kat2a</i> and <i>kat2b</i> KOs	145
3.3.12.	Rescue of the <i>kat2a</i> KO cardiac and fin phenotype by a <i>tbx5a</i> acetyl-mimic transcript.....	147
3.4.	Conclusions, discussion and perspectives for future work.....	150
4.	ANALYSIS OF CHD4 MUTATIONS FOUND IN SYNDROMIC-CHD PATIENTS	161
4.1.	Introduction	161
4.1.1.	The NuRD complex and its role in cardiac development.....	161
4.1.2.	CHD4 mutations found in syndromic CHD4 patients	164
4.1.3.	Hypothesis and aims of this project	168
4.2.	Results.....	168
4.2.1.	Analysis of CHD4 mutations found in syndromic CHD patients	168
4.2.2.	Generation and expression of CHD4 mutant variants in mammalian cells	170
4.2.3.	Physical interaction between CHD4 variants and the NuRD complex subunits	173
4.2.4.	Cellular localization of CHD4 mutant variants.....	177
4.2.5.	ATPase activity of CHD4 mutant variants (produced in COS7 cells)..	177
4.2.6.	CHD4 protein production in Sf9 insect cells	179
4.2.7.	ATPase assay using recombinant CHD4 produced in Sf9 cells.....	182
4.2.8.	DNA and nucleosome binding assays	183
4.2.9.	Nucleosome sliding assay	184
4.2.10.	CHD4 protein production in <i>Escherichia coli</i>	185
4.2.11.	ATPase assay using recombinant CHD4 produced in <i>E.coli</i>	186

4.3.	Conclusions, discussion and perspectives for future work.....	188
5.	<i>IN VIVO</i> CHARACTERIZATION OF THE ROLES OF CHD4 AND CDK13 IN CARDIAC DEVELOPMENT	196
5.1.	Introduction	196
5.1.1.	Cyclin-dependent kinase 13 (CDK13).....	196
5.1.2.	High-resolution episcopic microscopy	197
5.1.3.	Hypothesis and aims of this project	198
5.2.	Results.....	199
5.2.1.	<i>Chd4^{tm1b}</i> and <i>Cdk13^{tm1b}</i> genotyping	199
5.2.2.	Lethality assessment.....	200
5.2.3.	CHD4 and CDK13 expression during development	206
5.2.4.	HREM	208
5.3.	Conclusions, discussion and perspectives for future work.....	210
6.	SUMMARY OF FINDINGS AND IMPLICATIONS OF THIS RESEARCH	213
7.	REFERENCES.....	218

INDEX OF FIGURES

CHAPTER 1

Figure 1.1. Overview of cardiac valves development.....	20
Figure 1.2. Septation of the cardiac chambers.....	22
Figure 1.3. Timeline of cardiac development in mouse.	23
Figure 1.4. Anatomy of the adult heart and congenital heart defects.....	28
Figure 1.5. Secondary structure of TBX5.	30
Figure 1.6. Transcriptional networks involved in mammalian heart development. ...	37
Figure 1.7. Genome-editing technologies.	41
Figure 1.8. Stages of zebrafish cardiac development.	47

CHAPTER 2

Figure 2.1. Drawing showing the structure of a western blot “sandwich” prior to transfer.	59
Figure 2.2. Cloning-free method for the generation of sgRNA templates.	69
Figure 2.3. Drawing of a 48hpf zebrafish embryos indicating transection points.....	71
Figure 2.4. Representation of <i>kat2a</i> , <i>kat2b</i> and <i>tbx5a</i> morpholinos-binding site in their respective pre-mRNAs.	72
Figure 2.5. Setup used for the injection of 1-cell stage zebrafish embryos	73
Figure 2.6. Microinjection rig used for injection of 1-cell stage zebrafish embryos. ..	73
Figure 2.7. Representation on a microinjection into a 1-cell stage zebrafish embryo.	74
Figure 2.8. Diagram of the T7 endonuclease assay.	75
Figure 2.9. Graphical representation of how cardiac looping angle is measured.....	80
Figure 2.10. Structure of the tm1a and tm1b alleles.....	94
Figure 2.11. Image showing a properly embedded embryonic heart sample prior to imaging.	99

CHAPTER 3

Figure 3.1. Secondary structures of the KAT2A and KAT2B proteins.	104
Figure 3.2. TBX5 associates with, and is acetylated by KAT2B and KAT2A.....	108
Figure 3.3. KAT2A and KAT2B potentiate TBX5 transcriptional activity.	109
Figure 3.4. Prediction of the acetyl lysine residue in TBX5.	110
Figure 3.5. Finding the KAT2A/KAT2B acetylation motif.	112
Figure 3.6. Acetylation of TBX5 takes place at lysine 339.(A).....	114
Figure 3.7. Acetylation of TBX5 is required for its nuclear retention.....	117
Figure 3.8. CRM1 mediates nuclear export of non-acetylated TBX5.	118
Figure 3.9. Cell viability assay on COS7 cells treated with Garcinol.	119
Figure 3.10. Chemical inhibition of KAT2B impairs TBX5 cellular localization.	121
Figure 3.11. <i>kat2a</i> and <i>kat2b</i> genes in zebrafish.....	124

Figure 3.12. Titration of <i>kat2a</i> and <i>kat2b</i> translation blocking and splice blocking MOs.....	125
Figure 3.13. The effect of morpholino-mediated knockdown of <i>tbx5a</i> , <i>kat2a</i> or <i>kat2b</i> in 48hpf embryos.....	128
Figure 3.14. sgRNAs targeting the <i>kat2a</i> , <i>kat2b</i> and <i>tbx5a</i> genes.	130
Figure 3.15. sgRNA activity <i>in vitro</i> assays.....	131
Figure 3.16. Initial CRISPR/Cas9 system injections using Cas9 mRNA.....	132
Figure 3.17. CRISPR-Cas9-mediated knockouts of <i>kat2a</i> , <i>kat2b</i> and <i>tbx5a</i> in zebrafish present a similar phenotype than their knockdown counterparts.	136
Figure 3.18. Analysis of CRISPR/Cas9-injected embryos for the presence of mutations.	138
Figure 3.19. Fin development is also altered in <i>kat2a</i> and <i>kat2b</i> KO/KD embryos...	140
Figure 3.20. <i>In situ</i> hybridization in 48 hpf embryos using a <i>bmp4</i> probe.	141
Figure 3.21. Garcinol titration in zebrafish embryos.....	142
Figure 3.22. Chemical inhibition of acetyltransferase activity with Garcinol also impairs cardiac and fin development in zebrafish.....	144
Figure 3.23. Expression levels of <i>tbx5a</i> target genes in <i>kat2a</i> and <i>kat2b</i> KO embryos.	146
Figure 3.24. The expression of <i>fgf10</i> , <i>nppa</i> and <i>bmp4</i> via <i>in situ</i> hybridization.....	147
Figure 3.25. Rescue of the <i>kat2a</i> KO phenotype using an acetyl-mimic version of <i>tbx5a</i>	150
Figure 3.26. TBX5 activity regulation model.....	160

CHAPTER 4

Figure 4.1. Subunits of the NuRD complex.....	161
Figure 4.2. Secondary structure of the CHD4 protein.	162
Figure 4.3. Individuals with <i>de novo</i> CHD4 mutations.....	166
Figure 4.4. CHD4 mutations found in syndromic-CHD patients.	169
Figure 4.5. Verification of the generation of all five CHD4 mutant variants by Sanger sequencing.....	170
Figure 4.6. Expression of CHD4-FLAG variants in COS7 cells.....	171
Figure 4.7. Additional translation start sequences found in the human CHD4 cDNA sequence.....	173
Figure 4.8. Interaction between CHD4 variants and different NuRD complex subunits.	176
Figure 4.9. Interaction between CHD4 variants and TBX5.	177
Figure 4.10. Cellular localization of CHD4 mutant variants.	179
Figure 4.11. ATPase activity of CHD4 variants produced in COS7 cells.	179
Figure 4.12. Transfection of Sf9 insect cells with CHD4 WT, M1 or M2 bacmids.....	181
Figure 4.13. Large-scale expression and purification of CHD4 WT, M1 and M2 in Sf9 cells.	182
Figure 4.14. ATPase activity of CHD4 variants produced in Sf9 cells.....	183
Figure 4.15. The ability of CHD4 variants to bind DNA and nucleosomes.....	184
Figure 4.16. The ability of CHD4 variants to remodel nucleosomes.	185

Figure 4.17. Production and purification of CHD4-His variants in <i>E.coli</i>	187
Figure 4.18. ATPase activity of CHD4 variants produced in <i>E.coli</i>	188

CHAPTER 5

Figure 5.1. <i>Chd4</i> ^{tm1b} and <i>Cdk13</i> ^{tm1b} lines genotyping.	200
Figure 5.2. <i>Chd4</i> ^{tm1b} and <i>Cdk13</i> ^{tm1b} lethality assessment.	201
Figure 5.3. Preliminary phenotypic analysis of analysed <i>Chd4</i> ^{tm1b} embryos.	202
Figure 5.4. Gross morphological appearance of <i>Chd4</i> ^{+/+} (left panels) and representative <i>Chd4</i> ^{tm1b/+} (right panels) embryos and hearts at E11.5 and E13.5. ...	203
Figure 5.5. Preliminary phenotypic analysis of analysed <i>Cdk13</i> ^{tm1b} embryos.	204
Figure 5.6. Gross morphological appearance of <i>Cdk13</i> ^{+/+} (left panels) and representative <i>Cdk13</i> ^{tm1b/+} (right panels) embryos and hearts at E12.5 and E13.5. .	205

CHAPTER 6

Figure 6.1. Interaction maps for TBX5, KAT2A, KAT2B, CHD4, CDK13, PRKD1 and several cardiac transcriptor factors.	215
--	-----

INDEX OF TABLES

CHAPTER 2

Table 2.1. PCR reaction volumes for Taq, Phusion and Platinum polymerases	51
Table 2.2. PCR cycling conditions for Taq, Phusion and Platinum polymerases	52
Table 2.3. Restriction digestion volumes.....	53
Table 2.4. Ligation reaction volumes	54
Table 2.5. List of primary and secondary antibodies used in this thesis	60
Table 2.6. Stock and working concentrations for different antibiotics	65
Table 2.7. sgRNA template generation reaction	68
Table 2.8. Components of the sgRNA transcription reaction.....	68
Table 2.9. Primers used for reverse transcription PCR.....	70
Table 2.10. Morpholinos used for kat2a and kat2b knockdowns in zebrafish.....	71
Table 2.11. <i>In situ</i> hybridization probes used in this thesis and their respective primer sequences	76
Table 2.12. Components for the generation of DIG-labelled RNA probes	77
Table 2.13. Proteinase K treatment duration according to zebrafish developmental stage.....	78
Table 2.14. Primers used for relative quantitative PCR.....	81
Table 2.15. Primers used for the generation of CHD4 mutant variants	83
Table 2.16. PCR reaction for site-directed mutagenesis	83
Table 2.17. Cycling conditions used for site-directed mutagenesis	83
Table 2.18. Components for the ligation of CHD4 cDNA and the TOPO vector	88
Table 2.19. Components used for the nucleosome assembly reaction	91

Table 2.20. Preparation of the ATP+ADP standards.....	93
Table 2.21. Primers used for the genotyping of the CHD4 and CDK13 lines.....	95
Table 2.22. PCR cycling conditions for the genotyping of Chd4 ^{tm1b} and Cdk13 ^{tm1b} lines using Platinum polymerase	95
Table 2.23. JB-4 dye mix volumes.....	97

ABSTRACT

Congenital heart defects represent the most common form of congenital abnormality. These developmental disorders affect at least 4000 new-borns each year in the UK, and present an incidence of approximately one in every 100 babies born. The most common cause for CHDs are genetic mutations that are inherited or arise *de novo* during embryogenesis, and even though we have gained great insight into the CHD-causative genes during these last decades, we are still far from understanding the whole picture. In this PhD thesis, my main goal is to contribute to the genetics of congenital heart disorders by discovering and characterizing new factors involved in cardiac development and in the onset of CHD. During this thesis, I have worked on three different projects whose results have ultimately contributed to this field.

In the first part of this thesis, I report that acetyltransferases KAT2A and KAT2B associate with TBX5, a T-box transcription factor whose mutations are causative of Holt-Oram syndrome, a rare genetic condition that affects the development of the heart and upper limbs. KAT2A and KAT2B acetylate TBX5 at Lys339, and this acetylation potentiates its transcriptional activity and is required for TBX5 nuclear retention. Both morpholino-mediated knockdown and CRISPR-Cas-mediated knockout of *kat2a* and *kat2b* in zebrafish perturbs heart and limb development, mirroring thus the *tbx5a* loss-of-function phenotype.

In the second part, I performed a characterization of CHD4 mutations found in syndromic-CHD patients. Interactional studies revealed that none of these mutations affect CHD4 interaction with other NuRD subunits, and immunofluorescence staining showed that they localized properly to the nucleus. However, functional assays revealed that four of these mutants have a lower activity than their WT counterpart, whereas one of them appears to be a gain-of-function variant.

Last but not least, in the final chapter of this thesis I carried out the preliminary characterization of two transgenic mouse lines carrying null mutations for two genes which have been recently discovered to be involved in CHD: CHD4 and CDK13. These two lines were recently acquired by our laboratory, and following colony expansion and characterization, they were used for heart morphology studies using high-resolution episcopic microscopy.

All in all, the results presented in this thesis expand our understanding and knowledge of the genes and factors involved in the regulation of cardiac

development as well as in the onset of CHDs, and help defining the complete set of genes responsible for CHDs.

LIST OF ABBREVIATIONS

AA	Amino acids
ADP	Adenosine diphosphate
ALPM	Anterior lateral plate mesoderm
ASD	Atrial septal defect
ATP	Adenosine triphosphate
AV	Atrioventricular
AVC	Atrioventricular canal
AVSD	Atrioventricular septal defect
BCP	Bromochloropropane
BIIC	Baculovirus-infected insect cells
BPM	Beats per minute
BSA	Bovine serum albumin
CHD	Congenital heart defects
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
DAPI	4',6-diamidino-2-phenylindole
DMEM	Dulbecco's modified eagle medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide Triphosphate
DSB	Double-strand break
EDTA	Ethylenediaminetetraacetic acid
EMT	Endothelial-mesenchymal transition
HDR	Homology directed repair
HPf	Hours post fertilization
HREM	High-resolution episcopic microscopy

HOS	Holt-Oram Syndrome
HRP	Horseradish peroxidase
FHF	First heart field
HOS	Holt-Oram syndrome
IB	Immunoblot
IgG	Immunoglobulin G
IP	Immunoprecipitation
IPTG	Isopropyl β -d-1-thiogalactopyranoside
ISH	<i>In situ</i> hybridization
KD	Knockdown
KO	Knockout
LB	Luria-Bertani
LSCV	Left superior caval vein
MO	Morpholino
MRI	Magnetic resonance imaging
NCCs	Neural crest cells
NES	Nuclear export sequence
NHEJ	Non-homologous end joining
NLS	Nuclear localization sequence
OFT	Outflow tract
OPT	Optical projection tomography
PAGE	Polyacrylamide agarose gel electrophoresis
PAM	Protospacer adjacent motif
PBS	Phosphate buffered saline
PBS-T	PBS + 0.1% Tween-20
PCR	Polymerase chain reaction
PFA	Paraformaldehyde

PSC	Pluripotent stem cell
PTM	Post-translational modification
PTU	1-Phenyl-2-thiourea
P/S	Penicillin/streptomycin
qPCR	Quantitative PCR
RNA	Ribonucleic acid
RIN	RNA integrity number
RSCV	Right superior caval vein
RT	Room temperature
RT-PCR	Reverse transcription polymerase chain reaction
RVD	Repeat Variable Diresidue
SDS	Sodium dodecyl sulphate
sgRNA	Single guide ribonucleic acid
SHF	Secondary heart field
SSC	Saline sodium citrate
TALEN	Transcription activator-like effector nuclease
TBS	Tris-buffered saline
TBS-T	Tris-buffered saline with 1% Tween 20
TOF	Tetralogy of Fallot
TSA	Trichostatin A
VSD	Ventricular septal defect
WT	Wild-type
X-gal	X-Galactosidase

1. INTRODUCTION

The heart is the first major organ to form and function during vertebrate embryonic development¹. Forming a heart is not an easy task: many different and crucial morphological processes take place in parallel, and a fine and precise tuning of each of them is required in order to form a fully functioning cardiovascular machinery. A massive network of genes encoding many different proteins (e.g. transcription factors, signalling proteins and structural proteins, among many others) is involved in the regulation of each one of these processes^{2,3}, and perturbations in any of them can potentially lead to cardiac malformations and the onset of congenital heart disorders (CHDs)⁴. Thus, understanding the morphological and molecular mechanisms behind normal cardiogenesis is crucial in order to elucidate the pathogenesis of congenital heart disorders.

1.1. Cardiogenesis: a quick overview

The heart is comprised of 3 different layers: an inner endothelial layer (endocardium), a middle muscular layer (myocardium) and an outer connective layer (epicardium)⁵. Myocardial and endocardial cells have their origin in bilateral populations of mesodermal cells close to the midbrain-hindbrain boundary⁶. In the early heart, these two populations of cells expressing cardiac markers come together at the midline of the embryos to give rise to a linear heart tube, which shortly after begins to beat^{6,7}. Then, this linear tube expands and undergoes a series of morphological changes and contortions. This cardiac looping shifts the heart towards the right side of the embryo, the first sign of left-right asymmetry⁸. Once the loop is in place, major morphogenetic events take place to form the chambers of the heart: two atria and two ventricles. Following this, each pair of chambers undergo septation to isolate each chamber, and shortly after valves are developed, connecting the atria to the ventricles⁹. Final stages of heart maturation take place during the later stages of fetal life, until a fully functional heart is present at birth^{2,6}. The development of the heart has been extensively studied in several organisms, including human (Carnegie stages, CS)¹⁰, mouse (Embryonic days, E)¹¹ and chick (Hamburger and Hamilton stages, HH)^{1,12}. In this chapter, most of the information provided is based on human, chick and mouse studies. Zebrafish heart development will be discussed at a later point.

1.1.1. Gastrulation

Cardiac precursor cells can be first identified in the posterolateral region of the epiblast cell layer at the onset of the primitive streak formation¹³, a structure found in the blastula that will establish bilateral symmetry and determine the

site of gastrulation and initiate germ layer formation¹⁴. Gastrulation is one of the first morphogenic process that takes place during embryonic development, and it involves the migration of cells forming the blastula in order to establish the three embryonic germ layers: mesoderm, endoderm and ectoderm, each of which will give rise to specific tissues and organs in the developing embryo¹⁴. Migration of cells from the upper layer of the primitive streak will form the ectoderm, the outermost layer of the gastrula. The inward migration of cells from the inner layer of the primitive streak will give rise to the endoderm, whereas the mesoderm, which is situated between the ectoderm and endoderm, have its origin in some of the cells migrating to form the endoderm^{14,15}. While this process is taking place, the posterolateral epiblast, where the cardiac precursor cells are situated, interacts with the hypoblast in order to form the anterior lateral plate mesoderm, a structure that will eventually give rise to the heart¹⁶.

1.1.2. Cardiac crescent formation

The mammalian heart is derived from two cardiac progenitor cell populations: the first and secondary heart field¹⁷. The myocardial cells that constitute these fields are derived from the mesoderm, which emerges from the anterior region of the primitive streak during the gastrulation process¹⁸. Two opposite signals control the specification and differentiation of the cardiac progenitor cells; bone morphogenic protein (BMP) produced by the endoderm induces the cardiomyocyte fate whereas WNT-mediated signals released by the notochord suppress cardiomyocyte specification¹⁹. Cardiac progenitor cells migrate towards the anterior-lateral region of the embryo to give rise to two populations of cells on both the left and right side of the midline where cardiac markers are first detected^{17,20}. The first wave of cells to migrate (primary heart field) converge at the midline, where they form the cardiac crescent. The cells constituting the secondary heart field migrate later in a second wave, residing medial and anterior to the primary heart field^{17,20}. Then, embryonic folding promotes the fusion of the cardiac crescent at the midline, forming a primitive linear heart tube composed of cardiomyocytes and endothelial cells^{17,20}. As the heart develops, both heart fields are going to contribute to the formation of different cardiac structures; the primary heart field will give rise to the left ventricle as well as some parts of the atria, whilst the secondary heart field will give rise to the right ventricle, most parts of the left and right atria, and to the outflow tract^{21,22}.

1.1.3. Linear heart tube looping and chamber formation

Following cardiac crescent fusion at the midline, a linear heart tube is formed, which serves as a scaffold for subsequent heart growth. Cells migrating from

the secondary heart field contribute to the expansion of the heart tube, joining onto pre-specified ends and giving rise to the venous and arterial poles²³. At this stage, the developing heart is bilaterally symmetrical and it adopts an inverted-Y shape. In addition, thanks to several marking studies we now know that the heart tube is already patterned at this stage; the arms of the inverted-Y-shaped heart tube (venous pole) will eventually become the precursors of the atrial chambers and the atrioventricular canal (AVC)²³, whereas the stem of the inverted-Y is fated to become the left ventricle²³. Finally, the arterial pole contributes to the formation of the outflow tract and the right ventricle^{23,24}.

After expansion, the linear heart tube begins to bend towards the right side, marking the beginning of a process known as cardiac looping, the first visual evidence of loss of left-right symmetry in the developing embryo²⁵. The signalling pathway that controls and ensures that the loop is started and bent to the right is established earlier, during the gastrulation process²³. Although we know relatively little about it, it is thought that it involves a leftward flow of secreted proteins across the node²⁶, controlled by factors such as LEFTY, NODAL and PITX2²¹.

As the heart bends and starts to form a looped structure, different chambers begin to appear. First, the left ventricle begins to take its large, spherical structure, being exclusively derived from the primary heart field^{17,22}. Then, a right ventricle precursor also appears, derived from the secondary heart field^{22,27}. With looping progressing, the right ventricle becomes more obvious, and both chambers are rearranged: although the linear continuity of both ventricles is maintained, the looping brings them side by side²³. Ventricles continue growing downwards as looping proceeds. This leads to a rearrangement of the atria, which are positioned in the posterior part of the developing heart in a side by side configuration, and they also start to grow out²³. The atria are primarily constituted by cells derived from the secondary heart field, although there is also some contribution from primary heart field cells²². The stage is now set for formation of the definitive cardiac chambers.

1.1.4. Endocardial cushions formation

Once the looping process is completed and the cardiac chambers have acquired their characteristic morphology, two crucial cardiogenic processes will begin taking place: chamber septation and valves formation, both of which will contribute to form a 4-chambered heart in which the blood flows unidirectionally. Part of the septa that will divide the primitive ventricle and atrium into left and right chambers, as well as the valves that control blood flow, have their origin in a subset of cells which form structures known as endocardial cushions^{28,29}. Endocardial cushions arise from a subpopulation of

endothelial cells which delaminate and undergo epithelial-to-mesenchymal (EMT) transition after receiving signals from the myocardium^{29,30}. Following delamination, these endocardial cells migrate and invade the cardiac jelly, an extensive extracellular, gelatinous matrix which separates the outer layer of myocardium and the inner layer of endocardial cells in the developing heart^{29,30}. After invasion of the cardiac jelly, these cells will fully differentiate into mesenchymal cells that will produce and deposit collagens and glycoproteins in the cardiac jelly, thus promoting a “swelling” in the area. Continued migration of cells from the endocardium and swelling will form primitive valve-like structures that, through an extensive remodelling process, will eventually give rise to the cardiac valves and part of the atrioventricular septum^{29,30}. Many different genes and pathways are involved in the proliferation and EMT of endothelial cushion cells: myocardial BMP2 expression increases the deposition of hyaluronan and versican in cushion-forming regions, and it also promotes EMT of endothelial cushion cells^{31,32}. VEGF also promotes the proliferation of endothelial cells³³, whereas genes like TGFβ³², Notch1³⁴ and the Wnt/β-catenin³⁵ are all involved in the EMT process.

Cardiac valves

Cardiac valves are thin membranes attached to the heart walls which constantly open and close in order to regulate the unidirectional blood flow through the heart. The mammalian heart has four cardiac valves: the aortic and pulmonary valves (semilunar valves), which are located in the outflow tract and control blood flow out of the ventricles, the mitral and tricuspid valves (atrioventricular valves), which are situated in the atrioventricular canals and regulate blood flow to the ventricles.

As mentioned earlier, cardiac valves have their origin in endocardial cushions that form by the delamination, EMT and invasion of the cardiac jelly by endothelial cells. Once the valve primordia are formed, they undergo an extensive process of growth and remodelling which is regulated by many of the genes previously involved in the proliferation and EMT process, such as VEGF, TGFβ and the Wnt/β-catenin pathway³⁶. This remodelling process comprises an increase in the extracellular matrix organization and complexity, as well as a decrease in the proliferation of endocardial cushion cells. At the end of this process, both atrioventricular and semilunar valves are fully formed. The former are entirely derived from endocardial cushion tissue, whereas it is thought that the latter require the contribution of a population of neural crest cells (NCCs) derived from the branchial arches³⁷. An overview of cardiac valves development is represented in Fig. 1.1.

Even though both atrioventricular (AV) and semilunar valves derive from endocardial cushions, their formation differs in certain points. AV valves have their origin in the inferior atrioventricular cushion³⁸ (that will give rise to the anterior mitral leaflet and the septal tricuspid leaflet) and in two bilateral populations of mesenchymal cells situated in the lateral regions of each atrioventricular canal (giving rise to the posterior mitral leaflet and the anterior and posterior tricuspid leaflets)^{9,39}. Following endocardial cushion formation, these structures undergo a process of proliferation and remodelling that will give rise to the leaflets of the AV valves. In parallel, ventricular wall remodelling and trabeculation will lead to the formation of papillary muscles, which attach to the cusps of the AV valves through fibrous cords of connective tissue known as chordae tendineae^{39,40}. In the case of the semilunar valves, these are formed from four endocardial cushions situated in the outflow tract (OFT): two conotruncal (lateral and proximal) cushions and two intercalated (anterior/posterior and distal) cushions⁹. The conotruncal cushions will eventually fuse in order to septate the OFT, thus giving rise to a spiral septum which will divide the OFT into aorta and pulmonary trunk⁴¹ (in the truncal part) and into left and right ventricular outlets⁴² (in the conal part). This process will also form the aortic and pulmonary valves at the conotruncal junction. An important population of cells involved in this process are the cardiac neural crest cells, that migrate from the neural crest to the developing cushions of the OFT and subsequently contribute to semilunar valves formation and maturation^{41,43}.

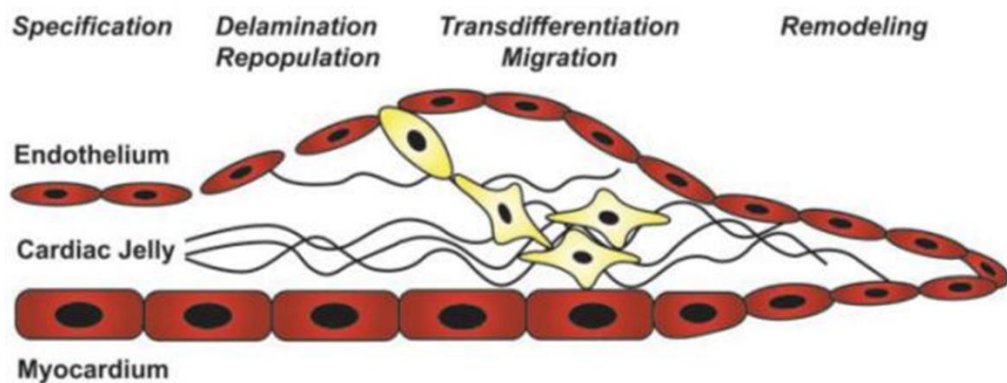


Figure 1.1. Overview of cardiac valves development. Cardiac valves formation is restricted to the atrioventricular canal and outflow tract regions of the looping heart. Initially, a subset of endothelial cells delaminates, differentiate and migrate into the cardiac jelly, a process called endothelial-mesenchymal transition (EMT). This process promotes an expansion and swelling of this extracellular matrix, leading to the formation of cardiac cushions. After this, these cushions undergo an extensive process of growing remodelling to finally give rise to the heart valves. *Adapted from Armstrong et al., 2004*²⁹

1.1.5. Cardiac chambers and outflow tract septation

After each chamber is positioned correctly, a final process is required to form a fully-functional 4-chambered heart; the septation (or separation) process, which takes place in the atria, ventricles and outflow tract (Fig. 1.2).

Around 34 days of human development, the convergent growth of two structures called “septa” separate the atria in two spatially independent chambers. The septum primum protrudes in a crescent shape from the ventral and posterior walls of the atrium, whilst the septum secundum forms a ridge on the dorsal and posterior walls of the atria^{28,44}. Then, by day 50 of development, the septum primum fuses with the endocardial cushions, and the septum secundum partly overlaps the septum primum^{28,44}. This leads to the formation of a septum in the middle of the atria, which is now separated in two chambers: the left atrium and the right atrium^{28,44}.

Septation of the ventricle into two independent chambers takes place by the upward growth of the cells that join the left and right primordial chambers. These two different populations of cells grow upwards without mixing towards the endocardial cushion, creating a septum^{28,44}. This septum will stop growing before the chambers are completely separated, and will remain so until later in development a new structure, the membranous septum, connects the septum with the endocardial cushion^{28,44}.

At day 40 of development, the outflow tract starts to rotate at its base, so that its original left side is now anterior and the right side posterior^{28,44}. At the same time, the outflow tract endocardium undergoes epithelial-mesenchymal transformation (EMT), forming internal outflow tract cushions, which converge towards each other. While the base of the outflow tract rotates clockwise, the length of the tract is reduced, and the internal outflow tract cushions rotate and converge at the midline^{28,44}. Once these processes are completed, the aorta and pulmonary outflow tracts are finally separated.

After completion of these processes and the heart valves are formed, the heart will achieve its final morphology. A timeline covering the main period of cardiac development (from E6.5 to E14.5 in mouse) is shown in Fig 1.3.

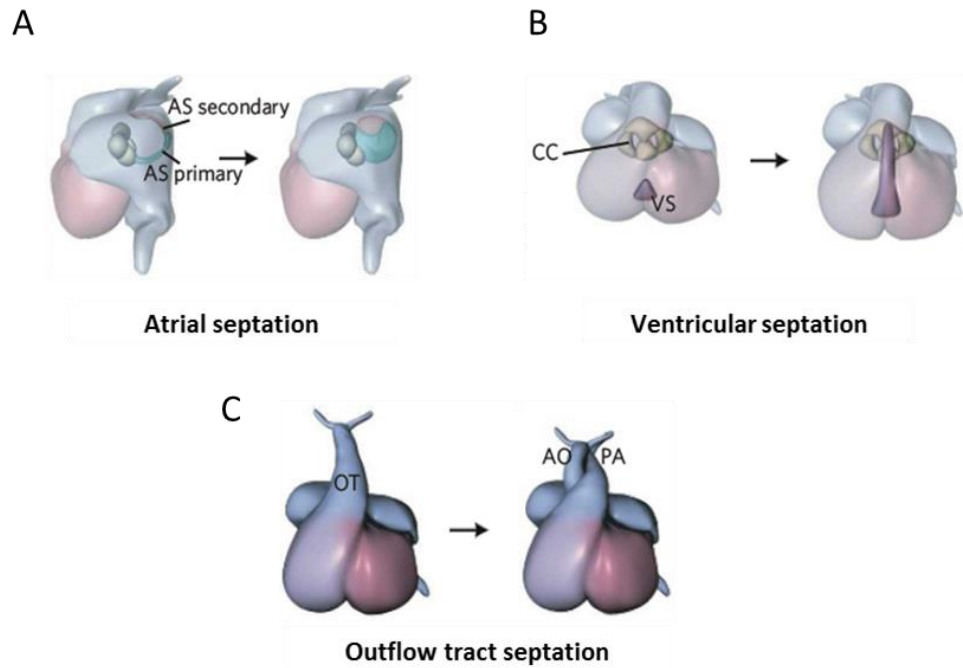


Figure 1.2. Septation of the cardiac chambers. Three different septation events take place at the latest stages of cardiogenesis. A) Atrial septation (AS) occurs by the growth of two septa: the primary septum (green) and the secondary septum (pink). B) Ventricular septum (VS) arises from the left and right ventricles, growing upwards until it fuses with the endocardial cushions (CC). C) Outflow tract septation divides the outflow tract (OT) into the aorta (AO) and the pulmonary artery (PA). *Figure from Bruneau, 2008⁷*

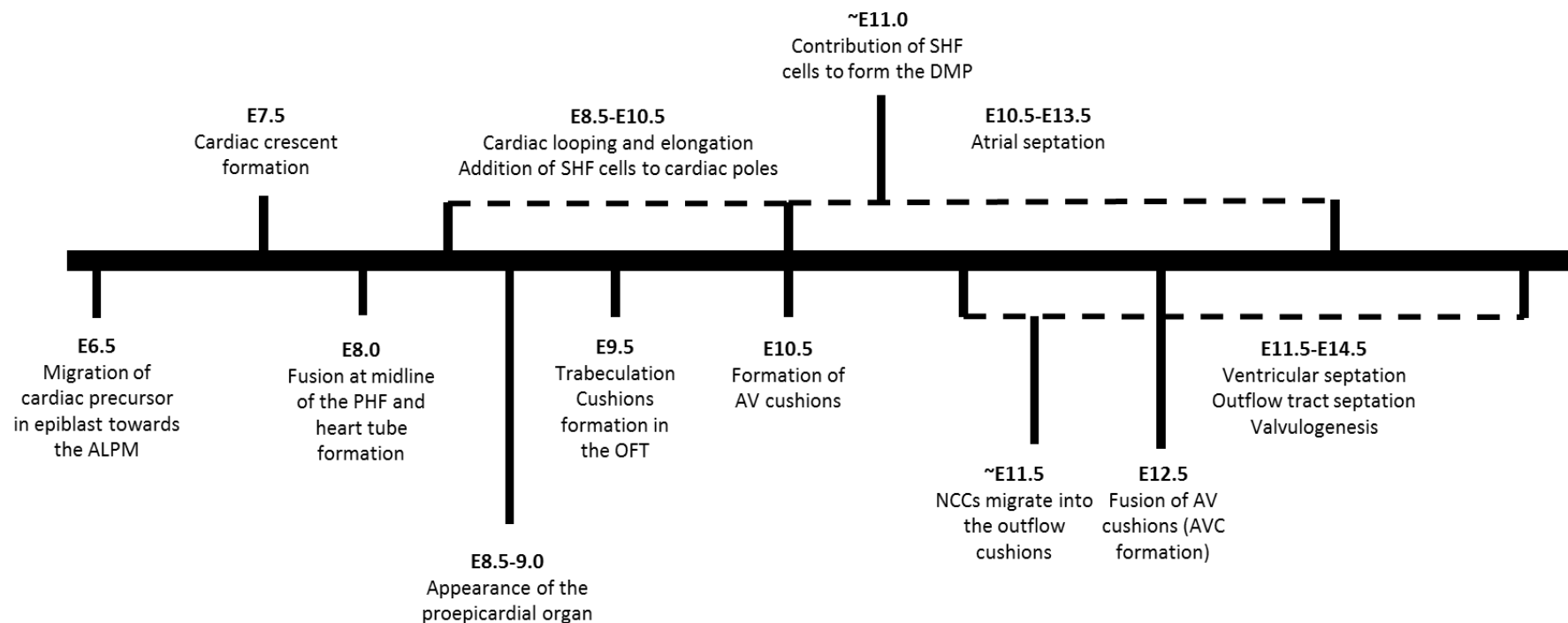


Figure 1.3. Timeline of cardiac development in mouse. Key development events and timing of addition of different progenitor populations (e.g. neural crest cells) are shown. ALPM, anterior lateral plate mesoderm; PHF, primary heart field; SHF, secondary heart field; OFT, outflow tract; AV, atrioventricular; DMP, dorsal mesocardial protrusion; NCCs, neural crest cells; AVC, atrioventricular canal. *Adapted from England, 2018³⁸; Chaudhry, 2014³, Bruneau, 2008⁷*

1.2. Congenital heart disease

Congenital heart disease (CHD), also known as congenital heart defects, is the most common congenital abnormality, and it usually refers to abnormalities in the heart's function and structure present at childbirth⁷. They occur often, affecting approximately 10 out of every 1000 live births⁴⁵, and the incidence is even higher if fetus that do not survive to term are taken into account⁴⁶. These cardiac abnormalities are present in very different forms: from small holes in the walls of the cardiac chambers, to more complex defects which can eventually lead to an early death⁴⁷.

1.2.1. Categories

Congenital heart defects can affect most parts of the developing heart, and they can be classified into 3 broad categories: cyanotic heart disease, left-sided obstructive lesions and septation defects. A representative figure of heart structures affected by CHD can be found in Fig. 1.4.

1.2.1.1. Cyanotic heart disease

Cyanotic heart disease is a condition that arises as a result of the mixing of oxygenated and deoxygenated blood, leading to a bluish hue to the skin of affected individuals⁴⁸. This type of disease can be caused by issues in the heart valves, a coarctation or complete interruption of the aorta or abnormalities in any of the large blood vessels. There are several heart defects that contribute to this condition. Tetralogy of Fallot (TOF) is the most common cause of cyanotic heart disease, and accounts for up to one-tenth of all congenital cardiac lesions⁴⁹. TOF is a condition characterized by the presence of four different heart defects: pulmonary stenosis (narrowing of the exit from the right ventricle), ventricular septal defect (presence of a hole between the ventricles), right ventricular hypertrophy (thickening of the right ventricular muscle) and overriding aorta (blood from both ventricles enter the aorta)⁴⁹. TOF is thought to be caused by a combination of environmental and genetic factors, although its precise causes are still unclear⁴⁹. However, thanks to the generation of animal models and knockout studies, there is growing evidence supporting the fact that the primary defects leading to TOF malformations are related to the alignment, elongation and septation of the outflow tract, processes in which SHF progenitor cells and NCCs play a key role⁵⁰. Many transcriptional regulators have been shown to be crucial for the proliferation and differentiation of SHF progenitor cells. One of them is *Tbx1*, which coordinates the addition of posterior second heart field progenitor cells to the arterial and venous poles of the heart⁵¹. Interestingly, the most common genetic cause leading to TOF in humans is haploinsufficiency for *Tbx1* (which accounts for 15% of TOF cases)⁵². *Tbx1* is also the major candidate gene for

DiGeorge syndrome, a condition characterized by the presence of TOF among other defects⁵³. In mouse, homozygous *Tbx1* mutations lead to common trunk with failure of subpulmonary myocardial addition⁵⁴, whereas mice carrying a hypomorphic *Tbx1* allele present different OFT alignment defects such as overriding aorta⁵⁴. *Nkx2-5* is another transcriptional regulator involved in SHF development, playing a crucial role in the elongation of the primitive heart tube⁵⁵. In a similar fashion to *Tbx1*, hypomorphic murine embryos for *Nkx2-5* display alignment defects such as overriding aorta as well as a shortened OFT⁵⁶. In addition to this, mutations in the *Nkx2-5* gene have been found in TOF patients⁵⁷. Intracellular signalling pathways such as the FGF pathway also play a key role in SHF development⁵⁸, and mouse embryos lacking components of this pathway such as *Fgfr1*⁵⁹, *Fgfr2*⁵⁹ and *Fgf15*⁶⁰ also present OFT shortening and alignment defects such as overriding aorta. Studies on different components of the Notch signalling pathway have also confirmed their role in SHF development regulation and potential implication in the onset of TOF: *Hey2* mutant mice display a range of cardiac malformations including TOF and ventricular septal defects⁶¹, whereas the *Jag1-Notch2* compound heterozygous mice present overriding aorta with pulmonary stenosis⁶². Last but not least, mutations in genes controlling cardiac NCCs development in mouse, such as *Pax3*⁶³, also lead to OFT septation and alignment defects.

Other conditions leading to cyanotic heart disease include the transposition of the great arteries (a condition in which the pulmonary artery and aorta have “swapped” places), pulmonary or tricuspid atresia (characterized by the abnormal development or absence of the pulmonary or tricuspid valves respectively) and persistent truncus arteriosus (the truncus arteriosus fails to properly divide into the pulmonary trunk and aorta) among others⁷.

1.2.1.2. Septal defects

The second category of CHDs is constituted by septal defects, which are typically characterized by a hole or opening in the walls that divide the right and left cardiac chambers. Septation defects constitute the second most common type of CHD⁶⁴. These defects can affect the septation of the atria (atrial septal defects, ASDs), septation of the ventricles (ventricular septal defects, VSDs) or formation of structures in the central part of the heart (atrioventricular septal defects, AVSDs).

ASDs

Atrial septal defects can be divided in ostium primum ASDs (if the defect affects the formation of the ostium primum) and secundum ASDs (if the defect affects the formation of the ostium secundum)⁶⁵. The latter are the most common type

of ASDs⁶⁶, and thanks to genetic studies and the generation of animal models, we now know that ASDs arise from genetic mutations in a large number of genes involved in atrial partitioning and septum secundum growth and regulation, including many of the component of the core regulatory network of cardiac development as well as sarcomeric proteins^{65,66}. In the first case, mouse models carrying null mutations in key regulators of cardiogenesis such as *Tbx5*⁶⁷ display ASD among a wide range of other cardiac defects, such as VSD or TOF. On top of this, mutations in the *Gata4*⁶⁸ and *Nkx2-5*⁶⁹ genes have also been found in patients with ASD among other cardiac defects. This pleiotropy exists probably because, as stated before, these genes act at the top of the cardiac regulatory network⁷⁰ and are expressed in different cardiac tissues, thus controlling the development of multiple cardiac structures. Studies using other animal models have also confirmed that *Tbx5* and *Gata4* are involved in an endocardial pathway that controls atrial septum formation⁷¹, and *Nkx2-5* regulates the proliferation of atrial myocytes⁷². Interestingly, certain sarcomeric genes such as *MYH6*⁷³ and *ACTC1*⁷⁴ are also associated with the onset of ASD in humans. In these patients, ASD is the only cardiac defect that can be found. *MYH6* is a cardiac-specific sarcomeric gene highly expressed in the developing atrial septum whose expression is regulated by *TBX5*⁷³, and morpholino-mediated knockdown of this gene in chick embryos lead to ASDs⁷³. A mouse model carrying a specific *ACTC1* mutation causes atrial-septal defects associated with late-onset dilated cardiomyopathy⁷⁵.

VSDs

Ventricular septation is an extremely complex process which involves a large number of genes, including many of the core regulatory factors which regulate many cardiogenic processes (such as the previously mentioned atrial septation)⁷⁶. The ventricular septum is composed of different parts with distinct origins⁷⁶, thus increasing the complexity of its study. However, thanks to *in vitro* and *in vivo* studies, we now know that in order to regulate ventricular septation and orchestrate the contribution from each one of these parts, these genes act on different populations of cells, such as myocardial cells derived from both the primary and secondary heart field, as well as the mesenchymal cells that constitute the endocardial cushions⁷⁶. The literature covering ventricular septation and VSDs formation is really extensive, so just a few selected examples of animal models that have helped us understand how this process takes place will be presented here. As mentioned previously, many of the transcription factors that orchestrate cardiogenesis such as *Tbx5*^{77,67}, *Gata4*^{78,79} and *Nkx2-5*^{80,81} (among others) play a key role in ventricular septation, as demonstrated by the fact that when absent in mice or in humans, the lack of these genes lead to VSDs but also to other septal defects such as

ASDs or atrioventricular septal defects (AVSDs). This could suggest that these genes act in both parts of the SHF: anterior and posterior⁷⁶. Apart from these core regulators of cardiogenesis, mutations in other genes also lead to different types of VSDs. For example, an experimental study using two different Mlc (myosin light chain) transgenic mouse lines showed that cells contributing to the trabeculated ventricular septum have their origin in both primary and secondary heart fields⁸². Also, mutations in genes that are expressed in neural crest cells such as *Hand1* lead to failure of interventricular septum formation among other defects⁸³.

1.2.1.3. Left-sided obstructive lesions

Left-sided obstructive lesions are the third main type of CHD, and they include defects that affect the left side of the heart and the aorta. This category encompasses conditions such as hypoplastic left heart syndrome, in which the left side of the heart is severely underdeveloped⁸⁴; aortic and mitral stenosis, characterized by the narrowing of the aortic⁸⁵ or mitral valves⁸⁶ respectively; and aortic coarctation.

1.2.1.4. Other congenital heart defects

In addition to the conditions comprehended in these three categories, there are other congenital heart diseases which do not properly fit in any of these categories. One of these conditions is bicuspid aortic valve (BAV), in which two of the leaflets constituting the aortic valve fuse during development, thus resulting in a two-leaflet valve (bicuspid) instead of the normal three-leaflet valve (tricuspid)⁸⁷. BAV is the most common form of congenital heart defect present at birth, affecting approximately a 1-2% of the population⁸⁷.

1.2.2. Quality of life

Improvements in diagnosis and treatment of CHDs have increased the survival rate of affected children into adulthood up to 85%⁸⁸. However, due to the extremely little regenerative ability of adult cardiomyocytes⁸⁹, cardiac defects found in CHD patients will persist unless dealt with surgically. These patients may need multiple surgeries to correct all the structural defects present in affected individuals, thus having a detrimental effect on their quality of life. In addition, children with CHDs often develop neurological disorders, suggesting an important secondary effect of congenital heart diseases *in utero*⁹⁰.

1.2.3. Etiology

Regarding CHD etiology, little was known about the causes of CHDs until relatively recently. Several studies have linked the likelihood of CHD at birth with different environmental factors such as smoking, infections, excessive alcohol use and diabetes⁹¹. However, these epidemiological studies have

mostly suggested risk factors rather than explaining the underlying disease mechanisms. Thanks to the progress in the field of genetics, we now know that the most common cause for CHDs are genetic mutations⁹². Human genetic studies have shown that most of the genes responsible for CHDs are transcription factors that regulate and control a wide range of events taking place during heart development, such as atrial and ventricular septation or great vessel formation^{92,93}. The first reported single-gene mutation leading to an inherited congenital heart disorder was in the gene *TBX5*, a member of the T-box transcription factors and the causative gene in Holt-Oram syndrome (HOS)^{94,77}. Following *TBX5*, mutations in other transcription factors have been connected to inheritable congenital heart disorders, including *NKX2-5*, linked to ASD and atrioventricular block⁹⁵; *GATA4*, associated with septation defects⁷⁸; and *MEF2C*⁹⁶ among others.

Mutations in genes other than transcription factors can also give rise to CHDs. A clear example is *MYH6*, a gene encoding cardiac α -myosin heavy chain, which is a contractile protein whose mutations are found in families with inherited ASD⁹⁷. In addition, different components of well-characterised signalling pathways such as Wnt⁹⁸ and Notch⁹⁹, as well as microRNAs¹⁰⁰ (small non-coding RNA molecules that regulate gene expression at mRNA level) have also been associated with CHDs.

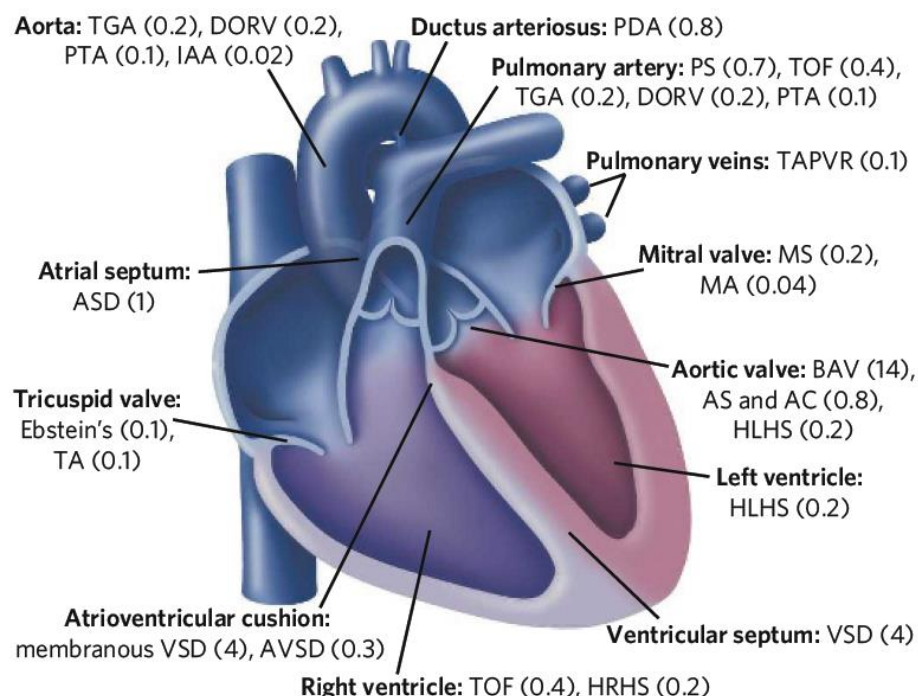


Figure 1.4. Anatomy of the adult heart and congenital heart defects. Diagram showing the anatomy of the adult human heart as well as the structures affected by CHDs. Numbers in parenthesis indicate the incidence rate of each disease per 1000 live births. AC, aortic coarctation; AS, aortic stenosis; ASD, atrial septal defect; AVSD, atrioventricular septal defect;

BAV, bicuspid aortic valve; DORV, double outlet right ventricle; Ebstein's, Ebstein's anomaly of the tricuspid valve; HLHS, hypoplastic left heart syndrome; HRHS, hypoplastic right heart; IAA, interrupted aortic arch; MA, mitral atresia; MS, mitral stenosis; PDA, patent ductus arteriosus; PS, pulmonary artery stenosis; PTA, persistent truncus arteriosus; TA, tricuspid atresia; TAPVR, total anomalous pulmonary venous return; TGA, transposition of the great arteries; TOF, tetralogy of Fallot; VSD, ventricular septal defect. *Figure from Bruneau, 2008*⁷

1.2.4. Potential treatments

Unfortunately, although the adult mammalian heart possesses a certain capacity for cardiomyocyte renewal, it is not sufficient to restore the lost cells and repair the structural defects present in CHD patients¹⁰¹. This translates into the need of multiple corrective surgeries, which carry significant risk to infants as they can develop secondary heart disease and eventually heart failure in later life^{7,102}.

In the last decades, a lot of effort and resources have been invested in the development of stem cell-based therapies that promote heart repair in damaged hearts or following an injury^{103,104}. Several studies in animal models and humans have shown that cell replacement strategies, in which stem cells are injected directly into the damaged heart or coronary circulation, have certain beneficial effects on cardiac function^{105,106}. In addition, it has been shown that cardiac fibroblasts can be differentiated into cardiomyocytes after transfection with cardiac transcription factors Tbx5, Gata4 and Mef2c¹⁰⁷. This finding could be also applied to promote cardiac repair not only in CHD patients but in other kinds of cardiomyopathy such as myocardial infarction¹⁰⁸.

However, the molecular basis of cardiac function improvement following administration of stem cells remain unknown, so the need remains for the development of new strategies to enhance cardiac regeneration and replace the lost cardiomyocytes in damaged hearts.

1.3. Cardiac transcription factors

Heart development is a complex process that requires the precise action and interaction of multiple transcription factors, signalling molecules and structural proteins to ensure it occurs properly. During the last decades, it has been shown that cardiac development is mainly governed by a core set of evolutionarily conserved transcription factors that through mutual cooperation, control the expression of thousands of genes, morphogenesis of cardiac structures and cell fate⁷⁰. In most cases, these cardiac transcription factors cooperate with each other in order to finely tune and reinforce the cardiac genetic program⁷⁰. These cardiac transcription factors are members of different protein families, including the T-box (TBX) family¹⁰⁹, GATA family zinc finger proteins¹¹⁰, MEF2 factors¹¹¹, NK-homeobox proteins¹¹² and the Hand

family¹¹³ (Fig. 1.5). Even though these are the key regulators of cardiogenesis, dozens of other transcription factors are also involved in cardiac development.

1.3.1. T-box family and TBX5

The T-box family is an ancient family of genes encoding similar transcription factors that are characterised by the presence of a highly conserved DNA-binding region, the T-box domain^{109,114}. This domain recognizes a specific DNA element, the T-half site, which can be found in promoters of a wide range of different genes. Thus, by binding to this T-half site, T-box family members mediate the transcriptional activation or repression of the downstream gene^{109,114}. In addition to its DNA-binding capacity, the T-box domain functions also as an interaction domain for other transcription factors, chromatin remodelling complexes or histone-modifying enzymes involved in transcriptional regulation^{109,114}.

Thanks to the progress in genome sequencing, we now have a better picture of the T-box family members in most animal species. In mouse and humans, 17 T-box genes are present and they are specifically expressed within progenitor fields and developing tissues/organs^{109,114}. If we focus specifically on the developing mammalian heart, 6 of the 17 T-box genes (TBX1, TBX2, TBX3, TBX5, TBX18 and TBX20) are expressed and required in different cardiac progenitor pools and in different developing compartments^{109,114}. In addition, knockout models of these genes exhibit cardiac defects during development^{67,115–117}, and *TBX1*, *TBX5* and *TBX20* mutations have been found in CHD patients^{77,118,119}.

Among all the members of the T-box family, TBX5 is perhaps the most relevant and well-characterized gene involved in heart development and CHD. In a similar fashion to other members of the family, TBX5 contains a T-box DNA binding domain in its N-terminal domain, and two nuclear localization sequences (NLS)¹²⁰ as well as one nuclear export sequence (NES)¹²¹ have been described for the protein (Fig. 1.5).

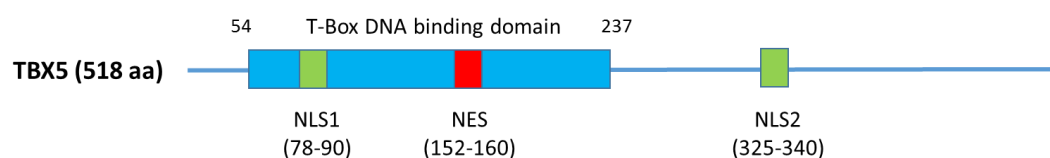


Figure 1.5. Secondary structure of TBX5. TBX5 is comprised of a single T-box DNA binding domain (blue), as well as two nuclear localization sequences (NLS1 and NLS2, green) and a nuclear export sequence (NES, red). The position (amino acids number) of each domain is also displayed. Aa = amino acids

In the murine heart, *Tbx5* can be first identified at E8.0 throughout the cardiac crescent¹²². Between E8.25 and E8.5, its expression becomes restricted to the

posterior pole of the primitive heart tube (corresponding to the sinus venosus and the primitive atria)^{122,123}. At this stage, a weaker expression level can also be found in the myocardium that encircles the endocardial tube¹²². During the looping process (E8.5-9.0), *Tbx5* expression expands anteriorly to encompass the future left ventricle¹²². Once the looping process progresses and the left and right ventricles present a left-right arrangement (around E9.0), *Tbx5* is significantly expressed in the left ventricle but not in the right one¹²². This left-restricted expression pattern persists throughout development: at E11.5, *Tbx5* is mainly expressed in the left ventricle, whereas just a really weak expression can be found in the right ventricle or outflow tract¹²². Additionally, atrial expression of *Tbx5* is even stronger than in the left ventricle, an expression pattern that persists throughout the chamber maturation and septation processes^{122,123}. At E13.5, *Tbx5* can also be found both in the left ventricle free wall and trabeculae, and in the right ventricle trabeculae, albeit weaker⁷⁰. At this stage, *Tbx5* mRNA is present in the developing atrial septa, whereas its expression in the ventricular septum is limited to the left side of it⁷⁰. *Tbx5* expression is also detected in the atrioventricular valves, the left and right superior vena cavae, and in the inferior vena cavae⁷⁰. In addition to its characteristic cardiac expression, *Tbx5* is also expressed in other non-cardiac tissues in mouse, such as the developing forelimb. Expression in this organ is maintained until E11.5, when it then becomes restricted to the proximal portion of the forelimb¹²⁴. *Tbx5* can also be found in the optic vesicle and the neural retina of the developing eye in mouse⁷¹.

During zebrafish development, *tbx5a* is expressed in the heart, the pectoral fins and the eye¹²⁵, mirroring the expression pattern found in mice. Focusing on the heart, *tbx5a* can be first found in the cardiac precursors residing in the anterior lateral plate mesoderm at the 6- or 7- somite stage¹²⁵. At the 10- and 15-somite stage, *tbx5a* expression expands mediolaterally and can be found in the bilateral heart tubes, which later will fuse (20-somite stage) resulting in a region of ring-like expression¹²⁵. At this stage, *tbx5a* is expressed in the myocardial precursors as well as in the posterior regions that will give rise to the fins¹²⁵. At 26 hpf, *tbx5a* expression is stronger in the primitive atrium than in the primitive ventricle¹²⁶. However, as cardiac looping progresses, cardiac *tbx5a* expression reverses in direction, and by 48 hpf, it can be mainly found in the ventricle, showing a weak expression in the atrium and completely absent in the outflow and inflow tracts¹²⁶.

In humans, *TBX5* was the first reported single-gene mutation associated to an inherited CHD: Holt-Oram syndrome^{94,77}. This syndrome is characterized by the presence of limb and heart abnormalities in patients, including septal defects and conduction disease¹²⁷. To date, more than 70 mutations in this gene have

been identified in HOS patients¹²⁸. These mutations lead to the translation of defective proteins in DNA-binding, transcriptional activity or protein-protein interactions¹²⁸.

Regarding animal models, both gain and loss-of-function experiments have addressed the precise functions of TBX5 in the developing heart. Mice lacking one copy of *Tbx5* mirror the Holt-Oram syndrome phenotype, as they present ASD, occasional VSD and AV block, as well as limb defects⁶⁷. Homozygous mutant mice for *Tbx5* do not survive past E10.5, and development appears to be arrested at E9.5⁶⁷. In addition, they establish a linear heart tube, showing high levels of hypoplasia as well as reduced expression of multiple genes⁶⁷. In contrast, overexpression of *Tbx5* in the mouse heart results in expansion of left ventricular identity, looping defects and loss of the ventricular septum¹²⁹.

Biochemical analyses have proved that TBX5 directly associates with other cardiac transcription factors, including the NK-type homeodomain factor NKX2-5 and the zinc finger factor GATA4^{78,130}. Acting in synergy with NKX2-5 and GATA4, these three transcription factors positively regulate the transcription of chamber-specific genes such as *GJA5* (which encodes connexin 40)⁶⁷ and *NPPA in vitro*¹³⁰. Moreover, CHD-mutations in any of these factors disrupt these interactions, and patients show overlapping clinical features^{78,130,131}.

1.3.2. NK-type homeodomain factor NKX2-5

NKX2-5 is a member of the NK homeobox gene family that is highly conserved during evolution and acts as a DNA-binding transcriptional activator¹¹². It was originally identified as a vertebrate homolog of the *Drosophila* gene *tinman*, responsible for heart formation in *Drosophila*¹³². Gene expression studies have shown that *Nkx2-5* is highly expressed in cardiac progenitor cells from both the first and secondary heart fields during mouse embryogenesis^{133,134}, and it also presents a heart-restricted expression pattern in other organisms such as humans or zebrafish¹³⁵.

NKX2-5 acts, in cooperation with other cardiac factors (such as GATA4 and TBX5), near the top of a large transcriptional cascade controlling the expression of multiple cardiac genes such as *IRX4*¹³⁶ and *JARID2*¹³⁷ and thus, it regulates many aspects of cardiac development, including the proliferation and differentiation of cardiac precursors¹³⁶ and outflow tract formation¹³⁷. The crucial role of NKX2-5 in heart development has been also deciphered by the generation of gain- and loss-of-function animal models. Overexpression of *Nkx2-5* in *Xenopus* and zebrafish embryos leads to large hyperplastic hearts due to increased myocardial cell number^{138,139}. In contrast, mice lacking *Nkx2-5* die

during embryonic development due to failure of cardiac looping and deficient myocardial differentiation, as well as cardiac conduction defects^{140,141}.

In humans, mutations in the *NKX2-5* gene have also been linked to CHD. Different mutations of *NKX2-5* have been found in CHD patients with a variety of structural abnormalities, including septation defects (ASD and VSD), subvalvular aortic stenosis and cardiac conduction defects⁹⁵.

1.3.3. GATA family of transcription factors

One of the crucial family of proteins for the regulation of heart development is the GATA family of transcription factors. Proteins belonging to this family are characterized by the presence of two highly conserved zinc finger DNA-binding domains, as well as two divergent transactivator domains^{142,143}. This family is comprised of six members: GATA1 to GATA6. These six proteins can be subdivided into two groups based on their spatiotemporal expression pattern: GATA1, 2 and 3 are mainly expressed in hematopoietic cell lineages¹⁴⁴ and are essential for the proliferation of hematopoietic stem cells and megakaryocyte differentiation among other functions^{145–147}. The second group is formed by GATA4, 5 and 6, which are primarily found in tissues of mesodermal and endodermal origin such as the heart¹⁴⁸.

1.3.3.1. GATA4

Gata4 is one the earliest transcription factors expressed in developing murine cardiac cells, and perhaps the most well-characterized member of the GATA members involved in cardiogenesis^{148,149}. Specifically, Gata4 can be detected as early as E7.5 in the precardiac mesoderm, and constant levels of both mRNA and protein can be found in the heart tube during the cardiac looping process^{148,149}.

In recent years, several studies have uncovered the precise roles of GATA4 during cardiac development. Dosage of Gata4 is responsible for correct regulation of cardiac morphogenesis¹⁵⁰, and it also acts as an important regulator of cardiomyocyte proliferation through transcriptional activation of cell cycle regulators¹⁵¹. In addition, like *NKX2-5*, GATA4 interacts and cooperate with other cardiac transcription factors to regulate proper heart development and cardiac gene program activation, including the previously described factors *TBX5* and *NKX2-5* as well as others such as *MEF2C*^{152–154}.

Loss-of-function studies have shown that mice lacking Gata4 present embryonic lethality at E10.5^{155,156}, and they show abnormal ventral folding, failure in heart tube formation and extensive endoderm defects^{155,156}. In humans, GATA4 is one of the main genes involved in the onset of CHD. Patients with mutations or deletions in this cardiac factor display a wide range of cardiac

disorders, including septation defects (ASD, VSD and AVSD), outflow tract alignment defects and pulmonary stenosis^{68,69,157}.

1.3.3.2. GATA5

Gata5 is expressed in the myocardium, endocardium and endocardial cushions in mouse embryos¹⁵⁸. In mice, deletion of both Gata5 isoforms leads to hypoplastic hearts and partially penetrant bicuspid aortic valve formation¹⁵⁹. In a different mouse model, embryos carrying a *Gata5* mutant allele lacking the zinc finger domains are viable; however, in a *Gata4*^{-/-} background, these embryos die at mid-gestation and present abnormalities in cardiomyocyte proliferation and cardiac chamber maturation¹⁶⁰. In humans, heterozygous mutations in the *GATA5* gene have been found in patients of different congenital heart disorders, including septal defects, tetralogy of Fallot and bicuspid aortic valve^{161,162}.

1.3.3.3. GATA6

In the case of Gata6, this transcription factor can be mainly found in vascular smooth muscle cells during murine development¹⁶³. Regarding mouse models, complete ablation of Gata6 in mouse embryos leads to early embryonic lethality (between E6.5-7.5) and defects in endoderm differentiation as well as widespread cell death within the embryonic ectoderm¹⁶⁴. Tissue-specific deletion of Gata6 in vascular smooth muscle cells¹⁶⁵ or in the neural crest¹⁶⁵ results in perinatal mortality from a spectrum of cardiovascular defects such as persistent truncus arteriosus and interrupted aortic arch¹⁶⁵, thus suggesting a role for Gata6 in the proper patterning of the cardiac outflow tract and aortic arch arteries¹⁶⁵. It has been also found that Gata6 regulates posterior cardiac development together with Wnt2 in a feedforward transcriptional loop¹⁶⁶.

In humans, and in similar fashion to GATA4 and GATA5, several mutations in the *GATA6* gene have been found in patients of different cardiac conditions such as tetralogy of Fallot, atrioventricular septal defects and persistent truncus arteriosus among others^{167,168}.

1.3.4. MEF2 family

The myocyte enhancer factor 2 family of transcription factors is characterized by the presence of a highly conserved N-terminal MADS box domain¹⁶⁹, which mediates the binding to DNA target sequences in muscle-specific genes, and a MEF2-specific domain which modulates MEF2 function and thus the activation or repression of its target genes¹⁶⁹. Nuclear localization signals and transcriptional activator domains can be found in the C-terminal domain of the members of this family, which is the main source of divergence between them¹⁶⁹. In vertebrates, the MEF2 family is comprised of four members:

MEF2A, MEF2B, MEF2C and MEF2D. These members of the MEF2 family present different, but overlapping, spatial and temporal expression pattern during developmental and adult stages. All four of them are primarily found in striated muscle and brain, but they can also be found in other tissues and cell types such as lymphocytes, neural crest and endothelium among others^{111,169,170}.

During these last decades, many studies have confirmed MEF2 factors as one of the core components of the muscle gene regulatory networks, and they are the only myogenic transcription factors that regulate the differentiation of all muscle cell types¹⁷¹. MEF2 cooperates and interacts with other core cardiac transcription factors such as TBX5¹⁷² and GATA4¹⁵⁴ in order to regulate contractile gene expression in cardiac muscle cells, whereas in skeletal muscle, MEF2 interacts with the MyoD family of transcription factors¹⁷³.

All MEF2 family members are expressed during cardiac development at different time points during murine development. Mef2b and Mef2c can be found as early as E7.5 in the cardiac mesoderm^{169,174}, with Mef2c being also expressed at this stage in the sinus venosus, which contributes to the cardiac atria^{169,174}. Around E8.5, Mef2a and Mef2d are expressed in the linear heart tube, and after this point, all four Mef2 members are expressed across the developing heart^{169,174}.

The generation of knockout models for each one of the MEF2 family members have confirmed their essential role during heart development. Complete Mef2a ablation in mice leads to an 80% of perinatal lethality¹⁷⁵. These pups also present dilation of the right ventricle, myofibrillar fragmentation, mitochondrial disorganization and activation of a fetal cardiac gene program¹⁷⁵. Surviving mice also shows deficiency of cardiac mitochondria and susceptibility to sudden death¹⁷⁵. Global knockout of Mef2c leads to embryonic lethality, defective cardiac looping morphogenesis and cardiovascular malformations^{176,177}. In contrast, knockout models of Mef2b¹⁷⁷ and Mef2d¹⁷⁹ are viable and they display a completely normal cardiac phenotype. Mef2 family members have also been overexpressed in mouse; overexpression of Mef2a or Mef2c in neonatal cardiomyocytes leads to dilated cardiomyopathy¹⁷⁹, whereas overexpression of Mef2d in cardiomyocytes produces extensive fibrosis and enlargement of the atria¹⁸⁰.

1.3.5. HAND1 and HAND2

HAND1 and HAND2 are members of the basic helix-loop-helix (bHLH) family of transcription factors. As all the other transcription factors mentioned in this chapter, HAND factors also regulate the transcription of cardiac genes by

cooperating with other core members of the cardiac regulatory network such as GATA4.¹⁸⁰

During the initial stages of cardiogenesis, Hand1 and Hand2 are expressed throughout the cardiac region; however, later on their expression becomes more restricted; Hand1 is predominantly expressed in the left ventricle, whereas Hand2 is mainly expressed in the right one^{181,182}. Regarding animal models, embryos homozygous for a *Hand1* null-allele present embryonic lethality at E9.5, showing deficiencies in extraembryonic mesoderm and abnormal heart development, as the heart did not progress beyond the cardiac-looping stage¹⁸³. In the case of the complete ablation of Hand2 in mouse embryos, this model dies at E10.5 and displays right ventricular hypoplasia as well as vascular malformations¹⁸².

Mutations in the *HAND1* and *HAND2* genes have been also found in CHD patients. *HAND1* has been shown to be linked to double outlet right ventricle and ventricular septal defects^{184,185}, whereas mutations in the *HAND2* gene have been discovered in patients with pulmonary stenosis, tetralogy of Fallot and ventricular septal defects^{186,187}.

1.3.6. Other cardiac developmental genes

Even though the previously described families of transcription factors constitute the core set of cardiac transcription factors implicated in heart development and in the onset of CHD, there are hundreds of other genes and factors involved in the control and regulation of cardiac development. Among them are other transcription factors, such as *SALL4*¹⁸⁸ or *FOG2*¹⁸⁹; and other types of genes, including sarcomeric proteins such as *MYH6*¹⁹⁰ or *ACTC1*⁷⁴. In addition, their relevance to cardiac development has been extensively proved as mutations in each have been found in CHD patients presenting a variety of conditions¹²⁸.

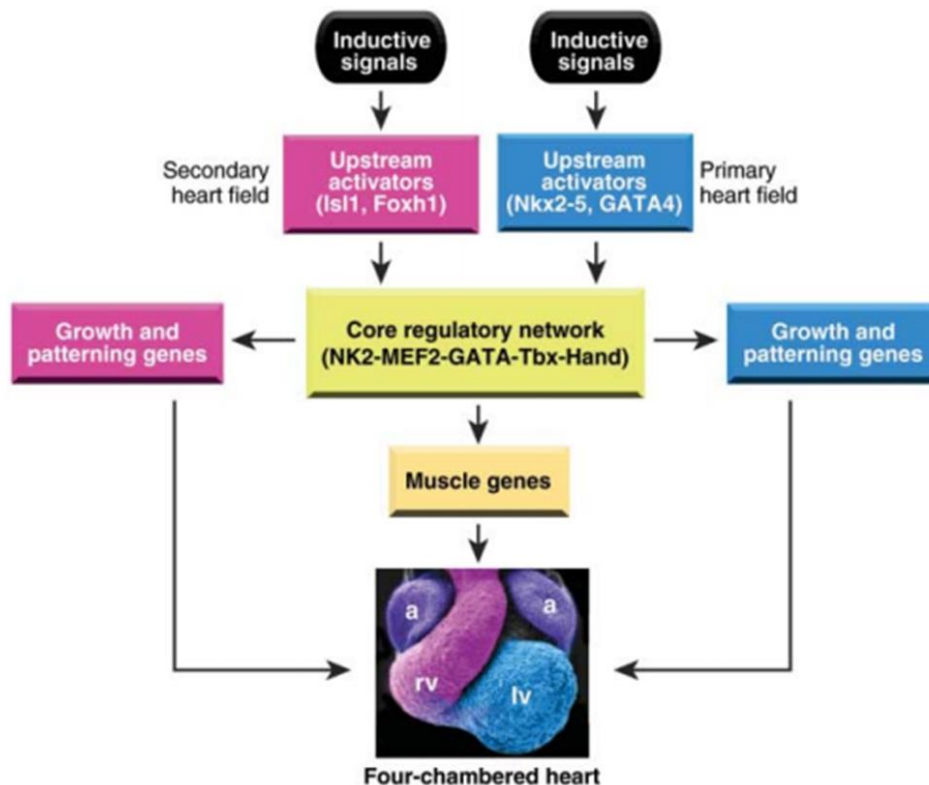


Figure 1.6. Transcriptional networks involved in mammalian heart development. At first, inductive signals activate a group of upstream activators both in the primary heart field (e.g. NKX2.5, GATA4) and in the secondary heart field (ISL1, FOXH1). These transcription factors then proceed to activate the expression of genes in the core regulatory network, which include members of the NKX, MEF2, GATA, TBX and HAND families. These core components have the ability to cross- and autoregulate their expression, and cooperate with each other in order to strengthen transcriptional activity. This way, these factors activate the expression of muscle-specific genes as well as the expression of growth and patterning genes of the primary and secondary heart fields. During cardiogenesis, the primary heart field will give rise to the left ventricle and parts of the atria, whereas the secondary heart field will give rise to the right ventricle, the outflow tract and the remaining portions of the atria. Rv: right ventricle; lv: left ventricle; a: atria. Primary heart field-derived structures are shown in blue, whereas secondary heart field-derived ones are shown in pink. The atria are shown in purple, as these structures are derived from both heart fields. *Adapted from Olson, 2006⁷⁰. Note: in the heart picture, it looks like a small area from the right ventricle is derived from the PHF. However, in the reference is indicated that the right ventricle is exclusively derived from the SHF, as previously indicated in this thesis.*

1.4. Genome-editing technologies in cardiovascular biology

Thanks to the recent advances in the field of genome-editing technologies, we now have many tools for rapidly modelling disease causative-mutations and studying new candidate genes of interest using both *in vitro* and *in vivo* models. Taking into account that genetic mutations are one of the main contributing factors to CHD, these technologies are quickly becoming one of the most useful

tools to ultimately understand the molecular mechanisms controlling cardiogenesis, to discover new disease causative-genes and also for their potential therapeutic applications.

1.4.1. Tools for genome editing

Nowadays, the most-prominent and widely-used genome-editing technologies are zinc-finger nucleases (ZFN), transcription activator-like effector nucleases (TALEN) and clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9¹⁹¹. Although different in many ways, these 3 technologies share the same principle of action, which consists in introducing directed changes in the genome by inducing double strand breaks (DSB) in the targeted region¹⁹¹. Upon cleavage, this DSB can be repaired through one of two endogenous DNA repair pathways: non-homologous end joining (NHEJ) or homology-directed repair (HDR). In the first case, broken ends are directly ligated without the need for a homologous template. Since this process is error-prone, it typically leads to the introduction of short insertions or deletions that can lead to frameshifts and premature stop codons¹⁹². In the second pathway, a homologous DNA sequence must be used to guide repair. After the endonuclease performs the DSB, the homologous DNA sequence recombines with the sequences flanking the DSB, thereby being incorporated into the genome at the target site position^{192,193}. Apart from ZNFs, TALENs and CRISPR/Cas9, other genome-editing tools such as meganucleases are also available¹⁹⁴. However, their low efficiency and adaptability have made of them a secondary option when compared to any of the three previously mentioned¹⁹⁵.

1.4.1.1. Zinc-finger nucleases (ZFNs)

Zinc-finger nucleases are engineered restriction enzymes which recognize DNA sequences flanking the genomic target site of interest. ZNFs are chimeric proteins, as they are created through the joining of two genes that originally coded for independent proteins: a zinc-finger DNA-binding protein, and a DNA cleavage domain which typically comprises the nuclease domain of the bacterial restriction endonuclease FokI (Fig. 1.6A)^{196,197}. The DNA-binding domains of ZNFs usually contain between three and six zinc-finger repeats. Each one of these repeats recognizes 3-4 base pairs (bp), thus accounting for a total recognition of 9-18 bp, which ultimately comprises the length of the DNA target sequence. Following binding of the DNA-binding domain to the target site flanking regions, the two DNA cleavage domains dimerize, producing a DSB which activates the DNA repairing machinery^{196,197}. ZNFs offer many advantages in terms of flexibility over other classic genome-editing technologies (e.g. RNAi, viral transgene expression); however, the design of constructs is highly time-consuming, laborious and potentially expensive. In

addition to this, ZFNs target site are rather limited, as generally they cannot be designed to target sequences with low guanine content¹⁹⁸.

1.4.1.2. Transcription activator-like effector nucleases (TALENs)

TALENs are conceptually and structurally similar to the already described ZFNs: they are comprised of a TAL-effector DNA binding domain fused to a DNA cleavage domain (bacterial FokI)¹⁹⁹. The main difference between TALENs and ZFNs resides in the DNA binding domain: in TALENs, this domain contains highly conserved 33-35 amino acids repeats (TAL repeats) with variable amino acids in the 12th and 13th position²⁰⁰. These two positions are referred to as the Repeat Variable Diresidue (RVD), which is able to bind specific nucleotides depending on the di-residue sequence. Since different RVD sequences can bind different nucleotides, TAL repeats can be engineered and modified in order to have specific RVDs which bind specific nucleotide sequences^{199–201} (Fig. 1.6B). One of the main advantages of TALENs in comparison to ZFNs is that they are much easier to engineer than the latter. In addition to this, TALENs do not have the target site limitations ZFNs show, as they have numerous dimeric target sequences that, on average, are found every 3 bp in the genome^{202,203}. In spite of this, engineering TALENs is still considered a time-consuming task when compared to newer genome-editing technologies like CRISPR/Cas9, and their relatively large size (TALEN cDNA is 3kb, ZFN cDNA is 1kb) makes them harder to deliver and express into cells, and subsequently less attractive for their use in therapeutic applications²⁰⁴.

1.4.1.3. Clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9

One of the genome-editing technologies that is gaining prominence very quickly is the CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)/Cas9 system, a defense mechanism system originally discovered in bacteria that works against invasion of foreign nucleic acids derived from virus or plasmids²⁰⁵. The system is comprised of three different components: an array of small CRISPR RNAs (crRNAs), an auxiliary trans-activating crRNA (tracrRNA) and a nuclease associated with the CRISPR locus (Cas9 protein)^{206,207}. The CRISPR locus carries short palindromic repeats flanking unique spacer sequences, which result from the incorporation of foreign nucleic acids into the CRISPR locus after infection of the bacterium by a virus or plasmid DNA^{206,207}. Upon transcription, these short sequences generate short RNAs (crRNAs), which recognize the original invading nucleic acid sequence by Watson-Crick complementarity^{206,207}. After binding of the sequence specific crRNA to the foreign nucleic acid, the tracrRNA is recruited and subsequently promotes the incorporation of the Cas9 protein to the crRNA/tracrRNA/foreign DNA complex. Then, Cas9 cleaves the foreign DNA, thereby inactivating the

invading pathogen^{206,207}. Recent *in vitro* work showed that the crRNA and tracrRNA can be fused to generate a single guide RNA (sgRNA) sufficient for Cas9 endonuclease-mediated cleavage of a target DNA sequence²⁰⁸. sgRNAs can be modified to include a 20-nt sequence complementary to a specific 20-nt sequence in the genome (target site) (Fig. 1.6C). The only constraint for the design of sgRNAs is that the genomic site chosen for targeting must be adjacent to a three nucleotide sequence called PAM (protospacer adjacent motif), which sequence is NGG and it is required for Cas9 activity. Thus, following sgRNA transcription, this binds to its complementary DNA sequence in the genome and recruits Cas9 endonuclease to this target site, resulting in target site cleavage 3-4 bp upstream of the PAM sequence²⁰⁸.

The main advantage of the CRISPR/Cas9, and perhaps the main reason for its exponential growth in popularity, is its ease of design and use. For each target site, all that is needed is the 20-nt sequence in the sgRNA sequence, which can be generated in a day. This allows for the possibility of screening several sgRNAs in a short period of time, greatly reducing optimization times and costs. Another of the great advantages of the system is its high efficiency in comparison to ZFNs or TALENs, with efficiency levels as high as 95% being reported^{209,210}. However, CRISPR/Cas9 system are not without their limitations. As mentioned previously, cleavage requires the presence of a PAM sequence in the target site, which only occurs every 8 bp in the genome on average²¹¹, a great limitation when a mutation is needed to be introduced in a specific position with no PAM sequence nearby. Also, in a similar fashion to other genome-editing technologies, the CRISPR/Cas9 system still suffers from off-target effects, a drawback that is currently being addressed by performing different alterations in the Cas9 protein, such as converting it to a nickase^{212,213}.

1.4.1.4. Uses of the genome-editing tools

In cardiovascular biology, as in many other fields, genome-editing technologies are routinely used for three different purposes: the generation of *in vitro* disease models, the generation of *in vivo* disease models and for gene therapy approaches.

1.4.1.4.1. Generation of *in vitro* disease models

Historically, the generation of *in vitro* disease models has involved the harvest of somatic cell from patients. These cells can be then reprogrammed into induced pluripotent stem cells (iPSC), which have the ability to differentiate into many different cell types, thus allowing for the functional analysis of this specific mutation^{214–216}. However, this approach is far from perfect, as each

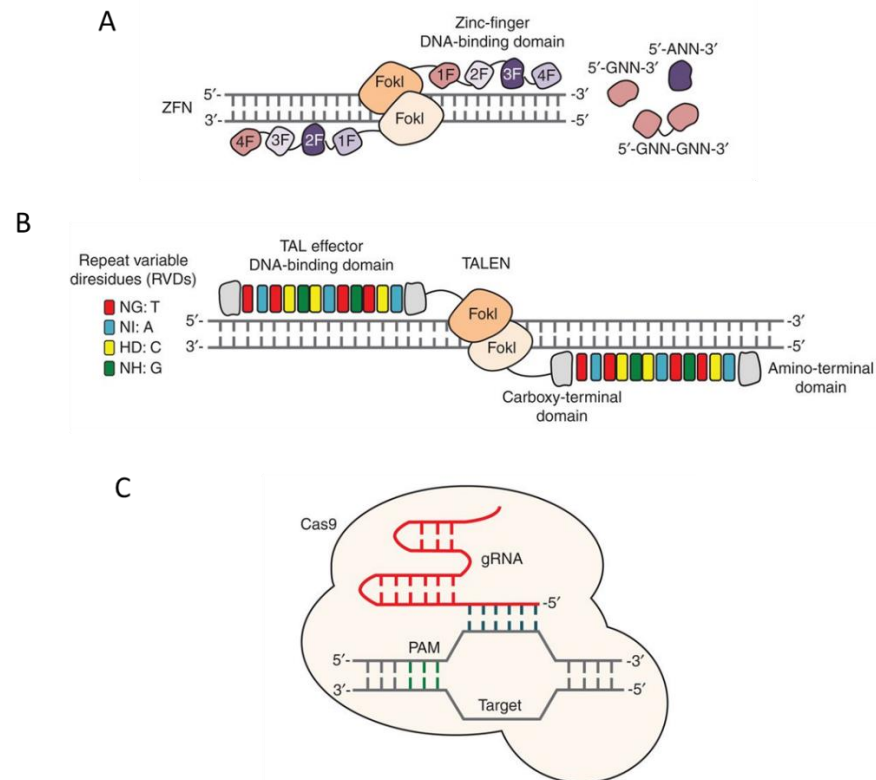


Figure 1.7. Genome-editing technologies. Illustration showing the structure and mechanism of action of (A) ZFNs, (B) TALENs and (C) CRISPR/Cas9. (A) ZFNs recognize a target site that is comprised of two zinc-finger binding domains flanking a central 5 to 7 bp spacer sequence. A double-strand break is performed by the FokI endonuclease domain in this spacer sequence following dimerization. (B) TALENs bind to target sites consisting of two TALE DNA-binding sites that flank a 12 to 20 bp spacer sequence recognized by the FokI endonuclease domain. (C) CRISPR/Cas9 system is comprised of a guide RNA (gRNA) and the Cas9 endonuclease. The gRNA recognizes a 20 nt sequence (target site) located immediately upstream of a protospacer adjacent motif (PAM), whereas the Cas9 endonuclease performs a double-strand break in that target site after binding. *Adapted from Gaj et al, 2016*²¹⁷

iPSC derived from patients present a different epigenetic profile²¹⁸ and genetic background²¹⁹, so the results obtained from one of them might not be applicable to a different patient. An alternative approach to avoid these problems consists on the introduction of patient-derived mutations in hPSCs by using genome-editing technologies. This allows for a more consistent, reproducible and rigorous characterization of the impact of the mutation. By using ZFNs, TALENs and CRISPR/Cas9, many cardiovascular diseases have been already modelled in hPSCs. For example, Wang et al. were the first to successfully model a cardiovascular disease in iPSCs by introducing mutations found in patients with Barth syndrome (a mitochondrial cardiomyopathy).²²⁰ In another study, a dilated cardiomyopathy model was generated in iPSCs by mutating the *Titin* gene²²¹, whereas valvular disease has also been modelled in iPSC-derived endothelial cells by introducing mutations in the *NOTCH1* gene²²².

1.4.1.4.2. *Generation of in vivo disease models*

Due to the ease of use and efficiency of the latest genome-editing technologies, strains of genetically modified animals can be generated in the course of several weeks. The most-widely used animal model for genetic modification is the mouse. Traditionally, transgenic mouse lines are generated by introducing the desired mutation in mouse embryonic stem cells²²³. Following appropriate selection, stem cells carrying the mutation of interest are subsequently injected into mouse blastocysts for the production of chimeric mice, and the stable transgenic line is finally obtained through breeding²²³. Nowadays, and with the development of the CRISPR/Cas9 system, a more direct approach is used: the genome-editing machinery is directly injected into 1-cell stage embryos to knock-out the targeted gene, or, if inserted with a piece of DNA with homologous arms, to produce a knock-in line. To date, hundreds of disease *in vivo* models have been generated, and cardiovascular diseases as well as cardiovascular development have been modelled and explored in many organisms, including mouse²²⁴, rat²²⁵, rabbit²²⁶ and zebrafish²²⁷ among others.

1.4.1.4.3. *Genetic therapy*

Many cardiovascular diseases have their cause in genetic mutations. Taking this into account, the ability to easily modify the genome offered by the latest genome-editing technologies entails an exciting new approach for the prevention and treatment of cardiovascular diseases²²⁸. Genome-editing technologies can be potentially used to disrupt genes that promote the onset of specific cardiovascular diseases, as well as to correct or repair mutations leading to disease²²⁸. In the last five years, the CRISPR/Cas9 system has been used to disrupt disease-promoting genes or to correct disease-causing DNA mutations related to cardiovascular diseases, ranging from a single nucleotide to large deletions. For example, the *PCSK9* gene has been successfully targeted in a “humanized” mouse model. When dysregulated, this factor promotes hypercholesterolemia; however, targeting of this gene using the CRISPR/Cas9 system leads to reduced levels of cholesterol and consequently the risk of myocardial infarction^{229,230}. CRISPR/Cas9 targeting of the *Fah* gene in adult mouse has been also used to treat tyrosinemia and prevent cardiovascular diseases²³¹; and Duchenne muscular dystrophy has also been a gene therapy target in a mouse model by restoring dystrophin expression²³². Even though the potential of these technologies for therapeutics application is substantial, they still have challenges and limitations, including potential off-target effects (that could lead to other diseases such as cancer), dysregulation of other biological processes and the difficulty of targeting specific tissues *in vivo*, among others.

1.5. Animal models for the study of cardiac development and CHD

Due to the extensive contribution of genetic causes to the onset of CHDs, model organism studies have become crucial in order to decipher the role of specific genes in CHD and cardiac development. Four organisms are used routinely for the study of vertebrate heart development and congenital heart disorders: mouse, zebrafish, chick and *Xenopus*. During this chapter, I will exclusively focus on mouse and zebrafish as these are the two animal models used during this thesis.

1.5.1. Mouse

1.5.1.1. Heart development in mouse

As in many other biological research fields, the mouse is perhaps the most popular animal model for the study of cardiogenesis and cardiac diseases. Murine heart development is extremely similar to human heart development²³³. This is one of the main reasons it has become the model organism of choice for studying human cardiogenesis. Most of the developmental processes explained in section 1.2 also take place during murine cardiogenesis; however, there are still certain differences between these species. The most prominent anatomical differences are found in the venous components of the atria, which reflect the differences in development of the venous tributaries that connect to the inferior atrial wall. In the human heart, four pulmonary veins connect to the left atrium, whereas in mouse these pulmonary veins join in a pulmonary confluency that joins the heart via a single orifice²³³. Differences can be also found in the venous drainage into the right atrium. Both in humans and mice, the left and right superior caval veins (LSCV and RSCV) drain into the sinus venosus during early cardiogenesis^{234,235}. In humans, as development proceeds, the left caval vein regresses and the remaining structure forms the coronary sinus; however, in mouse, the left caval vein does not regress and can be found in postnatal life²³⁶. In fact, the lack of left caval vein regression in humans lead to persistent LSCV, a congenital malformation with an incidence of 1:100²³⁶. Finally, another remarkable difference resides in the valve leaflets of the atrioventricular junction: in human, by CS23, these structures are thin and resemble what is seen in the postnatal heart; whereas in mouse the development and refinement of the valves continues through development (e.g. the septal leaflet of the tricuspid valve keeps maturing even after birth)²³⁷.

1.5.1.2. Advantages and disadvantages of using mouse as animal model

In spite of the differences described above, the sequence of events leading to the formation of a four-chambered heart is remarkably similar in mice and humans^{233,236}, thus constituting an excellent model for the study of cardiogenesis and cardiac disease. The mouse genome is very similar to the human one, with 99% of human genes having murine orthologs²³⁸, so potentially any human disease can be modelled in mouse. On top of this, mice are excellent models for studying complex biological systems like the endocrine system, and they can naturally develop diseases like diabetes²³⁹ and cancer²⁴⁰. Mice are also small rodents and relatively cheap to maintain. Adult mice can reproduce quite quickly (as often as every three-four weeks) and on average, litters are comprised of 5-8 pups, thus allowing for the quick growth of the colony. Amongst mammals, the mouse has one of the shortest gestation periods, giving birth to new pups in just 3 weeks approximately. The short lifespan of mice (around two years) is another great advantage of this model: one mouse year roughly equals to roughly 40 human years²⁴¹, a characteristic that allows for the study of the effect of diseases or mutations at an accelerated pace.

The mouse is indeed an invaluable animal model; however, as with every animal model, it has its drawbacks. The maintenance of mouse colonies is much more expensive than the maintenance of other animal models such as zebrafish, and this is still a limitation for many labs. When compared to this same model, mice gestation is intrauterine, so in order to perform developmental studies the mother needs to be culled. Also, unlike zebrafish, mouse embryos are not transparent, so the imaging of the cardiovascular system is rather challenging.

1.5.1.3. Genetic modification in mouse

Nowadays, the mouse is the golden standard for reverse genetics studies and for disease modelling. Since the creation of the first transgenic mouse in the early 1980s²⁴², each one of the genetic engineering techniques developed over the years have been successfully implemented in this model. In order to produce genetically modified mice, two different approaches are typically followed: in the first one, the genetic engineering system of choice is injected into 1-cell stage mouse embryos, where it will produce a genetic modification that will be present in all future cells derived from that one and eventually, in the transgenic animal. The second approach involves the modification of mouse embryonic stem cells with a DNA construct containing homologous sequences to the target gene, as well as a selection marker. These embryonic

stem cells are initially modified *in vitro*, and after selection, cells carrying the modification are injected into mouse blastocysts.

Nowadays, generation of transgenic mouse lines is performed using the three most prominent and widely-used genome-editing technologies: ZFNs, TALENs and CRISPR/Cas9. In the field of cardiovascular development, there are thousands of mouse models available that have allowed us to study the role of genes involved in cardiogenesis and to model human cardiac diseases²⁴³. Due to the massive number of models available nowadays, it would be practically impossible to list all of them in this thesis, so just a few selected examples are described next. For instance, the mouse has been successfully used to model many congenital heart disorders, the main topic of this thesis. Holt-Oram syndrome, a condition characterized by the presence of cardiac defects (such as ASD or VSD) and limb abnormalities, was successfully modelled by introducing mutations in the *Tbx5* gene in mouse^{77,94} using the ES cell approach. Another complex disease that has been partially modelled in mouse is the DiGeorge syndrome, also known as the 22q11.2 syndrome. This syndrome, which often includes congenital heart defects (interrupted aortic arch, truncus arteriosus and tetralogy of Fallot commonly) and thymus and parathyroid defects among others²⁴⁴, is produced by the deletion of a small section of the chromosome 22, which contains approximately 30 genes²⁴⁵. Between all these genes, studies have pointed out that haploinsufficiency of *Tbx1* is the cause of many of the symptoms found in this condition. Indeed, mutations in the *Tbx1* gene in mouse lead to several defects similar to those seen in humans, mainly affecting the development of the arteries^{246,247}. Other syndromic-CHD conditions for which mouse models have been generated include Marfan syndrome (mutations in the *FBN1* gene)²⁴⁸, Williams syndrome (microdeletion of a section of chromosome 7, which includes the *ELN* gene)²⁴⁹ and Alagille syndrome (mutation in the *JAG1* or *NOTCH2* genes)^{62,250}. In addition to these syndromic-CHD transgenic lines, mouse models of cardiomyopathies have also been generated, including hypertrophic²⁵¹, dilated²⁵² and noncompaction cardiomyopathy²⁵³; as well as models of arrhythmias and conduction system defects^{254,255}.

1.5.2. Zebrafish

1.5.2.1. Heart development in zebrafish

The zebrafish heart has some differences in comparison to the human and murine hearts. The main difference resides in its architecture, as this is much simpler than that of the mammalian models. It is composed of single ventricular and atrial chambers, which are separated by valve leaflets situated in the atrioventricular canal which prevent the backflow of blood²⁵⁶. In addition

to this, the zebrafish does not require separation of the cardiovascular system by atrial, ventricular or outflow tract septation as it lacks a proper pulmonary system²³⁷. Although the morphological and anatomical differences between the zebrafish and human heart may seem significant and substantial, many studies have demonstrated that the fish and mammalian hearts undergo similar morphogenetic processes. At the blastula stage (5 hours post-fertilization), two populations of cardiac progenitor cells (atrial precursors and ventricular precursors) are situated in the lateral marginal zone in a bilateral fashion²⁵⁷. During the gastrulation process, these two lateral populations of cardiac progenitors migrate towards the midline to end up in the anterior lateral plate mesoderm (ALPM), posterior to another bilateral population of cells which are the endocardial precursors²⁵⁸. At this point, the cardiogenic differentiation of the ventricular and atrial precursors to ventricular and atrial myocytes starts taking place²⁵⁹, and in parallel, the endocardial precursors start migrating from the ALPM towards the mid-line²⁶⁰. Atrial and ventricular myocardial cells start this migration process slightly later, and when the bilateral populations fuse in the mid-line, a disc-shaped cardiac structure is formed²⁶¹. In this cardiac disc, endocardial cells can be found within the hole at the centre, ventricular myocytes at the circumference and atrial myocytes right at the periphery of the disc. Then, morphological changes start taking place, and by a complex rotation and elongation perpendicular to the plane of the cardiac disc, this transforms into a cardiac tube^{262,263}. At 28 hpf, the linear heart tube is fully formed, with the arterial pole (ventricle) fixed at the mid-line and the venous pole (atrium) situated at the anterior left region of the embryo^{262,263}. At the arterial pole, a population of cells with similar characteristics to the secondary heart field, which are present in mammals, contribute with the addition of new cardiomyocytes to the cardiac tube²⁵⁹. At 36 hpf, cardiac looping begins. Through this process, the ventricle is displaced towards the mid-line and the heart tube forms an inverted S-shaped loop²⁶⁴. After this, the epicardium layer is formed by the re-localization of extra-cardiac pro-epicardial cells, which cover the myocardium²⁶⁵. A graphic overview of zebrafish heart development is shown in Fig. 1.7.

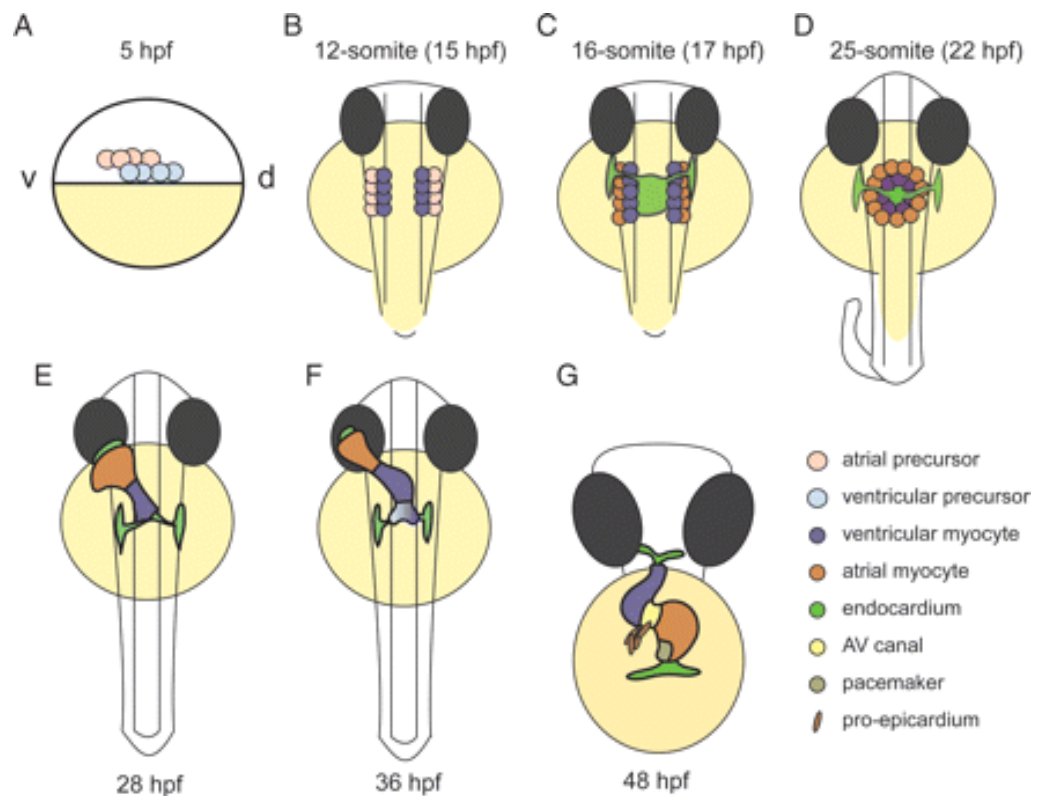


Figure 1.8. Stages of zebrafish cardiac development. (A) At 5 hours post fertilization (hpf), cardiac progenitor cells (atrial and ventricular precursors) are located bilaterally in the lateral marginal zone (pink and light blue). (B) By the 12-somite stage (15 hpf), these populations have already migrated towards the mid-line, sitting in a bilateral fashion. (C) At 17 hpf, endocardial precursors (green) situated anterior to the atrium and ventricular precursors starts migrating towards the mid-line, in between these two bilateral populations. Migration of the already differentiated cardiac cells (orange and purple) occurs slightly later. (D) By 22 hpf, the bilateral heart fields have already fused in the mid-line, forming a cardiac disc structure with the endocardial cells located in the centre. (E) Cardiac morphogenesis transforms the cardiac disc into a cardiac tube (28 hpf). (F) At 36 hpf, new cardiomyocytes are added to the arterial pole. Cardiac looping starts taking place as well, a process that is mostly completed at 48 hpf (G). At this stage, the atrioventricular canal is also visible, and the epicardium precursors start covering the myocardium. *Adapted from Bakkers, 2011*²⁶⁶

1.5.2.2. Advantages and disadvantages of using zebrafish as animal model

The zebrafish has emerged as one of the best options for the study of cardiovascular disease, as it offers several advantages as a genetic and embryonic model system. Zebrafish embryos are transparent, thereby facilitating the visualization of developmental processes (especially heart development, as it is one of the organs more easily visualized) and the use of fluorescent markers to tag specific cell types and visualize their location during embryogenesis. They also have a relatively fast developmental time, reaching the larval stage with most of the major organs fully developed 3 days post-fertilization. The zebrafish is one of the vertebrates models more closely

related to humans, with 70% of zebrafish genes having human counterparts²⁶⁷, and they are particularly well-suited to simulate human cardiac disorders as 82% of the disease-related genes listed in the Online Mendelian Inheritance in Man (OMIM) database have at least one zebrafish orthologue²⁶⁷. Another advantage is that zebrafish embryos do not completely depend on a functional cardiovascular system for their development. Thanks to their small embryonic size, oxygen can still enter embryos lacking blood circulation and reach all tissues by passive diffusion. This feature permits the embryos to overcome the initial phase of embryonic development and allows for the study of embryos with severe cardiovascular defects^{266,268,269}. Also, the zebrafish embryonic heart has a beating rate of 140-180 beats per minute, more in line with the human heart rate than other organisms like mouse (between 300-600 beats per minute)²⁷⁰; and the zebrafish conduction system is homologous to the human one²⁷¹. Last but not least, the zebrafish heart has unique regenerative capabilities, as it is able to regenerate myocardial tissue upon injury²⁷². This has made of the zebrafish an ideal model in order to identify novel therapeutic targets of interest for inducing heart regeneration in humans.

The zebrafish also has its disadvantages as an animal model. First and foremost, the main limitation of using zebrafish for the study of cardiogenesis resides in the morphological and anatomical differences with the human heart. As explained above, the zebrafish has a two-chambered heart instead of a four-chambered heart like mammals, and some cardiac developmental processes like cardiac septation do not take place^{273,274}. Another important thing to take into account is that the zebrafish has a duplicated genome, with many genes having two copies instead one²⁶⁷. This poses a real problem for genetic manipulation studies, as sometimes, it is necessary to target both of these copies in order to produce a complete ablation of its function. Finally, and even though the zebrafish is one of the most popular choices of animal models, there is still a limited number of biological reagents (e.g. antibodies) for this organism²⁷⁵.

1.5.2.3. Genetic modification in zebrafish

Due to the ease of delivery of morpholinos (MOs), RNA or DNA constructs via microinjection, genetic manipulation in zebrafish is pretty straightforward and easy to achieve, thereby making them an ideal animal model for reverse genetics approaches. During recent years, a number of tools have been developed to disrupt specific genes of interest or alter gene expression, including morpholinos²⁷⁶, TALEN²⁷⁷ and CRISPR/Cas9²⁷⁸.

Although not a genetic manipulation tool *per se*, morpholinos have been the standard knockdown tool used by the zebrafish community until recently.

Morpholinos are antisense oligos that bind to a target mRNA and thus, they prevent translation of the mRNA by steric blocking²⁷⁹. Shortly after their development, morpholinos were quickly and broadly adopted by the zebrafish research community to disrupt gene function^{276,280} due to their easiness of design and delivery and their ability to phenocopy already characterized mutants^{281,282}. However, a series of recent reports indicate that morpholinos have greater off-target effects than previously thought²⁸³. This, together with the discovery and exponential growth in use of genome-editing technologies like CRISPR/Cas9 has led to a general feeling of scepticism towards the use of morpholinos in the community, as it has been proven that MO-induced phenotypes can be different to those of the corresponding genetic mutants^{280,283}. In spite of all of this, morpholinos are still a valuable tool in gene function studies: MOs can quickly provide initial information and preliminary data about the function of a gene of interest, and when the MO-induced phenotype is similar to the mutant one, then the morphant can be considered an acceptable alternative for follow-up studies and certain experiments that cannot be performed using genome-editing technologies (e.g. generating a partial loss-of-function model using low doses of morpholino). In addition to this, recent evidence has emerged suggesting that the differences seen in morphant vs mutant phenotypes might be due to a genetic compensation effect induced by deleterious mutations but not gene knockdowns^{284,285}, thus providing a new potential explanation for these differences other than the morpholino-induced off-target effects. All in all, morpholinos are not a perfect tool, but when the correct experimental guidelines are followed^{280,286}, they can still provide useful information in gene function studies and serve as a valuable alternative to mutants. Taking into account the potential off-target effects coming from morpholinos and the genetic compensatory mechanisms that can be triggered by using genome-editing technologies, a properly designed reverse genetics study should always include the generation of both ZFN/TALEN/CRISPR-Cas9-mediated mutants and morpholino-mediated knockdowns²⁸⁰.

Even though morpholinos have been extensively used to analyse gene and protein function in zebrafish, this approach knocks down a specific gene but just for a relatively short period of time. In order to generate stable transgenic lines, most of the genome-editing technologies developed in the last decade have been successfully adapted to zebrafish as well, including ZFN²⁸⁷, TALEN²⁸⁸ and CRISPR/Cas9²⁸⁹. Thanks to this, the zebrafish is extensively used nowadays for reverse genetics studies in order to identify new genes involved in developmental processes (including cardiogenesis) or for the generation of stable transgenic lines of genes involved in development and disease. For

example, by using CRISPR/Cas9 ribocomplexes, Burger *et al.* generated *tbx5a* crispants²⁹⁰ whose cardiac and limb phenotype recapitulate that of the previously described *tbx5a* mutant heartstrings (*hst*)¹²⁶ and morpholino-mediated *tbx5a* knockdown¹²⁶. The phenotype of these mutants also recapitulates some of the phenotypic features found in Holt-Oram syndrome patients²⁹¹. Recently, and by using the CRISPR/Cas9 system, Tessadori *et al.* demonstrated the introduction of human cardiovascular disorder-causing mutations in the zebrafish orthologs of the *KCNJ8*, *SUR2* and *ABCC9* genes, which in humans are linked to Cantú syndrome (a syndromic-CHD condition)²⁹². These knockout lines recapitulated the phenotypic features found in human patients, including heart defects among others²⁹².

1.6. Hypothesis and aims

Although our knowledge of the molecular mechanisms underlying CHD and heart development have been greatly extended in recent years, we are still far from fully understanding them. Every year, novel factors regulating cardiac development and involved in CHD are discovered, but in order to have a clear picture of the complex network of genes and factors during cardiogenesis, we still need an even greater global effort.

In this PhD thesis, my main goal is to contribute to this field by discovering and characterizing novel factors implicated in cardiac development and linked to the onset of CHD. Throughout this thesis, I have worked on three different projects encompassed in the field of genetic studies of congenital heart disorders:

- The discovery and study of novel upstream regulators of TBX5, one of the key regulators of cardiac development and the main gene involved in Holt-Oram syndrome
- The *in vitro* characterization of novel CHD4 *de novo* mutations found in patients of syndromic-CHD
- The maintenance, setup and basic characterization of CHD4- and CDK13-null mouse models, which will serve as the basis for the characterization of congenital heart defects using a novel 3D imaging technique (high-resolution episcopic microscopy)

Further details about the background, hypotheses and aims of each one of these projects can be found in their respective chapters.

2. MATERIALS AND METHODS

2.1. Basic techniques

2.1.1. DNA and RNA manipulation techniques

2.1.1.1. Polymerase chain reaction

Several polymerases and PCR systems were used in this project depending on the application. For routine PCR amplification and genetic screening, regular NEB Taq polymerase was used (New England Biolabs, USA). For long-fragment amplification, screening of mutations or DNA amplification for sequencing, Phusion High-Fidelity DNA polymerase was used (New England Biolabs, USA). Finally, for mice genotyping I used Platinum Taq DNA Polymerase High Fidelity (ThermoFisher Scientific, USA) according to the protocol provided by the Wellcome Sanger Institute. Amplification reactions were carried out in 0.2-ml PCR tubes using a T100 Thermal-Cycler (Bio-Rad, USA) and using the reaction components and reagents indicated in Table 2.1. Cycling conditions are summarised in Table 2.2.

Table 2.1. PCR reaction volumes for Taq, Phusion and Platinum polymerases

	Taq polymerase (25 / 50 µl reaction)	Phusion polymerase (20 / 50 µl reaction)	Platinum polymerase (25 / 50 µl reaction)
Buffer	2.5/5 µl (10x)	4/10 µl (5x)	2.5/5 µl (10x)
10 mM dNTPs	0.5/1 µl	0.4/1 µl	0.5/1 µl
10 µM Forward Primer	0.5/1 µl	1/2.5 µl	0.5/1 µl
10 µM Reverse Primer	0.5/1 µl	1/2.5 µl	0.5/1 µl
Template DNA	Genomic DNA – 1 ng-1 µg Plasmid DNA – 1 pg – 1 ng	Genomic DNA – 50-250 ng Plasmid DNA – 1 pg – 10 ng	Genomic DNA – 1 ng-1 µg Plasmid DNA – 1 pg – 1 ng
Polymerase	0.125/0.25 µl	0.2/0.5 µl	0.1/0.2 µl
50 mM MgCl₂ or MgSO₄	-	-	1/2 µl
DMSO (optional)	-	0.6/1.5 µl	-
Nuclease-free H₂O	To 25/50 µl	To 20/50 µl	To 20/50 µl

Table 2.2. PCR cycling conditions for Taq, Phusion and Platinum polymerases

		Taq polymerase	Phusion polymerase	Platinum polymerase
Initial denaturation		94°C for 30 sec-2 min	98°C for 30 sec	94°C for 30 sec-2 min
<i>25-35 PCR cycles</i>	Denaturation	94°C for 15 sec	98°C for 10 sec	94°C for 15 sec
	Annealing	45-68°C for 30 sec	45-72°C for 30 sec	45-68°C for 30 sec
	Extension	68°C for 1 min/kb	72°C for 30 sec/kb	68°C for 1 min/kb
Final extension		68°C for 5 min	72°C for 10 min	68°C for 5 min
Hold		4°C indefinitely	4°C indefinitely	4°C indefinitely

2.1.1.2. Agarose gel electrophoresis

Agarose gels were made by dissolving the required amount of agarose powder (Sigma-Aldrich, USA) in 1x TAE buffer (40 mM Tris, 20 mM acetic acid, and 1 mM EDTA) and subsequently boiling the mixture in a microwave. Typically, 1.5% agarose gels (0.75 g of agarose in 50 ml 1xTAE buffer) were used, but in some instances, 2% agarose gels (1 g of agarose in 50 ml 2xTAE buffer) were also used in order to promote a better separation of DNA fragments. After cooling down, 2 µl of Ethidium Bromide 10 mg/mL solution (Sigma-Aldrich, USA) was added to the dissolved agarose, the mixture was then poured inside a gel cassette and left to solidify for 45 min. To facilitate the visualization of samples and the determination of fragment sizes, all samples were mixed with 10x DNA Loading Buffer (30% (v/v) glycerol, 0.25% (w/v) bromophenol blue and 0.25% (w/v) xylene cyanol FF) before loading and electrophoresis alongside a 100 bp or 1 kb DNA ladder. After loading, samples were typically run at 80-100V for 30 min-1 hour, or until the dye front is close to the bottom end of the gel. Following electrophoresis, DNA was visualized by exposing the gel to UV light inside a Gel Doc XR+ System (Bio-Rad, USA). Images were taken using the Quantity One software (Bio-Rad, USA).

2.1.1.3. DNA gel purification

In order to purify DNA from agarose gels, the gels were exposed to UV light and DNA-containing fragments excised from the gel using a scalpel. DNA was subsequently extracted from these fragments by using a DNA Gel Extraction Kit (QIAGEN, Germany). Agarose gel slices were weighted, and 3 volumes of Buffer QG was added to 1 volume gel (100 mg gel~100 µl). Gel slices were then incubated at 50°C until the gel slice had completely dissolved (~10 min typically), with brief vortexing every 2 min to help dissolve the gel. One gel volume of isopropanol was then added to the sample, followed by brief mixing. After this, the sample was applied to a QIAquick column (which was previously placed in a 2 ml collection tube) and centrifuged for 1 min to allow the sample to pass through the column. The flow-through was discarded, and 750 µl Buffer PE was added to the QIAquick column in order to wash the DNA. The column was centrifuged for 1 min and the flow-through was discarded again. The column was centrifuged again for 1 min to remove any residual wash buffer, and subsequently placed into a clean 1.5 ml microcentrifuge tube. To elute DNA, 25 µl of water was added to the center of the QIAquick membrane and the column was subsequently centrifuged for 1 min. Finally, DNA Concentration and purity was finally measured on a NanoDrop 2000 spectrophotometer (ThermoFisher Scientific, USA).

2.1.1.4. Restriction enzyme digestion

All restriction enzyme digestions in this project were performed with NEB restriction enzymes (NEB, UK), following the recommendations and protocol provided by the manufacturer. Typically, a 20 µl (screening purposes) or a 50 µl (cloning purposes) digestion reaction was set-up as indicated in Table 2.3. For screening purposes, reactions were usually incubated at 37°C for 1 hour in a heat block; whereas digestions for cloning purposes were incubated at 37°C for 4 hours to overnight to ensure that the template DNA was fully cleaved. Reactions were then run on an agarose gels as indicated in 2.1.1.2. for examination and to ensure that DNA digestion has taken place.

Table 2.3. Restriction digestion volumes

	20 µl reaction	50 µl reaction
Enzyme buffer (10x)	2 µl	5 µl
Restriction enzyme	1 µl	2.5 µl
Template	200-300 ng	3-4 µg
Nuclease-free H₂O	To 20 µl	To 50 µl

2.1.1.5. Ethanol precipitation of DNA

In order to purify and/or concentrate DNA, ethanol precipitation was routinely used²⁹³. 1/10 volume of 3M sodium acetate pH 5.2 and 3 volumes of 100% ethanol was added to the DNA sample, which was subsequently incubated at -80°C for 4 hours to overnight. Following incubation, DNA was precipitated by centrifugation at full speed in an Eppendorf Centrifuge 5415R (Eppendorf, Germany) at 4°C for 30 min. After centrifugation, DNA pellet was usually visible. Supernatant was discarded and 500 µl of 70% ethanol was added to the sample to wash the pellet. Following a second centrifugation, the supernatant was discarded and the pellet was left to dry at RT for 30 min. The DNA pellet was finally resuspended in 20 to 50 µl of H₂O (depending on the desired DNA concentration) and the concentration and purity was measured on a NanoDrop 2000 spectrophotometer (ThermoFisher Scientific, USA).

2.1.1.6. DNA ligation

Ligation of DNA fragments for cloning purposes was performed using T4 Ligase (Promega, USA). The amount of insert and vector needed to have a 1:1 or 1:3 ratio (these were the two insert:vector ratios typically used) was initially calculated by using a Promega webtool (<https://www.promega.com/a/apps/biomath/?calc=ratio>). Once calculated, both the insert and vector were incubated along with T4 Ligase and buffer as indicated in Table 2.4. Ligation reactions were usually incubated at RT for 1 hour; however, in the case a large insert was used, they were performed at 4°C overnight. Proper ligation was checked via agarose gel electrophoresis or via Sanger sequencing.

Table 2.4. Ligation reaction volumes

	10 µl reaction
Ligase buffer (10x)	1 µl
T4 Ligase	1 µl
Insert	10-50 ng (depends on length and ratio)
Vector	50 ng
Nuclease-free H₂O	To 10 µl

2.1.1.7. Plasmid DNA transformation

Following DNA ligation, these reactions were transformed into chemically competent bacteria (*E.coli DH5α*) using the heat shock method²⁹⁴. 50 µl-aliquots of competent cells were taken out of -80°C and thawed on ice for 10 min. After this, ligation reactions were mixed with 50 µl of competent cells and incubated on ice for 30 min. Competent cell/DNA mixture were then subjected to heat shock by placing it at 42°C for 1 min, and put back on ice for 2 min. 1 ml

of LB media was then added to the mixture and cells were allowed to recover for 1 hour at 37°C with agitation (200 rpm). Following recovery, cells were spun down at 300 rpm for 2 min on a microcentrifuge. Media was discarded and the cell pellet was then resuspended in 100 µl of fresh LB media. Finally, 100 µl of transformed cells were spread on LB-agar plates containing the appropriate antibiotic for selection and incubated at 37°C overnight to allow for colony formation. All bacterial-related procedures were performed next to a lit Bunsen burner used to create a sterile field.

2.1.1.8. Plasmid DNA isolation

Isolation of plasmid DNA transformed into chemically competent cells was achieved by using the GenElute Plasmid Miniprep Kit (Sigma-Aldrich, USA) for small-volume cultures (mini-prep) or the EndoFree Plasmid Maxi Kit (QIAGEN, Germany) for large-volume cultures (maxi-prep). First, cultures of clones containing the plasmid of interest were prepared by picking (using sterile toothpicks and next to a lit Bunsen burner) and inoculating transformed bacterial colonies into 5 ml of LB media with antibiotics. These cultures were grown overnight at 37°C with agitation (200 rpm). For mini-preps, the next morning 2 ml of each culture were spun down at 12000 rpm for 1 minute. Supernatant was then discarded and the pellet was resuspended in 200 µl of Resuspension Solution by pipetting up and down. Cells were then lysed by adding 200 µl of the Lysis Solution, followed by mixing by gentle inversion and a 5-min incubation. Cell debris were precipitated by adding 350 µl of the Neutralization/Binding solution and mixing by inversion. Then, these cell debris were spun down by centrifugation at 12000 rpm for 10 min. A GenElute Miniprep Binding Column was then inserted into a microcentrifuge tube, and 500 µl of the Column Preparation Solution was added to the column, which was then centrifuged at 12000 rpm for 1 min. Following discard of the flow-through, the cleared lysate obtained after the precipitation of cell debris was transferred to the column and centrifuged at 12000 rpm for 1 min. The flow-through was discarded and 750 µl of Wash Solution was added to the column, followed by centrifugation at 12000 rpm for 1 min (the flow-through was again discarded). Finally, the column was transferred to a new microcentrifuge tube and the plasmid DNA was eluted by adding 50 µl of nuclease-free water followed by centrifugation at 12000 rpm for 1 min. DNA concentration and purity was measured on a NanoDrop 2000 spectrophotometer (ThermoFisher Scientific, USA).

For maxi-preps, 1-2 ml of the initial 5 ml culture were inoculated in 100 ml of LB media with antibiotics and incubated overnight at 37°C with agitation (200 rpm). The next morning, 100 ml cultures were spun down at 4500 rpm for 20 min at 4°C, and the bacterial pellet was resuspended in 10 ml of Buffer P1 by

pipetting up and down. 10 ml of Buffer P2 was then added, followed by thoroughly mixing and a 5-min incubation at RT. After this, 10 ml of chilled Buffer P3 was added to the tube, followed by mixing by inversion (4-6 times). This lysate was then poured into the barrel of a QIAfilter Cartridge (included in the kit) and let to settle down for 10 min at RT. Following this incubation period, a plunger was inserted into the Cartridge and the cell lysate was filtered into a 50 ml Falcon tube. Once the lysate was filtered, 2.5 ml of Buffer ER was added and the tube was incubated on ice for 30 min. In the meantime, a QIAGEN-tip 500 (included in the kit) was equilibrated by applying 10 ml of Buffer QBT and by allowing the column to empty by gravity flow. After this, the filtered lysate was poured into the QIAGEN-tip and allowed to enter the tip. Then, it was washed twice with 30 ml of Buffer QC, and the plasmid DNA was eluted by adding 15 ml of Buffer QN, which was recovered in a 50 ml Falcon tube. In order to precipitate the DNA, 10.5 ml of isopropanol were added to the eluted DNA, followed by mixing and centrifugation at 5000 rpm for 1 h at 4°C. Finally, the supernatant was carefully discarded, and the air-dried DNA pellet (5-10 min) was resuspended in 300 µl of nuclease-free water. DNA concentration and purity was measured on a NanoDrop 2000 spectrophotometer (ThermoFisher Scientific, USA).

2.1.1.9. Sanger sequencing

DNA samples were typically sent to Source Bioscience (UK) for sequencing following the requirements provided by the company: 5-µl samples at a concentration of 100 ng/µl. 5-µl samples of sequencing primers at a concentration of 3.2 pmol/µl were also provided. Once received, sequencing data was analysed by using the Chromas software (Technelysium, Australia).

2.1.1.10. Total RNA extraction from zebrafish embryos

Total RNA was isolated from pooled zebrafish embryos using the TRI Reagent (Sigma-Aldrich, USA)²⁹⁵. Embryos pooled in microcentrifuge tubes (20 to 40) were first homogenized in 250 µl of TRI Reagent and then 750 µl of TRI Reagent was added to produce a final volume of 1 ml. Following a 5 min incubation at RT, homogenized embryos were centrifuged at 12000 g for 10 min at 4°C. Supernatant was then transferred to a fresh microcentrifuge tube and 100 µl of bromochloropropane (BCP) was added. After a 15 min incubation, tubes were centrifuged again at 12000 g for 10 min at 4°C leading to separation of phases inside the tube. The aqueous phase (top) was transferred to a new tube and after addition of 500 µl of isopropanol, vortexing, and incubation for 10 min tubes were subjected to centrifugation at 12000 g for 10 min at 4°C. Supernatant was discarded and 1 ml of 75% EtOH was added to each tube to wash the RNA pellet. Tubes were centrifuged at 7500 g for 5 min at 4°C and after discarding the ethanol, they were let to dry at RT for 20 to 30 min. Finally,

the RNA pellet was resuspended in 30 µl of nuclease-free H₂O and stored at -80°C.

2.1.1.11. RNA quality assessment using the Bioanalyzer

In order to assess the concentration, quality and integrity of total RNA or sgRNAs prior to use in downstream application, samples were analysed using a 2100 Bioanalyzer (Agilent Technologies, USA). Two different kits were used according to the source material: for total RNA, the Agilent 6000 Nano Kit was used (Agilent Technologies, USA) whereas for the analysis of sgRNAs the Agilent 6000 Small RNA Kit was employed (Agilent Technologies, USA). In both cases samples were analysed following the protocol provided by the manufacturer.

2.1.1.12. Reverse transcription and cDNA synthesis

First strand cDNA was synthesized from total RNA by using M-MLV Reverse Transcriptase (Promega, USA). To achieve this, 1 µg of total RNA and 1 µl of random primers in a total volume of 14 µl (nuclease-free H₂O) were mixed in a microcentrifuge tube and subsequently incubated at 70°C for 5 min in order to melt secondary structures within the template. Tubes were then put back on ice immediately in order to prevent secondary structure from reforming and to promote primer annealing to template. After this, 1 µl of M-MLV RetroTranscriptase (Promega, USA), 1 µl of RNAsin inhibitor (Promega, USA) and 4 µl of 5x ReT buffer (Promega, USA) was added to the tube (final volume of 20 µl) and after mixing, ReT reactions were vortexed and incubated for 60 min at 37°C. Typically, a RetroTranscriptase negative control was also prepared. Reactions were then further incubated for 20 min at 65°C to deactivate the retrotranscriptase and 80 µl of nuclease-free H₂O to bring up the final volume to 100 µl. Finally, cDNA was stored at -20°C.

2.1.1.13. Relative quantitative PCR

In order to measure the expression level of different genes in zebrafish embryos, relative quantitative PCR was performed on cDNA generated from 32 hpf zebrafish embryos. Initially, primers chosen for qPCR amplification were tested by performing a qPCR standard curve using different template DNA dilutions. To achieve this, eight 1:3 serial dilutions of cDNA template were prepared in H₂O. Then, a master mix was prepared by mixing 10 µl of SYBR Green 2X (Bio-Rad, USA), 0.5 µl of forward primer, 0.5 µl of reverse primer and 7 µl of nuclease-free water. On a 96-well plate (ThermoFisher Scientific, USA), 18 µl of master mix were mixed with 2 µl of each cDNA dilution on individual wells, performing three technical replicates per dilution. After all reactions were set up, the plate was sealed with a MicroAmp Optical Adhesive Film (ThermoFisher Scientific, USA) and the qPCR reaction was performed on a 7500 Real-Time PCR system (Applied Biosystems, USA) using the following cycling

conditions: 95°C for 10mins, [95°C for 15 seconds, 62°C for 30 seconds, 72°C for 40 seconds] (40 cycles), 95°C for 60 seconds. Finally, for each pair of primers, Ct values were plotted against the logarithm of the cDNA concentration used on Excel, and a standard curve was generated. The efficiency of the reaction was calculated by using the equation $E = -1 + 10^{(-1/\text{slope})}$, whereas the R^2 value was automatically calculated using Excel. Typically, desired amplification efficiencies range from 90% to 110% efficiency and a R^2 value ≥ 0.99 . Once primers were validated, qPCR analysis of actual samples was carried out by preparing the setting up the same reactions but substituting 2 μ l of cDNA dilution for 2 μ l of the actual sample. Typically, three biological replicates were used in each qPCR experiment, and three technical replicates were set up for each biological one. A retrotranscriptase negative control and a nuclease-free water negative control were also included. Cycling conditions were identical to the ones used for the standard curve generation. Finally, analysis of gene expression was performed by using the $2^{-\Delta\Delta CT}$ method²⁹⁶ in Excel.

2.1.2. Protein manipulation techniques

2.1.2.1. Protein extraction from cultured cells

In order to perform protein extraction on cultured cells²⁹⁷, cells grown on 100mm cell culture dishes (ThermoFisher Scientific, USA) were first scrapped in 1 ml PBS (ThermoFisher Scientific, USA) using the bottom of a 1 ml pipette tip. Cells were then collected in 1.5 ml Eppendorf tubes and spun down by centrifugation at 5000 rpm for 5 min at 4°C. PBS was then removed and the cell pellet was resuspended in 250 μ l of cell lysis buffer (150 mM NaCl, 20 mM TrisHCl pH 7.6, 0.5 mM EDTA pH 8.0 and 1% NP-40) with 2.5 μ l of 100x proteinase inhibitor cocktail (Cell Signaling Technology, USA) to avoid protein degradation. Cells were then left under ice for 3 hours (with vortexing every 20 min) to allow for efficient lysis to occur. Finally, cell debris was removed by centrifugation at 12000 rpm for 10 min at 4°C and the clear supernatant containing proteins was stored at -80°C until further use.

2.1.2.2. Determination of protein concentration

Concentration of protein samples was determined by using the DC Protein Assay (Bio-Rad, USA), which is based on the well-documented Lowry assay²⁹⁸. First, 20 μ l of reagent S was added to each ml of reagent A needed for the run to generate reagent A'. Then, six dilutions of a Bovine Serum Albumin (BSA; Sigma-Aldrich, USA) standard containing 0.125, 0.250, 0.5, 1, 1.5 and 2 mg/ml of protein were prepared by diluting BSA in protein buffer in order to generate a standard curve, which should be prepared each time the assay is performed.

Following this, 5 µl of both standards and samples was added to individual wells of a 96-well plate, and 25 µl of reagent A' and 200 µl of reagent B was added into each well. The plate was then gently agitated to mix the reagents and incubated for 15 min at RT. After this, absorbances were read at 750 nm using a microplate absorbance reader and samples concentration were calculated from the standard curve (which was generated by plotting the absorbance of each standard against the concentration).

2.1.2.3. Western blot

To detect the presence of specific proteins in whole cell lysates, western blot was routinely performed²⁹⁹. First, 20 µg of protein samples were mixed with 1/4 volume of NuPAGE LDS Sample Buffer 4X (ThermoFisher Scientific, USA) and 1/10 volume of NuPAGE Sample Reducing Reagent (ThermoFisher Scientific, USA) in a total volume of 15 µl and they were incubated for 5 min at 95°C to denature proteins. After this, they were loaded on a NuPAGE Novex 10% Bis-Tris Protein Gel (ThermoFisher Scientific, USA) alongside 10 µl of SeeBlue Plus2 Pre-stained Protein Standard (ThermoFisher Scientific, USA) and 5 µl of MagicMark™ XP Western Protein Standard (ThermoFisher Scientific, USA). Gels were then placed inside a XCell SureLock Mini-Cell (ThermoFisher Scientific, USA) filled with 800 ml of running buffer (prepared by mixing 40 ml of NuPAGE MOPS SDS Running Buffer 20X (ThermoFisher Scientific, USA) and 760 ml of H₂O) and they were run using a PowerEase 500 Power Supply (ThermoFisher Scientific, USA) at 200 V for 1 hour. Once the run was finished, proteins in the gel were transferred to an Invitrolon™ PVDF membrane by preparing a “sandwich” as indicated in Figure 2.1:

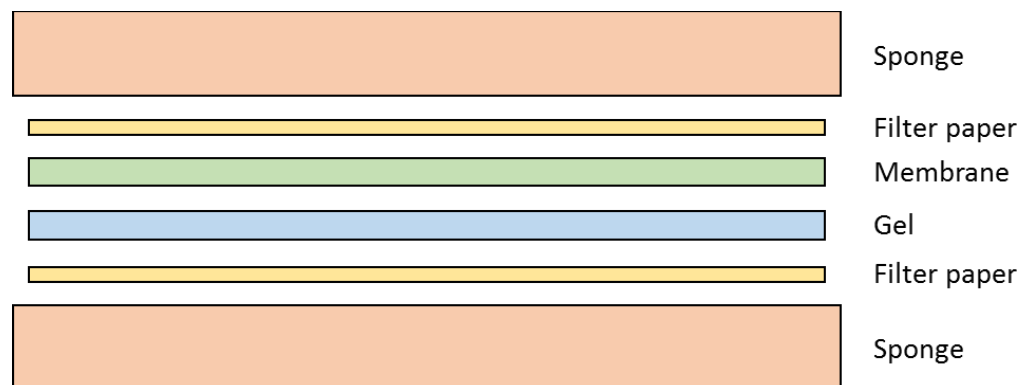


Figure 2.1. Drawing showing the structure of a western blot “sandwich” prior to transfer.

Prior to “sandwich” preparation, the PVDF membrane was incubated for 5 min with gentle agitation (40 rpm) in 100% MeOH, and both the sponges and filter papers were soaked in transfer buffer (prepared by mixing 40 ml of NuPAGE Transfer Buffer 20X (ThermoFisher Scientific, USA), 80 ml of 100% MeOH and 680 ml of H₂O). The sandwich was then placed in a XCell II Blot Module

(ThermoFisher Scientific, USA) filled with 800 ml of transfer buffer and the transfer was performed at 30 V for 1 hour. After transfer, the membrane was recovered, placed on a small tray with the protein-containing side looking upwards and washed twice with H₂O for 5 min, followed by blocking with 5% non-fat dry milk (Santa Cruz Biotechnology, USA) dissolved in TBS-T (20 mM Tris, 150 mM NaCl and 1% Tween-20) for 2 hours at room temperature. Membrane was then washed 3 times with TBS-T for 5 min and it was incubated overnight at 4°C in the presence of primary antibody diluted in TBS-T containing 5% BSA (primary antibody dilutions used in this thesis are indicated below). The next day, the membrane was washed again 3 times with TBS-T for 10 min prior to incubation for 1 hour at RT in the presence of secondary antibody diluted in TBS-T containing 5% BSA (secondary antibody dilutions used in this thesis are indicated below). After secondary antibody incubation, the membrane was washed 3 times in TBS-T for 10 minutes and bands were developed by incubating it with ECL Blotting Substrate (Bio-Rad). This substrate was prepared by mixing the reagents A and B in a 1:1 ratio, and 1.5 ml of the reagents A+B mix was then applied to the membrane, so it covered its whole surface. After a 5-minute incubation, the developing reagent was removed and bands were detected using a LAS-3000 Imaging System (Fujifilm, Japan) and analysed using the software AIDA Image Analyzer (Elysia-Raytest, Belgium). A list of the antibodies used in this thesis can be found in the following table:

Table 2.5. List of primary and secondary antibodies used in this thesis

Antibody	Company	SKU	Dilution used
Monoclonal ANTI-FLAG® M2 antibody produced in mouse	Sigma-Aldrich	F1804	1:1000 for WB, 1:500 for IF
ANTI-FLAG® antibody produced in rabbit	Sigma-Aldrich	F7425	1:500 for IF
Monoclonal Anti-HA antibody produced in mouse	Sigma-Aldrich	H9658	1:1000 for WB
Monoclonal Anti-c-Myc antibody produced in mouse	Sigma-Aldrich	M4439	1:1000 for WB
Mouse anti-human CRM1 Antibody (C-1)	Santa Cruz Biotechnology	sc-74454	1:250 for IF
Mouse monoclonal Anti-CHD4 antibody [3F2/4]	Abcam	ab70469	1:1000 for WB
Anti-CDK13 antibody produced in rabbit	Sigma-Aldrich	SAB2700810	1:1000 for WB

Anti-Mouse IgG (Fab specific)–Peroxidase antibody produced in goat	Sigma-Aldrich	A9917	1:10000 for WB
Anti-Rabbit IgG (whole molecule)–Peroxidase antibody produced in goat	Sigma-Aldrich	A0545	1:10000 for WB
Goat anti-mouse Alexa-488	Invitrogen	A-11001	1:5000 for IF
Goat anti-rabbit Alexa-594	Invitrogen	A-11005	1:5000 for IF

* WB = western blot, IF = immunofluorescence

2.1.2.4. Coomassie staining of proteins

Coomassie staining of proteins was routinely performed by using the SimplyBlue SafeStain reagent (ThermoFisher, USA). Following electrophoresis of proteins, the gel was placed in a tray and was rinsed 3 times for 5 min with 100 mL of water in order to remove SDS and buffer salts. The gel was then stained by adding enough SimplyBlue SafeStain reagent to cover the gel, and incubated for 1 hour at RT with gentle shaking (50 rpm). Following incubation, the stain was discarded and the gel was washed with 100 mL of water for 2 hours. Finally, a second wash with water for 1 hour was performed to obtain the clearest background for photography using a Gel Doc XR+ System.

2.1.3. Cell culture

2.1.3.1. Cell lines

Three different mammalian cell lines were used during this project: COS7, C2C12 and Sf9 cells. For pull-down assays, acetylation assays and immunofluorescence studies COS7 cells were primarily used, which are fibroblast-like cells derived from monkey kidney tissue and suitable for transfection by vectors requiring expression of the SV40 T antigen³⁰⁰. The C2C12 cell line was also used for immunofluorescence studies, which is an immortalized mouse myoblast cell line³⁰¹. Finally, for recombinant protein production using baculovirus the Sf9 cell line was employed, which is an insect cell line derived from ovaries of the *Spodoptera frugiperda* moth species³⁰². Both COS7 and C2C12 cell lines are adherent lines which were typically grown and maintained in Nunc T75 cm² flasks (ThermoFisher Scientific, USA) at 37°C in the presence of 5% CO₂, whereas the Sf9 cell line was maintained either adherent (T25 cm² flasks) or in suspension in flat-bottom glass flasks (with

agitation at 150 rpm) at 28°C. All cell culture-related work was carried out inside a laminar flow hood in sterile conditions.

2.1.3.2. Media and solutions

For the maintenance of COS7 and C2C12 cells, Dulbecco's Modified Eagle Medium (DMEM; ThermoFisher Scientific, USA) supplemented with 4.5 g/L D-Glucose, L-Glutamine, 10% Fetal Bovine Serum (Sigma-Aldrich, USA) and 5% Penicillin-Streptomycin solution (Sigma-Aldrich, USA) was used. For T75 flasks, cells were maintained in 20 ml of media, which was typically changed every 3-5 days. For the growth of Sf9 cells, Insect-XPRESS Protein-free Insect Cell Medium (Lonza, Switzerland) supplemented with L-glutamine, 10% Fetal Bovine Serum and 5% Penicillin-Streptomycin solution was employed. For Sf9 cells maintenance in adherent form, 10 ml of media was added to a T25 flask. Other solutions typically used for the maintenance of cells were Dulbecco's Phosphate Buffered Saline (ThermoFisher Scientific, USA) and 0.05% Trypsin-EDTA 1X solution (ThermoFisher Scientific, USA). All media and solutions used in cell culture were pre-warmed in a 37°C water bath for 10-15 min before any procedure with cells was carried out.

2.1.3.3. Passaging of cells

Typically, all three cell lines used in this project were sub-cultured after reaching 80% confluency, which was daily checked under the microscope. For adherent cells, media was first removed and cells were then washed with 5 ml of PBS. After this, 1.5 ml of 0.05% Trypsin-EDTA 1X solution was added to the cells and they were incubated for 5 min at 37°C in the presence of 5% CO₂ to let the cells detach from the surface of the flask. Disassociation of cells was confirmed under the microscope, and they were then homogenously dissolved in 10 ml of media by pipetting up and down. Finally, between 1.5-2 ml of cell suspension was added to a new T75 cm² and media was added to a final volume of 20 ml. In the case of Sf9 cells in suspension, cells were typically diluted to a density of 3-5 x 10⁵ cells/ml once they reached a density of 4-5 x 10⁶ cells/ml.

2.1.3.4. Cryopreservation of cells and storage

In order to cryopreserve cells for future use, DMSO was used as a cryopreserving agent. Briefly, cells were trypsinized and detached from the surface of the flask as indicated in 2.1.3.3. After this, cells were centrifuged for 5 min at 1000 rpm, the supernatant was discarded and the pellet was resuspended in 10 ml of media containing 10% DMSO. The cell suspension was then aliquoted into 1.8 ml Nunc CryoTubes (ThermoFisher Scientific, USA) and tubes were left at -80°C overnight. The next day, cryotubes were transferred to a liquid nitrogen container and the lab inventory was updated.

2.1.3.5. Revival of cells from liquid nitrogen

Cells stored in liquid nitrogen were typically revived by incubating the cryotube in which they were contained at 37°C (in a water bath) for 10 min. After this, 1 ml of pre-warmed media was added drop-wise into the cryotube, and the solution was then transferred to another tube containing 8 ml of pre-warmed media. The cell suspension was then centrifuged for 5 min at 1000 rpm, the supernatant was discarded and the cell pellet was resuspended in 10 ml of pre-warmed media. Cells were centrifuged again for 5 min at 1000 rpm and they were resuspended in 20 ml of pre-warmed media after discarding the supernatant. Finally, cell solution was transferred to a T75 cm² flask and they were allowed to recover at 37°C in the presence of 5% CO₂ for 4-5 days before starting experimental procedures.

2.1.3.6. Transfection of cells

Transfection of plasmid DNA into COS7 cells was achieved by using Polyfect Transfection Reagent (QIAGEN, Germany), whereas C2C12 cells were transfected using Amaxa Cell Line Nucleofector Kit V (Lonza, Switzerland).

COS7 cells (for protein extraction)

For COS7 cells transfection, 1.6 million cells (counted manually using a Neubauer chamber) were plated onto 100 mm plates in a total volume of 8 ml (for each construct to be transfected, two 100 mm plates were prepared). Plates were incubated O/N at 37°C in the presence of 5% CO₂. The next day, 12 µg of plasmid DNA (6 µg per plate, so 12 µg in total as two plates were prepared per construct) were mixed with 600 µl of free-serum media and 50 µl of Polyfect. These reagents were mixed by pipetting up and down and incubated for 15 min at RT. During the incubation, media was removed from the 100 mm plates and they were washed once with 1 ml of PBS, followed by addition of 7 ml of DMEM containing 10% FBS and 5% P/S. Then, 2 ml of this same medium was added to the DNA-polyfect mix, mixed by pipetting up and down and 1.3 ml was added to each 100 mm plate of the two plates prepared per construct. The plates were put back in the incubator and incubated 37°C in the presence of 5% CO₂ for 48 hours. After this, cells were subjected to protein extraction as indicated in 2.1.2.1.

COS7 cells (for immunofluorescence)

For immunofluorescence purposes, the transfection protocol followed was similar to the one used for protein extraction. In this case, 4x10⁵ cells were seeded onto rounded coverslips placed in the wells of a 6-well plate (in 3 ml of DMEM). The next day, 2 µg of plasmid DNA were mixed with 100 µl of free-serum media and 10 µl of Polyfect. During incubation, media was removed and

cells were washed with 0.5 ml of PBS, followed by addition of 1.5 ml of DMEM containing 10% FBS and 5% P/S. Then, 0.6 ml of this same medium was added to the DNA-polyfect mix, mixed by pipetting up and down and the mix was added to each corresponding well. The plates were put back in the incubator and incubated 37°C in the presence of 5% CO₂ for 24 hours. After this, cells were subjected to immunofluorescence analysis as indicated in 2.2.1.5.

C2C12 cells (for immunofluorescence)

Transfection of C2C12 cells was also carried out in 6-well plates. To achieve this, 1×10^6 cells grown originally in flasks were spun down at 90xg for 10 minutes at RT and resuspended in 100 µl of Nucleofector® Solution. This cell suspension was then mixed with 2-2.5 µg of DNA and transferred into one of the cuvettes included in the kit and ensuring that no air bubbles were formed. The cuvette was then closed with the cap and inserted into the cuvette holder of the Nucleofector® I Device. Program B-032 was selected and then applied by pressing the X-button. Following program application, the cuvette was taken out of the holder and 500 µl of DMEM was added to the DNA-cells mix, which was then gently transferred into a the well of a 6-well plate containing 1 ml of DMEM (final volume of 1.5 ml per well). Plates were then placed back in the incubator for 24 hours and cells were subjected to immunofluorescence analysis as indicated in 2.2.1.5.

2.1.3.7. MTT assay

In order to assess cell cytotoxicity following drug exposure, an MTT assay³⁰³ was carried out using a Cell Proliferation Kit I (Roche, USA). 5×10^4 COS7 cells were plated in the wells of a 96-well plate (100 µl total volume), and following drug exposure, media was removed and fresh media was added to each well. Then, 10 µl of the MTT labelling reagent was added to each well, and the plate was incubated at 37°C in the presence of 5% CO₂ for 4 hours. Following this, 100 µl of the Solubilization solution was added into each well, and the plate was allowed to stand O/N in the incubator. Finally, absorbance of each well (OD=590 nm) was measured in duplicates using a microplate absorbance reader.

2.1.4. Bacteriological techniques

2.1.4.1. Preparation of LB agar plates

For the preparation of LB agar plates, 300 ml of LB agar were initially heated up in the microwave until it was completely melted. Melted LB agar was then left to cool down at 55°C until the bottle was able to be picked up without a glove.

At this point, an appropriate amount of antibiotic was added to the LB agar to have a working concentration as indicated in Table 2.6:

Table 2.6. Stock and working concentrations for different antibiotics

	Stock concentration	Working concentration
Ampicilin	100 mg/ml	100 µg/ml
Chloramphenicol	25 mg/ml	25 µg/ml
Gentamycin	10 mg/ml	10 µg/ml
Kanamycin	50 mg/ml	50 µg/ml
Tetracyclin	10 mg/ml	10 µg/ml

LB agar solution was then poured into 100 mm dishes and left to cool at RT until the LB agar was completely solid again. Plates were stored at 4°C until further use for a period of up to 3 weeks.

2.2. ACETYLATION OF TBX5 AND ITS ROLE DURING CARDIAC DEVELOPMENT

2.2.1. In-vitro studies

2.2.1.1. Plasmids and constructs

Several plasmids and constructs generated by Dr. Tushar Ghosh were used for this part of the thesis. A FLAG tagged TBX5 PCR product was generated from plasmid pcDNA-TBX5 described previously³⁰⁴ and cloned into NheI and BamHI of pcDNA3.1 (Invitrogen, USA) to give rise to pcDNA-TBX5-FLAG. KAT2A-HA PCR product was generated from IMAGE clone 6137532 (Source Bioscience, UK). The PCR products were then digested and cloned into NheI and XhoI sites of pcDNA3.1 to generate pcDNA-KAT2A-HA. The construct pcDNA-KAT2B-HA was generated by subcloning KAT2B-HA from pBSSK-KAT2B-HA into XhoI and XbaI sites of pcDNA3.1. Generation of promoter-reporter plasmid pGL3-MYH6-I was reported previously^{73,304}.

2.2.1.2. Pull-down assays

For pull-down assays, TBX5-FLAG, p300-FLAG, KAT2A-HA and/or KAT2B-HA were transfected into COS7 cells as indicated in 2.1.3.6. and lysed 48 hours post-transfection as indicated in 2.1.2.1. Two plates were transfected per construct so a sufficient protein amount was obtained following cell lysis. Purification of FLAG-tagged proteins was performed by using Anti-FLAG M2 Affinity Beads (Sigma-Aldrich, USA) following manufacturer's protocol. Briefly, cell extracts were incubated overnight with 40 µl of washed beads (in wash

buffer, 20mM Tris (pH7.6), 150mM NaCl, 0.5% NP40 and 0.5mM EDTA) at 4°C with gentle agitation. The following day, non-specific proteins were removed by washing the beads 5 times in wash buffer, 5 min per wash. FLAG-tagged proteins and interacting partners were released from the beads by incubating them with 500 µg/ml of 3x FLAG peptide (Sigma-Aldrich, USA) diluted in wash buffer for 2 hours at 4°C with gentle agitation. Purified proteins were then fractionated on 10% Bis-Tris gel followed by western blotting as indicated in 2.1.2.3.. FLAG-tagged proteins were detected using a mouse monoclonal Anti-FLAG antibody (Sigma-Aldrich, F1894, 1:1000 dilution), whereas HA-tagged proteins were detected using a mouse monoclonal Anti-HA antibody (Sigma-Aldrich, H9658, 1:1000 dilution).

2.2.1.3. Luciferase-reporter assay

For reporter assays, COS7 cells were transfected as indicated in 2.1.3.6. with 1.5 µg of reporter plasmid pGL3-MYH6-I, 1.0 µg of pcDNA-TBX5 or pcDNA-TBX5K339R, 1.0 µg pcDNA-KAT2B or pcDNA-KAT2A, and 4 ng of pRL-TK as an internal control to normalize variation in transfection efficiency between the plates. 24 hours post-transfection, reporter assay was carried out by using the Dual-Luciferase Reporter Assay System (Promega) following manufacturer's instructions.

2.2.1.4. Immunofluorescence and cell imaging

For cell localization studies, COS7 and C2C12 cells were seeded on coverslips in 6-well plates and transfected with TBX5-WT-FLAG or TBX5-K339R-FLAG using Polyfect Transfection Reagent or Amaxa Cell Line Nucleofector Kit V respectively as indicated in 2.1.3.6. 24 hours post-transfected cells were then washed with 1 ml of PBS and fixed with 1 ml of 4% paraformaldehyde (PFA) for 10 mins, followed by three washes with 1 ml of PBS, 5 minutes each. Cells were then permeabilized with 1 ml of 0.25% Triton-X in PBS for 5 mins, washed 3 times in 1 ml of PBS and incubated in 1 ml of 5% BSA in PBS for 1 hour to suppress non-specific binding of IgG. Following blocking, cells were washed in 1 ml of PBS and incubated with 1 ml of mouse monoclonal Anti-FLAG antibody (Sigma-Aldrich, F1894, 1:250 dilution) or with rabbit anti-human CRM1 antibody (Santa Cruz Biotechnology, sc-74454, dilution 1:50) diluted with PBS containing 3% BSA for 1 hour at RT. Cells were then washed 3 times with 1 ml of PBS each for 5 minutes and incubated with 1 ml of goat anti-mouse Alexa-488 or goat anti-rabbit Alexa-594 secondary antibody (ThermoFisher Scientific, 1:500 dilution) diluted with PBS containing 3% BSA for 1 hour at RT. Cells were again washed 3 times with 1 ml of PBS and incubated with 1 ml of Hoechst solution (Sigma-Aldrich, USA, 1:1000 dilution) diluted with PBS for 10 min in the dark for nuclei staining. Cells were washed with 1 ml of PBS for 3 times, 5 min per wash, and coverslips were finally mounted on glass slides containing a small drop of

Vectashield mounting media without DAPI (Vector Laboratories, UK). After sealing the slides with nail polish, cell imaging was performed on a Zeiss LSM880F confocal microscope (Carl Zeiss, Germany) using a 63X 1.4 oil objective. Images were captured using the ZEN software (Carl Zeiss, Germany) and analysis was performed using the Fiji software (open source).

2.2.1.5. Drug treatment (cultured cells)

COS7 and C2C12 cells were treated with two different drugs in this project: leptomycin B (a potent inhibitor of the exportin CRM1³⁰⁵) and Garcinol (an inhibitor of the acetyl-transferase activity³⁰⁶). For leptomycin B, the concentration used for the treatment of COS7 (10 ng/ml) and C2C12 cells (10 ng/ml) was similar to those already reported elsewhere^{307,308}. In the case of Garcinol, a drug titration experiment was initially performed. To do this, 5x10⁴ COS7 cells were plated in 96-well plates and treated with 8 different concentrations (0.1 µM, 1 µM, 2.5 µM, 5 µM, 10 µM, 25 µM, 50 µM and 100 µM) of Garcinol diluted in DMEM for 16 hours. After this period, cell cytotoxicity was measured by performing an MTT assay as indicated in 2.1.3.7. A concentration of 20 µM was chosen for subsequent experiments, in which cells were typically treated for 16 hours with the required compound before fixation and subsequent immunofluorescence staining as indicated in 2.2.1.4.

2.2.2. In-vivo studies

2.2.2.1. Zebrafish maintenance and harvesting of embryos

Zebrafish used in this project were raised and maintained by the BioSupport Unit (BSU) of The University of Nottingham according to Home Office regulations and guidelines, local and national ethical guidelines and best animal welfare practices. In order to obtain zebrafish embryos, adult breeding pairs were placed in a small tank with a transparent divider between them in the afternoon before injection day. Next morning, the dividers were removed to allow the fish to breed. Eggs were then collected, subjected to microinjection as indicated in 2.2.2.7. and kept at 28°C in fish water until they reached the desired developmental stage.

2.2.2.2. sgRNA design

In order to generate *kat2a*, *kat2b* and *tbx5a* knockouts using the CRISPR-Cas system, sgRNAs for each target gene were designed using the web tool CHOPCHOP³⁰⁹ (<http://chopchop.cbu.uib.no/>). The cDNA sequence for each gene was uploaded to the software and the top 3 sgRNAs in terms of predicted activity were chosen for posterior DNA template synthesis and sgRNA transcription. Possible off-target effects for each sgRNA were further analysed using Cas-OFF finder webtool³¹⁰ (<http://www.rgenome.net/cas-offfinder/>) in order to make sure that none of them presented any potential off-target

effects. The sequences of each sgRNA generated in this thesis can be found in section 3.3.4.

2.2.2.3. sgRNA template generation and transcription

DNA templates for sgRNA transcription were generated using a cloning-free method previously described^{311,312} (Fig. 2.2.). Briefly, a 52-nt primer (sgRNA primer) containing the T7 promoter sequence (5'-TAATACGACTCACTATA-3'), the 20 nt target site without the PAM sequence and a constant 15-nt tail region (5'-GTTTGTAGAGCTAGAA-3') were annealed with an 80-nt constant oligonucleotide (5'-GTTTGTAGAGCTAGAAATAGCAAGTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT-3') to add the sgRNA invariable 3' end. For the annealing reaction, 1 µl of sgRNA primer 100 µM, 1 µl of constant oligonucleotide 100 µM and 8 µl of water were mixed and annealed in a thermocycler by using the following program:

- Heat to 95°C and maintain the temperature for 2 min
- Cool to 25°C over a period of 45 min
- Cool to 4 °C for temporary storage

After annealing, the ssDNA overhangs were then filled in with T4 DNA polymerase (New England Biolabs, USA) by setting up the following reaction:

Table 2.7. sgRNA template generation reaction

Component	Volume
Annealing reaction	10 µl
dNTPs (10 mM)	2.5 µl
10x NEB Buffer 2	2 µl
100x NEB BSA	0.2 µl
T4 NEB DNA polymerase	0.5 µl
Water	4.8 µl

Reactions were then incubated for 20 min at 12°C in a thermocycler, and the sgRNA template was run on an agarose gel in order to verify the presence of a product of the expected size (~120 bp), followed by DNA gel purification as indicated in 2.1.1.3. Once the template for sgRNA transcription was generated, sgRNA transcription was carried out using the MAXIscript T7 kit (ThermoFisher, USA) by setting up the following transcription reaction in a microcentrifuge tube:

Table 2.8. Components of the sgRNA transcription reaction

Component	Volume
ATP	1 µl

GTP	1 μ l
CTP	1 μ l
UTP	1 μ l
10x Buffer	2 μ l
Enzyme Mix	2 μ l
Nuclease-free water	2 μ l
Template	10 μ l

Transcription reactions were incubated for 4 hours at 37°C. After this, 1 μ l of TURBO DNase (included in the kit) was added to the reactions followed by 15 min incubation at 37°C. In order to precipitate and purify the sgRNA, 2 μ l of 3M ammonium acetate and 60 μ l of pre-chilled 100% EtOH was added to the reactions and these were incubated for 1 hour at -20°C to promote sgRNA aggregation and precipitation. After this, reactions were centrifuged for 30 min at 4°C, maximum speed. The supernatant was discarded and 1 ml of pre-chilled 70% EtOH was added to wash the sgRNA pellet. Tubes were centrifuged again for 10 min at 4°C, maximum speed, and after removing the supernatant, these were let to dry at RT for a few minutes to remove any residual EtOH. Finally, sgRNA was resuspended in 50 μ l of nuclease-free water, and sgRNA concentration and integrity was analysed in a Bioanalyzer as indicated in 2.1.1.11. sgRNAs were stored at -80°C in 5- μ l aliquots until further use.

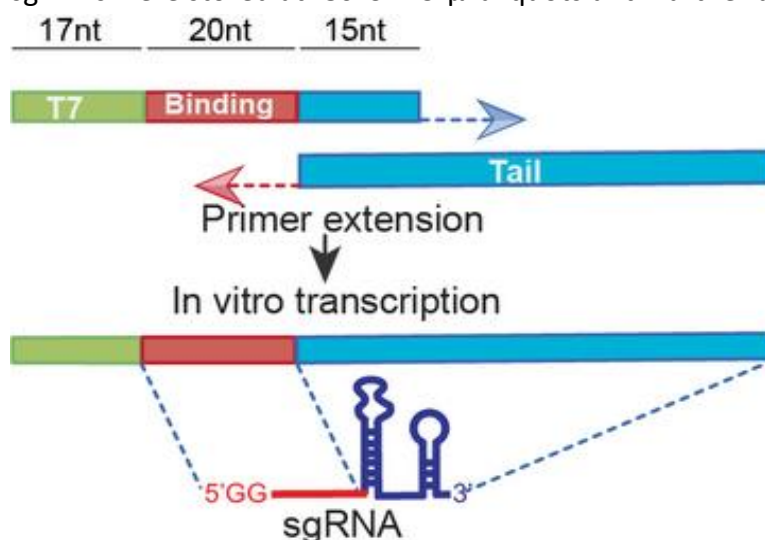


Figure 2.2. Cloning-free method for the generation of sgRNA templates. An oligonucleotide containing the T7 promoter (green), the 20 nt of the specific sgRNA DNA binding sequence (red) and a constant 15-nt tail for annealing is used in combination with an 80-nt reverse oligo (tail primer) to add the sgRNA constant 3' end (in blue). *Figure from Moreno-Mateos, 2015*³¹¹

2.2.2.4. Cas9 mRNA transcription

Cas9 mRNA for *in vivo* injection was generated from the pCS2-nCas9n plasmid (Chen lab, Addgene plasmid #47929). This plasmid contains a zebrafish codon-optimized Cas9 cDNA flanked by two nuclear localization sequences³¹³. pCS2-

nCas9n plasmid was first linearized using NotI-HF restriction enzyme as indicated in 2.1.1.4. Cas9 mRNA was then *in vitro* transcribed using a mMESSAGE mMACHINE SP6 Transcription Kit (ThermoFisher, USA), using a protocol remarkably similar to that of section 2.2.2.3. Cas9 mRNA precipitation was performed by using ethidium chloride and EtOH instead of ammonium acetate and EtOH.. The product of the Cas9 mRNA transcription was checked by running an aliquot on a 1% agarose diagnostic gel prior to storage at -80°C.

2.2.2.5. sgRNA *in vitro* activity assay

sgRNAs activity were checked *in vitro* in order to choose the best candidate for *in vivo* injection³¹⁴. Briefly, 30 nM sgRNA were mixed with 30 nM Cas9 Nuclease (New England Biolabs, USA) and 10x Cas9 nuclease reaction buffer in 20 µl total reaction volume. Following 10 min incubation at RT, 250 ng of a PCR product containing the sgRNA target site was added and the reaction was incubated for 1 hour at 37°C. Reactions were then run on a 2% agarose gel for fragment analysis as indicated in 2.1.1.2. The presence of smaller bands besides the original PCR product band indicate that the sgRNA is active.

2.2.2.6. RT-PCR analysis

Wild type zebrafish embryos were isolated and collected at various developmental stages; they were subsequently dissected using surgical microlances to produce tail and cardiac sections (Fig. 2.3). Due to the difficulty of dissecting the zebrafish heart, some cardiac tissue dissections inevitably contained tissue from other body parts like the lower part of the head or the yolk. Dissected sections were collected on dry ice and immediately frozen for storage prior to processing. Total RNA was isolated from pools of 20-40 zebrafish embryos of the respective section using TRI-Reagent according to as indicated in 2.1.1.10. DNase treatment was also carried out at 37°C for 30 minutes. First strand cDNAs were synthesised from 1 µg of total RNA as indicated in 2.1.1.12. Primers shown in Table 2.9 were used to detect the expression of *kat2a* and *kat2b* and a reference control gene *β-actin1* in zebrafish via PCR amplification using Taq polymerase and cDNA as template (as indicated in 2.1.1.1.)

Table 2.9. Primers used for reverse transcription PCR

Primer name	Sequence
kat2aF	CCTAAGCCAACAGCAATGGT
kat2aR	GCTTGGTGTCCGTGTCTTCT
kat2bF	GGTACGAGGCCACACAGATT
kat2bR	GTAGGAAACATCCGGAAGCA
β-actin1F	ACATCAGCATGGCTTCTGCT
β-actin1R	GTGTACAGAGACACCCTGGC

Following PCR amplification, products were resolved by electrophoresis on 1.5% agarose gels and visualized with ethidium bromide.

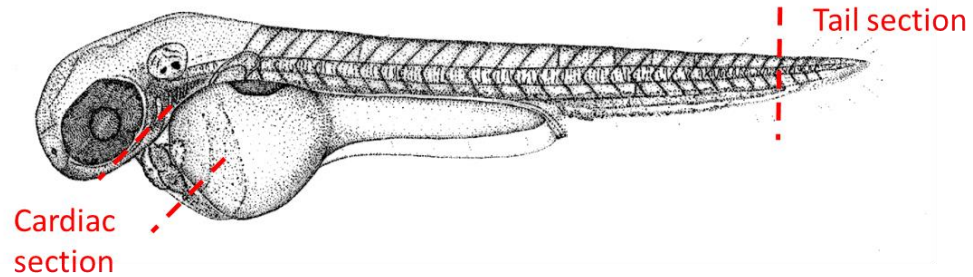


Figure 2.3. Drawing of a 48hpf zebrafish embryos indicating transection points. Transections are represented by red dashed lines. Zebrafish drawing adapted from https://zfin.org/zf_info/zfbook/stages/figs/fig1.html

2.2.2.7. Injection of morpholinos and CRISPR-Cas system in zebrafish embryos

Initially, to assess the function/phenotype of *kat2a* and *kat2b*, antisense morpholino oligonucleotides were used. A translation blocking morpholino as well as a splice blocking morpholino targeting *kat2a* or *kat2b* intron1-exon2 splice sites were designed and made by Gene Tools, USA. A translation-blocking morpholino targeting *tbx5a* previously described was also used as a positive control¹²⁶. A 5-bp mismatch morpholino was also used as a negative control. The sequence of each one of these morpholinos can be found in Table 2.10. A graphical representation of the binding site for the translation and splice blocking morpholinos is shown in Fig. 2.4:

Table 2.10. Morpholinos used for *kat2a* and *kat2b* knockdowns in zebrafish

Morpholino name	Sequence
kat2a trans blocking	TCCGCCATAACAGCAGCCACACTTC
kat2a splice	ATTCGCCTGAGATAAGAAGAGTGCA
kat2b trans blocking	GATCCCCGTGCTTCCGACATCGAA
kat2b splice	TCAGCCTGTGTGAAAGAAAGAAGAGTGCA
tbx5a trans blocking	GAAAGGTGTCTTCACTGTCCGCCAT
Mismatch morpholino	TCCGGCATATCAGAAGCTACAGTTC

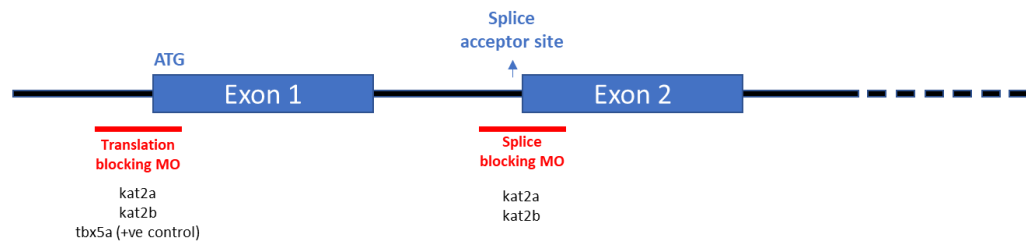


Figure 2.4. Representation of *kat2a*, *kat2b* and *tbx5a* morpholinos-binding site in their respective pre-mRNAs. Translation blocking morpholinos for *kat2a*, *kat2b* and *tbx5a* bind to the translation start site of their respective pre-mRNAs, whereas splice-blocking morpholinos for *kat2a* and *kat2b* bind to the splice acceptor site situated in the boundary of the intron 1-exon in both genes.

For morpholino experiments, a morpholino dose-titration was initially performed. Pools of 20 zebrafish embryos were injected with increasing doses of morpholinos, ranging from 0.5 ng to 8 ng (0.5 ng, 1 ng, 2 ng, 4 ng and 8 ng), until a dose which produced a consistent and reproducible phenotype at 48-72 hpf without inducing toxicity or extremely severe development defects was found (4 ng for *kat2a* and *kat2b* splice MOs; both translation blocking MOs led to toxicity and severe development defects across all MO amounts tested. For the *tbx5a* translation blocking MO, 1.7 ng were used as reported in the original publication¹²⁶). For injections, several morpholino dilutions containing 1/10 of phenol red dye were prepared in DEPC water right before injection so the desired amount morpholino was injected during the procedure. In order to generate CRISPR-Cas knockouts of *kat2a*, *kat2b* and *tbx5a*, 100 ng/μl of active sgRNA targeting each one of these genes were mixed with 500 ng/μl of Cas9 mRNA or Cas9 protein (New England Biolabs, USA), 1/10 of phenol red dye for visualization of the injection and 300 mM KCl when using Cas9 protein (to stabilize the ribonucleoprotein complex). For both morpholino and CRISPR-Cas injection, 1-cell stage zebrafish embryos were collected and lined up on a 90 mm dish with the help of a microscope slide (see Fig. 2.5.).

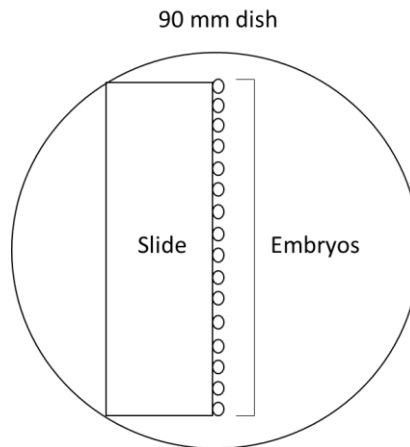


Figure 2.5. Setup used for the injection of 1-cell stage zebrafish embryos

Injection mix was then loaded into a microinjection needle by using a Microloader pipette tip (Eppendorf, Germany) and the needle was then coupled to a microinjection rig as shown in Fig. 2.6. The microinjection rig is connected to a Picospritzer III Injection System (Parker, USA), which is also connected to a CO₂ bottle and a foot pedal to initiate the pulse. The duration of the pulse was set up so the injection volume per pulse was approximately 1 nl.

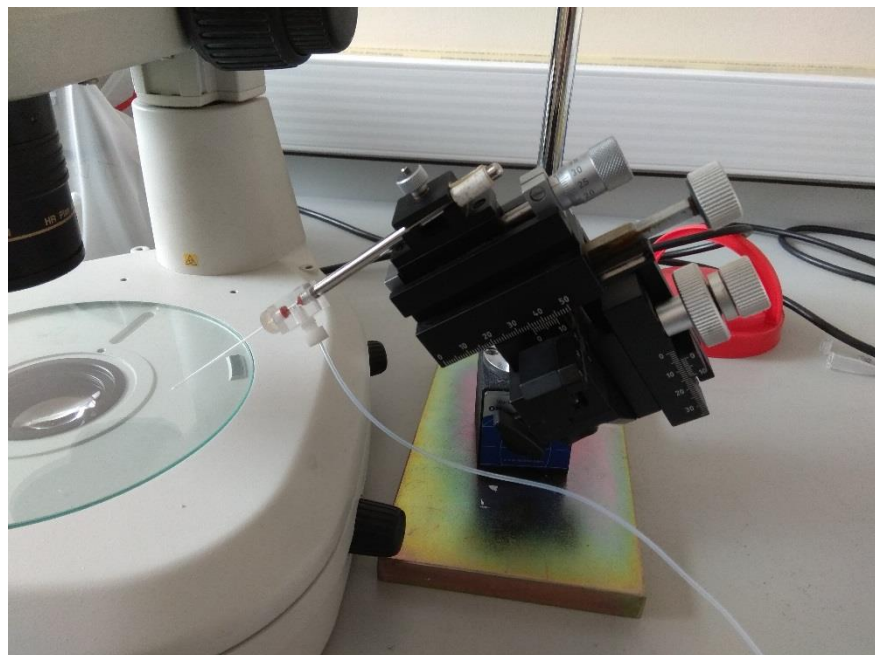


Figure 2.6. Microinjection rig used for injection of 1-cell stage zebrafish embryos.

It is important to insert the needle inside the embryo cell and avoid injecting into the yolk sac to reach maximum morpholino/CRISPR-Cas efficiency (see Figure 2.7). Approximately 50 to 60 embryos were injected per batch, and

following injection, embryos were transferred to a new 90 mm dish containing fish facility water and incubated at 28°C for 48-72 hours.

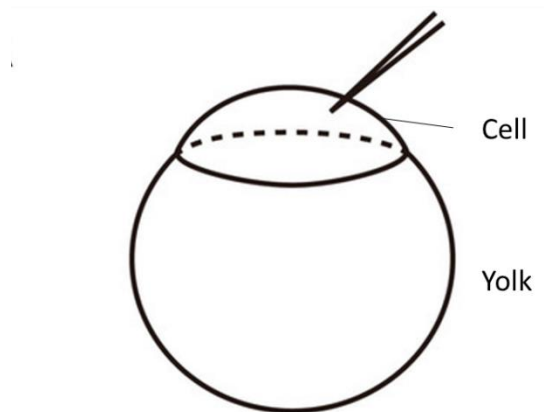


Figure 2.7. Representation on a microinjection into a 1-cell stage zebrafish embryo.

2.2.2.8. Zebrafish embryos screening

Injected embryos were screened 48 or 72 hpf for the presence of cardiac or limb defects, respectively. Following addition of 1 ml of an anaesthetic compound (tricaine) to the plate were embryos were maintained, screening and capturing of images of live zebrafish embryos were performed using a Zeiss Stereo Lumar V12 microscope with a camera. Images were then analysed using ImageJ software.

2.2.2.9. Zebrafish genomic DNA extraction

Upon screening, genomic DNA was extracted from injected zebrafish in order to check for the presence of mutations. Individual embryos were incubated in microcentrifuge tubes with 50 µl of DNA extraction buffer (10 mM Tris-HCl pH 8.2, 10 mM EDTA, 200 mM NaCl, 0.5%SDS, 200 µg/ml proteinase K) for 4 hours at 55°C. Genomic DNA was subsequently precipitated with ethanol as indicated in 2.1.1.5. and the final DNA concentration was measured using a Nanodrop spectrophotometer.

2.2.2.10. T7 endonuclease assay

The presence of mutations in the CRISPR-Cas9 target site was analysed using the T7 endonuclease assay³¹⁵ (Fig. 2.8). T7 endonuclease I recognises and cleaves non-perfectly matched DNA (heteroduplexes). Due to zebrafish biallelism and mosaicism induced by the CRISPR-Cas9 system, PCR products of mutant embryos contain mixtures of molecules with different mutations in the target site. This leads to the formation of heteroduplexes once these PCR products are denatured and re-annealed. Following incubation with T7 endonuclease, heteroduplexes derived from mutant embryos will be digested, therefore producing smaller fragments that can be subjected to gel analysis.

For the T7 endonuclease assay, a 700-800bp PCR product harbouring the sgRNA target site was generated from previously extracted genomic DNA by setting up 50 µl PCR reactions using Phusion high-fidelity polymerase (as described in section 2.1.1.1.). Following the cycling reaction, the whole PCR reaction volume was then run on a 1.5% agarose gel as described in 2.1.1.2. in order to verify the generation of the correct PCR product. PCR products were then purified via agarose gel purification as described in 2.1.1.3., and concentration was measured using a Nanodrop spectrophotometer (typically, 30-70 ng/µl in 20 µl total volume). For denaturation and re-annealing, 200 ng of PCR products were mixed with 2 µl of 10x NEBuffer 2 and water up to a total volume of 19 µl. This reaction was then subjected to the following temperature cycles on a thermocycler:

- Heat to 95°C and maintain the temperature for 2 min
- Cool to 25°C over a period of 45 min
- Cool to 4 °C for temporary storage

After re-annealing, 1 µl of T7 endonuclease was added to each reaction and these were incubated for 1 hour at 37°C. Finally, samples were ran on a 2% agarose gel and visualized using ethidium bromide. Embryos harbouring mutations should present two or more bands depending on where the T7 is cutting (target site). WT embryos only present 1 band corresponding to the uncut PCR product

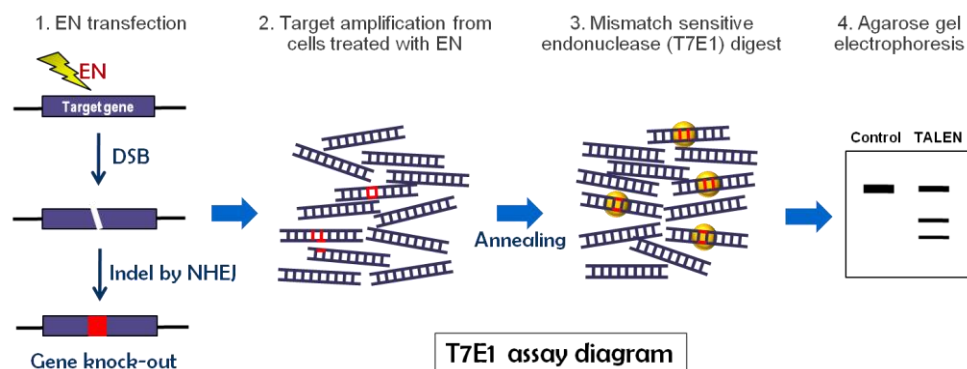


Figure 2.8. Diagram of the T7 endonuclease assay. Following gene knockout by the introduction of indels, the target site is amplified from zebrafish embryo cells. PCR products harbouring different mutations (red) are then denatured and re-annealed, therefore leading to the generation of heteroduplexes (yellow spheres). T7 endonuclease will recognise and cleave these heteroduplexes, and small fragments will be produced.

2.2.2.11. Individual alleles isolation

In order to analyse and sequence individual mutations in the target site, PCR products used for T7 endonuclease were isolated and sent for Sanger sequencing. Briefly, amplicons were ligated into a pGEM-T vector using a

pGEM®-T Easy Vector System (Promega, USA) as indicated in 2.1.1.6. and subsequently transformed into chemically competent *E.coli* as indicated in 2.1.1.7. After bacterial growth, ligated products were extracted using a GenElute Plasmid Miniprep Kit (Sigma-Aldrich, USA) as indicated in 2.1.1.8. Five clones from each individual embryo were then sent for sequencing and sequences were analysed for the presence of insertions, mutations or substitutions in the target site using the Chromas software (Technelysium, Australia).

2.2.2.12. In-situ hybridization

In order to study heart and fin morphology, injected embryos were fixed overnight in 4% PFA solution in PBS and subjected to *in situ* hybridization (ISH) according to a method previously described³¹⁶.

Probe design and PCR template generation

Initially, in order to generate both antisense and sense RNA probes for the detection of gene expression, 500-1000 bp PCR products containing a fragment of the cDNA sequence of the gene of interest were generated. The primers used for the generation of these products contained either the T7 promoter sequence (F primer, 5'-TAATACGACTCACTATAGGG-3') or the T3 promoter sequence (R primer, 5'-CATTAACCCTCACTAAAGGGAA-3'), thus allowing the generation of both sense and antisense probes from the same PCR product. The full list of probes generated in this thesis and their corresponding pair of primers can be found in Table 2.11:

Table 2.11. *In situ* hybridization probes used in this thesis and their respective primer sequences

Probe	Probe Size	Primer sequences
kat2a	600 bp	F: TAATACGACTCACTATAGGGCAACATAGAGCAGGGGGTGT R: CATTAACCCTCACTAAAGGGAAACCGATTGTTGGTTCAGAGG
kat2b	620 bp	F: TAATACGACTCACTATAGGGTGGCTGGAAGAGTCAAAACC R: CATTAACCCTCACTAAAGGGAAAAACACAGAGCGCAGGAAAAT
cmlc2	410 bp	F: TAATACGACTCACTATAGGGCTCTTCCAATGTCTTCTCC R: CATTAACCCTCACTAAAGGGAATATTTCCAGCCACGTCTA
bmp4	1000 bp	F: TAATACGACTCACTATAGGGTGCCAAGTCCTACTGGGAC R: CATTAACCCTCACTAAAGGGAACGTGATTGGTGGAGTTGAG
fgf10a	533 bp	F: TAATACGACTCACTATAGGGTGCTTCTGTTCTGTGTTTCG R: CATTAACCCTCACTAAAGGGAACACGATAGGAATGGGGAGAA
nppa	328 bp	F: TAATACGACTCACTATAGGGACACGTTGAGCAGACACAGC R: CATTAACCCTCACTAAAGGGAAAGGGTGCTGGAAGACCCTAT

In order to generate the PCR templates for probe generation, a 50-µl PCR reaction using zebrafish cDNA as template and Taq polymerase was set up as

indicated in 2.1.1.1. PCR reactions were then run on an 1.5% agarose gel in order to confirm the generation of the expected PCR products, and these were subsequently purified as indicated in 2.1.1.3.

Synthesis of the antisense and sense RNA probes

To generate the antisense or sense RNA probes, the following reactions were prepared in a microcentrifuge tube:

Table 2.12. Components for the generation of DIG-labelled RNA probes

Component	Amount
Template DNA	150 ng
5x Transcription buffer	1 µl
DTT (0.1 M)	0.5 µl
DIG-RNA labeling mix (10x)	0.5 µl
RNAse in (40 U/ml)	0.25 µl
T7 or T3 RNA polymerase (20 U/ml)	0.25 µl
RNAse-free water	Up to 5 µl total volume

Reactions were incubated for 2 hours at 37°C. After this, 2 µl of RNAse-free DNase (Roche, Switzerland) was added and reactions were incubated for 30 min at 37°C. Reactions were finally stopped by adding 1 µl of 0.5M EDTA and 9 µl of RNAse-free water. For the purification of RNA probes, Sigma spin purification columns (Sigma-Aldrich, USA) were used. Columns were placed on top of a microcentrifuge tube and centrifuged for 15 s at 750g. Then, the base of the column was broken and the lid was discarded, followed by centrifugation at 2 min for 750 g. Reactions were then applied to the centre of the column (on top of the resin) and centrifuged for 4 min at 750g. The column was discarded, and probes were then stored in 5-µl aliquots (to avoid repeated freeze-thaw cycles) at -80°C. Typically, an aliquot was also saved in order to measure probe concentration on a Nanodrop spectrophotometer and to check its integrity on a 1% agarose gel (at 230V for 30 min). Good probes should appear as one or two discrete bands of the gel.

Removal of pigmentation and embryo fixation

If post-gastrulation stages are going to be examined via *in-situ* hybridization, the formation of melanin pigments needs to be prevented. To achieve this, embryos were kept in a 0.0045% solution of 1-Phenyl-2-thiourea (PTU) (Sigma-Aldrich, USA) in zebrafish water starting from the end of gastrulation (around 10 hpf) onwards. This medium was changed once a day until the desired developmental stage was reached. After reaching this stage, embryos were dechorionated using a pair of forceps and fixed on freshly prepared 4% PFA

solution in 1xPBS overnight at 4°C (typically, they were placed on 6-well plates for fixation and subsequent washes steps. For all washes steps indicated in this protocol, 1 ml of each solution was used unless stated otherwise). The following morning, embryos were dehydrated by incubating them in 100% MeOH for 15 min at RT. At this point, embryos can be stored in 100% MeOH at -20°C for several months.

Permeabilization and hybridization of the embryos

Embryos were rehydrated by incubating them in successive dilutions of MeOH in 1xPBS at RT: 5 min in 75% (vol/vol) MeOH, 5 min in 50% (vol/vol) MeOH, and 5 min in 25% (vol/vol) MeOH. Then, embryos were washed four times in 1x PBS-T (PBS and 0.1% of Tween-20 (Sigma-Aldrich, USA)) for 5 min each wash, and permeabilization was achieved by digesting the embryos with a 10 µg/ml proteinase K (Roche, Switzerland) solution in 1xPBS-T at RT for the time indicated in the table below:

Table 2.13. Proteinase K treatment duration according to zebrafish developmental stage

Developmental stage	Duration of Prot K treatment
1 somite stage	30 sec
1-8 somite stage	1 min
9-18 somite stage	3 min
18 somite stage – 24 hpf	10 min
36 hpf – 5 days	30 min

Proteinase K digestion was stopped by incubating the embryos in 4% PFA in 1xPBS for 20 min at RT, being followed by four washes, 5 min per wash, in 1xPBS-T to remove any residual PFA (at this point, embryos can be pooled according to the probe which will be used for hybridization). Pools of 15 embryos were prehybridized by incubating them in 1 ml of Hybridization Mix (HM) for 5 hours at 70°C, a solution made of:

- 50% deionized formamide (Sigma-Aldrich, USA)
- 5xSaline Sodium Citrate (SSC): 20x stock solution is prepared by dissolving 175.3 g of NaCl (Sigma-Aldrich, USA) and 88.2g of Citric acid trisodium (Sigma-Aldrich, USA) in 1L of water
- 0.1% Tween-20
- 50 µg/ml of heparin (Sigma-Aldrich, USA)
- 500 µg/ml of RNase-free tRNA (Sigma-Aldrich, USA)
- 460 µl of 1M citric acid for 50 ml of HM to adjust pH (6.0)

These prehybridized embryos can be stored in HM at -20°C for several weeks. After prehybridization, HM was discarded and replaced with 200 µl of HM containing 50 ng of antisense (or sense) probe, and embryos were hybridized overnight at 70°C.

Washes and antibody incubation

Following hybridization, HM was gradually changed to 2xSSC by performing a series of washes in HM (without heparin nor tRNA) diluted with 2xSSC: 75% HM, 50% HM, 25% HM and 100% 2xSSC. All these washes were performed for 10 min at 70°C with gentle agitation. Embryos were then washed twice, for 30 min per wash, in 0.2xSSC at 70°C with gentle agitation. Then, 0.2xSSC was progressively replaced with 1xPBS-T through a series of 10 min washes in 0.2xSSC diluted in PBS-T: 75% 0.2xSSC, 50% 0.2xSSC, 25% 0.2xSSC and 1xPBS-T. These washes were performed at RT with gentle agitation (50 rpm). Once all washes were performed, embryos were incubated in blocking buffer (1xPBS-T, 2% sheep serum (vol/vol) (Sigma-Aldrich, USA), 2 mg/ml BSA (Sigma-Aldrich, USA)) for 4 hours at RT. After blocking, embryos were incubated in 500 µl of anti-DIG antibody (Roche, 11093274910, 1:10000 dilution) diluted in blocking buffer overnight at 4°C with gentle agitation (50 rpm).

Washes and staining

The following morning, the antibody solution was discarded and embryos were washed 6 times, for 15 min each wash, in PBS-T with gentle agitation (50 rpm) at RT. Then, embryos were incubated in alkaline Tris buffer (100 mM Tris HCl, pH 9.5, 50 mM MgCl₂, 100 mM NaCl and 0.1% Tween 20) three times, 5 min per wash, with gentle agitation (50 rpm) at RT. After the washes, alkaline Tris buffer was replaced with 1 ml of staining solution prepared fresh and protected from light (staining solution was prepared by diluting 225 µl of 50 mg/ml nitro blue tetrazolium (NBT; Sigma-Aldrich, USA) and 175 µl of 50 mg/ml 5-bromo 4-chloro 3-indolyl phosphate (BCIP; Sigma-Aldrich, USA) with 50 ml of alkaline Tris Buffer). Color reaction was then monitored periodically (every 10 min) under a dissecting scope, keeping the embryos protected from light between checks. Once the desired staining intensity was reached (based on my experience, this time may vary from 20 minutes to 3 hours depending of the gene analysed), the staining reaction was stopped by removing the staining solution and incubating the embryos with 1 ml of stop solution (1x PBS, pH 5.5, 1 mM EDTA, 0.1% Tween-20) three times, 15 min each incubation, with gentle agitation (50 rpm) at RT.

Mounting and visualizing embryos

Using a pipette and a 1-ml pipette tip cut to get a wider opening, embryos were transferred, in the minimum possible volume of stop solution, to a new 6-well plate containing 100% glycerol. Following this, embryos were incubated overnight in the dark with gentle agitation (50 rpm) at RT. This step allows the exchange of water and glycerol, which is a mild clearing agent that imparts a degree of transparency to the biological material it impregnates. Next day, embryos visualization and images acquisition were performed using a Zeiss Stereo Lumar V12 microscope with a camera. Images were then analysed using ImageJ software.

2.2.2.13. Cardiac looping measurement

Cardiac looping angle was determined by measuring the angle formed by the plane of the atrioventricular junction (AVJ) relative to the anterior/posterior axis of the embryo as described elsewhere³¹⁷. Fig. 2.9. shows a graphical representation of how cardiac looping is measured. Angle of looping was calculated using either ImageJ or Photoshop software.

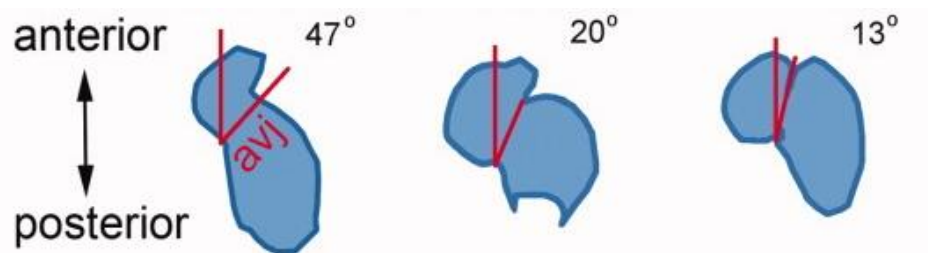


Figure 2.9. Graphical representation of how cardiac looping angle is measured. The anterior/posterior axis of the embryo is used as reference (vertical red line), and the angle formed by the plane of the atrioventricular junction (avj) relative to the anterior/posterior axis is measured. Cardiac looping angle in unlooped hearts (left, 47°) is greater than in looped hearts (right, 13°). Adapted from Chernyavskaya et al, 2012³¹⁷

2.2.2.14. Heart rate measurement

Zebrafish heart rate measurement was performed manually at room temperature using a manual mechanical clicker. Embryos were removed from the 28°C incubator, and heart rate measurement was performed by counting the heart rate manually under a Zeiss Stereo Lumar V12 microscope. One-minute measurements were performed in all cases.

2.2.2.15. Drug treatment (zebrafish embryos)

In order to chemically inhibit the acetyltransferase activity, Garcinol was used (Santa-Cruz Biotechnology, USA). First, pools of 10-15 embryos were treated using a range of different concentrations (0.1 µM, 1 µM, 5 µM, 10 µM, 50 µM, 100 µM, 250 µM and 500 µM) of Garcinol diluted with zebrafish water. Treatments were carried out for 24 hours starting at 20 hpf (initial stages of cardiogenesis). Embryos were washed at 48 hpf to remove the drug and then

embryo toxicity was calculated by counting the number of dead embryos or embryos showing severe developmental defects/delay at each concentration. For subsequent drug treatments, Garcinol 100 μ M was used as at this concentration, embryos showed a consistent cardiac phenotype without compromising viability in excess (<30% of embryo toxicity).

2.2.2.16. qPCR analysis of *tbx5a* target genes

Expression levels of *tbx5a* and *tbx5a* target genes were measured using a 7500 Real-Time PCR system (Applied Biosystems, USA). Total RNA was extracted from 32 hpf *kat2a*, *kat2b* and *tbx5a* KO embryos for posterior cDNA synthesis and qPCR analysis as indicated in 2.1.1.13. Primers shown in Table 2.14, which had been already described and validated elsewhere^{318,319}, were used to detect the expression levels of *tbx5a* and its target genes, as well as the expression levels of two *tbx5a* unrelated genes which have been demonstrated to be expressed during zebrafish development (*akt1*³²⁰ and *igf3*³²¹). Two well established zebrafish housekeeping genes, *ef1a*³²² and *lsm12b*³²³, were used for normalization. Typically, three biological replicates and three technical replicates were included in each qPCR experiment.

CGCATAATTCAATCAAAGTCCG
GCTGCTCCAGGTTTGCCTATGT

Table 2.14. Primers used for relative quantitative PCR

Primer name	Sequence
nppaF	CAGACACAGCTCTGACAGCAACAT
nppaR	CTCTGTGTGTCAAATCCATCCGAG
fgf10F	CTCATCGTCTGCCGTGGTG
fgf10R	GTCAATGCCGAAATCCCTC
bmp4F	CACAGTATCTGCTCGACCTCTA
bmp4R	GATATGAGTTCGTCCTCTGGGATG
hey2F	GAAAGAAGCGGAGAGGGATCATTG
hey2R	AGAAGTCCATGGCCAGAGAATGAG
tbx2bF	GTCCCTTTCCCTTTCATCTGTCTC
tbx2bR	CTGGGAGCTGATAAGGGTTGAATC
tbx5aF	ACACCTTTCGGCTCCAAACT
tbx5bR	CTTTGTCCACAGCTCTCGCTC
akt1F	TCGGCAGGTGTCTTCTCAAT
akt1R	ACCCATTGCCATACCACGAG
igf3F	CGCATAATTCAATCAAAGTCCG
igf3R	GCTGCTCCAGGTTTGCCTATGT
ef1aF	CGGTGACAACATGCTGGAGG
ef1aR	ACCAGTCTCCACACGACCA
lsm12bF	AGTTGTCCCAAGCCTATGCAATCAG
lsm12bR	CCACTCAGGAGGATAAAGACGAGTC

2.2.2.17. Rescue experiment

For rescue experiments, capped mRNA transcripts were synthesized from the SP6 promoter of linearized plasmid pCS2-TBX5-K339Q, which was generated via cloning by myself. For the generation of this construct, the cDNA sequence of the WT Tbx5a gene was cloned into the expression vector pCS2, which was then used as template for site-directed mutagenesis as indicated in 2.2.1.2. The primers for the generation of the K339Q variant are F: 5'-CATACGGCTTCTGATAGGGGTGTTCTCCTGCAGG-3' and R: 5'-CCTGCAGGAGAACACCCCTATCAGAAGCCGTATG-3'. Then, capped TBX5-K339Q mRNA was *in vitro* transcribed using a SP6 mMessage mMachine kit (ThermoFisher, USA) as indicated in 2.2.2.4. 300 pg of TBX5 K339Q mRNA were co-injected with 100 pg of sgRNA and 500 pg of Cas9 protein (New England Biolabs, USA) into 1-cell stage zebrafish embryos. Injected embryos were screened 48 and 72 hpf for the presence of cardiac and fin defects

2.3. ANALYSIS OF CHD4 MUTATIONS FOUND IN SYNDROMIC CHD PATIENTS

2.3.1. Plasmids and constructs

FLAG-tagged CHD4, MYC-tagged HDAC1, MYC-tagged HDAC2, MYC-tagged MTA2 and MYC-tagged RBBP4 PCR products were generated from IMAGE clones 5528023, 4976514, 5271899, 5531965 and 6526200 respectively (Source Bioscience, UK). PCR products as well as the mammalian expression vector pcDNA3.1 were digested with KpnI-HF and NotI-HF (CHD4, HDAC1 and HDAC2) or EcoRI-HF and NotI-HF (MTA2 and RBBP4) restriction enzymes, and they were subsequently ligated together using T4 Ligase to generate pcDNA3.1-CHD4-FLAG and pcDNA3.1-HDAC1/HDAC2/MTA2/RBBP4-MYC. Ligation products were transformed into chemically competent E.coli. After bacterial growth, ligated products were extracted using a GenElute Plasmid Miniprep Kit (Sigma-Aldrich, USA) and verified via Sanger sequencing. The expression vector pcDNA3.1-TBX5-MYC was kindly provided by Dr. Tushar Ghosh. Protocols followed for the ligation, transformation and purification of these constructs can be found in sections 2.1.1.4 to 2.1.1.8. For the generation of FLAG-tagged CHD4 mutant variants (M1 to M5), QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies, USA) was carried out as indicated in 2.3.1.1. Introduction of specific mutations was finally confirmed via Sanger sequencing.

2.3.1.1. Site-directed mutagenesis

Site directed mutagenesis was carried out using the QuickChange Site Directed Mutagenesis Kit (Agilent Technologies, USA) in order to generate the mutants

CHD4 M1 to M5 (pcDNA-CHD4-FLAG was used as a template). First, mutagenic oligonucleotide primers were designed according to the desired mutation using the QuickChange Primer Design web tool (<https://www.agilent.com/store/primerDesignProgram.jsp>). In this web tool, the sequence of the original DNA template is introduced, as well as the desired deletions, insertions or substitutions to be introduced, in order for the tool to generate a pair of mutagenic primers for the generation of mutant variants. Primers used for the generation of the five CHD4 mutant variants are indicated in Table 2.15.

Table 2.15. Primers used for the generation of CHD4 mutant variants

Mutant variant	Primers for mutagenesis
CHD4 C467Y (M4)	F: CTGCTCTGCTGTGATACCTATCCTTCTTCCTACCA R: TGGTAGGAAGAAGGATAGGTATCACAGCAGAGCAG
CHD4 S851Y (M3)	F: ATTCCATGTGCTGCTGACATACTATGAATTGATCACCATTG R: CAATGGTGATCAATTCATAGTATGTCAGCAGCACATGGAAT
CHD4 L1009_V1011 del (M5)	F: CTTAAGATCCATCACCAGAGACACCTGGTTGCCA R: TGGCAACCAGGTGTCTCTGGTGATGGATCTTAAG
CHD4 R1068H (M2)	F: TAAGGAGGGTGGGCATCATGTACTCATCTTTTCCC R: GGGAAAAGATGAGTACATGATGCCACCCCTCCTTA
CHD4 V1608I (M1)	F: CAGAGGATGAAAAGGTCATTGTTGAACCCCTGAG R: CTCAGGGGGTTCAACAATGACCTTTTCATCCTCTG

Once the mutagenic primers were generated and commercially synthesized (Thermofisher, USA), the PCR reactions shown in Table 2.16. were set up. A control reaction using a pWhitescript 4.5-kb control plasmid (included in the kit) was also prepared for each mutagenesis experiment. Cycling conditions are shown in Table 2.17.:

Table 2.16. PCR reaction for site-directed mutagenesis

Component	Amount
10x Reaction buffer	5 µl
DNA template	10 ng
F primer	125 ng
R primer	125 ng
dNTP mix	1 µl
H ₂ O	To 50 µl

Table 2.17. Cycling conditions used for site-directed mutagenesis

Segment	Cycles	Temperature	Time
1	1	95°C	30 sec

2	12-18	95°C	30 sec
		55°C	1 min
		68°C	1 min/kb of plasmid length

Segment number 2 of the cycling reaction was adjusted according to the type of mutation introduced in the DNA template as follows:

- Point mutations: 12 cycles
- Single aminoacid changes: 16 cycles
- Deletions or insertions: 18 cycles

Once the cycling reaction was finished, mutagenesis reactions were placed on ice for 2 min. After cooling down, 1 µl of *DpnI* restriction enzyme (included in the kit) was added to each reaction. Following thorough mixing by pipetting the solution up and down several times, reactions were spun down for 1 minute and subsequently incubated at 37°C for 1 hour in order to digest the parental, non-mutated DNA.

Transformation of XL1-Blue Supercompetent Cells

XL-1 Blue supercompetent cells (included in the kit) were gently thaw on ice for 15 min. For each reaction, a 50 µl sample was aliquoted, and 1 µl of *Dpn I*-treated DNA was transformed into XL-1 Blue cells as indicated in section 2.1.1.7. After transformation, 50 µl of each transformation reaction were plated on agar plates containing the appropriate antibiotic for the plasmid vector (ampicillin in the case of the pcDNA3.1 vector). For the pWhitescript mutagenesis control, cells were plated on LB-ampicilling agar plates containing 80 µg/ml X-gal and 20 mM IPTG. Plates were finally incubated overnight at 37°C and the next day, plates were checked for colony formation. Successful mutagenesis reactions was indicated by the presence of 10-100 colonies in actual samples and 50-800 in the control sample. In addition to this, greater than 80% of the colonies in the control reaction should appear as blue colonies on agar plates containing IPTG and X-gal. Finally, mutated DNA vectors were purified as indicated in 2.1.1.8., and the presence of the desired mutation in the vector plasmid was verified via Sanger sequencing.

2.3.2. Expression of CHD4 mutant variants in mammalian cells

Expression vectors for each CHD4 variant were transfected into COS7 cells using Polyfect reagent (QIAGEN, Germany) as indicated in 2.1.3.6. Cells were harvested from two 100 mm plates in 1 ml of PBS, pelleted down and

resuspended in 200 µl of lysis buffer (150 mM NaCl, 20 mM TrisHCl pH 7.6, 0.5 mM EDTA pH 8.0, 1% NP-40 with 2.5 µl of 100x proteinase inhibitor cocktail). The cell suspensions were then placed under ice for 1h followed by brief sonication. Cells were then centrifuged and the supernatant (whole protein lysate) was stored at -80 °C. In order to check CHD4-FLAG expression, 20 µg of total protein lysate were fractionated on NuPAGE 10% Bis-Tris gels (ThermoFisher, USA) followed by western blotting as indicated in 2.1.2.3. CHD4-FLAG expression was detected using a mouse monoclonal Anti-FLAG antibody (Sigma-Aldrich, F1894, 1:1000 dilution).

2.3.3. Pull-down assays

For pull-down assays, CHD4-FLAG and HDAC1-MYC, HDAC2-MYC, MTA2-MYC, RBBP4-MYC or TBX5-MYC were co-transfected into COS7 cells and lysed 48 hours post-transfection (see 2.1.3.6. and 2.1.2.1.). Purification of CHD4-FLAG was performed by using Anti-FLAG M2 Affinity Beads (Sigma-Aldrich, USA) following manufacturer's protocol. Briefly, cell extracts were incubated overnight with 40 µl of washed beads (in wash buffer, 20mM Tris (pH7.6), 150mM NaCl, 0.5% NP40 and 0.5mM EDTA) at 4°C with gentle agitation. The following day, non-specific proteins were removed by washing the beads 5 times in wash buffer, 5 min per wash. CHD4-FLAG and its interacting proteins were released from the beads by incubating them with 500 µg/ml of 3x FLAG peptide (Sigma-Aldrich, USA) diluted in wash buffer for 2 hours at 4°C with gentle agitation. Purified proteins were then fractionated on 10% Bis-Tris gel followed by western blotting as indicated in 2.1.2.3. CHD4-FLAG expression was detected using a mouse monoclonal Anti-FLAG antibody (Sigma-Aldrich, F1894, 1:1000 dilution), whereas HDAC1-MYC, HDAC2-MYC, MTA2-MYC, RBBP4-MYC and TBX5-MYC expression was detected using a mouse monoclonal Anti-MYC antibody (Sigma-Aldrich, M4439, 1:1000 dilution).

2.3.4. Immunofluorescence and cell imaging

For cell localization studies, COS7 cells were grown on coverslips in 6-well plates and co-transfected with CHD4-FLAG and HDAC1-mCherry using Polyfect Transfection Reagent. 24 hours post-transfection, cells were washed with 1 ml of PBS and fixed with 1 ml of 4% paraformaldehyde (PFA) for 10 mins, followed by three washes with 1 ml PBS, 5 minutes each. Cells were then permeabilized with 1 ml of 0.25% Triton-X in PBS for 5 mins, washed 3 times with 1 ml of PBS and incubated in 1 ml of 5% BSA in PBS for 1 hour to suppress non-specific binding of IgG. Following blocking, cells were washed with 1 ml of PBS and incubated with 1 ml of mouse monoclonal Anti-FLAG antibody (Sigma-Aldrich, F1894, 1:250 dilution) diluted with PBS containing 3% BSA for 1 hour at RT. Cells were then washed 3 times with 1 ml of PBS each for 5 minutes and incubated

with 1 ml of goat anti-mouse Alexa-488 (ThermoFisher Scientific, 1:500 dilution) diluted with PBS containing 3% BSA for 1 hour at RT. Cells were again washed 3 times with 1 ml of PBS and incubated with 1 ml of Hoechst solution (Sigma-Aldrich, USA, 1:1000 dilution) diluted with PBS for 10 min in the dark for nuclei staining. Cells were washed with 1 ml of PBS, 3 times for 5 min, and coverslips were finally mounted on glass slides containing a small drop of Vectashield mounting media without DAPI (Vector Laboratories, UK). After sealing the slides with nail polish, cell imaging was performed on a Zeiss LSM880F confocal microscope (Carl Zeiss, Germany) using a 63X 1.4 oil objective. Images were captured using the ZEN software (Carl Zeiss, Germany) and analysis was performed using the Fiji software (open source).

2.3.5. Recombinant CHD4 production in E.coli

2.3.5.1. Plasmids and constructs

In order to perform a functional characterization of the CHD4 mutant variants, recombinant CHD4 protein was produced and purified in *Escherichia coli* first. A truncated version of CHD4 cDNA (covering residues 359 to 1213) was amplified via PCR from IMAGE clone 5528023 (Source Bioscience, UK) and subsequently cloned into the NcoI and SalI restriction sites of the bacterial expression vector pET28a, which also contains a 6xHis-tag for purification. Ligation products were transformed into chemically competent *E.coli*. After bacterial growth, ligated products were extracted using a GenElute Plasmid Miniprep Kit (Sigma-Aldrich, USA) and the construct pET28a-CHD4-6xHIS was verified via Sanger sequencing. Protocols followed for the ligation, transformation and purification of these constructs can be found in sections 2.1.1.4 to 2.1.1.8. For the generation of the mutant variants M2 to M5, QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies, USA) was used as indicated in 2.2.1.2. Primers used for mutagenesis were those used in 2.3.1. Introduction of specific mutations was finally confirmed via Sanger sequencing.

2.3.5.2. Transformation and production in E.coli

For expression purposes, pET28A-CHD4-6xHis or its respective mutant variants were transformed into chemically competent *E.coli* BL21 DE3 using the heat shock procedure²⁹⁴ (section 2.1.1.7.) and plated on antibiotic (ampicillin) selection plates. Following overnight incubation, a single colony was resuspended in 10 ml LB media with antibiotics and incubated at 37°C with agitation until OD600 reached 0.5-0.7. This 10 ml culture was then inoculated into a 500 ml LB culture with antibiotics, and protein expression was induced by adding 0.7 mM IPTG followed by overnight incubation at 20°C.

2.3.5.3. Protein extraction

Following protein expression induction, cells were harvested by centrifugation at 4500 rpm at 4°C for 30 min and the pellet was then frozen at -20°C in order to increase lysis efficiency. For lysis and protein extraction, the frozen cell pellet was thawed at RT and resuspended in 4 ml/g of cell pellet of B-PER Bacterial Protein Extraction Reagent (ThermoFisher, USA) until the solution was completely homogenous. Following incubation for 30 min at RT, lysate was centrifuged at 15000g for 5 min to separate soluble proteins from the insoluble ones. 10 µl samples of both the soluble and insoluble fractions (before and after induction) were analysed by fractionation on NuPAGE 10% Bis-Tris gels (ThermoFisher, USA) followed by protein staining using SimplyBlue SafeStain (ThermoFisher, USA) in order to check for the induction of CHD4-6xHis expression. Protocols for fractionation and protein staining can be found in sections 2.1.2.3. and 2.1.2.4., respectively.

2.3.5.4. Purification of CHD4-6xHis

In order to purify CHD4-6xHis protein (or any of its mutant variants), the whole protein lysate was applied to a TALON CellThru Resin Kit (Takara Bio, USA) and purification was performed according to manufacturer's protocol.

Resin equilibration

First, the TALON resin was thoroughly resuspended and the required amount of resin that will accommodate 15 times the resin bed volume was transfer to one of the provided sterile tubes. The resin was centrifuged for 2 min at 700 x g in order to pellet it, and the supernatant was then removed and discarded. Subsequently, 10 bed volumes of Equilibration Buffer was added and mixed by inversion to pre-equilibrate the resin, followed by recentrifugation for 2 min at 700 x g. The supernatant was discarded again, and the resin was washed once more with Equilibration Buffer.

Sample application

After resin pre-equilibration, the whole protein lysate obtained in 2.3.5.3. was applied to the resin, which was gently agitated on ice for 30 min on a platform shaker to allow binding of the His-tagged protein to the resin. After this incubation, the resin was centrifuged at 700 x g for 5 min, and as much supernatant as possible was removed without disturbing the pellet.

Washing

The resin was then washed twice by adding 15 bed volumes of Equilibration Buffer and incubating the suspension on ice for 10 min on a platform shaker, followed by centrifugation at 700 x g for 5 min and removal of the supernatant.

One bed volume of Equilibration Buffer was added to the resin, and this was resuspended by vortexing. The resin was then transferred to a 2 ml gravity-flow column (provided with the kit). The end-cap of the column was removed, and the buffer was allowed to drain until it reached the top of the resin bed. The column was then washed once with 5 bed volumes of Wash Buffer.

Elution

The His-tagged protein was eluted by adding 5 bed volumes of Elution Buffer to the column, and the eluate was collected in 500 µl fractions (a total of 6-8 fractions were collected typically). The presence of purified CHD4-6xHis in the eluate was detected by fractionation on NuPAGE 10% Bis-Tris gels (ThermoFisher, USA) followed by protein staining using SimplyBlue SafeStain as indicated in 2.1.2.4.

2.3.6. Recombinant CHD4 production using the baculovirus-insect cell expression system

2.3.6.1. Plasmids and constructs

In order to generate recombinant bacmid DNA for the production of CHD4 protein in insect cells, the Bac-to-Bac TOPO Cloning Kit (ThermoFisher, USA) was used following manufacturer's protocol. First, a blunt-end PCR product containing CHD4 cDNA was amplified from IMAGE clone 5528023 (Source Bioscience, UK), whereas the blunt-end PCR products containing CHD4 M1 and M2 cDNA variants were amplified from vectors pcDNA-CHD4-M1 and pcDNA-CHD4-M2 respectively. These cDNA fragments were cloned into the pFastBac/CT-TOPO vector by setting up the following reaction:

Table 2.18. Components for the ligation of CHD4 cDNA and the TOPO vector

Component	Volume
PCR product	2 µl
Salt solution	1 µl
TOPO vector	1 µl
Sterile water	2 µl

Components were mixed by pipetting up and down and the reaction was incubated for 5 min at RT. Following incubation, reactions were placed on ice and subsequently transformed into One Shot® Mach1™ T1R Chemically Competent *E. coli* by following the same protocol described in section 2.1.1.7. For the selection of colonies carrying the ligated pFastBac/CT-TOPO vector-CHD4 construct, LB-agar plates containing ampicillin were used. Plasmid DNA isolation (maxi-prep) was carried out as indicated in 2.1.1.8.

After this, the pFastBac-CHD4 constructs were transformed into the DH10Bac *E. coli* strain (as indicated in 2.1.1.7), which contains a bacmid for insect cell infection. By transposition, the CHD4 sequence contained in the pFastBac vector is incorporated into the bacmid, which will be used for posterior transfection into insect cells, virus generation and production of recombinant CHD4 protein. In this case, in order to select only the colonies carrying the CHD4 bacmid, LB agar plates containing 30 µg/ml kanamycin, 7 µg/ml gentamicin, 10 µg/ml tetracycline, 100 µg/ml X-gal and 40 µg/ml IPTG were used. Only the large, white colonies contain the recombinant bacmid, whereas blue colonies, or white colonies with a blue centre do not. Several white candidates and one blue control were picked and streaked to fresh plates (followed by overnight incubation at 37°C) in order to verify their phenotype. A small (5 ml) overnight culture in LB containing 30 µg/ml kanamycin, 7 µg/ml gentamicin and 10 µg/ml tetracycline was also started from each restreaked colony. Next day, these cultures were used to prepare glycerol stocks (500 µl of bacterial culture + 500 µl of 50% glycerol) of clones verified as white upon restreaking. Finally, glycerol stocks were stored at -80°C

2.3.6.2. Transfection of Sf9 insect cells

Bacmid DNA for transfection into insect cells was typically purified the day before transfection by using the PureLink HiPure Plasmid DNA Miniprep Kit (ThermoFisher, USA) following manufacturer's instructions. For transfection, 1×10^6 Sf9 cells were transferred into one well of a 6-well plate and left to adhere for 1 hour. Between 2 to 3 µg of bacmid DNA were diluted in 500 µl of Xpress media (serum and P/S free) for each transfection, and this bacmid DNA solution was then mixed with 5 µl of Escort IV Transfection Reagent (Sigma-Aldrich, USA) diluted in 500 µl of Xpress media (serum and P/S free). Escort-DNA mix was incubated at RT for 45 minutes to allow for complex formation. After washing the plated cells with fresh Xpress media, transfection mix was added to the cells and these were incubated for 5 hours at 28°C. Transfection mix was then removed and replaced with 3 ml of Xpress media (containing 2% FBS and 2% P/S), and plates were incubated for at least 3 days. Cells were screened after 48-72 hours for signs of successful transfection and subsequent virus infection (cells stop growing and proliferating, their diameter is increased by 20-30% and they have irregular morphology, among others). Virus was usually harvested 5 to 7 days after transfection by spinning down the transfection media (1300 rpm for 5 min) and keeping the supernatant. Virus stocks (passage 1) were protected from light with aluminium foil and stored at 4°C until further use (they are stable for 6 months at 4°C). A 0.5 ml aliquot was also stored at -80°C for long-term storage.

2.3.6.3. Generation of baculovirus Infected Insect Cells (BIIC) stocks

The classic protocol for baculovirus-mediated protein expression consists of expanding the virus stock from passage 1 to passage 3 and then use the passage 3 for large-scale protein expression. However, this procedure is quite time-consuming and the virus loses infectivity over time when stored at 4°C. In this project we used the previously established baculovirus-infected insect cells (BIIC) method³²⁴, in which cells are infected using the passage 1 virus and frozen before they lyse. Such BIIC stocks are then used for large-scale protein expression. Briefly, 50 ml of 0.5×10^6 Sf9 cells/ml were infected with 1 ml of passage 1 (P1) virus stock. Cell diameter were then monitored using the Moxi Z Mini Automated Cell Counter (Orflo, USA) at 24 and 36 hours post-infection. As soon as the mean diameter of the infected cells reached 17-18 µm, BIIC stocks were prepared by harvesting the cells at 1500 rpm for 5 min. Cell pellet was resuspended in 2.5 ml of Xpress media supplemented with 10% DMSO and 2% P/S, and the cell solution (BIIC stock) was aliquoted in cryotubes and frozen down at -80°C prior to transfer to liquid nitrogen for long-term storage.

2.3.6.4. Large-scale CHD4 protein expression and purification

For large-scale production of recombinant CHD4, 100 µl of BIIC stock was added to 300 ml of 1.5×10^6 Sf9 cells/ml diluted in Xpress media supplemented with 2% P/S and cells were incubated at 28°C with shaking (150 rpm) for 64 hours. After this, cells were spun down by centrifugation at 4000 rpm for 10 min at RT, and the cell pellet was resuspended in 2 ml/g wet weight of cell lysis buffer (20 mM HEPES pH 7.5, 200 mM KCl, 0.1% NP-40 and 10% glycerol) followed by incubation on a roller at 4°C for 1 hour. Cell lysate was then centrifuged at 20000 rpm for 1 hour at RT, and the supernatant was incubated with 1 ml of previously washed (three times in ultra-pure water) Ni-NTA Agarose Beads (QIAGEN, Germany) on a roller at 4°C for 1 hour. The cell lysate-beads suspension was transferred to a Poly-Prep Chromatography Column and the supernatant was allowed to flow-through (FT) the column and was collected in a tube situated beneath the column. The column was then washed 3 times with 5 ml of 40 mM imidazole Ni-affinity buffer (20 mM HEPES pH 7.5, 200 mM KCl, 0.1% NP-40, 10% glycerol and 40 mM imidazole) discarding the flow-through (WASH) every time. Finally, recombinant CHD4 protein was eluted by adding 2.5 ml of 200 mM imidazole Ni-affinity buffer to the column and collecting the flow-through (ELUTE) in 8-10 fractions of ~300 µl each. For storage, protein tubes were snap-frozen in liquid nitrogen and subsequently store at -80°C until further use. 10 µl aliquots of the FT, WASH and ELUTE fractions were run on NuPAGE 10% Bis-Tris gels (ThermoFisher, USA) followed by protein staining using SimplyBlue SafeStain (ThermoFisher, USA) as indicated in 2.1.2.4.

2.3.7. Nucleosome assembly

In order to perform nucleosome binding assays and nucleosome sliding assays, the Epimark Nucleosome Assembly Kit (New England Biolabs, USA) was used according to manufacturer's protocol. This kit is based on the well-known salt gradient method for nucleosome assembly³²⁵, by which histones H3-H4 tetramer and H2A-H2B dimer progressively bind to a linear fragment of DNA by decreasing the salt concentration and ionic strength subsequently. Finally, at a concentration of 0.25 M salt, the H3-H4 tetramer and H2A-H2B dimer are stably associated to the DNA, forming the nucleosome structure. For this project, nucleosomes were assembled on a 194 bp long DNA fragment which contains the '601' nucleosome position sequence, which produces end-positioned mononucleosomes with a 47 bp overhang that allow for nucleosome repositioning after remodelling by the CHD4 protein (0-601-47 mononucleosome). This 194 bp DNA fragment was amplified by PCR from the pGEM-3z/601 plasmid, which was a gift from Jonathan Widom (Addgene plasmid # 26656). Nucleosome assembly reaction was then prepared in the following order (a positive control using supplied DNA was also prepared):

Table 2.19. Components used for the nucleosome assembly reaction

Component	Volume
Water	0 to 7 µl
5M NaCl	4 µl
DNA	50 pmol
20 µM Dimer (supplied in 2M NaCl)	5 µl
10 µM Tetramer (supplied in 2M NaCl)	5 µl

Reactions were incubated for 30 minutes at RT. Then, 7 µl of room temperature Dilution buffer (10 mM Tris, pH 8.0) was added to each reaction followed by incubation at RT for 30 min. Subsequently, 13 µl of room temperature Dilution buffer was added to each reaction followed by incubation at RT for 30 min. Next, 27 µl of room temperature Dilution Buffer was added again to each reaction and these were incubated at RT for 30 min. Finally, 93 µl of room temperature Dilution Buffer was added to each reaction to bring up the final volume to 160 µl and the NaCl concentration to 0.25M. After this procedure, proper nucleosome assembly was checked on a 1.5% agarose gel. Nucleosomes were stored at 4°C until further use.

2.3.8. Nucleosome/DNA binding assay

In order to exert its nucleosome remodelling activity, CHD4 physically binds to and interacts with nucleosomes before carrying out their remodelling. This binding of CHD4 to the nucleosomes can be assessed by performing a

nucleosome binding assay, which works in a similar way to DNA binding assays: purified CHD4 protein is incubated alongside *in vitro* assembled nucleosomes (as indicated in 2.3.7) and if binding does take place, this can be visualized on an acrylamide DNA retardation gel as a shift of the nucleosome-only band to a higher position. The nucleosome and DNA-binding ability of recombinant CHD4 variants was analysed by using the 0-601-47 nucleosome and 0-601-47 free DNA respectively. A typical reaction was performed in 15 μ l total volume containing 1.5 μ l of 10x CHD4 buffer (0.5M Tris-HCl pH 7.5, 0.5M NaCl, 30 mM MgCl₂), 300 nM of 0-601-47 mononucleosome or DNA, varying concentrations of recombinant CHD4 protein (50, 100, 300 and 600 nM) and nuclease-free water up to 15 μ l. Reactions were incubated at 30°C for 30 minutes and then mixed with 4 μ l of 5x Novex Hi-Density TBE Sample Buffer (ThermoFisher Scientific, USA) before loading on 6% DNA Retardation Gels (ThermoFisher Scientific, USA). Gels were run inside a XCell SureLock Mini-Cell (ThermoFisher Scientific, USA) filled with 800 ml of 0.5X TBE at 100 V and 15 mA for 90 minutes using a PowerEase 500 Power Supply (ThermoFisher Scientific, USA). Gels were stained with 1x SYBR Gold (ThermoFisher Scientific, USA) for 40 minutes followed by two 5 min washes in water. Nucleosomes/DNA were visualized by exposing the gel to UV light inside a Gel Doc XR+ System (Bio-Rad, USA). Images were taken using the Quantity One software (Bio-Rad, USA).

2.3.9. Nucleosome sliding assay

Nucleosome sliding assays were performed under the same conditions as the nucleosome-binding assay, but since the ability of CHD4 to remodel the nucleosome was monitored, 1.5 μ l of 10 mM ATP solution (ThermoFisher Scientific, USA) was added to the reaction. In this case, reactions were pre-incubated at 30°C for 5 minutes before the addition of ATP and further incubation at 30°C for 45 minutes. After this, reactions were stopped by adding 1 μ l of competitor plasmid DNA (this is, a plasmid DNA containing the 601 nucleosome positioning sequence in order to compete away the recombinant CHD4 protein from the nucleosome) at a concentration of 200 ng/ μ l and incubation on ice for 15 minutes. Reactions were analysed on 6% DNA Retardation Gels as indicated in 2.3.8.

2.3.10. ATPase activity assay

ATPase activity of CHD4 variants was measured using an ADP-Glo™ Kinase Assay Kit (Promega, USA) following manufacturer's protocol. Reactions were carried out in 96-well plates, and three technical replicates were performed per experiment. Initially, a standard curve for conversion of ATP to ADP was generated. To do this, 1ml of 1 mM ATP and 500 μ l of 1 mM ADP were prepared by diluting the supplied Ultra Pure ATP and ADP in 1x kinase reaction buffer

(40mM Tris pH 7.5, 20mM MgCl₂, 0.1mg/ml BSA). Then, the 1 mM ATP and 1 mM ADP solutions were combined in individual wells of the plate as indicated in Table 2.20 to simulate increasing rates of ATP to ADP conversion:

Table 2.20. Preparation of the ATP+ADP standards

Well number	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
1 mM ADP (μl)	100	80	60	40	20	10	5	4	3	2	1	0
1 mM ATP (μl)	0	20	40	60	80	90	95	96	97	98	99	100

Then, 25-μl ATPase reactions were set up in individual wells of the plate by mixing 600 nM of CHD4 protein, 1 mM of ATP, 300 nM of nucleosome/free DNA and 1x kinase buffer, followed by incubation at 37°C for 30 min. After the incubation, 25 μl of ADP-Glo reagent was added to remove excess ATP and terminate the ATP-to-ADP reaction. Reactions were incubated with ADP-Glo reagents for 40 min at room temperature. Finally, 50 μl of Kinase Detection Reagents was added to each well and reactions were incubated for 60 min at room temperature to convert ADP into ATP and to introduce luciferin and luciferase. Luminescence was quantified using a Tecan Infinite 200 PRO plate reader (Tecan, Switzerland), and the relative ATPase activity of samples was calculated from the ATP-to-ADP conversion standard curve.

2.4. MOUSE WORK

2.4.1. Mouse lines

Two different transgenic mouse lines were used in this project: *Chd4*^{tm1b(EUCOMM)Wtsi} and *Cdk13*^{tm1b(EUCOMM)Hmgu}, which were generated by the Wellcome Trust Sanger Institute (UK). Both of these lines carry a LacZ tagged null allele (tm1b), which is considered a true knockout as skipping over of the LacZ cassette will no longer restore gene expression. In addition to this, the knock-in cassette expresses LacZ in tissues where the gene of interest is knocked out, which can be used to follow the tissue expression of the targeted gene following β-galactosidase staining. In the case of the *Chd4*^{tm1b} transgenic line, the missing exons are 11 and 12 (240 bp and 204 bp, respectively), whereas in the *Cdk13*^{tm1b} line the missing exons are 3 and 4 (171 bp and 140 bp, respectively). Fig. 2.10A shows the structure of the *Chd4*^{tm1a} and *Cdk13*^{tm1a} alleles (right before Cre-loxp recombination), whereas Fig. 2.10B shows the structure of the tm1b allele and how it is generated.

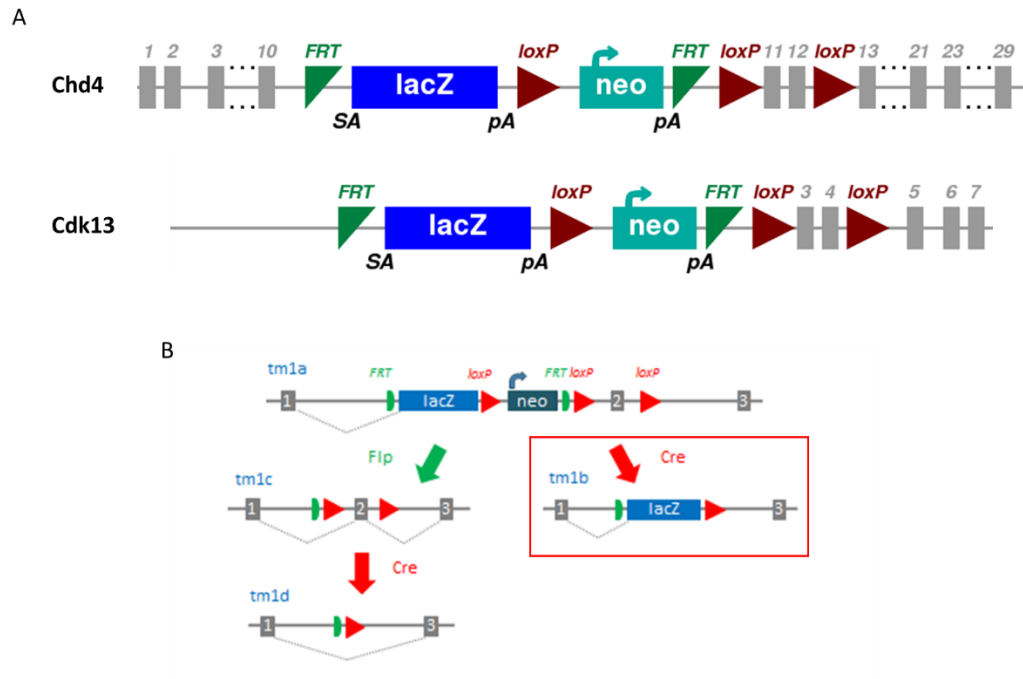


Figure 2.10. Structure of the *tm1a* and *tm1b* alleles. (A) Structure of the *Chd4*^{tm1a} and *Cdk13*^{tm1a} alleles. (B) The *tm1b* allele is generated from the *tm1a* allele after Cre recombinase-mediated excision of the critical exon and the neomycin cassette.

2.4.2. Mice maintenance and welfare

The mice studied in this project were raised and maintained by the BioSupport Unit (BSU) of The University of Nottingham according to Home Office regulations and guidelines, local and national ethical guidelines and best animal welfare practices. The breeding of genetically modified mice and relevant experimental procedures were performed in accordance with the guidelines and regulations of the Animal Procedures Act, under license number P375A76FE (Nottingham), issued 2018-08-22, and granted by The Home Office.

2.4.3. DNA extraction for genotyping

Genotyping of pups or embryos was performed by using ear notches (collected by the personnel of the Animal Facility) or the embryo carcass. In an eppendorf tube, 500 µl of lysis buffer (100mM Tris-HCl pH 8.8, 5mM EDTA pH 8.0, 0.2% SDS, 200mM NaCl) and 25 µl of 10 mg/ml proteinase K (MP-Bio, UK) was added to the ear notch/embryo carcass, and tubes were incubated in a water bath at 55°C overnight. The next day, tubes were vortexed briefly and tissue debris was spun down at 10000 rpm for 10 min. The supernatant was transferred to a clean Eppendorf tube, and an equal volume of isopropanol (500 µl) was added to precipitate the DNA. After mixing by inversion, tubes were centrifuged at 13000 rpm for 10 min and the supernatant was discarded. The genomic DNA pellet was washed with 500 µl of 70% EtOH and after discarding it, the pellet was left to air dry at room temperature for 30 minutes. Genomic DNA was

finally resuspended in 30 µl of nuclease-free water and its concentration was measured on a NanoDrop 2000 spectrophotometer (ThermoFisher Scientific, USA).

2.4.4. Genotyping

After DNA extraction, mouse samples were genotyped by PCR amplification of the targeted allele using the Platinum Taq DNA Polymerase High Fidelity (ThermoFisher Scientific, USA), following the protocol provided by the Wellcome Sanger Institute. PCR reaction setup using Platinum Taq polymerase is indicated in section 2.1.1.1., whereas the primer sequences used for each line are indicated in Table 2.21. PCR cycling conditions are summarised in Table 2.22:

Table 2.21. Primers used for the genotyping of the CHD4 and CDK13 lines

Line	WT reaction	Mutant reaction	LacZ reaction
Chd4 ^{tm1b}	F: GTCACCACCCAAGTTCCTG R: CCCTCTTCTGTCACGACATC	F: GTCACCACCCAAGTTCCTG R: TCGTGGTATCGTTATGCGCC	F: GGTAAACTGGCTCGGATTAGGG R: TTGACTGTAGCGGCTGATGTTG
Cdk13 ^{tm1b}	F: GCTCTAAGGGCAACCTTGAA R: AGCTGGGAAGATGGTGTGT	F: GCTCTAAGGGCAACCTTGAA R: TCGTGGTATCGTTATGCGCC	F: GGTAAACTGGCTCGGATTAGGG R: TTGACTGTAGCGGCTGATGTTG

Table 2.22. PCR cycling conditions for the genotyping of Chd4^{tm1b} and Cdk13^{tm1b} lines using Platinum polymerase

		Chd4 ^{tm1b}	Cdk13 ^{tm1b}
Initial denaturation		94°C for 5 min	94°C for 5 min
35 PCR cycles	Denaturation	94°C for 30 sec	94°C for 30 sec
	Annealing	58°C for 30 sec	58°C for 30 sec
	Extension	72°C for 45 sec	72°C for 45 sec
Final extension		72°C for 5 min	72°C for 5 min
Hold		4°C indefinitely	4°C indefinitely

After amplification, PCR reactions were run on a 2% agarose gel for genotype analysis, as indicated in 2.1.1.2.

2.4.5. Embryonic heart dissection

For embryonic heart dissection, pregnant females were culled by cervical dislocation. The mouse was then placed with its belly facing upwards and after spraying it with 70% ethanol, the skin was lifted over the belly and a small horizontal incision performed using surgical scissors. The incision was then enlarged and the string of embryos was pulled out and transferred to a Petri

dish containing warm PBS. Embryos were dissected out of the uterus by using a pair of surgical forceps, making sure that all membranes were properly removed. For E14.5 embryos or older, decapitation was also performed to ensure the death of the embryo. The umbilical cord was also removed in order to promote exsanguination. When the downstream application required the full removal of blood (e.g. HREM), embryos were placed inside a 15 ml tube containing warm PBS and washed for 10-15 min with gentle agitation. Embryos were then transferred to a container with fresh, warm PBS and heart dissection was performed using a pair of surgical forceps. Briefly, the thoracic cavity was opened by cutting the costal cartilages at their point of union with the bone, starting from the last rib and proceeding until the sternum was reached. With the help of forceps, both sides of the thoracic cavity were pinned to expose the heart, and this was removed by pinching right underneath it and pulling upwards. Any lung tissue still attached to it was then carefully removed, and the heart was snap-frozen in liquid nitrogen or placed in a tube with warm PBS depending on the downstream application. The carcass was routinely saved for genotyping purposes.

2.4.6. β -Galactosidase staining of whole embryos/organs

β -Galactosidase staining of whole embryos or individual organs was performed on samples from the *Chd4*^{tm1b} and *Cdk13*^{tm1b} lines, as they both carry a lacZ reporter gene into the knockout construct. Staining was typically performed on whole embryos up to embryonic day 12.5-13.5, after which is advisable to isolate organs or perform sectioning as the penetration of fixative and staining reagents becomes much more difficult. The protocol used in this thesis is based on one previously published³²⁶. After embryo harvesting, these were placed on individual wells of a 24-well plate and washed in PBS three times at RT for 10 minutes each. Embryos were then fixed in cold LacZ fixative solution (1% PFA, 0.2% glutaraldehyde, 2 mM MgCl₂, 5 mM EGTA pH 8.0 and 0.2% NP-40 in 1X PBS) for 5 to 15 min (depending on embryo stage) and subsequently washed in LacZ wash solution (2 mM MgCl₂, 0.01% sodium deoxycholate and 0.02% NP40 in 1X PBS) 3 times for 30 minutes each at RT. Staining of the embryos was performed by incubating them in LacZ staining solution (5 mM potassium ferrocyanide, 5 mM potassium ferricyanide, 2 mM MgCl₂, 0.01% sodium deoxycholate, 0.02% NP40 and 1 mg/ml X-gal) overnight at 37°C. The next day, embryos were washed 3 times in PBS at RT for 20 minutes each and post-fixed in 4% PFA (in 1x PBS) overnight at 4°C prior to imaging, which was performed using a Zeiss Stereo Lumar V12 microscope with a camera. All washing steps were carrying out with 1 ml of the corresponding solution.

2.4.7. HREM Sample Preparation

In preparation for embedding and HREM imaging, samples are initially dehydrated and subsequently infiltrated with a JB-4 resin-dye mix. The exact procedure, incubation times and conditions depends on the size and tissue density of the samples processed. The methods described below serve as a general guideline of the whole procedure for embryonic and adult mouse hearts: however, some adjustments might be required for the processing of other samples.

2.4.7.1. JB-4 dye mix preparation

For the preparation of the JB-4 dye mix, JB4 solution A (Sigma-Aldrich, USA), benzoyl peroxide (Sigma-Aldrich, USA), eosin B (Sigma-Aldrich, USA) and acridine orange (Sigma-Aldrich, USA) were mixed as indicated in Table 2.23:

Table 2.23. JB-4 dye mix volumes

Final volume	100 ml	200 ml	300 ml
Solution A (ml)	100	200	300
Benzoyl peroxide (g)	1.25	2.50	3.75
Eosin B (g)	0.275	0.550	0.825
Acridine Orange (g)	0.056	0.113	0.169

Solution A was measured and placed on a stirrer. While mixing vigorously, benzoyl peroxide was slowly added to avoid the formation of lumps. Eosin B and acridine orange dyes were then added slowly, and the solution was left to stir overnight at room temperature. The next day, the mix was filtered through a 0.22 µm PES membrane (Sigma-Aldrich, USA) to remove any dust and undissolved dye. After this, the JB-4 dye mix was stored at 4°C for 3 weeks.

2.4.7.2. Preparing mouse embryos hearts for HREM

For imaging purposes, the amount of blood inside mouse embryos hearts which will undergo HREM analysis has to be reduced to a minimum. In order to achieve this, hearts removed from the body cavity were kept immersed in PBS at 37°C (with occasional flipping over to avoid pooling of blood on one side of the embryo) until the heart stops beating. Hearts were then fixed in freshly diluted 4% PFA for 30 min at RT, and subsequently washed extensively in tap water to remove PFA. Hearts were then placed on 15 ml tubes filled with water and incubated on roller bars for 1 hour, or until all blood was lysed (indicated by the white coloration of the heart). Typically, the water was changed 3 or 4 times during the 1-hour incubation. Finally, hearts samples were transferred back to 4% PFA for overnight fixation.

2.4.7.3. Dehydration of samples

Heart samples fixed in 4% PFA were first dehydrated before embedding in the JB-4 dye mix. To achieve this, hearts were washed 4 times for 30 minutes (embryonic hearts) or 2 hours (adult hearts) in PBS, followed by 45 minutes (embryo) or 1 hour (adult) washes in a series of MeOH/PBS solutions (10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% and 100% MeOH). After the MeOH washes, hearts were incubated overnight at 4°C in a 50% MeOH/50% JB-4 dye mix solution. The next day, they were transferred into 100% JB4 dye mix and incubated overnight (embryonic heart) or during 3-4 days (adult heart) at 4°C to allow for dye infiltration.

2.4.7.4. JB-4 embedding

Following infiltration with JB4 dye mix, samples were embedded prior to sectioning and imaging. At room temperature, 0.6 ml of Solution B (Sigma-Aldrich, USA), which catalyzes the polymerization of the JB-4 dye mix, was added to 10 ml of cold JB-4 dye mix and mixed vigorously. One well of the mould (Electron Microscopy Sciences, USA) was filled with the polymerising mix, and the heart sample was placed into the well immediately afterwards, letting it sink into the JB-4 polymerising mix. Usually we fitted 3 embryonic hearts or one adult heart in one well. Using glass capillaries, hearts were repositioned so the heart apex points up and the remnants of the great vessels and tissue sits on the bottom of the well. This had to be done in a 5-10 minutes time window, as after this the viscosity of the JB-4 mix greatly increases. Once the hearts were correctly positioned and were held in place, a plastic chuck (Indigo Scientific, UK) was placed on top of the well, ensuring that enough mixture is present to fill up the central hole in the chuck. A drop of mineral oil was also applied to the central hole in the chuck after the JB-4 mixture had been left to set for an hour or two. Samples were then left overnight at RT to allow for the complete polymerisation of the JB-4 mix and formation of the block.

2.4.7.5. Preparing JB-4 blocks for imaging

The next day, the mould was inverted onto paper towels to remove any unpolymerised JB-4 mix. Blocks were then removed from the mould by gripping the plastic chuck with an adjustable wrench and pulling each block vertically, and the excess resin around the outside of the chuck was removed with a blunt knife. Finally, blocks were baked at 95°C overnight and then cooled down in a fridge for several hours/days to ensure that they were completely hard before sectioning.

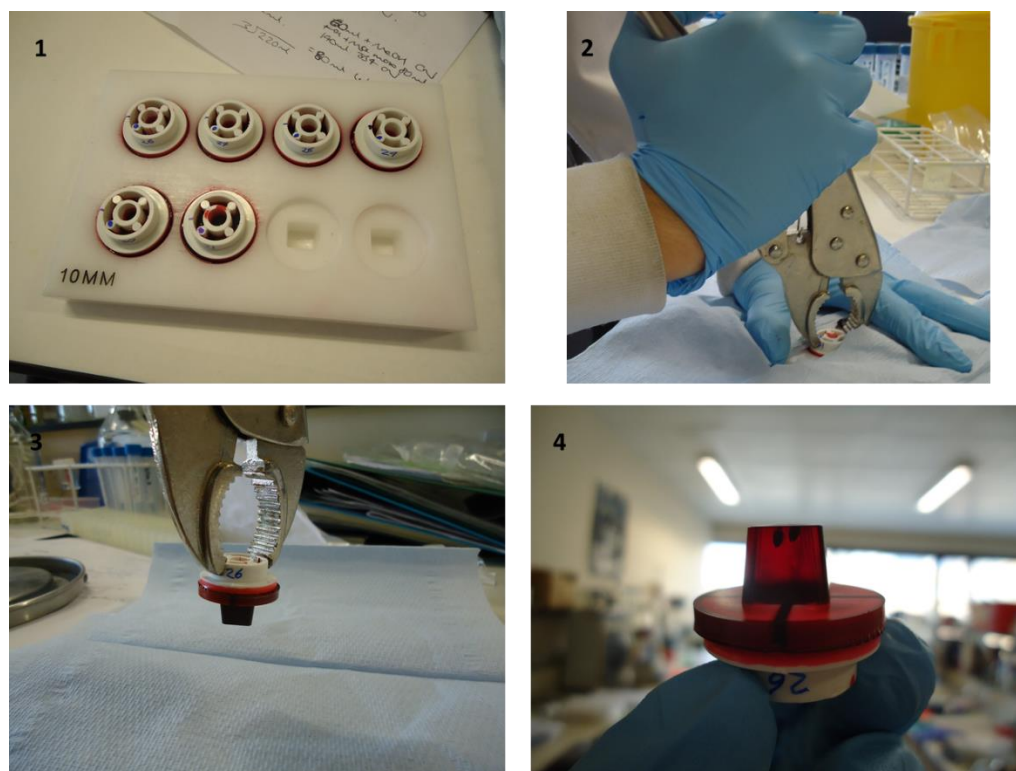


Figure 2.11. Image showing a properly embedded embryonic heart sample prior to imaging. Samples are embedded in 8-well moulds and removed using an adjustable wrench. The block shown in 4 contains two embryonic hearts.

2.4.8. High-resolution episcopic microscopy

Sectioning and imaging of the blocks were performed in Prof. Tim Mohun's laboratory (The Francis Crick Institute, London) using a HREM apparatus (Indigo Scientific, UK), which is comprised of a sectioning device (typically a microtome), a fluorescence optics system, a device for focusing and adjusting the field of view and a digital camera, which is connected to a PC running data generation software (Optical-HREM) which controls sectioning, image capturing and image storage. Briefly, the JB-4 resin blocks were mounted in the HREM apparatus and sectioned (3 μM /section). After every section, a digital image of the surface of the block was captured by the digital camera. This sectioning procedure was performed in a room at 20° Celsius, and on average it took around 8 hours per block to be completed. Data processing, visualization and 3D reconstructions were performed by Dr. Siobhan Loughna, Dr. Anna Wilsdon and Qazi Waheed (The University of Nottingham).

2.5. STATISTICAL ANALYSIS

To determine the statistical significance of the results obtained in this thesis, two different statistical analysis were carried out using Microsoft Excel or

GraphPad: two-tailed, unpaired T-test for analysis of two independent conditions with equal variance (in order to assess the statistical significance of the difference between two sample means) ; or chi-squared test (to assess whether there is a statistically significant difference between the expected frequencies and the observed frequencies). A p-value of <0.05 was considered to be significant. Results are expressed as mean $\pm$ standard deviation (SD). All results presented are calculated from at least three independent experiments unless stated otherwise.

3. ACETYLATION OF TBX5 AND ITS ROLE DURING CARDIAC AND FIN DEVELOPMENT

3.1. Introduction

3.1.1. Acetylation as a post-translational modification

Forming a fully-functional heart that pumps blood for approximately 80 years is a really difficult and challenging task. Thus, heart development is a precisely regulated process where multiple specific gene programmes must be tightly modulated in response to physiological or pathological signalling^{327,328}. This translates into the need for multiple levels of transcriptional control, which occur through the modification of chromatin structure that in turn regulates the accessibility of DNA to transcriptional regulatory factors³²⁹. The main factor influencing chromatin structure and the control of gene expression is the extent to which lysine residues within the N-terminal tails of the histone core are acetylated and deacetylated³³⁰. Histones are a family of basic proteins found in the cell nucleus that associate with and package DNA to help condense it into chromatin³³¹. Their tails are subjected to reversible post-translational modifications (PTM), which modulate their structure and thus, chromatin structure and gene transcription³³².

One of the most widespread post-translational modifications of histones is acetylation/deacetylation, which refers to the process of introducing or removing an acetyl group of the N-terminal end of the protein³³⁰. Acetylation within the histone tails leads to the neutralization of their natural positive charge, thereby relaxing chromatin structure and increasing the accessibility of transcriptional factors and regulators that will activate gene expression in most cases. Conversely, deacetylation of histone tails favours a condensed chromatin state, thereby promoting transcriptional repression^{333,334}.

3.1.2. HDACs and HATs in cardiac development

Histone acetylation is a highly dynamic process regulated by two families of antagonistic enzymes, histone acetyltransferases (HATs) and histone deacetylases (HDACs), which catalyse the acetylation/deacetylation of histone tails and function as transcriptional activators and repressors respectively³³⁰. The balance of the acetylated/deacetylated status of histones is a crucial mechanism to maintain correct gene expression during developmental processes, such as muscle differentiation and cardiac hypertrophy^{335,336}.

3.1.2.1. Histone deacetylases (HDACs)

HDACs are a large protein family comprised of 11 different proteins in mammals, and all of them share a highly conserved deacetylase domain³³⁷.

HDACs are classified into four subfamilies (Class I, IIa, IIb and IV) based on their structure, enzymatic function, subcellular localization and expression pattern. Classes I and IIa are the subfamilies that have been most extensively studied and characterized. The class I HDAC family is comprised of HDAC1, 2, 3 and 8. They are expressed ubiquitously and localized predominantly inside the cell nucleus^{337,338}. HDAC1 and 2 are usually found together in repressive complexes such as NuRD (nucleosome remodelling deacetylase) or Swi-independent³³⁹, whilst HDAC3 generally associates with N-CoR and SMRT to also form a potent repressive complex³⁴⁰. No complex has been described for HDAC8.

Class II HDACs are subdivided into class IIa (HDAC4, 5, 7 and 9) and class IIb (HDAC6 and 10). Class IIa HDACs have large N-terminal domains with highly conserved binding sites for the transcription factor MEF2 (myocyte enhancer factor 2) and the chaperone protein 14-3-3^{337,338}. Furthermore, their activity and binding capacity is regulated through phosphorylation by kinases such as calcium/calmodulin-dependent protein kinase (CAMK)³⁴¹ and protein kinase D (PKD)³³¹. The regulated phosphorylation of this family of HDACs provides a mechanism for linking extracellular signals with transcriptional regulation, having a crucial role in several tissues during development and disease. Regarding expression patterns, unlike class I HDACs, class IIa HDACs are only expressed in certain tissues. Hdac5 and Hdac9 are highly enriched in muscles, heart and brain^{343,344}; Hdac4 is mainly expressed in brain, heart and growth plates of the skeleton³⁴⁵; and Hdac7 is enriched in endothelial cells and thymocytes³⁴⁶.

Biochemical studies, as well as the generation of loss-of-function models, have given insight into the role of HDACs during cardiac development. Myocardial deletion of Hdac1 or Hdac2 in mice results in no apparent cardiac defects³⁴⁷. However, deletion of both genes together in the myocardium leads to severe cardiac phenotype, suggesting a high level of redundancy of these HDACs during development³⁴⁷. More specifically, conditional deletion of Hdac1 and Hdac2 in the heart results in neonatal lethality accompanied by cardiac arrhythmias, dilated cardiomyopathy and altered gene expression of cardiac genes³⁴⁷. Regarding other class I HDACs, it was also found that deletion of Hdac3 in cardiomyocytes results in lethality around 3-4 months after birth. These mice show cardiac hypertrophy, fibrosis, abnormal fatty acid metabolism and lipid accumulation in heart muscle³⁴⁸. Mouse embryos lacking Hdac3 in cardiac progenitor cells exhibit precocious cardiomyocyte differentiation, severe cardiac developmental defects, upregulation of Tbx5 target genes and embryonic lethality³⁴⁹. In contrast, overexpression of Hdac3 in the heart leads to increased thickness of the myocardium due to hyperplasia³⁵⁰.

Class II HDACs seem to have a role in the control of cardiovascular growth and function. Single Hdac5- or Hdac9-null mice present no apparent cardiac phenotype; however, when these mice are exposed to cardiac stress, they developed increased cardiac hypertrophy³⁴⁴. When both Hdac5 and Hdac9 are depleted together, these compound mutant mice die at E15.5 and they present frequent VSD and thin-walled myocardium, which usually arise from defects in growth and maturation of cardiomyocytes³⁴⁴. Given the fact that class II HDACs interact with MEF2, and that this factor plays a central role in cardiomyocyte differentiation, the cardiac defects found in these mice could be due to an overexpression and super activation of Mef2. Loss-of-function models for the remaining class II HDACs have proved that only Hdac7 is involved in cardiovascular development, as knockout mice for this deacetylase show embryonic lethality as well as loss of endothelial-cell interaction, rupture of blood vessels and haemorrhaging³⁴⁶.

3.1.2.2. Histone acetyltransferases (HATs)

The proteins that act in an antagonistic way to HDACs are the histone acetyltransferases (HAT), which add acetyl groups to histone tails and thus activate gene expression. HATs are divided into five different families based on structural and functional similarity of their catalytic domains: the GCN5 N-acetyltransferases (GNATs), the MYST-related HATs, the KAT3 acetyltransferases; the general transcription factor HATs; and the nuclear hormone-related HATs SRC1 and ACTR³⁵¹. The two most predominant families are GNAT and MYST; however, certain proteins of the KAT3 family (such as p300) also play important roles during developmental processes.

GNAT family

The GNAT family consists of acetyltransferases that are similar in structure and sequence to GCN5. Proteins that belong to this family share several function domains, including an acetyltransferase domain, a C-terminal bromodomain that interacts with lysine residues, and an N-terminal domain of variable length³⁵¹. This family includes several proteins such as GCN5 (or KAT2A), PCAF (or KAT2B), HAT1, ELP3 and NUT1 among others³⁵¹.

The two most well-characterized members of this family, and two of the major mammalian acetylases, are GCN5 and PCAF (from now on, they will be referred to as KAT2A and KAT2B respectively). Both KAT2A and KAT2B present the typical secondary structure of the members of the GNAT family, including a N-acetyltransferase domain and a bromodomain (Fig. 3.1). They both share a high degree of homology, presenting 73% amino acid sequence identity and conserved domains³⁵². Many publications have reported the physiological roles

of these two acetyltransferases, showing that they are involved in many different biological processes such as cancer³⁵³, genome stability³⁵⁴, embryonic development³⁵⁵ and brown tissue adipogenesis³⁵⁶ among others. During development, these two acetyltransferases are also part of different signalling pathways such as Notch and retinoic acid pathways^{357,358}. In the adult mouse, Kat2a and Kat2b are expressed ubiquitously (including heart, lung, liver and brain among others) but in complementary amounts^{355,359}. During development, Kat2a is expressed as early as E7.5, and it can be detected all the way up to E16.5. In contrast, Kat2b expression is relatively low during development, but it is highly expressed after birth and in adult tissues. In terms of spatial expression pattern, the information available is quite limited, as only certain time points during development have been checked: Kat2a is expressed uniformly throughout the embryo from E7.5 to E12.5, whereas Kat2b is first detected at E12.5 in a widespread manner as well, including cardiac, liver and skeletal muscle tissues^{355,359}.

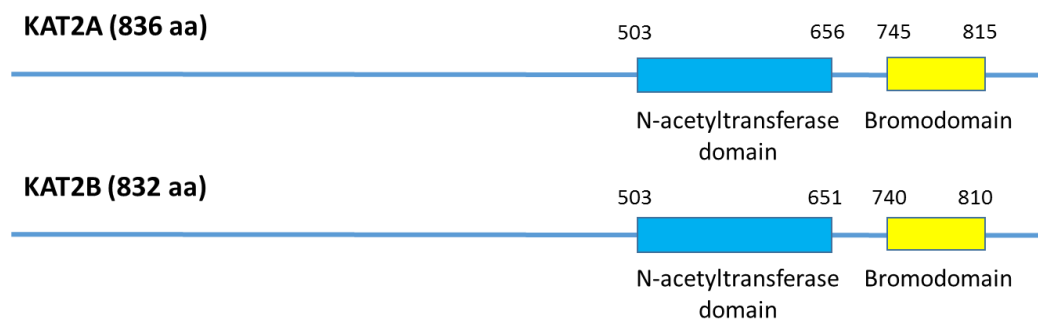


Figure 3.1. Secondary structures of the KAT2A and KAT2B proteins. Both KAT2 proteins are comprised of two domains: a N-acetyltransferase domain (blue) and a bromodomain (yellow). The position (amino acids number) of each domain in both proteins are also displayed. Aa = amino acids

Regarding mouse models, two studies in the early 2000s demonstrated that *Kat2a* knockout mice die during embryogenesis around E10.5 due to a failure in the formation of dorsal mesoderm lineages³⁵⁵, whereas *Kat2b* knockout mice are completely normal and survive until adulthood without apparent defects³⁵⁹. When both *Kat2a* and *Kat2b* are ablated together, embryos die around E7.5 instead, and the developmental defects found in embryos are much more aggravated³⁵⁵, thus suggesting the functional redundancy and overlapping functions of these two acetyltransferases during development. A different *Kat2a* knockout model carrying two point mutations in the HAT domain of the protein leads to embryonic lethality at E16.5 and cranial neural tube defects, suggesting that the HAT activity is crucial during development³⁶⁰.

MYST family

The MYST family of HATs is named after four of its members: MOZ (or KAT6A), YBF2 (or SAS3), SAS2 and TIP60³⁶¹. These HATs are characterized by the presence of a highly conserved MYST domain composed of an acetyl-CoA binding motif and a zinc finger, whereas some of the family members also have additional domains such as chromodomains or PHD-type zinc fingers³⁶¹. This family of proteins is involved in many different biological processes such as DNA repair, cell-cycle control, stem cell homeostasis and development^{362,363}, and it has also been associated with human diseases including cancer³⁶⁴. Regarding cardiac development, there are a few studies linking some MYST acetyltransferases to this process. For example, Tip60 is highly expressed in the embryonic myocardium between E8.5 and E11.5, and it has been demonstrated that it interacts with cardiac transcription factors such as Tbx5 and Srf to activate target genes³⁶⁵. In addition to this, it has also been shown that the chronic ablation of Tip60 from the ventricular myocardium beginning at the early stages of neonatal heart development leads to cardiomyocyte death after 8 weeks³⁶⁶. Another member of this family involved in cardiac development is MOZ. Voss et al. reported that Moz is crucial for the activation of the cardiac transcription factors Tbx1 and Tbx5³⁶⁷, and more importantly, *Moz* homozygous mutant mice partially phenocopy DiGeorge syndrome, characterized by the presence of CHDs, craniofacial defects and absent of thymus³⁶⁷. In-depth analysis of these mice revealed that mesodermal expression of Moz is necessary for cardiac septum development³⁶⁸.

P300 and CBP

Among all the remaining HATs reported to date, P300 and its closely related family member CBP (KAT3 family of HATs) are also tightly linked to cardiac development. Expression studies on these genes demonstrated that both are expressed in the heart throughout development, with p300 being expressed at early stages and CBP at later ones³⁶⁹. Homozygous deletion of the *p300* gene in mouse leads to embryonic lethality between E9.5-E11.5 and defects in cardiac muscle differentiation and trabeculation among others³⁷⁰, whereas complete ablation of *Cbp* also results in embryonic lethality and developmental defects similar to those of the *p300* knockout³⁷⁰. In a different study, it was shown that transgenic expression of p300 in the adult heart results in cardiac hypertrophy and heart failure³⁷¹. It has been also shown that P300 is able to regulate the transcriptional activity of multiple cardiac transcription factors such as TBX5³⁴⁹, GATA4³⁷² and MEF2C³⁷³ via acetylation.

3.1.2.3. Acetylation of non-histone proteins

Although HATs and HDACs are primarily known for their histone-acetylation and deacetylation capacity, they are also capable of regulating non-histone protein activity. This has caused HATs to be reclassified as lysine acetyltransferases (KATs). Thanks to *in vitro* studies, several non-histone substrates have been identified as targets for KATs, including transcriptional activators such as E2F1, p53, EKLF and HNF-4; or structural proteins such as tubulin, polyamines or members of the importin- α family³⁷⁴. Furthermore, KATs also acetylate cardiac transcription factors such as GATA4 and MEF2C, thereby regulating myocardial cell hypertrophy and cardiomyocytes differentiation^{375,376}, and it has also been shown that acetylation of GATA4 triggers the differentiation of embryonic stem cells into cardiomyocytes³⁷⁷. Recent studies also suggest that p300 and Hdac3 regulate early cardiogenesis by governing Tbx5 activity³⁴⁹, and a possible interaction between Kat2b, p300 and Tbx5 have also been proposed³⁴⁹.

3.1.3. Hypothesis and aims of this project

One of the main cardiac factors governing cardiac development is TBX5, mutations in which cause Holt-Oram Syndrome in humans⁷⁷. However, how TBX5 activity is regulated during cardiogenesis remains largely unknown. Thus, the main goal of this project is to identify TBX5 upstream and downstream genes, and proteins with which it interacts to elucidate its precise role in cardiac development.

The initial experiments for this project were initially undertaken by Dr. Tushar Ghosh (post-doc in Prof. David Brook's lab), who generated promising preliminary data suggesting that acetyltransferases KAT2A and KAT2B physically interact with and acetylate TBX5 *in vitro* (unpublished data at the time). These preliminary results served as the starting point for this project. Taking them into account, as well as the growing evidence suggesting that acetylation/deacetylation of cardiac transcription factors participate in the regulation of cardiogenesis, we hypothesize that KAT2A and KAT2B acetylases regulate heart development through their interaction with TBX5. In order to unravel this potential relationship, we established the following aims:

1. To further confirm the physical interaction and acetylation of TBX5 by KAT2A/KAT2B acetylases
2. To study the effect of acetylation on TBX5 transcriptional activity
3. To map the acetylation residue in the TBX5 protein, and based on this, to try to decipher the TBX5 regulation model via acetylation

4. To knockdown/knockout *kat2a* and/or *kat2b* using morpholinos and the CRISPR/Cas9 system in an *in vivo* context (zebrafish)
5. To study and characterize the cardiac and limb phenotype of *kat2a/kat2b* KD/KO zebrafish embryos

3.2. Preliminary results & data

Preliminary experiments in this project were performed by Dr. Tushar Ghosh (David Brook's lab), who performed an initial screening of the potential interaction between TBX5 and different mammalian acetylases. These experiments showed promising data regarding the potential acetylation of TBX5 by KAT2A, KAT2B and P300, and this is why these three acetylases were chosen for further experiments in this project. The experiments showed in sections 3.2.1., 3.2.2. and 3.2.3. were originally designed and carried out mainly by Dr. Tushar Ghosh and Sarah Buxton and were already on-going when I joined the lab. However, they will be shown in this section (permission was granted by Dr. Tushar Ghosh and Sarah Buxton) as they constitute the basis of the experiments that I carried out built upon, they serve to tell a more complete and coherent story, and as part of my initial training period, I was involved in reagents preparation, results interpretation, technical support and in carrying out some of the replicates in some cases (my specific contribution is indicated for each one of the experiments). The experiments showed in section 3.3 ("Results") were planned, designed and carried out by myself.

3.2.1. KAT2A and KAT2B interact with and acetylate TBX5

In view of the promising results obtained in the preliminary experiments, as well as the growing evidence supporting an important role for acetylation/deacetylation of cardiac transcription factors in transcriptional regulation^{372,376,378}, our first goal was to confirm whether histone acetyltransferases KAT2A, KAT2B and P300 are able to interact physically with TBX5 and acetylate it. To do this, Dr. Tushar Ghosh carried out pulldown experiments (three independent experiments were carried out in total) in COS7 cells using plasmids encoding the acetylases KAT2B (KAT2B-HA), KAT2A (KAT2A-HA), P300 (P300-FLAG) and TBX5 (TBX5-FLAG). Following pulldown and western blot analysis using an anti-acetyl lysine antibody, it was shown that TBX5 was specifically acetylated by KAT2B (Fig. 3.2A) and KAT2A (Fig. 3.2C), but not by P300 (Fig. 3.2A). Treatment of the cells with Trichostatin A (TSA) also potentiated the acetylation of TBX5 by KAT2B (Fig. 3.2A). By performing another set of pulldown experiments in COS7 cells (three independent experiments), it was also shown that both KAT2B and KAT2A are able to physically interact with TBX5, as demonstrated by the fact that when

transfected together, TBX5 is able to pulldown both acetylases from the lysate (Fig. 3.2B and 3.2D).

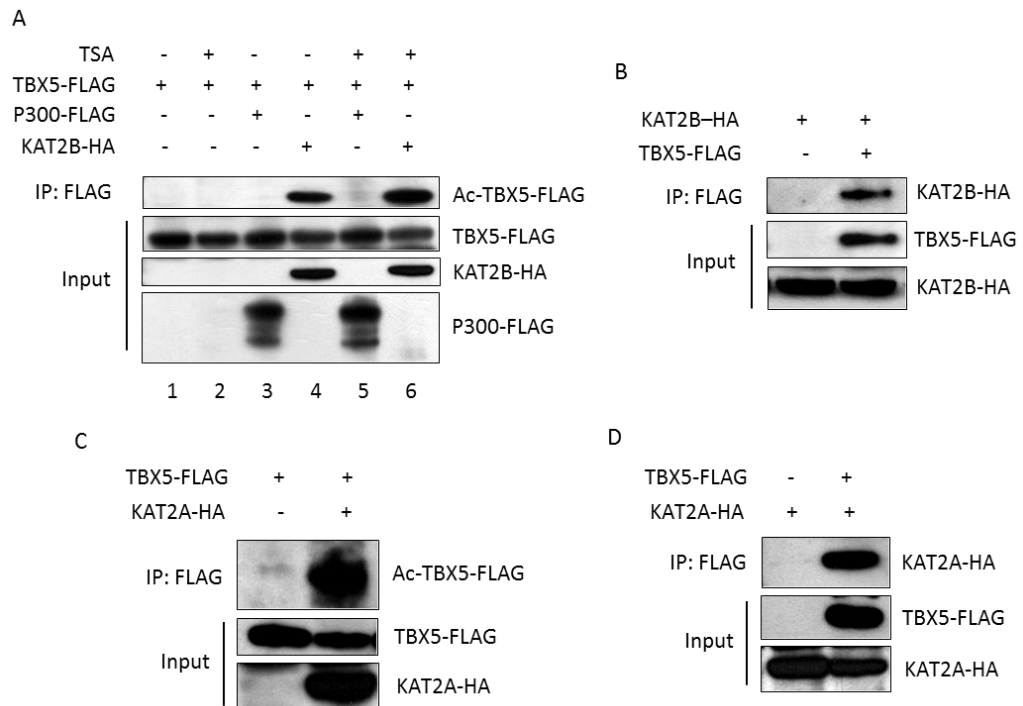


Figure 3.2. TBX5 associates with, and is acetylated by KAT2B and KAT2A. (A-D) Western blots of pull down assays show that (A) KAT2B acetylates TBX5 (lane 4 and 6), whereas p300 does not (lanes 3 and 5). TSA enhances the acetylation level of TBX5 (lane 6). (B) TBX5 and KAT2B physically associate. (C) TBX5 is acetylated by KAT2A. (D) Physical association of TBX5 and KAT2A. IP- immunoprecipitation, TSA-Tricostatin A. Three independent experiments were performed. Experiments were designed and carried out by Dr. Tushar Ghosh, and I was involved in reagents preparation, technical support, and data interpretation. Figure images and legend were generated by Dr. Tushar Ghosh and shown here with his permission

3.2.2. TBX5 acetylation potentiates its transcriptional activity

After confirming that KAT2B and KAT2A are able to interact with and acetylate TBX5, we next wondered whether TBX5 acetylation has an impact on its transcriptional activity. In order to achieve this, luciferase-reporter assays were performed on C2C12 cells by transfecting plasmids encoding KAT2B and TBX5 as well as the reporter plasmid pGL3-MYH6-I, which harbours the luciferase cDNA under the control of the MYH6 promoter, which is a well-characterized target of TBX5^{73,379}. Three independent experiments were performed (two of them performed by Dr. Tushar Ghosh and one by myself). Following luciferase activity measurement and pooled data analysis, it was demonstrated that in the presence of either KAT2B or KAT2A, TBX5 transcriptional activity on the MYH6 promoter was significantly increased in comparison to the TBX5-only condition (10.3-fold activation in the presence of KAT2B and 10.8-fold

activation in the presence of KAT2A vs 6.2-fold activation in the TBX5-only condition; $p < 0.05$ in both cases) (Fig. 3.3). This potentiation of TBX5 activity was indeed due to acetylation, as transfection of KAT2B-only or KAT2A-only did not significantly enhanced the activation of the MYH6-promoter in comparison to basal conditions (1.4-fold activation in the KAT2B-only condition and 1.5-fold activation in the KAT2A-only condition vs 1-fold activation in basal conditions; $p > 0.05$ in both cases) (Fig. 3.3).

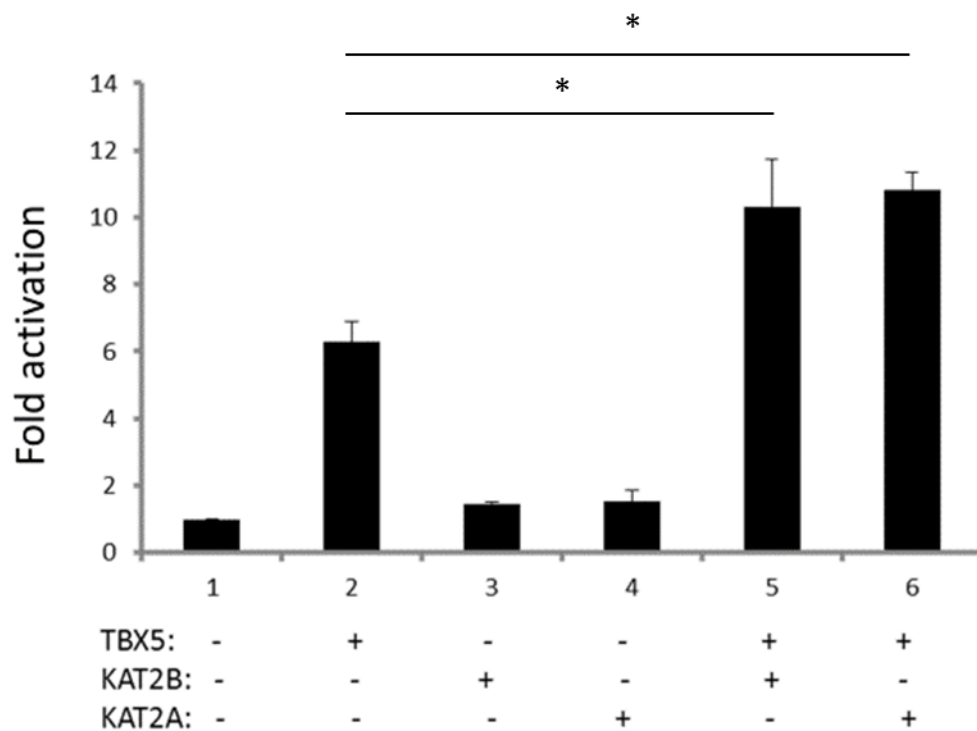


Figure 3.3. KAT2A and KAT2B potentiate TBX5 transcriptional activity. Reporter assays show that in the presence of KAT2B or KAT2A, TBX5-mediated transcription on a MYH6 promoter is significantly enhanced. No significant enhancement of the MYH6 promoter activation is found in the presence of KAT2B or KAT2A only. Results are from three independent experiments, two of them performed by Dr. Tushar Ghosh and one by myself. Mean $\pm$ SD is represented, * $P < 0.05$ (Student's t test).

3.2.3. Mapping the acetylation residue in TBX5

Following confirmation that TBX5 is acetylated by KAT2B and KAT2A, and that acetylation potentiates TBX5 transcriptional activity on the MYH6 promoter, we next wanted to map the acetylation residue in the TBX5 protein. In order to achieve this, the TBX5 protein sequence was analysed on three different web-based acetylation prediction sites: PREDMOD (<http://ds9.rockefeller.edu/basu/predmod.html>), PAIL (using a medium stringency performance mode) (<http://bdmpail.biocuckoo.org/>) and PHOSIDA (<http://141.61.102.18/phosida/index.aspx>). As shown in Fig 3.3A-C, each web-based tool picked up different potential acetylation residues in the TBX5

protein sequence; however, lysine 234 was the only one picked up by all three sites, thus suggesting that this is the acetylation residue in the TBX5 protein. In order to confirm this, Dr. Tushar Ghosh carried out a conservative substitution of lysine 234 to arginine, and subsequently checked via western blot whether this mutant variant could still be acetylated by KAT2B. As shown in Fig. 3.4D, the mutant variant K234R was acetylated by KAT2B, thus indicating that lysine 234 is not the acetylation residue in the TBX5 protein.

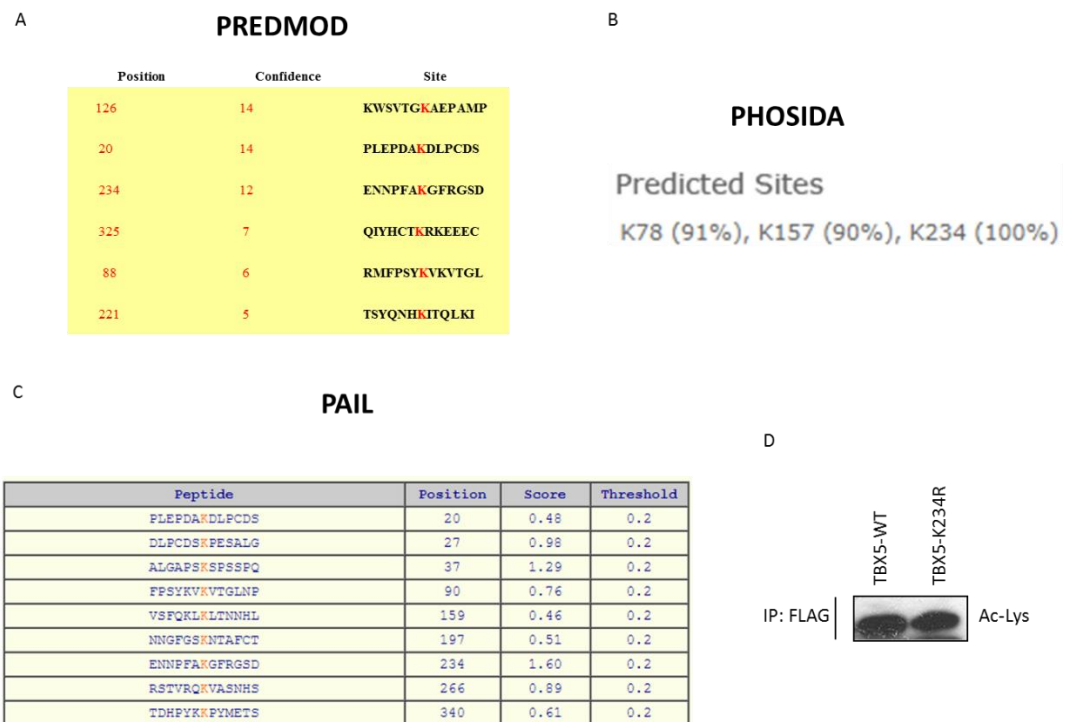


Figure 3.4. Prediction of the acetyl lysine residue in TBX5. Web-based tools PredMod (A), PHOSIDA (B) and PAIL (C) were used in order to predict potential TBX5 acetylation residues. As shown in the figures, lysine 234 was picked up by all three sites. (D) Western blot analysis showing acetylation of both TBX5 and mutant TBX5 K234R, suggesting that Lys234 is not the target for acetylation. The western blot showed in D was performed by Tushar Ghosh (shown here with his permission).

In view of these results, a different approach was followed: by searching in the literature for proteins that are acetylated by KAT2A and KAT2B, short amino acid sequences including six amino acids either side of the key acetylated residue were extracted, thus leading to 13 amino acid sequences. In total, 25 acetylation sequences were extracted by Dr. Tushar Ghosh and Sarah Buxton (Fig. 3.5A), and they were aligned in order to perform a frequency plot analysis that would allow us to generate a KAT2A-KAT2B acetylation consensus motif. Frequency plot analysis using the web-tool WebLogo (<https://weblogo.berkeley.edu/logo.cgi>) suggested an enrichment of a lysine (K) residue at position -3, a lysine (K) or arginine (R) at position +1, and a proline (P), leucine (L) or lysine (K) at position +3. Overall, this frequency plot analysis

suggested a KAT2B-KAT2A consensus motif K-X-X-K-K/R-P/L/K (where X indicates any amino acid) (Fig.3.5B). In order to confirm that this motif is indeed specific to KAT2B-KAT2A acetylation, Dr. Tushar Ghosh and Sarah Buxton randomly picked up short sequences of similar length around lysine residues from 25 non-acetylated proteins in the UniProt database. Following similar frequency plot analysis using WebLogo, no enrichment of specific aminoacids was found around the key lysine residue (Fig. 3.5C). On close scrutiny, it was found that 5 out of the 25 proteins (20%) included in the frequency plot analysis presented the full-length motif around the key lysine residue (Fig. 3.5D). However, it was revealed that around this lysine, many of them harboured one of two partial motifs in which the full-length one can be divided into: K-X-X-K (6/25 proteins, 24%) and K-K/R-P/L/K (5/25 proteins, 20%). In total, 16/25 acetylated proteins contained either the full-length motif or one of the partial motifs around the key lysine residue for acetylation (Fig. 3.5D). In order to check how frequently these two partial motifs appear in the genome by chance, the number of control, non-acetylated protein sequences that included any of them was also analysed (in this case, they were analysed taking into account the whole protein sequence as there is no key lysine residue of reference). As indicated in Fig. 3.5D, the full-length motif was not found in any of the proteins analysed (0/25, 0%). The partial motif K-X-X-K was contained in just 1 out of the 25 sequences analysed (4%), whereas the partial motif K-K/R-P/L/K was found in 2/25 sequences analysed (8%). Even though this analysis is quite preliminary and limited (a greater number of sequences should be analysed), it suggests that the two partial motifs derived from the full-length one might be enriched in KAT2B-KAT2A acetylated proteins and do not appear in other non-acetylated with the same frequency. Interestingly, these two partial motifs also overlap with the already described recognition sequences obtained for the acetyltransferases P300 and Tetrahymena GCN5^{380,381}.

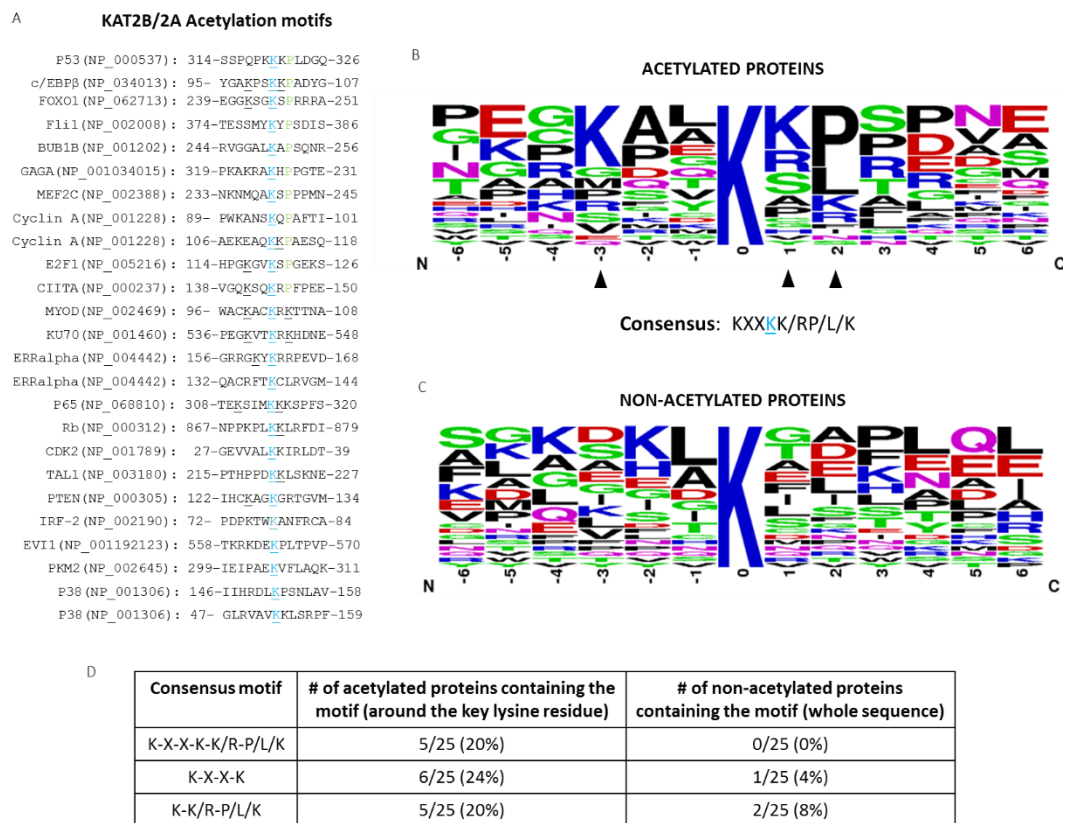


Figure 3.5. Finding the KAT2A/KAT2B acetylation motif. (A) Peptide sequences used from acetylated proteins to derive the motif. (B) Frequency plot analysis using WebLogo showing the frequencies of amino acids at positions -6 to +6 with reference to the acetylated lysine residue. Arrows show the positions of enrichment. (C) Sequence logo plot for non-acetylated proteins indicating no enrichment of specific amino acids around the key lysine residue. The stretches of peptides were randomly selected from non-acetylated proteins (Uniprot database). (D) Number of both acetylated and non-acetylated (control) sequences that contain either the full-length motif or any of the two partial motifs. The 25 acetylated and non-acetylated protein picked up were chosen by Dr. Tushar Ghosh and Sarah Buxton. The analysis showed in D was carried out by myself.

Analyzing the TBX5 protein sequence, we were not able to find the KAT2B-KAT2B consensus acetylation motif; however, we did find two of the partial motifs: K-R-K at positions 325-327, and K-K-P at positions 339-341 (Fig. 3.6A). As revealed by protein alignment of TBX5 orthologs in different species (human, chimpanzee, mouse, dog, chicken, xenopus and zebrafish), the aminoacids constituting those motifs are extremely well-conserved across species, thus supporting their relevance in TBX5 structure and/or function (Fig. 3.6A). Next, in order to identify whether any of these residues are acetylated by KAT2B or KAT2A, conservative substitutions of lysines 325, 327, 339 and 341 to arginine were performed and subsequently analysed, via transfection in COS7 cells and western blotting, to check whether they are susceptible to acetylation. Mutant variants generation and acetylation experiments were designed and carried out by Tushar Ghosh, whereas I was involved in reagents

preparation, technical support and data interpretation. As shown in Fig. 3.6B and 3.6C, conservative substitution of lysines 325 (K325R-FLAG), 327 (K327R-FLAG) and 340 (K340R-FLAG) did not abolish TBX5 acetylation by KAT2B; however, conservative substitution of lysine 339 to arginine did abolish TBX5 acetylation by both KAT2B (Fig. 3.6C) and KAT2A (Fig. 3.6E), thus indicating that lysine 339 is the key residue for acetylation in the TBX5 protein. In order to determine whether the transcriptional activity of the acetyl-deficient variant TBX5 K339R can be potentiated via acetylation, luciferase-reporter assays were carried out using the pGL3-MYH6-l reporter plasmid. Three independent experiments were performed (two of them performed by Dr. Tushar Ghosh and one by myself), and for each one of them, three technical replicates (three wells of the plate were transfected with the same combinations of plasmids) were carried out. As shown in Fig. 3.6D, transfection of KAT2B alongside the acetyl-deficient variant TBX5-K339R did not potentiate its transcriptional activity on the MYH6 promoter (4.8-fold activation in the TBX5-K339R-only condition vs 4.8-fold activation in the TBX5-K339R + KAT2B condition; $p>0.05$), thus suggesting that K339R is required for the potentiation of TBX5 transcriptional activity via acetylation. As expected, and in similar fashion to the results shown in Fig. 3.3, KAT2B was able to significantly increase the transcriptional activity of WT TBX5 on the MYH6 promoter (10.1-fold activation in the presence of KAT2B vs 7.2-fold activation in its absence; $p<0.05$). Interestingly, we also noted that even though the mutant variant K339R cannot be acetylated by KAT2B, it still retains transcriptional activity on the MYH6 promoter, albeit significantly reduced to that of TBX5 WT (4.8-fold activation by TBX5 K339R vs 7.2-fold activation by TBX5 WT; $p<0.05$) (Fig. 3.6D).

Overall, these experiments suggest that TBX5 is acetylated at lysine 339, and that this residue is critical to potentiate its transcriptional activity.

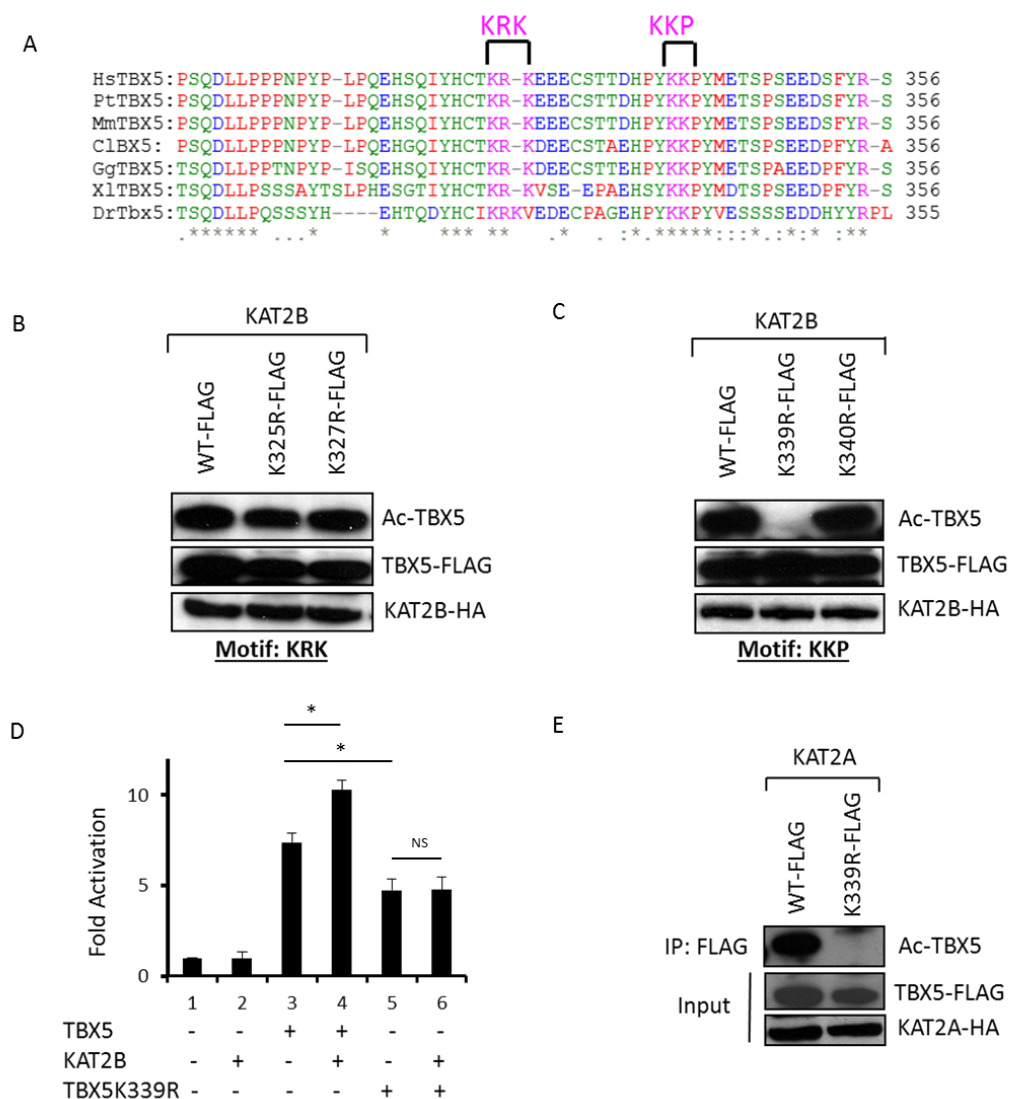


Figure 3.6. Acetylation of TBX5 takes place at lysine 339. (A) Alignment of TBX5 protein sequences showing conservation of the KRK and KKP motifs. Hs-Human, Pt-chimpanzee, Mm-mouse, Cl-Dog, Gg-chicken, Xl-Xenopus and Dr-zebrafish. (B and C) Pull-down assays to map the key lysine residue for acetylation in KRK (B) and KKP (C) motifs. Only when the lysine K339 is mutated, TBX5 acetylation is eliminated. (D) Reporter assay shows KAT2B failed to potentiate the transcriptional activity of the acetyl-deficient mutant protein TBX5 K339R on a MYH6 promoter. Results are from three independent experiments, two of them performed by Dr. Tushar Ghosh and one by myself. Mean $\pm$ SD is represented, *P < 0.05, NS = not significant (Student's t test). (E) Pull-down assay showing that KAT2A does not acetylate TBX5K339R. For B, C and E, three independent experiments were performed. Experiments were designed and carried out by Dr. Tushar Ghosh, and I was involved in reagents preparation, technical support and data interpretation. Figure images and legend were generated by Dr. Tushar Ghosh and shown here with his permission.

3.3. Results

3.3.1. Acetylation of TBX5 and nucleocytoplasmic shuttling

TBX5 is a transcription factor that can shuttle between the nucleus and the cytoplasm^{382,383}, presenting two nuclear localization signal domains (NLS1, position 78-90; and NLS2, position 325-340) that act in a cooperative way in order to regulate TBX5 nuclear translocation¹²⁰. In addition to these two NLSs, a nuclear export sequence (NES) can also be found in the TBX5 protein sequence (positions 152-160), which is recognized by the CRM1 exportin and mediates TBX5 translocation outside of the nucleus¹²¹ (Fig. 1.5). As shown in the previous experiments, TBX5 acetylation by KAT2B and KAT2A takes place at lysine 339, which is one of the residues contained in TBX5 NLS2. Taking this into account, I wanted to gain insight on whether TBX5 acetylation has an impact on its cellular distribution. To achieve this, I carried out immunofluorescence experiments on both COS7 and C2C12 cells using expression plasmids for TBX5 WT (pcDNA-TBX5-WT-FLAG) and the acetyl-deficient mutant TBX5 K339R (pcDNA-TBX5-K339R-FLAG). For each one of the cell lines, three independent experiments were performed. One hundred cells were counted per well/cover slip, so a total of 300 cells (100 cells x 3 independent experiments) were counted in total per condition (WT or K339R) and per cell line (COS7 and C2C12 cells). In COS7 cells transfected with TBX5-WT, TBX5 showed a nuclear localization in 95% of the cells analysed (n=97/100, 93/100 and 95/100 for each independent experiment) (Fig. 3.7A and C). In contrast, in COS7 cells transfected with the acetyl deficient mutant TBX5-K339R, just 24% of the cells (n=25/100, 20/100 and 27/100) presented a nuclear TBX5 localization, whereas 76% of the cells counted (n=75/100, 80/100 and 73/100) presented a cytoplasmic distribution of some sort (either nuclear and cytoplasmic or predominantly cytoplasmic) (Fig. 3.7A and C). In most cells showing a cytoplasmic distribution of TBX5-K339R, accumulation of TBX5-K339R around the outer nuclear membrane was also seen (Fig. 3.7A). The results obtained in C2C12 cells were in concordance with those obtained in COS7 cells: when transfected with TBX5-WT-FLAG, 94% of the cells analysed (n=94/100, 93/100 and 96/100) displayed a nuclear distribution of TBX5, whereas just in 6% of the cells TBX5-WT was also found in the cytoplasm (n=6/100, 7/100 and 4/100) (Fig. 3.7A and C). When the acetyl-deficient mutant was transfected, only in 17% of the cells analysed (n=12/100, 21/100 and 19/100) TBX5 was found exclusively in the nucleus, 37% displayed a nuclear and cytoplasmic TBX5 distribution (n=38/100, 33/100 and 41/100) and 46% presented a predominantly cytoplasmic distribution of TBX5 (n=52/100, 46/100 and 40/100) (Fig. 3.7A and C).

This impairment in cellular localization showed by the acetyl-deficient mutant TBX5 K339R suggests that acetylation of TBX5 is necessary for its correct localization inside cells. This impaired cellular localization can be due to a defect in two different processes: either TBX5 nuclear import or TBX5 nuclear retention. To ascertain which process is TBX5 acetylation required for, COS7 and C2C12 cells transfected with TBX5-WT or TBX5-K339R were treated with 10 ng/ml of Leptomycin B (a potent inhibitor of exportin CRM1, which mediates TBX5 nuclear export¹²¹) for 16 hours and subjected to immunofluorescence. When transfected in COS7 cells treated with Leptomycin B, TBX5-K339R was found exclusively in the nucleus in 96% of the cells counted and analysed (n=95/100, 98/100 and 96/100), in contrast to the 24% previously found in untreated cells (Fig. 3.7B and C). In treated C2C12 cells, there was a dramatic shift in TBX5-K339R cellular distribution as well: the acetyl-deficient mutant displayed a nuclear localization in 85% of the cells analysed (n=78/100, 87/100, 89/100), in contrast to the previous 17% described in non-treated cells (Fig. 3.7B and C). 11% of these cells showed a nuclear and cytoplasmic distribution (n=17/100, 8/100 and 9/100) whereas 4% showed a cytoplasmic distribution (n=5/100, 5/100 and 2/100). Leptomycin B treatment of COS7 or C2C12 cells transfected with TBX5-WT did not have any significant effect on its normal nuclear distribution (in COS7 cells, 98% of cells analysed showed a TBX5 WT nuclear distribution (n=97/100, 97/100 and 100/100); in C2C12 cells, 97% of them (n=96/100, 97/100 and 99/100) (Fig. 3.7B and C).

These results indicate that when CRM1 activity is inhibited, the acetyl-deficient mutant can remain inside the nucleus. Taking this into account, I next aimed to study the potential interaction and co-localization of TBX5-K339R and endogenous CRM1 in COS7 cells. To achieve this, cells transfected with TBX5-WT-FLAG or TBX5-K339R-FLAG were subjected to immunofluorescence analysis using anti-FLAG and anti-CRM1 antibodies (n=3 individual experiments). This analysis revealed that TBX5-K339R co-localize with CRM1 right outside of the nucleus, mainly around the outer nuclear membrane (Fig. 3.8A). In the case of TBX5-WT, or when cells treated with leptomycin B are transfected with TBX5-K339R, no co-localization was seen outside of the nucleus, and both TBX5 and CRM1 were found strictly inside the nucleus (Fig. 3.8A). To gain further insight into the potential TBX5-K339R and CRM1 interaction, Tushar Ghosh carried out *in vitro* pull-down assays in COS7 cells followed by western blot analysis (three independent experiments were performed). Densitometry analysis revealed that the interaction between TBX5-K339R and CRM1 is significantly stronger than that of TBX5 WT and CRM1 (1.54 times stronger according to densitometry of the western blot bands, $p < 0.05$) (Fig. 3.8B and 3.8C).

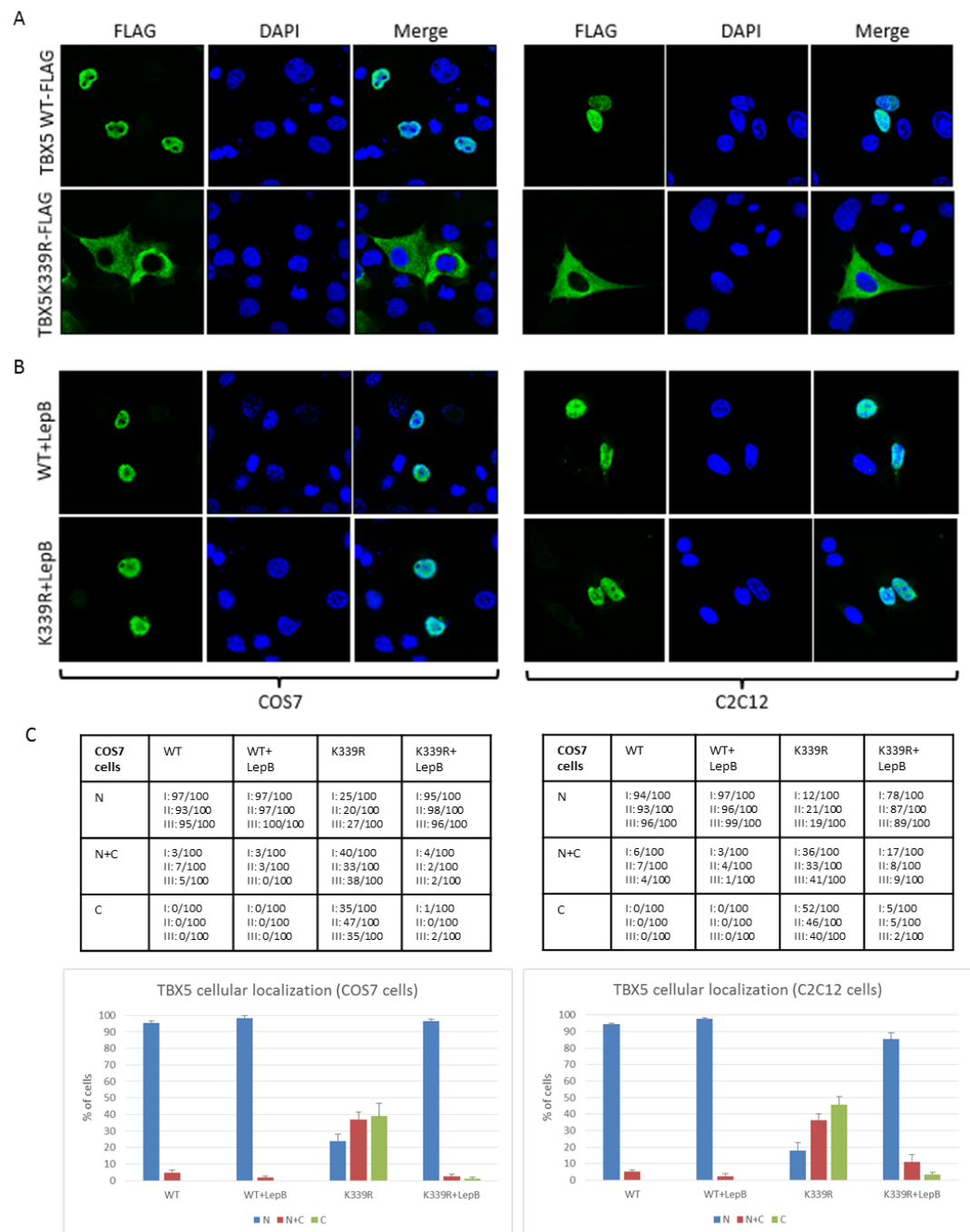


Figure 3.7. Acetylation of TBX5 is required for its nuclear retention. (A) Representative images of the cellular distribution of TBX5-WT and TBX5-K339R in both COS7 and C2C12 cells. The acetyl-deficient mutant TBX5-K339R is found in the cytoplasm in many of the cells analysed. (B) Representative images showcasing the effect of leptomycin B on the nuclear/cytoplasmic distribution of TBX5-WT and TBX5-K339R in both COS7 and C2C12 cells. (C) Number of cells displaying a nuclear (N), nuclear and cytoplasmic (N+C) or a cytoplasmic (C) TBX5 localization in the different conditions analysed, both in COS7 and C2C12 cells. I, II and III represent the three independent experiments performed (300 cells were counted in total per experiment and per condition). Below, graphs showcasing the percentage of cells displaying a certain TBX5 cellular localization in WT, WT+LepB, K399R and K339R+LepB conditions are shown. N-nuclear, C-cytoplasmic. Data are shown for COS7 cells (left) and C2C12 cells (right). Results are mean $\pm$ SD from three individual experiments.

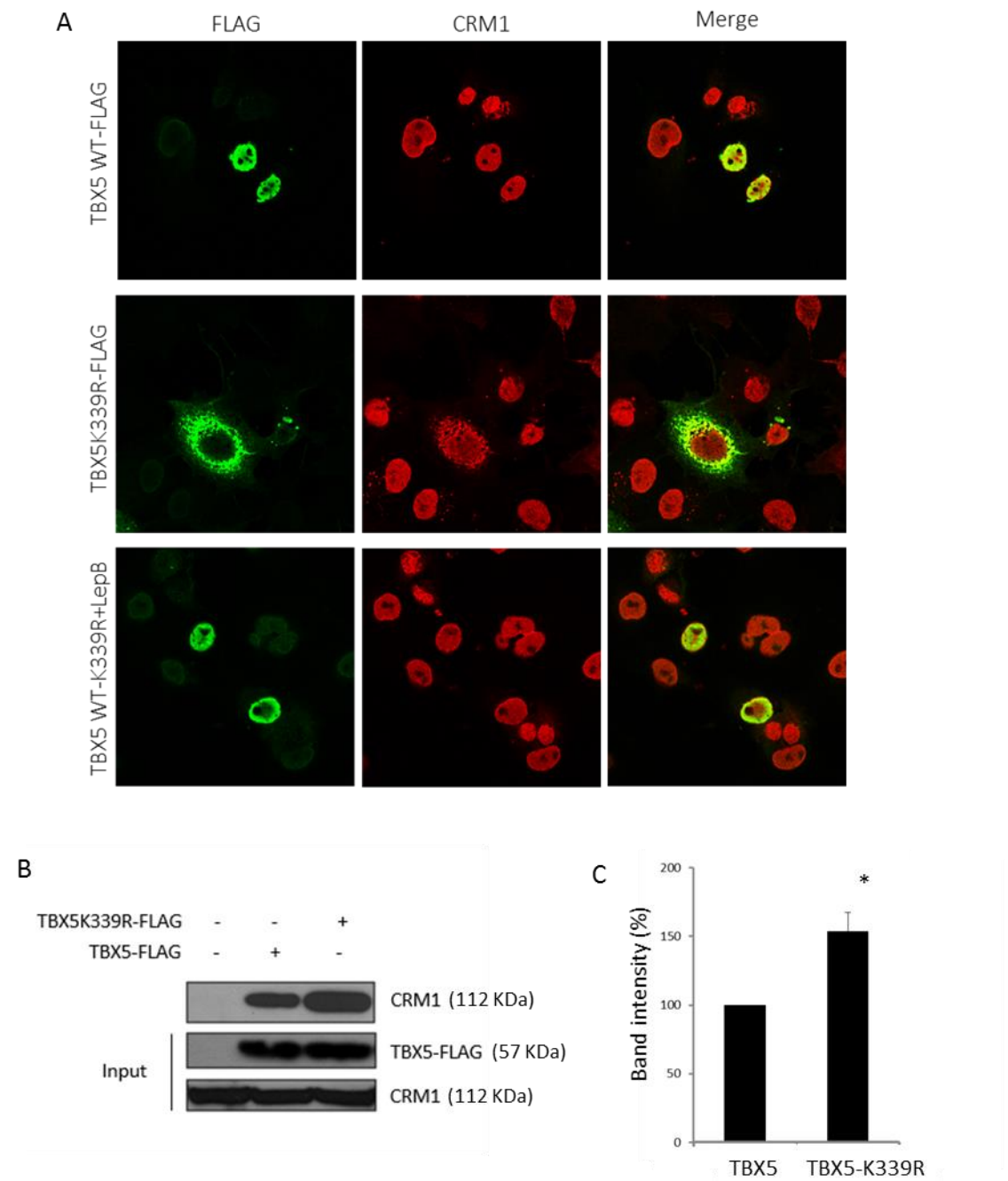


Figure 3.8. CRM1 mediates nuclear export of non-acetylated TBX5. (A) Representative images of the co-localization between TBX5-K339R and endogenous CRM1 outside the nucleus in Cos7 cells, right next to the outer nuclear membrane. TBX5-WT and CRM1 do not co-localize outside of the nucleus. (B) Pull-down assay shows the increased affinity of mutant protein TBX5K339R for CRM1. (C) Quantification of CRM1 protein pull-down by wildtype (TBX5-FLAG) and acetyl-deficient mutant protein (TBX5 K339R-FLAG). Experiments B and C were carried out by Dr. Tushar Ghosh. Results are mean $\pm$ SD from three individual experiments, * $P < 0.05$ (Student's *t* test.)

Even though these results suggest that the lack of acetylation promotes TBX5 nuclear export via interaction with the exportin CRM1, the possibility that the results obtained in my experiments are due to the conservative substitution in

the residue K339, and not due to the lack of acetylation, cannot be completely ruled out. In order to confirm that acetylation of this residue is responsible for proper TBX5 nuclear localization, I treated COS7 cells transfected with either TBX5-WT or TBX5-K339R with Garcinol, an inhibitor of acetyltransferases KAT2B and p300 that inhibits the former more efficiently than the latter^{306,384}. A Garcinol titration in COS7 cells has been already described elsewhere³⁸⁴, reporting a working concentration of 6 μM . In spite of this, I also performed a Garcinol titration in COS7 cells. To do this, COS7 cells were treated for 16 hours in 96-well plates with 8 different concentrations of Garcinol (0.1 μM , 1 μM , 2.5 μM , 5 μM , 10 μM , 25 μM , 50 μM and 100 μM) in triplicates (three independent experiments were performed) and cell viability was assayed by carrying out an MTT assay. Fig. 3.9 shows the results of the cell viability assay, which indicates that COS7 viability is not affected by using Garcinol concentrations going from 0.1 μM to 5 μM . However, starting at 10 μM , COS7 cells viability was affected, going down to 83% in cells treated with 10 μM of Garcinol, 62% in cells treated with 25 μM , 28% in cells treated with 50 μM and 8% in cells treated with 100 μM . In view of these results, I decided to choose a concentration of 20 μM for further Garcinol experiments, as according to the assay, cell viability would still be around 75% when using this concentration.

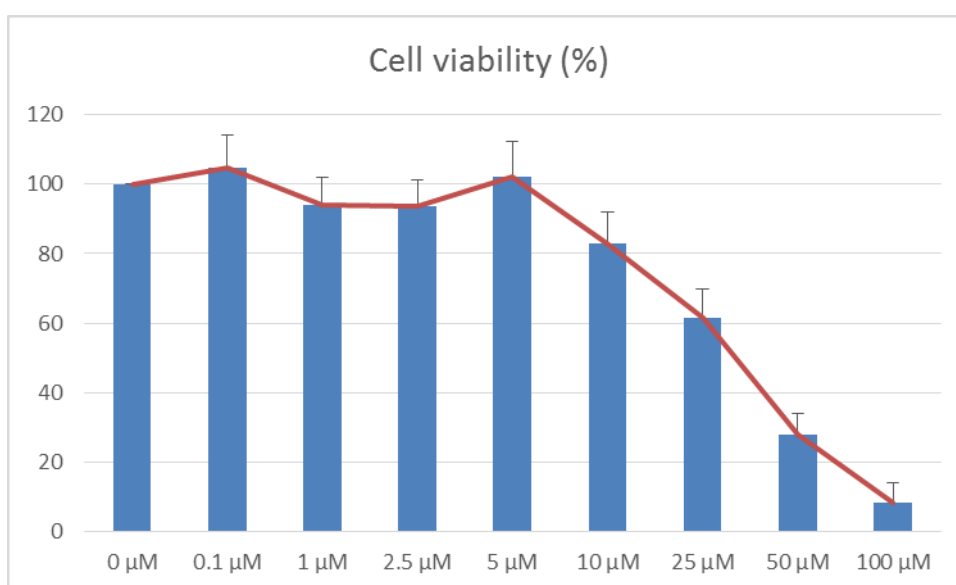
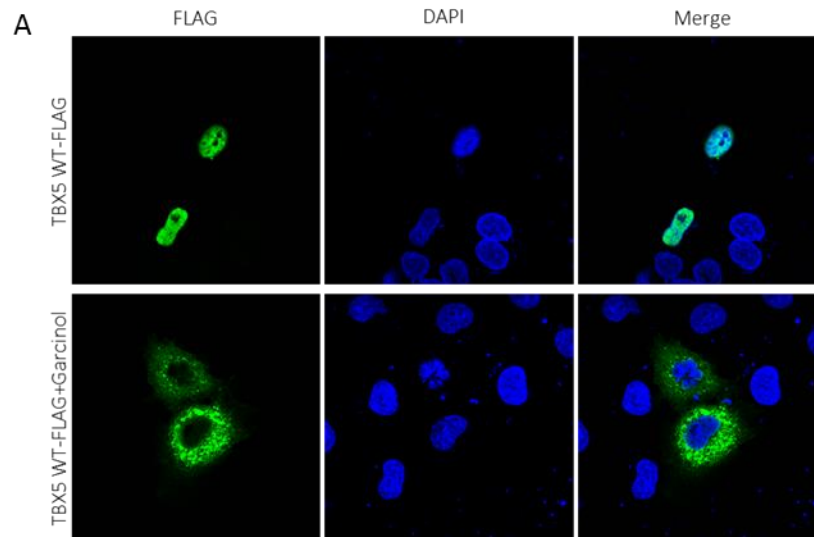


Figure 3.9. Cell viability assay on COS7 cells treated with Garcinol. Cell viability (%) is not affected when using Garcinol concentrations ranging from 0.1 μM to 5 μM . However, starting at 10 μM , cell viability is progressively reduced with increasing concentrations of Garcinol. Results are mean + SD from three independent experiments.

Once a working concentration for Garcinol was established, COS7 cells transfected with either TBX5-WT or TBX5-K339R were treated with 20 μM of Garcinol for 16 hours and then they were subjected to immunofluorescence analysis (as previously, three independent experiment were carried out. One

hundred cells were counted per group and per experiment). When treated with Garcinol, TBX5-WT localization was completely nuclear in 50% of the cells analysed (n=48/100, 44/100 and 59/100), in contrast to the 96% of cells (n=95/100, 98/100 and 94/100) displaying a nuclear localization in untreated cells (Fig. 3.10A and B). 27% of the treated cells analysed showed a nuclear and cytoplasmic TBX5 WT localization (n=25/100, 35/100 and 20/100), whereas TBX5 WT was found predominantly in the cytoplasm in 23% of the cells analysed (n=27/100, 21/100 and 21/100) (4% and 0% for untreated cells, respectively) (Fig. 3.10A and B).

Overall, these results indicate that chemical inhibition of acetylation in COS7 cells also lead to an impaired TBX5 cellular localization, thus supporting the fact that acetylation of this transcription factor is required for its retention inside the nucleus and to avoid nuclear export via the CRM1 pathway.



B

COS7 cells	WT	WT+ Garcinol
N	I: 95/100 II: 98/100 III: 94/100	I: 48/100 II: 44/100 III: 59/100
N+C	I: 5/100 II: 2/100 III: 5/100	I: 25/100 II: 35/100 III: 20/100
C	I: 0/100 II: 0/100 III: 1/100	I: 27/100 II: 21/100 III: 21/100

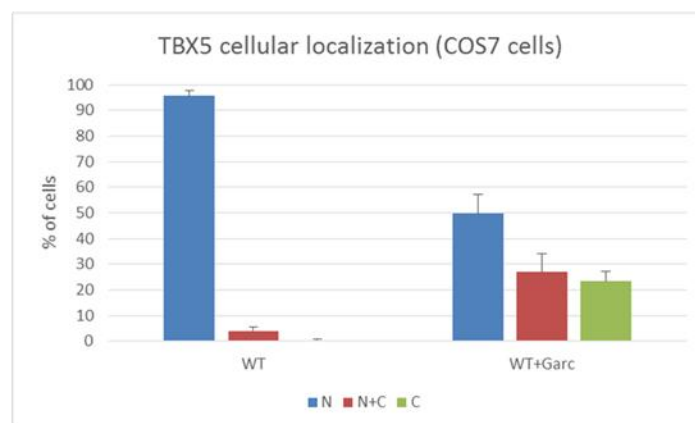


Figure 3.10. Chemical inhibition of KAT2B impairs TBX5 cellular localization. (A) Representative images displaying the cellular distribution of TBX5-WT in Garcinol-treated cells. When treated with Garcinol, TBX5-WT is partially relocated to the cytoplasm. (B) Number of cells displaying a nuclear (N), nuclear and cytoplasmic (N+C) or a cytoplasmic (C) TBX5 localization in untreated or treated COS7 cells. I, II and III represent the three independent experiments performed (300 cells were counted in total per experiment and per condition). Below, graphs showcasing the percentage of cells displaying a certain TBX5 cellular localization in untreated or treated conditions are shown. N-nuclear, C-cytoplasmic. Results are mean $\pm$ SD from three individual experiments.

3.3.2. *kat2a* and *kat2b* expression in zebrafish

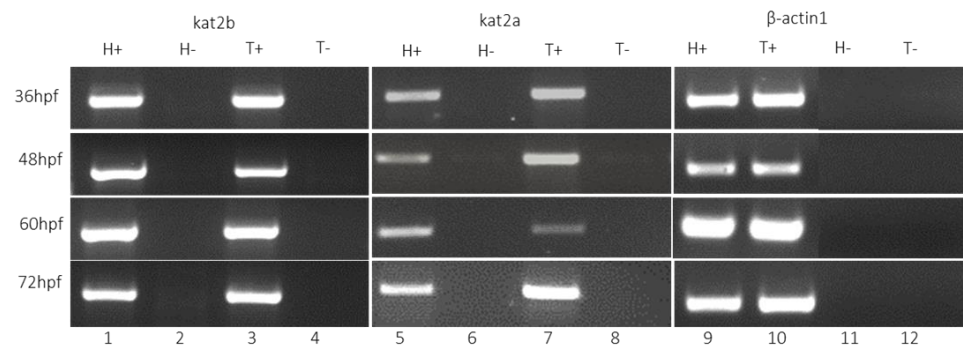
After studying the interaction between KAT2A and KAT2B in an *in vitro* context, we then decided to characterize the potential *in vivo* link between these factors and study whether the perturbation of these two acetylases in zebrafish has an impact on heart and limb development, as it happens with *tbx5a*¹²⁶. One characteristic of the zebrafish genome is that many genes are duplicated²⁶⁷, so I first checked out whether this was the case for *kat2a* and *kat2b*. Upon searching in the Ensembl Genome Browser (<https://www.ensembl.org/index.html>), only one copy of the *kat2a* and *kat2b* genes were found, thus indicating that these genes are not duplicated in the zebrafish genome. The similarity between the KAT2A and KAT2B proteins in different species such as zebrafish, mouse and human was also analysed by using the ClustalW2 software (Fig. 3.11A). This analysis revealed that these KAT proteins share a high degree of homology between them: zebrafish *kat2a* share an 84% and 85% of homology with their human and murine counterparts respectively, whereas zebrafish *kat2b* share an 81% and 80% of homology respectively. Analysis also revealed that the zebrafish *kat2a* and *kat2b* proteins share a 73% of homology, thus suggesting a potential redundancy between them (Fig. 3.11A).

Next, I moved onto the study of *kat2a* and *kat2b* expression in zebrafish by using two different techniques: RT-PCR and *in situ* hybridization. For the RT-PCR, sections of zebrafish head-cardiac and tail regions were obtained as explain in 2.2.2.6 (pools of 20-40 embryos were used for RNA extraction). Three different developmental stages were analysed: 36 hpf, 48 hpf, 60 hpf and 72 hpf. Following RNA extraction from corresponding tissues, cDNA was generated and *kat2a* and *kat2b* expression was analysed via gel agarose analysis. As shown in Fig. 3.11B, both *kat2a* and *kat2b* show a strong expression throughout the different stages analysed both in the cardiac region and in the tail. When compared to the housekeeping gene *β-actin 1*, *kat2b* expression seemed to be slightly stronger than that of *kat2a* in both tissues at 36, 48 and 60 hpf. *In situ* hybridization analysis using *kat2a* and *kat2b* RNA probes in 48 hpf zebrafish embryos was also carried out (n=10 embryos were subjected to ISH per probe). As shown in Fig. 3.11C, both *kat2a* and *kat2b* display a remarkably similar expression pattern: both genes are predominantly expressed in the cranial area, including the brain/head, retina and heart. As showed in the dorsal view, *kat2a* and *kat2b* expression was also found in the spinal cord area and in the pectoral fins, albeit weakly, whereas the expression in the tail was much weaker than expected in comparison to what was found in the RT-PCR analysis. Higher magnification images showcasing both the heart and the pectoral fins are shown as well (Fig. 3.11C)

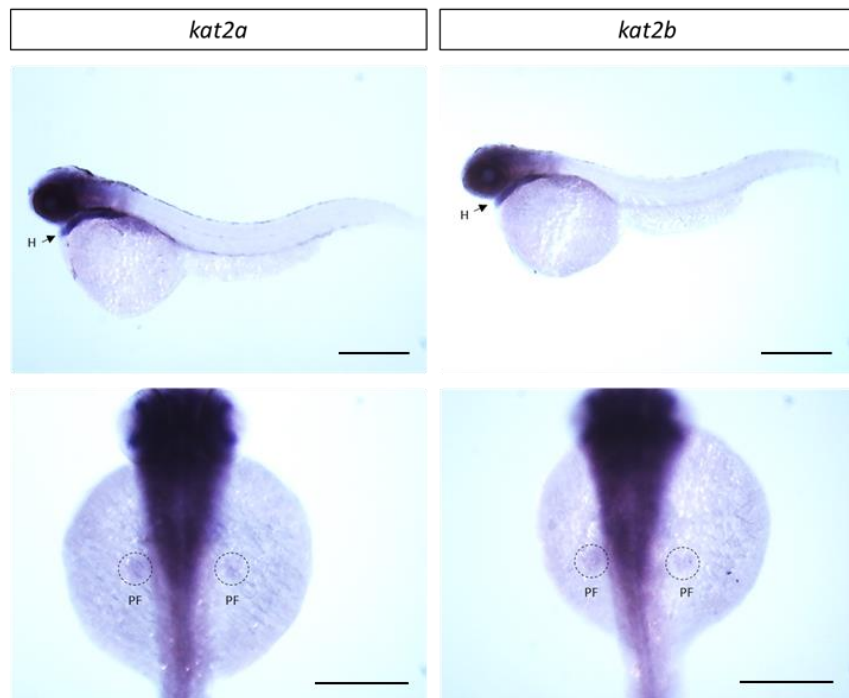
A

	zK2A	hK2A	mK2A	zK2B	hK2B	mK2B
zK2A		84.20	84.99	73.07	73.60	72.48
hK2A	84.20		98.07	73.11	70.54	71.09
mK2A	84.99	98.07		73.77	71.31	71.75
zK2B	73.07	73.11	73.77		80.96	79.97
hK2B	73.60	70.54	71.31	80.96		93.71
mK2B	72.48	71.09	71.75	79.97	93.71	

B



C



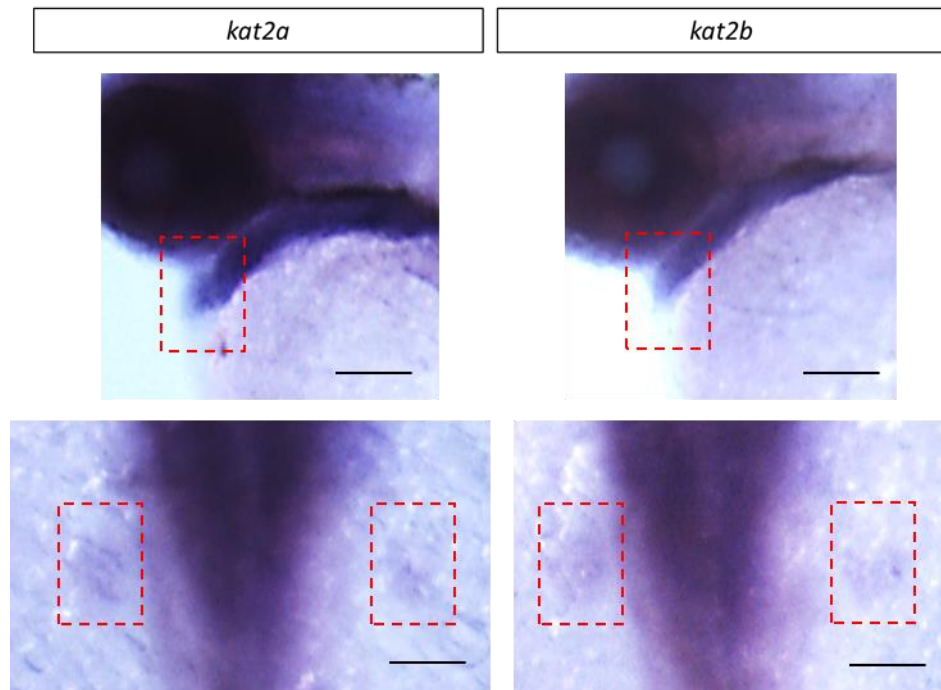


Figure 3.11. *kat2a* and *kat2b* genes in zebrafish. (A) Amino acid alignments of human, murine and zebrafish KAT proteins were performed to study their percentage of similarity using ClustalW2. (B) RT-PCR expression analysis of *kat2b* and *kat2a* in zebrafish. The analyses were carried out on 36, 48, 60 and 72h of post fertilized embryos. H and T denote heart and tail regions and + and – denote +RT (retrotranscriptase) and –RT samples respectively. β -actin1 is used as a control. The agarose gels show the ubiquitous expression pattern of both *kat2b* and *kat2a*. For the analysis, tissue sections derived from pools of 20-40 zebrafish embryos were used. (C) *In situ* hybridization using *kat2a* and *kat2b* probes was performed on 48 hpf zebrafish embryos. Expression in the cardiac region is indicated with a black arrow (H=heart), whereas expression in the pectoral fins (PF) is highlighted by dashed circles. Below, higher magnifications of the cardiac and fin regions are shown (heart and fins are highlighted with dashed, red rectangles) (n=10 embryos stained per probe). Scale bar = 400 μ m for lower magnification images, 100 μ m for higher magnification images.

3.3.3. Morpholino-mediated knockdown of *kat2a*, *kat2b* and *tbx5a*

In order to study what is the role of *kat2a* and *kat2b* in cardiac development in zebrafish, morpholino-mediated knockdown of these genes was carried out and the cardiac phenotype was analysed 48 hpf. Translation blocking morpholinos, splice blocking morpholinos and a mismatch control morpholino were designed by Gene Tools (USA). First, a titration experiment for each one the morpholinos was carried out (n=3 independent experiments). For each morpholino, pools of 20 embryos (for a total of 60 embryos taking into account all three experiments) were injected with increasing morpholino doses, ranging from 0.5 ng to 8 ng. An injection control (no morpholino) was also carried out. Embryos were then incubated for 48 hours at 28°C, and then, the percentage

of dead embryos or embryos that presented extremely severe developmental defects in comparison to the total number of embryos injected was calculated. As shown in Fig. 3.12, regarding the translation blocking morpholinos, even at the lowest dose (0.5 ng) they led to a high proportion of dead or severely affected embryos (53% (n=32/60) and 48% (n=29/60) for kat2a and kat2b translation-blocking MOs, respectively). At higher doses, this percentage is even higher, reaching a 90% (n=54/60) and 92% (n=55/60) of dead/affected embryos at the highest dose (8 ng). In the case of the splice blocking morpholinos, the proportion of dead/severely affected embryos was not remarkably high even at a dose of 4 ng both for kat2a and kat2b MOs (22% (n=13/60) and 27% (n=16/60) respectively); however, at a dose of 8 ng, the proportion of dead/severely affected embryos increased until 87% (n=52/60) for the kat2a splice MO and 67% (n=40/60) for the kat2b splice MO. Taking all these data into account, it looks like both translation blocking MOs induced a high level of toxicity in zebrafish embryos even at the lowest tested dose; whereas both splice blocking MOs are relatively well-tolerated up to a dose of 4 ng. Based on this data, I decided to carry out further MO experiments using just the splice blocking MOs, as the translation blocking MOs seem to induce a high level of toxicity even at low doses.

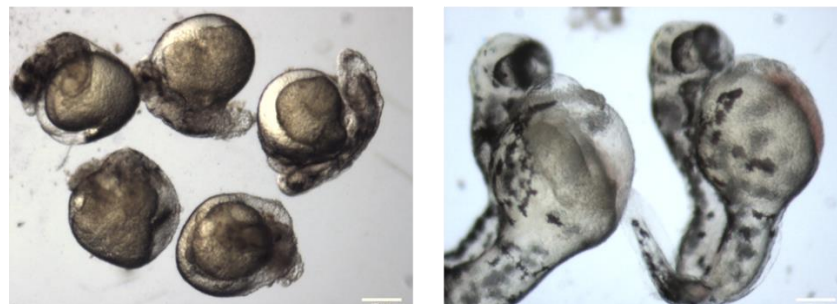
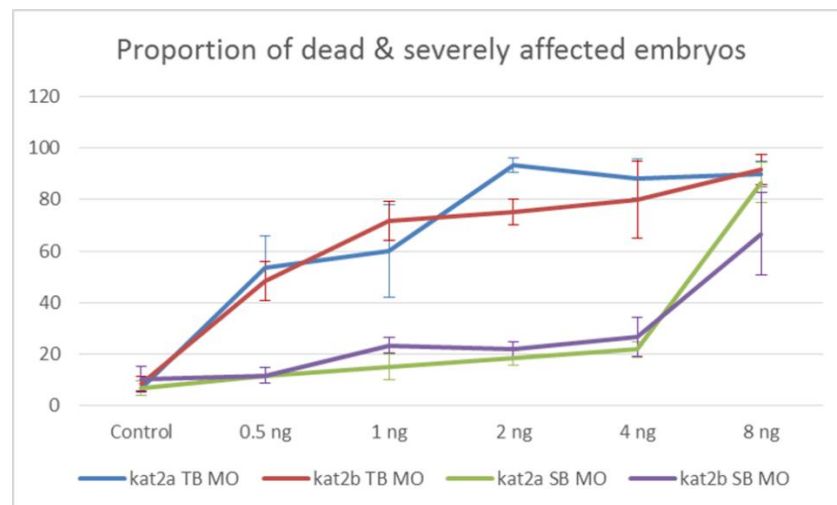


Figure 3.12. Titration of kat2a and kat2b translation blocking and splice blocking MOs. Batches of 25 embryos were injected with different concentrations of each morpholino and

the proportion of dead embryos + embryos presenting severe developmental defects was calculated. Below, representative images of necrotic/dead embryos as well as embryos showing severe developmental defects are shown. Results are from three independent experiments. Mean + SD is represented. TB = translation blocking, SB = splice blocking, MO = morpholino.

Once the MO dose was titrated, I proceeded to perform knockdown experiments on 1-cell stage zebrafish embryos using *kat2a* splice blocking MO (4 ng), *kat2b* splice blocking MO (4 ng) and a *tbx5a* translation blocking MO already used elsewhere¹²⁶ which serves as a positive control (1.7 ng as described in the cited publication). For negative controls, a mismatch MO was used (more information about its sequence can be found in 2.2.2.7). Three independent experiments were performed: in each one of them, two batches comprised of 30-35 embryos were injected with each morpholino so a total of 200 embryos (70, 70 and 60 in total for each independent experiment) were injected per morpholino taking into account all three experiments. Following injection, zebrafish embryos were allowed to develop for 48 hpf at 28°C, and surviving embryos were analysed for the presence of cardiac defects or abnormalities. According to the severity of the cardiac phenotype, embryos were assigned to one of three possible categories: normal, mild or severe. In normal embryos, both the atrium and the ventricle were clearly defined and tightly packed into a figure-eight configuration. In the mild phenotype the atrium and ventricle were clearly distinguishable, but looping was noticeably relaxed and slight pericardial oedema was usually seen, as well as a slight reduction in the rate of heart beat. In the severe phenotype the atrium and ventricle were hardly distinguishable and they were irregular in shape and/or size, and the heart in general was stretched to a thin tube due to the complete loss of cardiac looping. Also, a moderate-large pericardial oedema was typically observed and the heart rate was notably slower.

Following injection and incubation, embryos were analysed. As shown in Fig. 3.13A, an average of 72% (n=50/70; 55/70; 39/60), 76% (n=47/70; 58/70; 47/60) and 77% (n=55/70; 51/70; 48/60) of the embryos injected with *kat2a*, *kat2b* or *tbx5a* MOs survived and were analyzed, whereas a 80% (n=53/70; 58/70; 49/60) of embryos injected with the mismatch MO survived. Overall, *kat2a* and *kat2b* splice blocking MOs produced remarkably similar developmental defects than those observed with the *tbx5a* translation blocking MO (Fig. 3.13B). In both *kat2a* and *kat2b* morphants, heart looping was relaxed or totally interrupted, and pericardial oedema as well as alteration of cardiac chambers morphology was detected. In *kat2a* and *kat2b* single knockdowns, 44% (n=63/144) and 45% (n=68/152) of analyzed embryos respectively presented what it was described as a mild cardiac phenotype. In contrast, just

a 12% (n=18/144) and 20% (n=30/152) respectively showed a severe cardiac phenotype (the remaining 44% (n=63/144) and 35% (n=53/152) displayed a phenotype that was undistinguishable from their negative control counterparts) (Fig. 3.13C). In the case of the *tbx5a* morphants, 57% (n=88/154) of analysed embryos showed a severe cardiac phenotype, whereas a 19% (n=29/154) displayed a mild cardiac phenotype (the remaining 24% (n=37/154) were undistinguishable from their negative control counterparts) (Fig. 3.13B and C). The effect of knocking down both *kat2a* and *kat2b* concomitantly was also assessed by injecting embryos with 2 ng of *kat2a* splice blocking MO + 2 ng of *kat2b* splice blocking MO. Out of the 200 embryos injected in total, 64% of the embryos survived for at least 48 hours (n=46/70; 39/70; 43/60) and were analysed (Fig. 3.13A). Out of these surviving embryos, 29% (n=37/128) of them showed a severe cardiac phenotype, whereas 48% (n=61/128) of them presented a mild cardiac phenotype, thus supporting a potential compensatory and redundant effect between *kat2a* and *kat2b* (Fig. 3.13B and C). In some of the embryos, mild defects or slight development delay was also seen in other organs areas such as the head or the tail; unfortunately, no pictures of the whole embryos were taken as I focused on obtaining high-magnification and high-resolution images of the heart, so these defects initially seen in other parts of the zebrafish body could not be properly assessed.

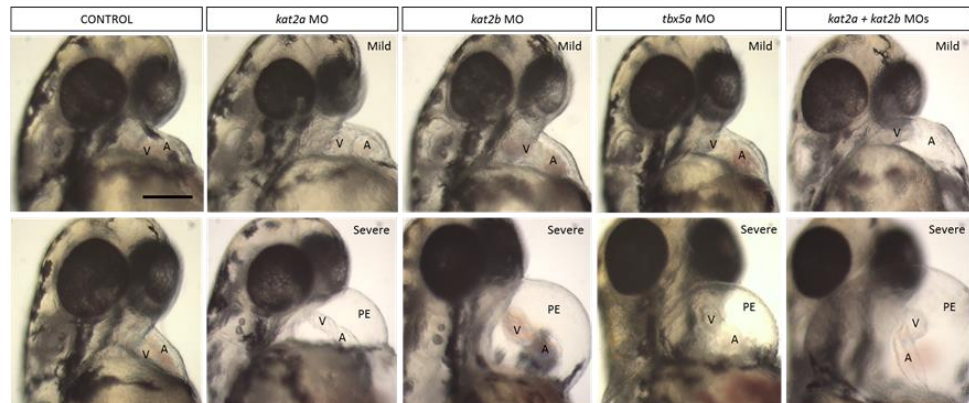
In order to confirm that these embryos showing cardiac defects were indeed *kat2a* and *kat2b* morphants, RT-PCR was performed on RNA samples obtained from them. Following RNA extraction, cDNA generation and PCR amplification, samples were analysed by agarose gel electrophoresis. As shown in Fig. 3.13D., splicing on exon 2 was altered in the *kat2a* gene as indicated by the presence of a normal size allele (345 bp) and an additional, shorter isoform lacking this exon (221 bp). Results for *kat2b* are similar: in control embryos, a single allele of 342 bp can be detected, whereas morphants show an additional allele of 204 bp. Expression of the housekeeping gene *gapdh* in both control and morphant embryos is also shown (Fig. 3.13D).

Overall, these results indicate that the knockdown of *kat2a* or *kat2b* in zebrafish lead to cardiac defects which are similar to those found in *tbx5a* morphants, thus suggesting a link between these two acetylases and *tbx5a* in an *in vivo* context.

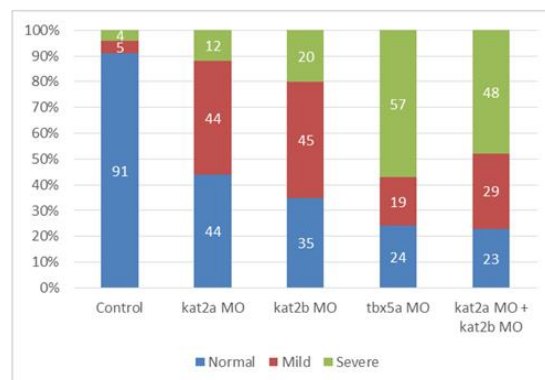
A

	# of surviving embryos 48 hpf	Avg % of survival
kat2a splice MO	I: 50/70; II: 55/70; III: 39/60	72%
kat2b splice MO	I: 47/70; II: 58/70; III: 47/60	76%
tbx5a translation MO	I: 55/70; II: 51/70; III: 48/60	77%
kat2a+kat2b splice MOs	I: 46/70; II: 39/70; III: 43/60	64%
Mismatch MO	I: 53/70; II: 58/70; III: 49/60	80%

B



C



D

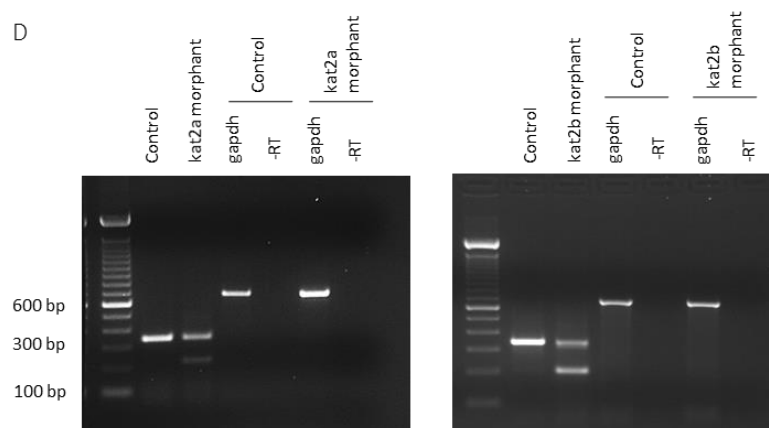


Figure 3.13. The effect of morpholino-mediated knockdown of *tbx5a*, *kat2a* or *kat2b* in 48hpf embryos. (A) Table showing the number of embryos which survived for at least 48 hpf following morpholino injections, as well as the average percentage of survival (60-70 embryos were injected per experiment, three independent experiments were carried out). These

surviving embryos were the ones analysed for the presence of cardiac defects. (B) Representative pictures of the cardiac phenotype found in zebrafish embryos injected with *tbx5a*, *kat2a* and *kat2b* morpholinos. These embryos were categorized as “mild” or “severe” based on the severity of the cardiac phenotype (V=ventricle, A=atrium, PE=pericardial oedema). Scale bar = 200 µm. (C) Distribution of cardiac phenotypes based on severity in analysed embryos. (D) Agarose gels showing the effect of *kat2a* (left) and *kat2b* (right) mRNA on splice-blocking morpholinos. An additional allele is seen in both cases.

3.3.4. sgRNA design and synthesis

To validate the *kat2a*, *kat2b* and *tbx5a* loss-of-function phenotypes generated via morpholino injection, I also aimed to generate F0 mutants for each of these genes using the CRISPR/Cas9 genome-editing technology. Although the CRISPR/Cas9 technology provides a highly-efficient and powerful system for genome engineering, the variable activity of different sgRNAs still results in inconsistent CRISPR/Cas9 activity. Several studies have shown that sgRNAs stability, loading and targeting *in vivo* depend on several features, such as nucleotide composition and relative position^{311,385,386}. This has led to the construction of different predictive models of sgRNA activity to improve sgRNA design for gene editing purposes. To design sgRNAs targeting the *kat2a*, *kat2b* and *tbx5a* genes I used the free webtool CHOPCHOP (<http://chopchop.cbu.uib.no/>). Fig. 3.14A shows the target sequences and features of nine sgRNAs designed using CHOPCHOP that have relatively high on-target scores and low off-target effects. All of these sgRNAs were synthesized *in vitro* using a free-cloning approach (see 2.2.2.3. for further details) and used for subsequent *in vitro* activity assays.

3.3.5. Checking target sequences presence in the zebrafish genome

The population of zebrafish in the aquarium are continually interbred to produce new fish and they are not regularly restocked. This could eventually lead to the generation of new and random alleles in the zebrafish population, so it becomes necessary to check that our target sites are present in the zebrafish genome. Following genomic DNA extraction, PCR products harbouring the expected target sites for each sgRNA were generated and sequenced. As shown in Fig. 3.14B, all nine target sequences are present in the population of zebrafish we routinely use, showing no evidence of mutations or variations (n=3 biological replicates were analyzed).

A

Name	Target sequence	Target position (cDNA)	On-target score	# of off-targets
kat2a sgRNA 1	CCAGCGGCCCGACCGGCTCCAAC	+78	92	0
kat2a sgRNA 2	CCCAAGTCCGTTTCGTTCCGCGG	+155	88	0
kat2a sgRNA 3	GGCTGTTATGGCGGACCCGGCGG	+10	82	0
kat2b sgRNA 1	CCCAGTTCTGTGACAGTCTCCC	+785	89	0
kat2b sgRNA 2	AGTTCTGTGACAGTCTCCCAAGG	+780	87	0
kat2b sgRNA 3	CCGCTGTCAGCCTAATGGAGCCA	+297	82	0
tbx5a sgRNA 1	CGGACAGTGAAGACACCTTTCCG	+5	84	0
tbx5a sgRNA 2	CCCGTCATCGCAGACTACATAC	+116	90	0
tbx5a sgRNA 3	CTGAGCCTGCGATGCCCGGACGG	+377	87	0

B

kat2a TS 1 TTCAGCAAGCGCAGTCCAGCGGCCCGACCGGCTCCAACTCGAACCCCGGAGCC
kat2a TS 2 CCAGCCAGAAGAAAGCCCAAGTCCGTTTCGTTCCGCGGCGCAAAAGCTGGAG
kat2a TS 3 CAGGGGAAGTGTGGCTGCTGTTATGGCGGACCCGGCGGCACAGAGCTCCGCGG
kat2b TS 1 TGTTATTGTAACGTGCCAGTTCTGTGACAGTCTCCCAGGTACGAGGCCAC
kat2b TS 2 ATTGTAACGTGCCCCAGTTCTGTGACAGTCTCCCAAGGTACGAGGCCACAG
kat2b TS 3 GGGCTGAGCAACCAACCGCTGTCAGCCTAATGGAGCCATGCAGAAGCTGCATT
tbx5a TS 1 GCCTCTCCGACATGGCGGACAGTGAAGACACCTTTCCGCTCCAAAACCTCTCCC
tbx5a TS 2 CAGCCGTCAGCAAAATCCCGTCATCGCAGACTACATACATTCAACAGGTAACG
tbx5a TS 3 CTGTGACGGGGAAAGCTGAGCCTGCGATGCCCGGACGGCTTTATGTCCATCCA

Figure 3.14. sgRNAs targeting the *kat2a*, *kat2b* and *tbx5a* genes. (A) Features of the sgRNAs designed for the targeting of *kat2a*, *kat2b* and *tbx5a*. Target sequence, position in the cDNA, on-target score and number of potential off-target effects are shown. (B) PCR products were generated from adult zebrafish genomic DNA (n=3) and subjected to Sanger sequencing. All target sites (highlighted in yellow, PAM sequence in red) are present in the zebrafish genome.

3.3.6. sgRNA activity *in vitro* assay

Following sgRNA *in vitro* synthesis, their real targeting activity was tested by performing an *in vitro* activity assay using each of the sgRNAs and commercially available Cas9 purified protein (NEB) (n=1 independent experiment was carried out). Both the sgRNA and the Cas9 protein were incubated with a 500-900 bp PCR product harbouring its respective target sequence. Restriction analysis following digestion of the PCR products with each sgRNA+Cas9 showed that *kat2a* sgRNA3, *kat2b* sgRNA1, *kat2b* sgRNA2 and *tbx5a* sgRNA2 presented *in vitro* activity, as they generated the expected pattern of bands after cleavage of their respective PCR products (Fig. 3.15A and B). The presence of single bands for the remaining sgRNAs indicates that Cas9 cleavage is not taking place, and therefore, sgRNA activity is low or non-existent. Thus, the generation of at least one active sgRNA targeting *kat2a* (sg3), *kat2b* (sg1 and

sg2) or *tbx5a* (sg2) was confirmed (their cleavage site in each of their targeting genes are shown in Fig. 3.15C). To confirm the integrity and length of these active sgRNAs, they were analyzed on a Bioanalyzer 2100. As shown in Fig. 3.15D, all three active sgRNAs produce a band of ~90 bp, which is the expected size of a typical sgRNA. A band of ~200 bp was also detected for all of them, which corresponds to secondary structures of the sgRNA.

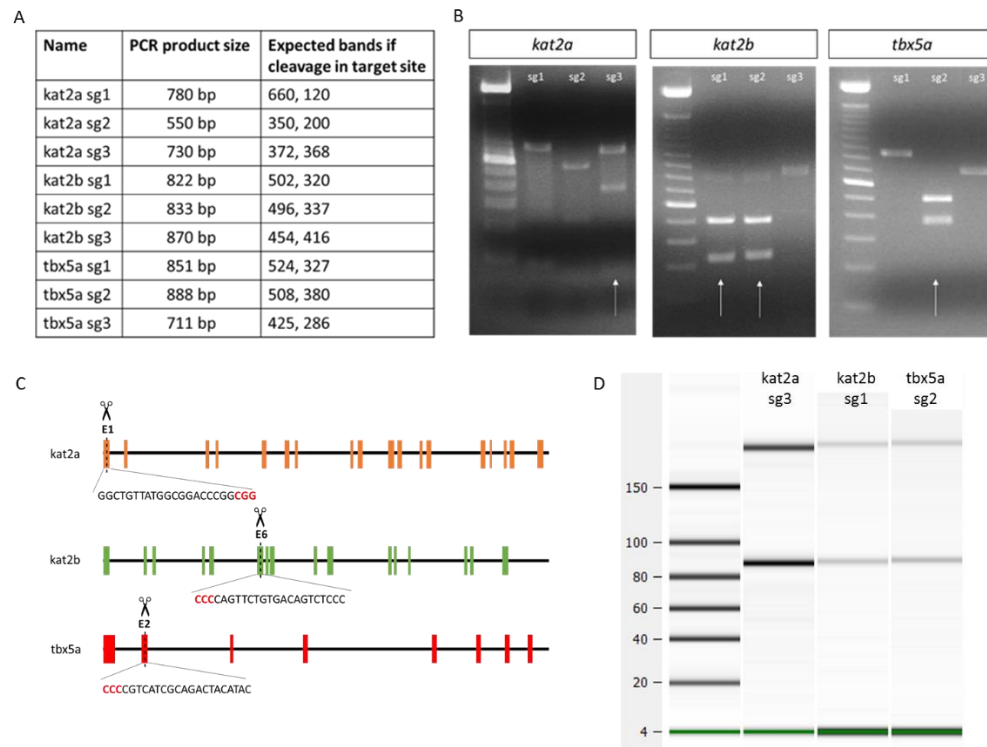


Figure 3.15. sgRNA activity *in vitro* assays. (A) Table showing the PCR products size used for the *in vitro* assays, as well as the expected fragments if Cas9 cleaves in the target site. (B) Gel analysis of the *in vitro* assays. *Kat2a* sgRNA3, *kat2b* sgRNAs 1 and 3, and *tbx5a* sgRNA5 present a high *in vitro* activity (white arrows), as indicated by the presence of multiple bands of the expected size. (C) Schematic representation of the CRISPR/Cas9 target sites for *kat2a* (exon 1, + strand), *kat2b* (exon 6, - strand) and *tbx5a* (exon 2, - strand). The PAM sequence is shown in red bolded. (D) Active sgRNAs analysis on a Bioanalyzer 2100. One band of ~90 bp (sgRNA) and another one of ~200 bp (secondary structure of the sgRNA) were detected in all three sgRNAs analysed.

3.3.7. CRISPR/Cas9-mediated knockout of *kat2a*, *kat2b* and *tbx5a*

After finding at least one active sgRNA for each of the *kat2a*, *kat2b* and *tbx5a* genes, each one of them was injected alongside Cas9 mRNA or protein into 1-cell stage zebrafish embryos (Cas9-only injected embryos were used as negative controls). For the initial injection attempts, *in vitro* synthesized Cas9 mRNA was used and two batches of embryos were injected per sgRNA (so a total of 83-110 embryos were injected per sgRNA). As shown in Fig. 3.16A, it

was found that just a 7% (n=8/110), 5% (n=4/83) and 5% (n=5/104) of injected embryos with Cas9 mRNA + *kat2a*, *kat2b* or *tbx5a* sgRNA respectively had any kind of phenotypic abnormality (e.g. defects in the development of any organs or tissues). In order to check whether these defects were indeed produced by the CRISPR/Cas9 system, DNA of those few abnormal embryos was extracted and Sanger sequencing was performed in order to check for the presence of mutations. None of the embryos that were analysed (n=3 embryos per sgRNA injection) carried mutations on the target sequence (Fig 3.16B), suggesting this that the CRISPR/Cas9 system did not work and that those defects were probably spontaneous abnormalities that usually arise during development in a small proportion of zebrafish embryos or injection artefacts.



Figure 3.16. Initial CRISPR/Cas9 system injections using Cas9 mRNA. (A) Number and percentage of Cas9 mRNA+sgRNA injected embryos showing a normal or abnormal phenotype

48 hpf. Two batches of embryos were injected per sgRNA so a total of 83-110 embryos were injected per sgRNA. (B) Sanger sequenced revealed that the embryos showing an abnormal phenotype did not harbour any mutations in the target site for *kat2a*, *kat2b* or *tbx5a*. Target site is highlighted in yellow, and PAM sequence is red underlined (n=3 embryos analyzed per sgRNA injected)

These results suggested that the CRISPR/Cas9 system was not working due to the fact that Cas9 mRNA was being used, which is susceptible to be degraded once it is injected into the 1-cell stage embryo, and it needs to be translated into a protein in order to exert its function. In order to overcome these problems, I decided to use purified Cas9 protein (New England Biolabs, UK) instead for the next injection experiments. Three independent experiments were performed: in each one of them, two batches comprised of 30-35 embryos were injected with Cas9 protein + the corresponding sgRNA so a total of 200 embryos (70, 70 and 60 in total for each independent experiment) were injected per sgRNA taking into account all three experiments. A Cas9-only negative control was also included. Following injection, zebrafish embryos were allowed to develop for 48 hpf at 28°C and the number of surviving embryos was analysed. As shown in Fig. 3.17A, an average of 71% (n=46/70; 55/70; 41/60), 66% (n=51/70; 41/70; 39/60) and 68% (n=42/70; 50/70; 43/60) of the embryos injected with *kat2a*, *kat2b* or *tbx5a* sgRNA+Cas9 protein survived and were analyzed, whereas 78% (n=59/70; 52/70; 45/60) of embryos injected with the Cas9 protein-only control survived. Surviving embryos were analysed for the presence of cardiac defects or abnormalities. In embryos injected with Cas9 and either *kat2a* sgRNA or *kat2b* sgRNA, 43% (n=61/142) and 39% (n=51/131) of embryos respectively displayed a mild phenotype remarkably similar to that found in *kat2a* and *kat2b* morphants, which included the presence of relaxed cardiac looping and slight pericardial oedema. In contrast, just a 6% (n=9/142) and 5% (n=7/131) of embryos injected with *kat2a* or *kat2b* sgRNA showed what it was described as a severe cardiac phenotype (no looping, severe pericardial oedema) (Fig. 3.17B and C). 51% (n=72/142) of *kat2a*-injected embryos and 56% (n=73/131) of *kat2b*-injected embryos presented a normal cardiac phenotype which was indistinguishable from their control counterparts. In the case of embryos injected with Cas9 + *tbx5a* sgRNA, 17% (n=23/135) of embryos injected presented a mild cardiac phenotype, whereas 43% (n=58/135) showed a severe one and 40% (n=54/135) displayed a completely normal cardiac phenotype (Fig. 3.17B and C).

In order to study more in-depth to which extent cardiac looping is affected in these embryos and to gain further insight into the cardiac phenotype of these F0 crispants embryos, *in-situ* hybridization using a *cm/c2* RNA probe (which is a typical cardiac marker in zebrafish^{387,388}) was carried out in pools of 15 mild and

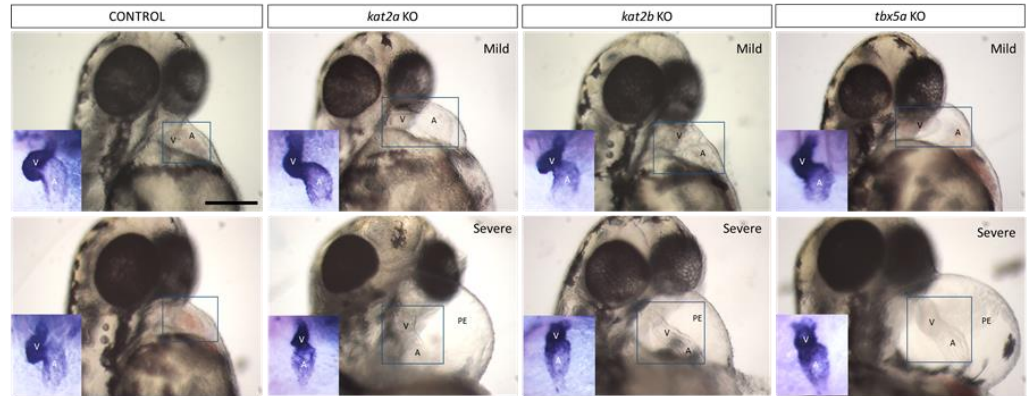
severe embryos for each one of the targeted genes. This procedure revealed that in mild embryos, cardiac chamber morphology as well as cardiac looping were slightly altered in comparison to controls; whereas in severe embryos, the heart is no more than a stretch tube displaying complete lack of looping and impaired cardiac chamber morphology. Cardiac looping angle measurement was also performed in F0 crispants as described in 2.2.2.13. (n=10 mild phenotype embryos and n=5 severe phenotype embryos due to the low number of severe embryos obtained in the *kat2a* and *kat2b* injections). Following analysis, it was revealed that the mean looping angle of *kat2a*, *kat2b* and *tbx5a* mild embryos was significantly higher than that of control embryos (*kat2a* mild $28.8^{\circ} \pm 5.2$; *kat2b* mild $25.6^{\circ} \pm 5.2$; *tbx5a* mild $36.4^{\circ} \pm 7.5$ vs control $16.4^{\circ} \pm 3.1$; $p < 0.05$ for all three comparisons) (Fig. 3.17E). In crispant embryos presenting a severe cardiac phenotype, the relaxation in cardiac looping was even more aggravated, as demonstrated by the presence of a greater cardiac looping angle in all three crispant groups (*kat2a* severe $55.8^{\circ} \pm 6.0$; *kat2b* severe $54.4^{\circ} \pm 4.8$; *tbx5a* severe $64.8^{\circ} \pm 4.0$ vs control $16.4^{\circ} \pm 3.1$; $p < 0.005$ for all three comparisons) (Fig. 3.17E). Besides cardiac looping, I also wanted to determine whether heart beating rate was affected in *kat2a*, *kat2b* and *tbx5a* crispant embryos in comparison to their control counterparts. Heart rate per minute (measured in beats per minute, or bpm) was measured as described in 2.2.2.14. (n=20 embryos for mild phenotypes and n=5 per severe phenotype due to the low number of severe embryos obtained in the *kat2a* and *kat2b* injections). As shown in Fig. 3.17D, heart beating rate in *kat2a*, *kat2b* and *tbx5a* mild embryos was significantly reduced in comparison to control embryos (*kat2a* mild $118.9 \text{ bpm} \pm 4.4$; *kat2b* mild $110.7 \text{ bpm} \pm 11.2$; *tbx5a* mild $109.5 \text{ bpm} \pm 7.3$ vs control $134.6 \text{ bpm} \pm 6.8$; $p < 0.05$ for all three comparisons). In *kat2a*, *kat2b* and *tbx5a* severe crispant embryos, heart beating rate was also severely diminished in comparison to control embryos (*kat2a* severe $89.8 \text{ bpm} \pm 4.9$; *kat2b* severe $92.6 \text{ bpm} \pm 7.6$; *tbx5a* severe $87.2 \text{ bpm} \pm 5.9$ vs control $134.6 \text{ bpm} \pm 6.8$; $p < 0.005$ for all three comparisons)

Overall, these results demonstrate that F0 *kat2a*, *kat2b* and *tbx5a* crispants embryos phenocopy the cardiac phenotype obtained in MO-mediated knockdowns of these genes, which is mainly characterized by the partial or complete lack of cardiac looping, the presence of pericardial oedema and a reduction in heart beating rate.

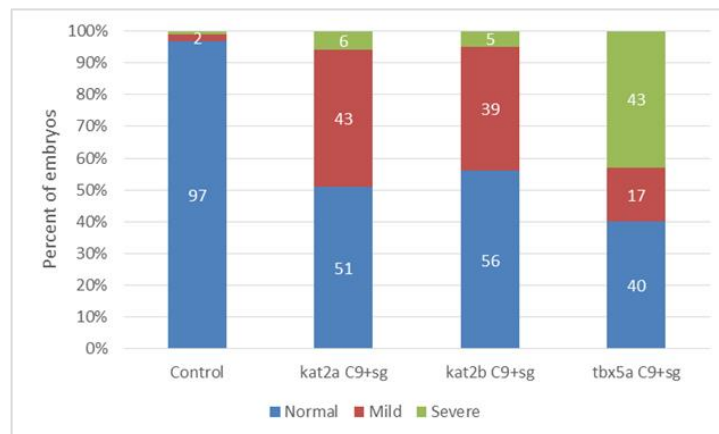
A

	# of surviving embryos 48 hpf	Avg % of survival
Cas9 prot + kat2a sgRNA	I: 46/70; II: 55/70; III: 41/60	71%
Cas9 prot + kat2b sgRNA	I: 51/70; II: 41/70; III: 39/60	66%
Cas9 prot + tbx5a sgRNA	I: 42/70; II: 50/70; III: 43/60	68%
Cas9 prot-only control	I: 59/70; II: 52/70; III: 45/60	78%

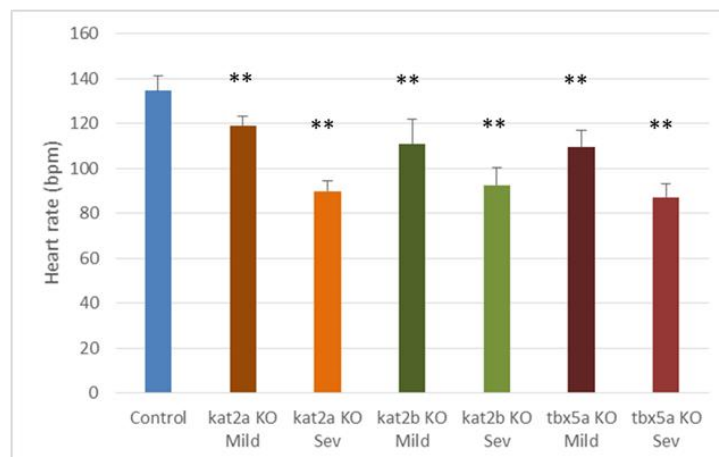
B



C



D



E

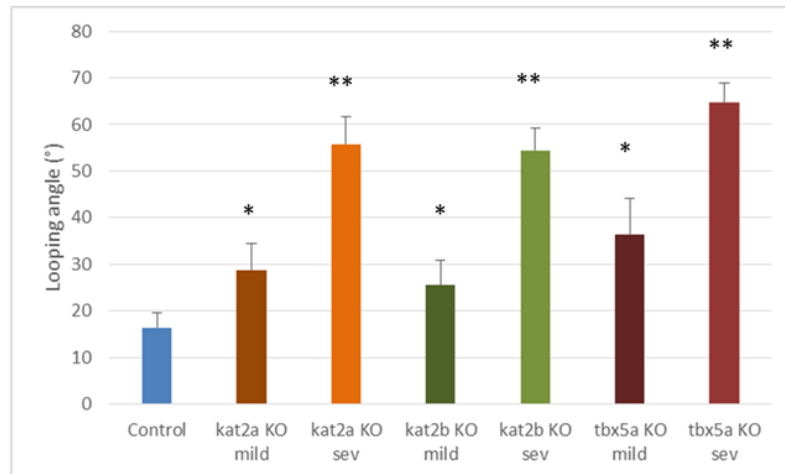


Figure 3.17. CRISPR-Cas9-mediated knockouts of *kat2a*, *kat2b* and *tbx5a* in zebrafish present a similar phenotype than their knockdown counterparts. (A) Table showing the number of embryos which survived for at least 48 hpf following CRISPR-Cas9 injections, as well as the average percentage of survival (60-70 embryos were injected per experiment, three independent experiments were carried out). These surviving embryos were the ones analysed for the presence of cardiac defects. (B) Representative images of 48 hpf control embryos (injected with Cas9 only) and embryos injected with Cas9 and a *kat2a*, *kat2b* or *tbx5a* sgRNA, respectively. Based on the severity of the observed cardiac phenotype, embryos were categorized as “mild” or “severe”. In the images, the heart is highlighted with a blue rectangle. In the bottom left corner, representative *in situ* hybridization images using a *cmhc2* probe are shown for each phenotype (bottom left corner, V=ventricle, A=atrium, PE=pericardial oedema). Scale bar = 200 μ m. (C) Histogram showing the distribution of cardiac phenotypes based on severity in injected embryos which survived for 48 hpf. (D) Heart rate was measured in 48 hpf embryos. This analysis revealed the presence of bradycardia in *kat2a*, *kat2b* and *tbx5a* KOs (n=20 mild phenotype embryos, 5 severe phenotype embryos). (E) Average looping angle was measured in 48 hpf control and KO embryos, revealing that this was significantly higher in mild and severe phenotype embryos for all three targeted genes (n=10 mild phenotype embryos, 5 severe phenotype embryos). For (D) and (E), mean $\pm$ SD is represented. Relative to controls, KO embryos were statistically significant (**P < 0.005, or *P < 0.05, Student’s t-test).

3.3.8. Analysis of mutations in CRISPR/Cas9-mediated KOs

In order to confirm that the F0 crispant embryos analyzed harboured mutations in the targeted genes, two different approaches were followed: T7 endonuclease assay and Sanger sequencing. For the T7 endonuclease assay, PCR products harbouring the CRISPR-Cas-targeted region in the *kat2a*, *kat2b* or *tbx5a* genes were generated using genomic DNA extracted from mild and severe embryos as a template. These products were re-annealed and subjected to T7 endonuclease digestion. Gel agarose analysis revealed that 95% (n=38/40) of *kat2a* crispants analyzed, 98% (n=39/40) of *kat2b* crispants analyzed and 100% (n=40/40) of *tbx5a* crispants harboured mutations in their respective target sites (n=40 embryos analyzed per targeted gene), as indicated by the presence of multiple bands of the expected size in the gel (Fig. 3.18A).

In control embryos analyzed (n=20 per gene), mutations were detected in just a 5%, 10% and 5% of them respectively. Apart from the PCR product and cleavage product bands, additional bands were also detected in the T7 analysis of all three knockouts. These bands are originated due to the presence of polymorphic sequences in the intronic regions included in the PCR products used for the assay. Following T7 endonuclease assay, individual alleles derived from positive samples for this assay were subjected to Sanger sequencing in order to confirm the presence of mutations and gain further information about them (n=8 samples per gene and 5 alleles were analyzed per sample, for a total of 40 alleles analyzed per gene). Sequencing confirmed that injected embryos displaying a mild or severe phenotype did harbour different mutations (insertions, deletions or substitutions) in their respective CRISPR-Cas target site (Fig. 3.18B). As expected, different mutations were found in an individual sample, confirming the presence of mosaicism in this F0 crispant embryos. A preliminary CRISPR-Cas9 system rate of mutagenesis was also calculated by dividing the number of mutant alleles found per gene by the total number of alleles analyzed per gene (5 alleles analyzed per embryos, 8 embryos analyzed per gene). This analysis indicated that the rates of mutagenesis for Cas9 + *kat2a* sgRNA, *kat2b* sgRNA or *tbx5a* sgRNA were 88% (n=35/40), 78% (n=31/40) and 82% (n=33/40) respectively, thus suggesting the presence of bi-allelic mutations in a substantial number of cells in these embryos (Fig. 3.18C).

Overall, these results strongly suggest that the mild and cardiac phenotype found in *kat2a*, *kat2b* and *tbx5a* injected embryos is indeed due to the presence of disruptive and, in most cells, bi-allelic mutations in the targeted sites of these genes.

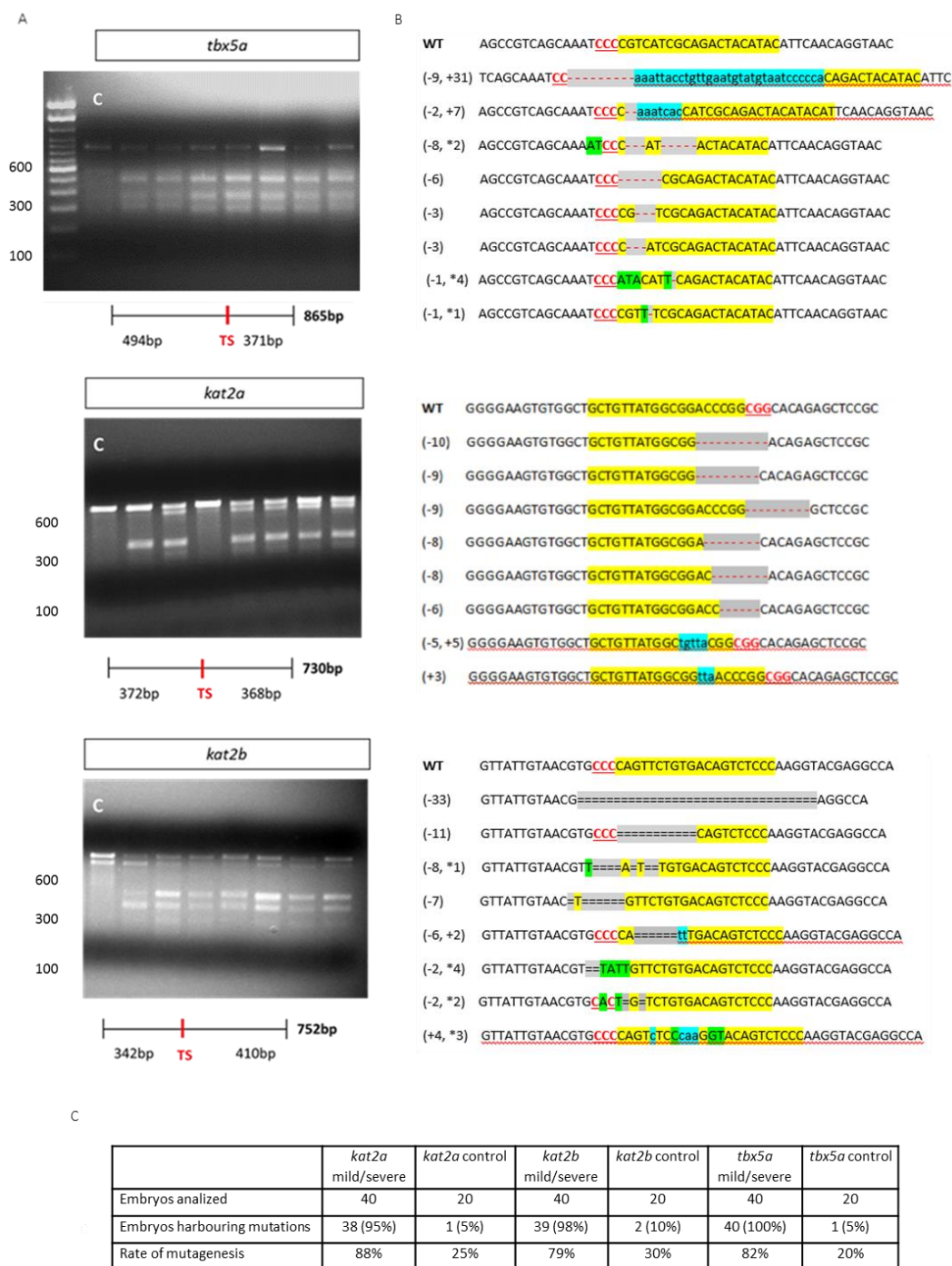


Figure 3.18. Analysis of CRISPR/Cas9-injected embryos for the presence of mutations. (A) T7 assay confirmed the presence of indels in the *kat2a*, *kat2b* and *tbx5a* genes in injected embryos. For *kat2a*, 6/7 analysed embryos gave a positive result (presence of extra bands of expected size), whereas 7/7 gave a positive result for both *kat2b* and *tbx5a* injected embryos (C = control). A diagram of the PCR products used for the assay is also shown (TS = target site). In total, 40 embryos were analyzed per gene. (B) A list of some of the mutations found in *kat2a*-, *kat2b*- and *tbx5a*-injected embryos is shown under the relevant embryos. Target site is highlighted in yellow, and PAM sequence is red underlined. Deletions are highlighted in grey, insertions in blue and substitutions in green. (C) Table indicating the number of embryos harbouring any mutation in the *kat2a*, *kat2b* or *tbx5a* genes in Cas9+sgRNA injected embryos.

Rate of mutagenesis is defined as the percentage of mutant alleles in all alleles analysed for each individual gene (n=5 alleles analyzed per embryo, 8 embryos analyzed in total per gene).

Taking all these results together, the cardiac phenotype found in *kat2a*, *kat2b* and *tbx5a* knockdown embryos was validated in embryos harbouring mutations generated by the CRISPR/Cas9 genome-editing system. Furthermore, analysis of the cardiac phenotype in KO embryos confirmed the similarity between *kat2a/kat2b* and *tbx5a* mutants in terms of cardiac morphology, looping degree and rate of heart beat.

3.3.9. *kat2a* and *kat2b* KDs/KOs present abnormal fin development

It has been demonstrated that during zebrafish development, *tbx5a* also play a crucial role in fin development^{126,389}. In order to study whether *kat2a* and *kat2b* are implicated in the regulation of this process as well, both MO-mediated knockdown and CRISPR-Cas-mediated knockout embryos for *kat2a*, *kat2b* and *tbx5a* were screened for the presence of abnormal fin development at 72 hpf. Embryos analysed for the presence of cardiac defects were the ones included in the fin phenotype analysis, except those which were used for other purposes such as ISH or did not survive up to 72 hpf. In the case of CRISPR-Cas-mediated knockouts, 33% (n=39/117), 36% (n=38/106) and 18% (n=20/110) of *kat2a*, *kat2b* and *tbx5a* crispants analyzed presented stumpy, shortened fins (mild phenotype) in comparison to their control counterparts, respectively (Fig. 3.19A and B). In *kat2a* and *kat2b* crispants, a low proportion of embryos (5% (n=6/117) and 8% (n=8/106) respectively) displayed a more severe fin phenotype in which they completely lacked both the left and right pectoral fins (Fig. 3.19A and B). In *tbx5a* crispants, the proportion of embryos showing this severe fin phenotype was much higher than in *kat2a* or *kat2b* crispants (42%, n=46/110). Similar results were obtained when *kat2a*, *kat2b* and *tbx5a* knockdown embryos were analysed: 39% (n=56/144), 44% (n=67/152) and 15% (n=23/154) of embryos displayed shortened, stumpy fins, whereas 9% (n=13/144), 6% (n=9/152) and 55% (n=85/154) of them had no fins at all. In double knockdown embryos for both *kat2a* and *kat2b*, the proportion of embryos with a severe fin phenotype was higher than their single knockdown counterparts (36%, n=46/128) and much closer to the proportion of *tbx5a* embryos showing this phenotype (55%, n=70/128), whereas 17% (n=22/128) had stumpy fins (Fig. 3.19A and B).

In order to better visualize the fins and to further confirm the similarities between the fin phenotype found in *kat2a*, *kat2b* and *tbx5a* knockouts, *in situ* hybridization using a *bmp4* RNA probe was performed on 48 hpf *kat2a*, *kat2b* and *tbx5a* crispant embryos (n=10 embryos subjected to ISH per gene). *Bmp4*

is typically used as a fin marker in zebrafish, and it has also been shown to be one of *tbx5a* target genes in this animal model³¹⁸. As shown in Fig. 3.20, *bmp4* signal can still be detected in the fins of those *kat2a*, *kat2b* and *tbx5a* crispants with shortened and stumpy fins; however, the signal is much smaller and fainter than the *bmp4* signal detected in control embryos. In embryos presenting a severe fin phenotype (lack of both fins), no signal in the pectoral fins area could be detected at all.

Altogether, these results indicate that both *kat2a*/*kat2b* morphants and crispants show an abnormal fin phenotype that resemble that of *tbx5a* morphants and crispants, thus supporting the existence of an *in vivo* link between them

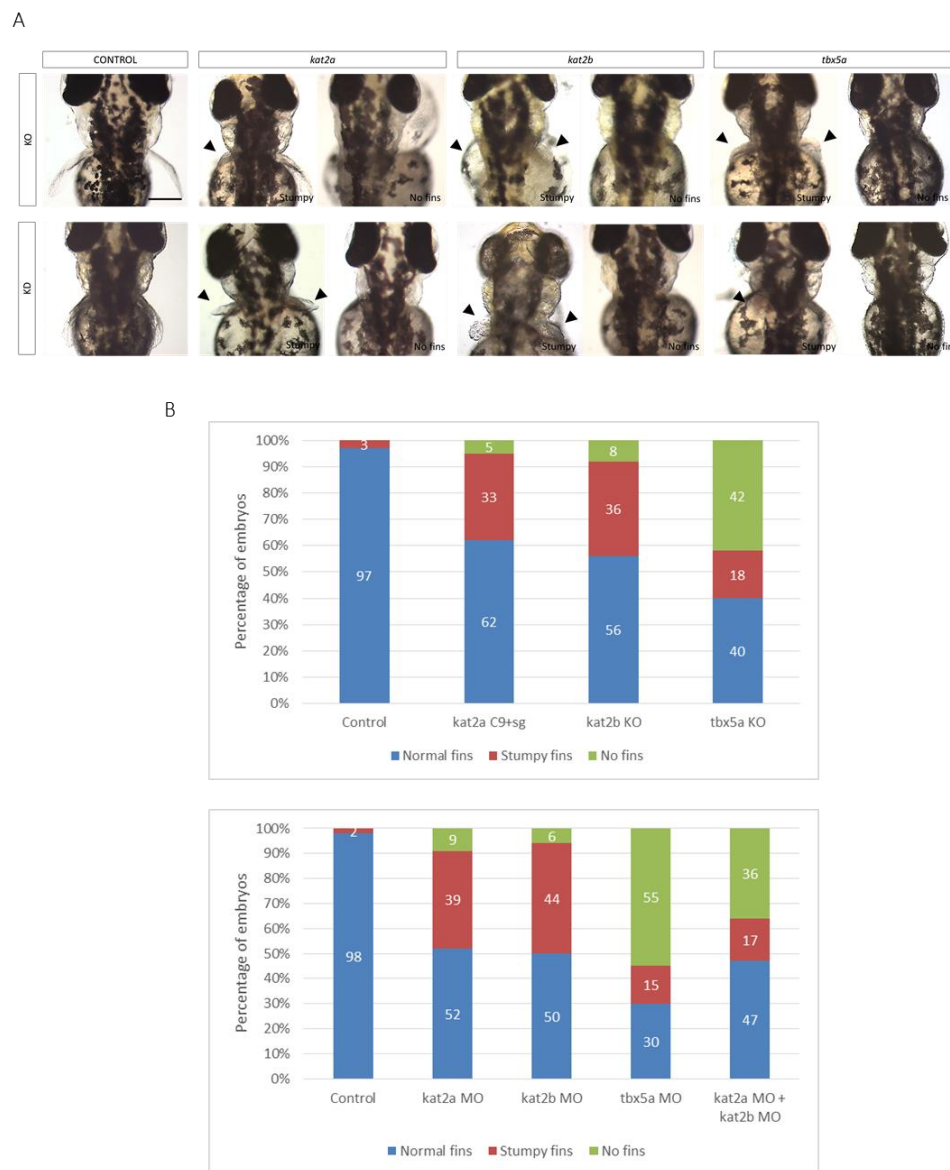


Figure 3.19. Fin development is also altered in *kat2a* and *kat2b* KO/KD embryos. Representative images of 72 hpf control embryos showing fully-developed pectoral fins and

kat2a, *kat2b* and *tbx5a* crispant/morphants embryos in which fin development is affected. For both knockout and knockdown embryos, representative pictures of samples with stumpy fins or no fins are shown. Black arrows indicate the presence of shortened, stumpy fins. Scale bar = 200 μ m (B) Graphs showing the distribution of fin phenotypes in Cas9+sgRNA-injected embryos and in morpholino-injected embryos according to severity. The total number of embryos analysed for each group is indicated in the main text.

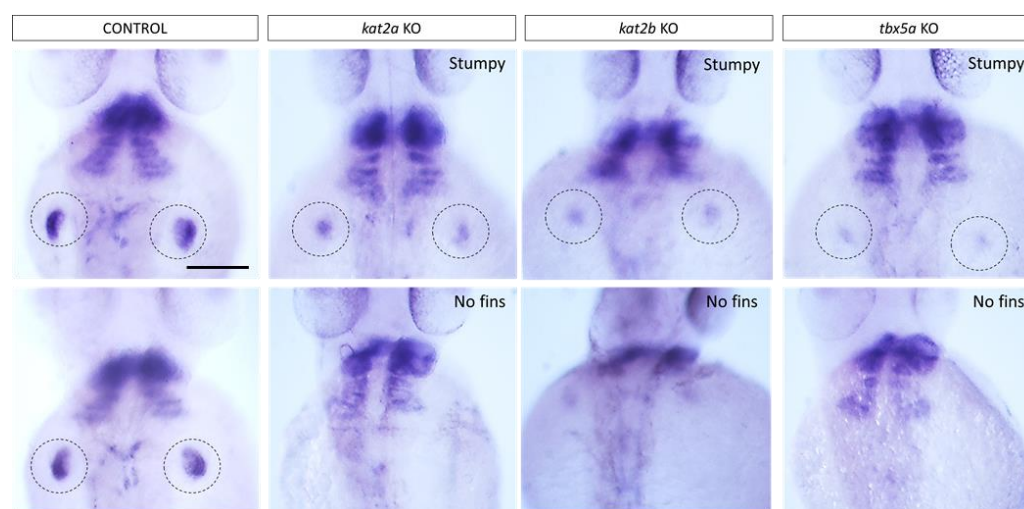


Figure 3.20. *In situ* hybridization in 48 hpf embryos using a *bmp4* probe. As shown in the pictures, in embryos with stumpy fins these are smaller in size and the *bmp4* signal is reduced when compared to the signal found in control embryos. In contrast, no *bmp4* signal can be found in embryos which completely lack pectoral fins. In control and stumpy embryos, fins are highlighted with dashed blue circles. Scale bar = 200 μ m. (n=10 embryos subjected to ISH per group).

3.3.10. HAT inhibitor treatment of zebrafish embryos during development

Following the generation of *kat2a*, *kat2b* and *tbx5a* morphants and crispants using both MOs and the CRISPR-Cas9 system, I wanted to further demonstrate that the phenotype seen in these embryos is indeed due to the lack of acetylation, and not due to other reasons (e.g. off-target effects). In order to achieve this, zebrafish embryos were treated with Garcinol, an inhibitor of both KAT2B and p300 which presents a higher affinity for the former^{306,384}. Initially, I carried out a Garcinol titration in zebrafish embryos by exposing them (n=10-15 embryos per concentration, three independent experiments were carried out) to increasing doses of the drug for 24 hours starting at 20 hpf, and then the proportion of dead embryos or embryos presenting severe developmental defects was calculated. Fig. 3.21 shows the results of the Garcinol titration. Up to a concentration of 100 μ M, the proportion of dead/severe embryos was below 30% for all concentrations analysed (10% at 0 μ M (n=4/40 taking into account all three experiments), 15% at 0.1 μ M (n=6/40), 18% at 1 μ M (n=7/40),

20% at 5 μM ($n=8/40$), 23% at 10 μM ($n=9/40$), 25% at 50 μM ($n=10/40$) and 28% at 100 μM ($n=11/40$). However, at 250 μM and 500 μM this proportion dramatically increased until 60% ($n=24/40$) and 85% ($n=34/40$) respectively. Taking into account that in both knockdown and knockout experiments the proportion of dead/severe embryos following injection oscillates between 25% and 35%, I decided to choose a concentration of 100 μM of Garcinol for further experiments, as this one leads to a survival rate of 73%, similar to the survival rates found in both MO and CRISPR-Cas9 experiments.

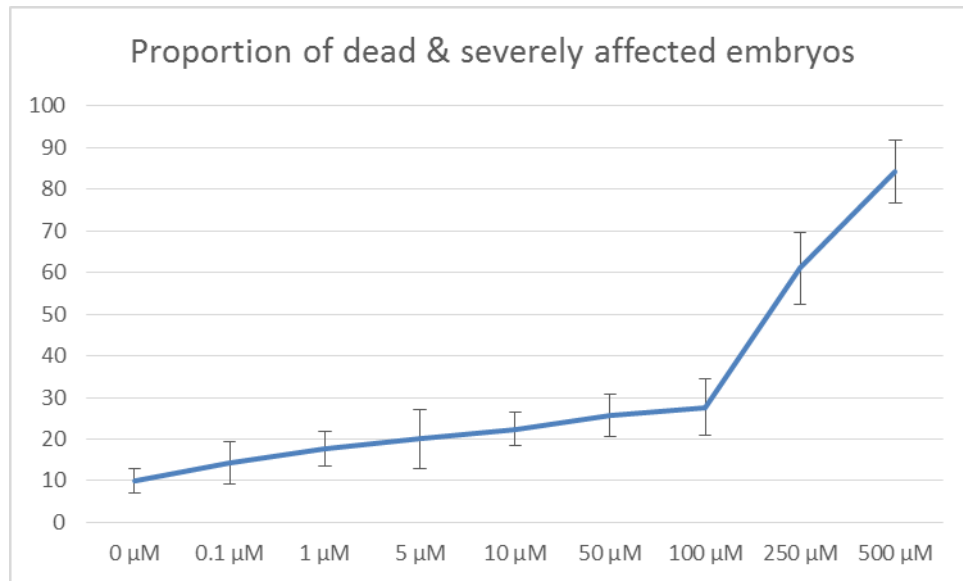


Figure 3.21. Garcinol titration in zebrafish embryos. Embryos were treated for 24 hours starting at 20 hpf, and the proportion of dead embryos or embryos showing severe developmental defects was calculated. Result are mean $\pm$ SD ($n=3$ independent experiments, 10-15 embryos were analysed per experiment and per concentration)

Once a working concentration for Garcinol was established, WT zebrafish embryos were treated with 100 μM of Garcinol for 24 hours starting at 20 hpf ($n=60-70$ embryos treated per group (control and Garcinol) and per experiment for a total of 200 embryos treated per group. Three independent experiments were performed). As shown in Fig. 3.22A, an average of 69% ($n=47/70$; 52/70; 39/60), of the embryos treated with Garcinol 100 μM survived until they reached 48 hpf and were analysed, whereas 82% ($n=58/70$; 54/70; 51/60) of control embryos survived. Analysis of surviving embryos at 48 hpf revealed that Garcinol-treated embryos showed cardiac and limb defects similar to those found in *kat2a*, *kat2b* and *tbx5a* mild embryos, including a slight pericardial oedema, relaxation in cardiac looping (as revealed by *cm1c2* ISH), and the presence of shorter pectoral fins in comparison to their control counterparts (Fig. 3.22B). 84% ($n=116/138$) of them showed a mild cardiac phenotype, whereas just a 3% ($n=4/138$) showed a severe cardiac phenotype (in non-treated, control embryos, 94% ($n=153/163$) of them displayed a completely

normal cardiac phenotype) (Fig. 3.22C). Regarding fin development, 75% (n=104/138) of surviving Garcinol-treated embryos showed a mild fin phenotype, characterized by the presence of shortened fins when compared to controls (Fig. 3.22B). In order to gain further insight into the cardiac phenotype of these embryos, cardiac looping angle and heart beating rate was calculated in a similar fashion to how it was done in F0 crispants. Measurement of looping angle showed that this is significantly increased in Garcinol treated embryos in comparison to control embryos ($28.4 \pm 5.6^\circ$ vs $16.2 \pm 2.4^\circ$, $p < 0.05$, n=10 embryos per group), thus confirming the relaxation of cardiac looping (Fig. 3.22E) Heart rate measurement also confirmed the presence of significant bradycardia in treated embryos: in these, heart rate was 111.2 ± 6.9 bpm, whereas in control embryos this rate reached 134.8 ± 7.1 bpm ($p < 0.05$, n=20 embryos per group) (Fig. 3.22D).

Overall, these results indicate that treatment of zebrafish embryos with an acetyltransferase inhibitor such as Garcinol also lead to a cardiac and fin phenotype resembling that of *kat2a* and *kat2b* morphants and crispants, thus further confirming the relevance of acetylation in an *in vivo* context.

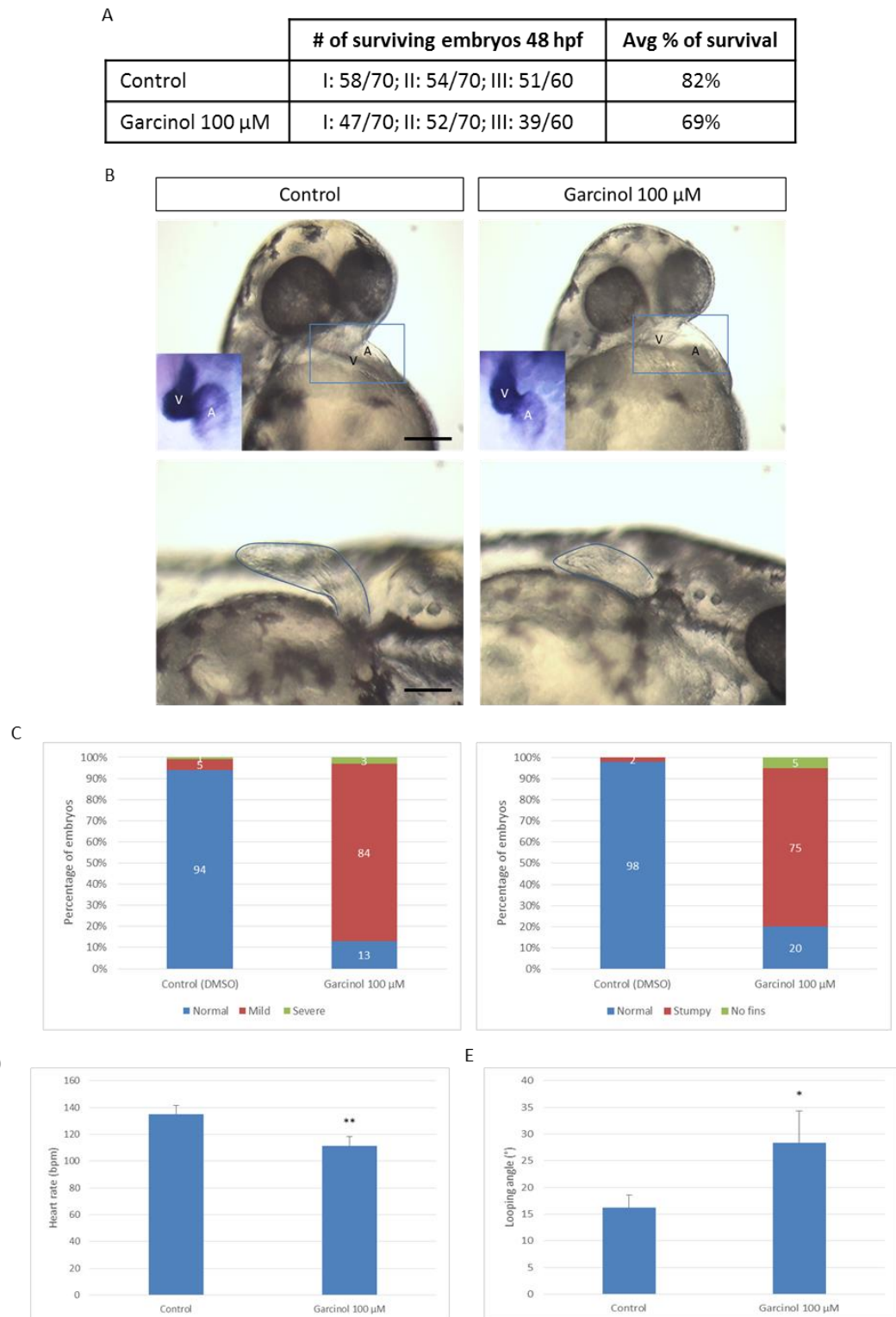


Figure 3.22. Chemical inhibition of acetyltransferase activity with Garcinol also impairs cardiac and fin development in zebrafish. (A) Table showing the number of embryos which survived for at least 48 hpf following treatment with Garcinol 100 μ M, as well as the average percentage of survival (60-70 embryos were treated per experiment, three independent experiments were carried out). These surviving embryos were the ones analysed for the presence of cardiac or fin defects. (B) Representative images of 48 hpf control embryos treated with Garcinol 100 μ M. The cardiac phenotype is shown on the upper row: the heart is

highlighted with a blue square, and representative *in situ* hybridization images using a *cmlc2* probe are shown for each phenotype (bottom left corner of each image, V=ventricle, A=atrium). In the bottom row, representative images of the fin phenotype are also shown. The size and shape of the pectoral fins are highlighted in blue. (B) Distribution of cardiac and fin phenotypes based on severity in treated and untreated embryos. (C) Heart rate measurement indicated the presence of bradycardia in 48 hpf embryos treated with Garcinol (n=20 embryos analysed per group). (D) Average looping angle in 48 hpf control and treated embryos (n=10 embryos per group). For (C) and (D), mean $\pm$ SD from three individual experiments is represented. Relative to controls, treated embryos were statistically significant (**P < 0.005, or *P < 0.05, Student's t-test). Scale bar = 200 μ m

3.3.11. Expression levels of *Tbx5a* target genes in *kat2a* and *kat2b* KOs

The results obtained up to this point indicate that the loss of *kat2a* and *kat2b* during zebrafish development lead to a cardiac and fin phenotype remarkably similar to that produced by the lack of *tbx5a*. However, no direct link between these factors in zebrafish has been established yet. In order to address this, I carried out qPCR analysis on 32 hpf *kat2a* and *kat2b* KO embryos to check the potential misregulation of *tbx5a* activity, and thus, of several of its target genes such as *fgf10*, *bmp4*, *nppa*, *hey2* and *tbx2b*. The expression of two genes unrelated to *tbx5a* which have been demonstrated to be expressed during zebrafish development (*akt1*³²⁰ and *igf3*³²¹) was also checked, and two well-established housekeeping zebrafish genes, *ef1a* and *lsm12*, were used for normalization (n=3 biological replicates; 3 technical replicates per biological replicate and per gene analysed were also included). Following RNA extraction, cDNA synthesis and qPCR analysis, it was revealed that all five *tbx5a* target genes analysed were significantly downregulated both in *kat2a* and *kat2b* crispant embryos 32 hpf in comparison to their WT counterparts (p<0.05 for all of them both in *kat2a* and *kat2b* crispants except for *bmp4* in *kat2b* KOs, which is close to reach significance as p = 0.05). In the case of *akt1* and *igf3*, the expression of these two genes did not significantly change in *kat2a* or *kat2b* crispants (p>0.05 in both), thus giving confidence that the changes observed reflect the downregulation of *tbx5a* activity (Fig. 3.23). In order to check if the loss of *kat2a* or *kat2b* has any effect on *tbx5a* expression, the relative expression level of *tbx5a* was also measured in these crispants. As shown in Fig. 3.23, *tbx5a* relative expression was slightly reduced both in *kat2a* and *kat2b* crispants; however, no significant differences were found in both cases (p>0.05).

To further confirm the down-regulation of *tbx5a* target genes observed in the qPCR analysis, three of these genes (*fgf10*, a fin marker; *bmp4* and *nppa*, two

cardiac markers) were chosen for ISH on 32 hpf *kat2a* and *kat2b* crispant embryos (n=5 *kat2a* and 5 *kat2b* crispant embryos were subjected to ISH per gene analysed). ISH using a *fgf10* RNA probe revealed the down-regulation of this gene in both the right and left pectoral fins of *kat2a* and *kat2b* crispant embryos, as well as in other areas of the embryo such as the head (Fig. 3.24). In the case of *nppa*, a strong reduction in its signal was detected in the heart of *kat2* crispants in comparison to controls, both in the ventricle and in the atrium. Similar results were obtained for *bmp4*: a reduction in its expression was detected in the heart of both *kat2a* and *kat2b* crispants, although these down-regulation was more subtle than that observed for *nppa* (Fig. 3.24).

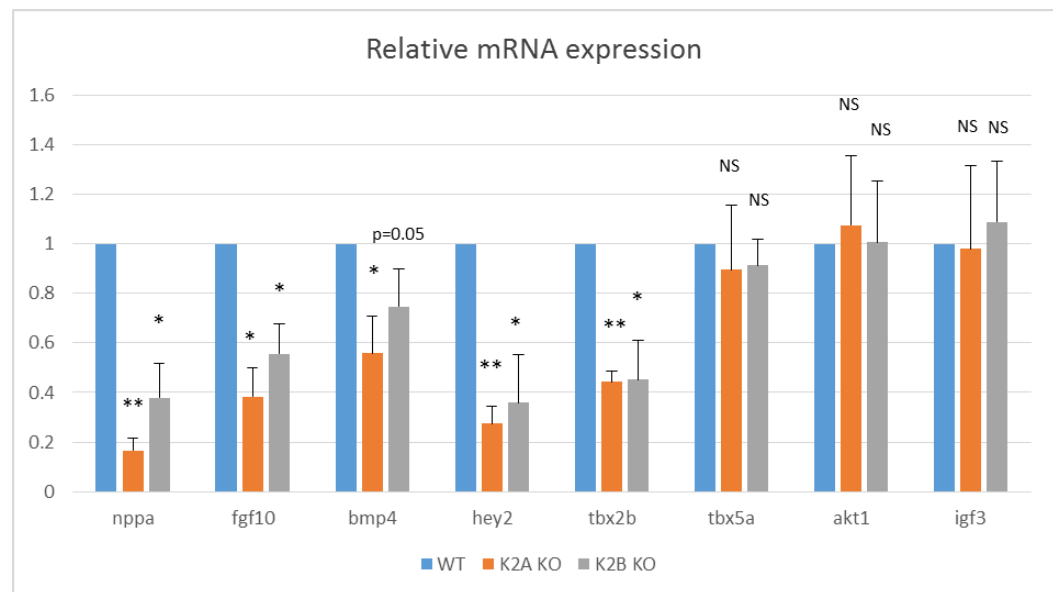


Figure 3.23. Expression levels of *tbx5a* target genes in *kat2a* and *kat2b* KO embryos. mRNA levels of *nppa*, *fgf10*, *bmp4*, *hey2* and *tbx2b* were assayed by quantitative PCR at 32 hpf in *kat2a* and *kat2b* KO embryos. Relative expression of *tbx5a* and unrelated genes *akt1* and *igf3* was also measured. Mean $\pm$ SD is represented (n=3 biological replicates; 3 technical replicates per biological replicate and per gene analysed were also included). **P < 0.005, or *P < 0.05, Student's t-test. NS = not significant

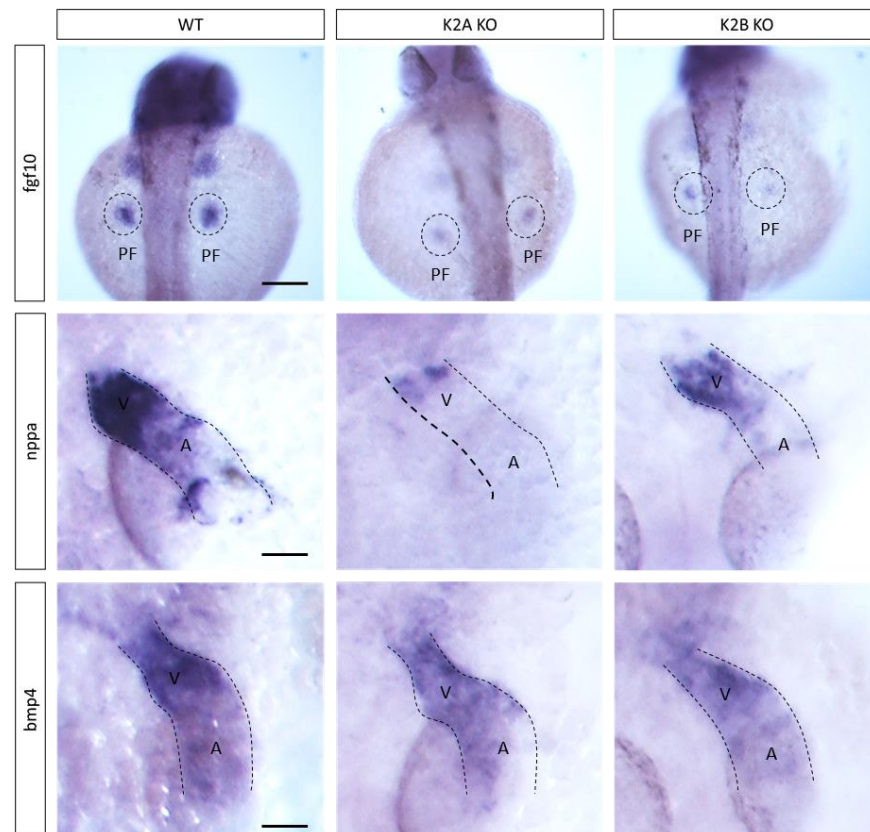


Figure 3.24. The expression of *fgf10*, *nppa* and *bmp4* via *in situ* hybridization. The experiment was performed on 32 hpf *kat2a* or *kat2b* KO embryos (n=5 *kat2a* and 5 *kat2b* crispant embryos were subjected to ISH per gene analysed). Pectoral fins (PF) are highlighted by a dashed circle, and the heart is highlighted by dashed line (V=ventricle, A=atrium). Scale bar = 200 μ m for *fgf10* and 100 μ m for *nppa* and *bmp4*

3.3.12. Rescue of the *kat2a* KO cardiac and fin phenotype by a *tbx5a* acetyl-mimic transcript

The qPCR and ISH analysis on *tbx5a* target genes provided evidence that there is an *in vivo* link between *kat2* acetylases and *tbx5a*. In order to further confirm the relationship between these factors in zebrafish, I performed a rescue experiment of the *kat2a* crispant phenotype using an acetyl-mimic version of *tbx5a* (*tbx5a*-K339Q). The plasmid carrying this acetyl-mimic version of the protein was designed and produced by myself as indicated in 2.2.2.17. Following generation of the mRNA transcript, this was co-injected alongside Cas9 protein+*kat2a* sgRNA in 1-cell stage zebrafish embryos. Three independent experiments were performed: in each one of them, two batches comprised of 30-35 embryos were injected with Cas9 protein-only (control), Cas9 + *kat2a* sgRNA (knockout) or Cas9 + *kat2a* sgRNA + *tbx5a*-K339Q mRNA (rescue) so a total of 200 embryos (70, 70 and 60 in total in each independent experiment) were injected per group taking into account all three experiments.

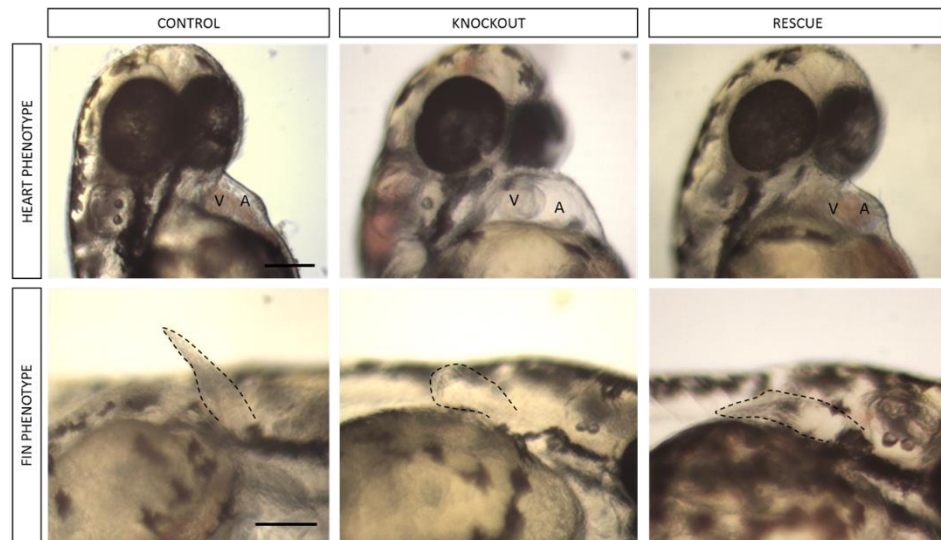
Following injections, embryos were let to develop for 72 hpf, and the cardiac and fin phenotypes of surviving embryos were analysed at 48 and 72 hpf respectively. As shown in Fig. 3.25A, an average of 67% (n=48/70; 40/70; 46/60) and 70% (n=46/70; 52/70; 41/60) of the knockout and rescue embryos survived 48 hpf, whereas 75% (n=53/70; 49/70; 48/60) of control embryos did. In surviving embryos, injection of Cas9 + *kat2a* sgRNA (knockout group) produce a mild cardiac phenotype in 42% (n=56/134) of them at 48 hpf, whereas just 6% (n=8/134) displayed a severe cardiac phenotype and 52% (n=70/134) showed a normal cardiac phenotype. However, when *tbx5a* K339Q mRNA was injected alongside the CRISPR-Cas9 system (rescue group), the proportion of surviving embryos showing a mild cardiac phenotype was smaller (25%, n=35/139), whereas the percentage of embryos displaying a completely normal cardiac phenotype was increased up to 72% (n=100/139) (Fig. 3.25B and C). This increase in the percentage of embryos with a normal cardiac phenotype in the rescue group was also seen with regards to fin phenotypes. In the knockout group, 34% (n=46/134) of analysed embryos showed the previously described mild fin phenotype (stumpy fins), whereas 59% (n=79/134) of them presented normal fins. In contrast, in the rescue group, the proportion of embryos with a normal fin phenotype was 75% (n=104/139), and 23% (n=32/139) displayed a mild fin phenotype (Fig. 3.25B and C).

These results indicate that the acetyl-mimic version of *tbx5a* (*tbx5a* K339Q) is able to partially rescue the cardiac and fin phenotype shown by *kat2a* crispant embryos, thus supporting the fact that the lack of *tbx5a* acetylation is the responsible for the phenotype and strengthening the *in vivo* link between *kat2* acetylases and *tbx5a*.

A

	# of surviving embryos 48 hpf	Avg % of survival
Control	I: 53/70; II: 49/70; III: 48/60	75%
Knockout	I: 48/70; II: 40/70; III: 46/60	67%
Rescue	I: 46/70; II: 52/70; III: 41/60	70%

B



C

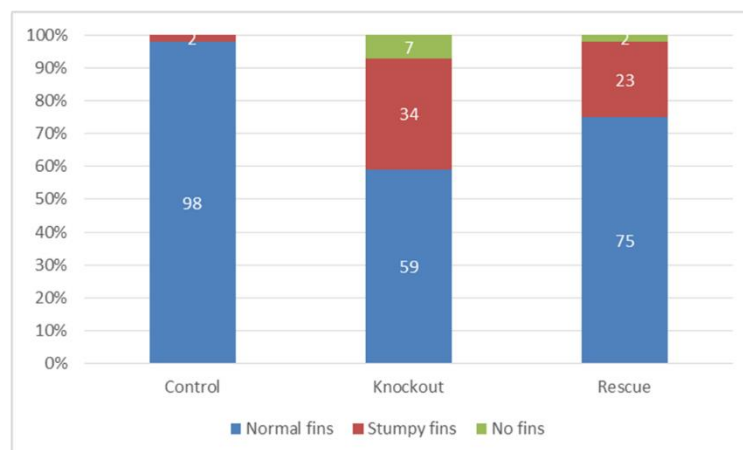
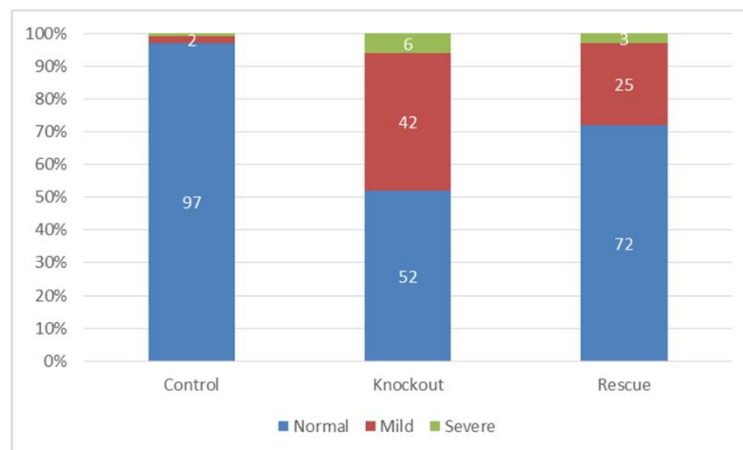


Figure 3.25. Rescue of the *kat2a* KO phenotype using an acetyl-mimic version of *tbx5a*. (A) Table showing the number of embryos which survived for at least 48 hpf following CRISPR-Cas9 injections, as well as the average percentage of survival (60-70 embryos were injected per experiment, three independent experiments were carried out). These surviving embryos were the ones analysed for the presence of cardiac or fin defects. (B) Representative images of 48 hpf control embryos (injected with Cas9 only) and embryos injected with Cas9 + *kat2a* sgRNA (knockout) or Cas9 + *kat2a* sgRNA + *tbx5a* K339Q mRNA (rescue). Both the main heart and fin phenotypes found in each group are showed. In the rescue group, a rescue of the mild phenotype found in knockout embryos is seen in some of the embryos. Scale bar = 200 μ m. (C) Distribution of cardiac and fin phenotypes based on severity in control, knockout and rescue embryos.

3.4. Conclusions, discussion and perspectives for future work

TBX5 is one of the key regulators of cardiac development, controlling the expression of multiples genes during cardiogenesis³⁹⁰. Several studies have demonstrated that the transcriptional activity of TBX5 is modulated by post-translational-modifications such as acetylation exerted by the acetyltransferase p300³⁴⁹, and PIAS1-mediated sumoylation³⁹¹. In this project, further insight into the upstream regulators of TBX5 is provided, as we show that TBX5 transcriptional activity is potentiated by KAT2A and KAT2B-mediated acetylation and that these two factors are required for normal cardiac and limb development in zebrafish.

The initial *in vitro* experiments which laid the foundations of this project were primarily designed and carried out by Dr. Tushar Ghosh, although I had the chance to participate in them during my initial training period in the lab. These experiments demonstrated that both KAT2A and KAT2B are able to physically interact and acetylate TBX5 *in vitro*, and that this acetylation leads to an enhanced TBX5 transcriptional activity. These results are in line with earlier reports that demonstrate that acetylation and deacetylation by p300 and Hdac3 regulate Tbx5-mediated gene transcription during early cardiogenesis³⁴⁹. It has also been demonstrated that TBX5 interacts with Tip60 (a member of the MYST family of histone acetyltransferases) in order to activate transcription on the SRF promoter³⁹², and that acetylation of other cardiac transcription factors such as GATA4³⁷⁷ and MEF2C³⁷⁶ is required for proper cardiomyocyte differentiation. In contrast to what was reported by Lewandowski *et al.*³⁴⁹, no acetylation of TBX5 by p300 was found in these initial experiments. A possible explanation for this disparity could be the fact that different cell lines were used in these two experiments, or that p300 promotes

a low-level of acetylation that was not detected in the experiments presented here.

In order to identify the KAT2 key acetylation residue in the TBX5 protein, three different online prediction tools were used: PredMod, PHOSIDA and PAIL. These three webtools identified the lysine 234 as one of the potential acetylation residues in the TBX5 protein; however, as demonstrated by Tushar Ghosh, the conservative substitution of this residue does not abolish TBX5 acetylation. In view of these results, we decided to derive a KAT2A/KAT2B acetylation motif from the sequences of several of their acetylation targets reported in the literature. The alignment of these sequences around the key lysine residue for acetylation generated the consensus motif K-X-X-K-K/R-P/L/K. This full-length motif was not found in the TBX5 sequence; however, two partial motifs KRK and KKP were found at positions 325-327 and 339-341 respectively. Out of these four lysine residues, only the conservative substitution of lysine 339 abolished TBX5 acetylation, thus confirming that this is the KAT2A/KAT2B acetylation residue in this protein. Even though these two partial motifs are quite short in length, they overlap with the recognition sequences of acetylases p300 and GCN5^{380,381}. In spite of this, the possibility that these motifs might appear in proteins which are not susceptible to acetylation by random chance should be considered. With regards to this, it could have been of great interest to study how often they appear in the genome in order to provide further confirmation of whether they are meaningful in the context of acetylation; however, this constitutes a quite time-consuming task that is beyond the scope of this thesis. In this context, a preliminary and quick analysis on the sequence of 25 proteins that, to date, have not been reported to be susceptible to acetylation demonstrated that just three of them included either of these motifs in their full sequence, thus suggesting that these two acetylation motifs might show up more frequently in proteins susceptible to be acetylated. In spite of this, this analysis was performed on a very small sample, and further data is needed in order to confirm that these motifs are indeed meaningful in the context of acetylation.

TBX5 acetylation takes place at lysine 339, which is included in one of the two nuclear localization sequences found in TBX5: NLS1 (aa 78-90) and NLS2 (aa 325-340)¹²⁰. Considering this, I next aimed to study whether acetylation has an impact on the already characterized TBX5 nuclear localization¹²⁰. Immunofluorescence experiments performed both in COS7 and C2C12 cells confirmed that the acetyl-deficient variant TBX5-K339R is relocated to the cytoplasm in many of the cells analysed from both lines, in contrast to the exclusively nuclear localization displayed by its WT counterpart. This impaired

TBX5 cellular distribution was also observed when TBX5-WT-transfected cells were treated with Garcinol, an inhibitor of acetyltransferase activity. In view of these results, I hypothesized that non-acetylated TBX5 can be relocated to the cytoplasm due to two different mechanisms: either by an impaired TBX5 nuclear import, or by a defect in its nuclear retention. To ascertain what the correct mechanism was, TBX5-transfected cells were treated with Leptomycin B, a potent inhibitor of the exportin CRM1, which has been described to mediate TBX5 nuclear export through interaction with a nuclear export sequence localized in the TBX5 protein (NES, aa 152-160)¹²¹. When treated with Leptomycin B, the acetyl-deficient variant TBX5-K339R was found in the nucleus in most cells analysed, in contrast to the cytoplasmic localization found in untreated cells. In addition to this, immunofluorescence analysis revealed that TBX5-K339R co-localize with CRM1 right outside of the nucleus. These results strongly suggest that TBX5 is able to get inside the nucleus and, when acetylated, is retained inside; however, when this acetylation is abolished, TBX5 is still able to be translocated inside the nucleus, but shortly after CRM1 brings it outside thus preventing its nuclear retention. Pull-down assays performed by Dr. Tushar Ghosh also demonstrated that TBX5-K339R presents a stronger affinity for CRM1 than its WT counterpart, thus supporting this mechanism. Interestingly, TBX5 acetylation takes place in a residue localized inside the NLS2; however, the experiments showed here indicate that acetylation plays a role in TBX5 nuclear export, which is mediated by a NES. This kind of mechanism has been already described in other proteins: acetylation of HNF-4³⁹³, POP-1³⁹⁴ or CtBP2³⁹⁵ within their respective NLS leads to their nuclear retention by the inhibition of CRM1-mediated export. How a modification in the NLS2 has an effect in the NES is still unclear and would require of further experiments (such as 3D structure analysis); however, it is possible that acetylation of the NLS promotes a conformational change in the protein that leads to an enhanced nuclear export.

A different hypothesis as to how acetylation affects TBX5 cellular localization involves the kinetics of TBX5 movement in and out of the nucleus. As described elsewhere^{121,382}, and in concordance with the results presented in this thesis, TBX5 is a dynamic protein which is able to move between the nucleus and the cytoplasm. It should be considered the possibility that, in its non-acetylated state, TBX5 moves into the nucleus at a very slow rate but when acetylated, the rate at which TBX5 moves into the nucleus is dramatically increased. This process would tip the dynamic equilibrium in favour of nuclear localization and would increase the functional concentration of TBX5 inside the nucleus, thus leading to a higher transcriptional activity. In order to gain further insight into this hypothesis, it would be necessary to compare the dynamics of TBX5

movement when acetylated and when not acetylated (for example, by performing cell-live imaging). However, the experiments required to address this point are beyond the scope of our study at present.

After studying the role of TBX5 acetylation in an *in vitro* context, I then moved onto the study of the *in vivo* relevance of KAT2A and KAT2B using zebrafish as animal model. First, I checked whether these two acetyltransferases are duplicated in the zebrafish genome, as it happens with many other genes such as *tbx5* (*tbx5a* and *tbx5b*). The entries retrieved by using the Ensembl Genome Browser webtool confirms that both *kat2a* and *kat2b* appear as single genes and are not duplicated in the zebrafish genome. After confirming this, the expression of both *kat2a* and *kat2b* genes was analysed in zebrafish by using two different approaches: RT-PCR and ISH. Both approaches revealed a strong expression of both *kat2a* and *kat2b* in the cranial region of zebrafish and in the heart; however, different results were obtained for other zebrafish organs such the tail, where a strong expression was found in RT-PCR but this was not replicated in the ISH analysis. Even though the experimental approach followed here provide new information regarding the expression of these two genes in zebrafish, this analysis could have been improved in different ways in order to generate more robust and consistent data. For example, regarding the *in situ* hybridization analysis, additional developmental stages could have been analysed to study whether the expression pattern of *kat2a* and *kat2b* changes throughout development. Also, in order to gain further insight into the cardiac structures and cell types in which both *kat* genes are expressed in the heart, cardiac tissue sections could have been prepared and subjected to ISH as well. Last but not least, other techniques such as qPCR or western blot analysis could have been carried out in order to support the data already obtained and to gain information about the protein expression pattern of these two genes.

To examine the *in vivo* role of these two genes, I aimed to perform gene knockdown and knockout experiments in zebrafish at early stages of development and examine the phenotypic consequences in heart and limb development. For knockdown approaches morpholinos were used. As discussed in section 1.5.2.3, morpholinos were rapidly adopted by the zebrafish community soon after their development due to their easiness of use among other factors; however, several publications demonstrated that the phenotype produced by MOs can be due to off-target effects²⁸³, so concerns about their use arose in the community. In response to this, a few set of guidelines were published that, if followed, would help distinguish specific phenotypes from off-target effects. One of them was published a couple of years ago by Stainier

*et al.*²⁸⁰, which postulate some guidelines that should be followed in order to carry out a proper MO experiment:

- Multiple MOs (such as ATG and splice blocking), or MOs and another approach (such as CRISPR-Cas9) should be used to target individual genes and their efficiency should be assessed whenever possible to minimize the amount of MO injected.
- Rescue experiments should be attempted for the approaches listed above, and if rescue is successful, control experiments should be conducted
- An injection control MO should be used to account for developmental delay
- Essential routine procedures include a dose response curve and the examination of statistically meaningful numbers of control and experimental animals, among others.

In this thesis, most of these guidelines were properly followed. First, I initially planned to use both translation and splice blocking MOs targeting both *kat2a* and *kat2b*. However, the generation of titration curves for each of them revealed that both translation blocking MOs induced a high level toxicity even at low doses, so experiments with just the splice blocking MOs were performed instead. In spite of not being able to use multiple MOs targeting the same genes, I used splice blocking MOs as well as two other approaches (CRISPR-Cas9 and chemical inhibition) in order to abolish the activity of my target genes. The efficiency of MOs was assessed by RT-PCR, whereas the efficiency of the CRISPR-Cas9 system was assessed by calculating the average rate of mutagenesis. This aspect could have been improved by studying via qPCR or western blot what is the impact of each one of these approaches on the mRNA and protein levels of zebrafish *kat2a* and *kat2b*, although the latter would have been extremely challenging due to the lack of commercial *kat2a* and *kat2b* zebrafish antibodies. Regarding the second point, a rescue experiment was also carried out by using an acetyl-mimic version of *tbx5a*, but just on crispant embryos. Even though this experiment provided evidence that the phenotype seen in crispant embryos is due to the lack of acetylation, it could have been definitely improved by performing a control rescue experiment (using mutant mRNA, for example) or by carrying it out in morphants as well. As demonstrated in the knockdown experiments, an injection control MO using a mismatch control MO was also performed, thus fulfilling another of the guidelines suggested by Stainier. Only one mismatch control MO was used though, so these experiments could have been strengthened by using additional mismatch MOs. Last but not least, a dose response curve showing the effect on mortality and the induction of severe developmental defects was generated for all four MOs, and a meaningful number of embryos (a total of

200 per MO taking into account all batches and experiments) were injected and analysed in my experiments. In spite of fulfilling many of these guidelines, the MO experiments performed in this thesis could have been improved in different ways. On top of the suggestions mentioned in the lines above, the effect of lower doses of MOs (e.g. 0.5 ng or 1 ng) on the cardiac and fin phenotype should have been assessed in order to check whether there is a relationship between dose and severity of the phenotype. Also, no blinding strategies were used as I in charge of preparing the MO/CRISPR-Cas9 samples for injection, the actual injections and the posterior analysis, so I was aware at any moment of which samples had been injected with each MO/sgRNA.

For the knock-out experiments, I decided to use the recently developed CRISPR/Cas9 genome-editing technology. This system had been successfully used in zebrafish for the first time just two years before the start of this thesis^{210,278}, so setting it up in our lab was one of the main challenges I had to overcome in this project. Even though the CRISPR/Cas9 system has gained much prominence in the last years due to its ease of use, several groups have reported that the individual mutagenesis efficiency when injecting sgRNA+Cas9 mRNA varies between 30-60% depending on the sgRNA used, far from the ideal 100% efficiency that would lead to an easily reproducible and significant phenotype penetrance even in the F0 generation^{278,311,312}. This problem can be solved by the generation of stable mutant strains (F1 generation); however, this is still a great limitation due to the time requirement and cost of animal facilities, and in our lab, we did not have the necessary animal license in place in order to raise zebrafish transgenic lines at that time. A few years ago, a new solution to this problem emerged, as several groups reported that by injecting sgRNA:Cas9 ribonucleoprotein complexes (this is, sgRNA+Cas9 protein) instead of Cas9 mRNA, mutation efficiencies as high as 100% can be achieved, thus reducing phenotype variance and generating complete somatic mutants that reproduce the phenotype found in loss-of-function models previously established^{227,290,396}. Taking this into account, I injected *kat2a*, *kat2b* and *tbx5a* sgRNAs:Cas9 complexes into 1-cell stage zebrafish embryos and following T7 assay and Sanger sequencing, I confirmed the introduction of mutations on these three genes with 80-90% mutagenesis efficacy, thus suggesting the presence of bi-allelic mutations in a substantial number of cells. In addition to this, the *tbx5a* crispants generated in this thesis also recapitulate the cardiac and limb defects found in an already established *tbx5a* mutant transgenic line (heartstrings)¹²⁶, thus suggesting that this approach can successfully recapitulate the phenotype found in established transgenic lines.

As demonstrated in this project, the loss of *kat2a* or *kat2b* in zebrafish during development, both via morpholino-mediated knockdown or CRISPR-Cas9-mediated knockout, lead to a mild cardiac and fin phenotype mainly characterized by the relaxation of cardiac looping, reduction of heart beating rate and the presence of stumpy, shortened pectoral fins. These phenotypic features are similar to those displayed by *tbx5a* morphants and crispants, thus suggesting that this cardiac factor and *kat2* acetylases may work in a common pathway regulating heart and fin development in zebrafish. The main difference between the phenotypes showed by *kat2* and *tbx5a* morphants/crispant resides in their severity: whereas loss of *kat2a* or *kat2b* leads to a mild cardiac and fin phenotype in the embryos analyzed, a more severe cardiac (no looping, severe bradycardia) and fin (lack of pectoral fins) phenotype was found in a large proportion of *tbx5a* morphants and crispants. Interestingly, when embryos were injected with MOs targeting both *kat2a* and *kat2b* to generate double knockdowns, the proportion of embryos showing a severe cardiac and fin phenotype was increased and much closer to that found in *tbx5a* morphants. These results suggest the presence of redundancy between *kat* proteins, by which the lack of one of them could still be partially compensated by the presence of the second one. This compensatory effect between *kat2a* and *kat2b* have already been described elsewhere^{397,398}, and it has been demonstrated that mice lacking both *Kat2a* and *Kat2b* die at an earlier developmental stage than their single KO counterparts³⁵⁵. Apart from these defects found in the heart and fins, mild defects in other organs or slight developmental delay were also seen (source: visual observation) in some of the embryos analysed in other organs such as the head; unfortunately, no representative pictures of the whole embryos were taken as I focused on obtaining high-magnification and high-resolution images of the heart, so these defects initially seen in other parts of the zebrafish body could not be properly assessed. These whole-body pictures would have allowed me to perform a more in-depth assessment of other organs, thus characterizing more in detail the phenotype of these embryos.

The role of acetylation during heart and fin development in zebrafish was further substantiated by performing a chemical inhibition of *kat2b*. This was achieved by treating zebrafish embryos with Garcinol, an inhibitor of the acetyltransferase activity which has been shown to inhibit KAT2B with high specificity^{306,384}. Embryos treated with this compound showed a mild cardiac and fin phenotype extremely similar to that displayed by *kat2b* morphants and crispants, with relaxed cardiac looping and shortened fins among other defects. Most of the treated embryos that were analysed presented a mild phenotype

instead of a severe one, thus suggesting that this compound only inhibits *kat2b* activity and not *kat2a* activity as well.

To sum up, all three approaches to abolish the activity of *kat2a* or *kat2b* led to similar cardiac and fin phenotypes, thus suggesting that the defects found in these embryos are indeed due to the loss of activity of either of these acetylases and not due to off-target effects and/or injection artefacts. However, in spite of the weight of evidence provided by these three approaches, it would have been far more preferable to breed my F0 crispants to successive generations in order to generate “clean”, stable *kat2a* and *kat2b* transgenic lines which would have allowed me to fully confirm that the phenotype seen in this crispants is indeed due to the lack of acetylase activity. In addition to this, the breeding of these F0 crispants to further generations would have eliminated other causes of concern such as the presence of different mutations in the same sample (mosaicism), and more importantly, it would have led to possible off-target mutations being “diluted” over generations. In addition to this, the possibility that the cardiac and limb phenotype observed in *kat2a* and *kat2b* morphants, crispants and embryos treated with Garcinol is due to developmental delay cannot be completely ruled out, especially in those embryos categorized as having a mild cardiac/fin phenotype. As stated previously, at the time of generation of these models our laboratory did not own the necessary animal license in place in order to raise zebrafish transgenic lines, so we could keep zebrafish embryos for a maximum of 5 days post-fertilization. The best method to assess whether these phenotypes are indeed due to the lack of *kat2* acetylases or due to developmental delay would have been to allow these embryos to reach the adult stage and then breed them in order to create an stable transgenic line, thus overcoming potential developmental delays as well as other intrinsic problem of using F0 crispants mentioned before such as the presence of mosaicism or potential off-target effects.

qPCR and *in situ* hybridization analysis of the expression of several *tbx5a* target genes in *kat2a* and *kat2b* crispants also provide further evidence of the *in vivo* relationship between these factors. For the qPCR experiment, the relative mRNA levels of five different *tbx5a* target genes were analysed at 32 hpf (a stage at which the heart tube is undergoing heavy remodelling): *nppa*, *fgf10*, *bmp4*, *hey2* and *tbx2b*. All these five genes have been shown to be regulated by *tbx5a* in zebrafish in the past^{318,319,399}. In my experiments, all these genes were shown to be significantly downregulated in both *kat2a* and *kat2b* crispant embryos. The expression level of two totally unrelated genes to *tbx5a* (*akt1* and *igf3*) was also checked, and both of them did not show any significant

differences in expression in comparison to WT embryos, thus providing confidence that the changes observed with regards to *tbx5a* target genes are indeed due to the downregulation of the transcriptional activity of this factor. The downregulation of *tbx5a* target genes in *kat2a* and *kat2b* knockouts was further confirmed by carrying out *in situ* hybridization using *bmp4*, *nppa* and *fgf10* probes. Even though these results strongly suggest that the lack of *kat2a* or *kat2b* is responsible for the downregulation of *tbx5a* activity, these results could have been further substantiated by analysing the expression of these genes at other developmental stages or by increasing the number of targets analysed. Last but not least, the *in vivo* link between *kat2* acetylases and *tbx5a* was also studied by performing a rescue experiment of the *kat2a* KO phenotype using a *tbx5a* K339Q acetyl-mimic variant. The lysine-to-glutamine substitution was used as an acetyl-mimic version of *tbx5a* as these two residues share a similar charge and chemical structure, and it has been shown that glutamine can function as acetyl-lysine in an *in vivo* context^{400,401}. The rescue experiment showed that, when *tbx5a* K339Q mRNA is injected alongside Cas9 protein + *kat2a* sgRNA, a partial rescue of both the cardiac and fin phenotypes of these embryos is achieved, thus suggesting that these phenotypes are produced by the lack of *tbx5a* acetylation. In order to further substantiate these results, a “control rescue” group could have been included by injecting a mismatch, mutated version of the *tbx5a* K339Q mRNA instead.

The results presented in this thesis indicate that two of the main developmental processes affected in *kat2a* and *kat2b* morphants and crispants are cardiac looping and fin development. As shown in this thesis as well as in other publications over the last couple of decades^{126,318,402}, the loss of *tbx5a* activity in zebrafish also leads to a cardiac phenotype mainly characterized by defects in cardiac looping and the presence of pericardial oedema, as well as the development of stumpy fins or complete lack of them. In mouse, a heterozygous transgenic line lacking one copy of *Tbx5* mirror the Holt-Oram syndrome phenotype, as they present ASD, occasional VSD and AV block, as well as forelimb abnormalities⁶⁷. More interestingly, a homozygous transgenic line lacking both copies of *Tbx5* do not survive past E10.5, and development is arrested at E9.5⁶⁷. In these embryos, cardiac looping did not occur, and the heart is a deformed, large tubular structure showing signs of hypoplasia. This cardiac phenotype is remarkably similar to that of the *kat2a* and *kat2b* crispants generated in this thesis. Generation of F0 crispants typically lead to the complete loss of the targeted gene, and mutagenesis analysis of these crispants suggested the presence of biallelic mutations in them, although further and a more in-depth analysis (for example, by increasing the number of samples and alleles analysed) would be required to fully confirm this. In spite

of this, these data seem to indicate that the lack of both copies of *kat2a* or *kat2b* in zebrafish lead to remarkably similar cardiac defects than those produced by the lack of both copies of *Tbx5* in mouse, thus strengthening the relevance of these experiments. In humans, mutations in the *TBX5* gene lead to Holt-Oram syndrome⁷⁷, which is characterized by the presence of congenital heart defects (most common ones are ASD and VSD) and radial defects. Regarding cardiac defects, one of the main limitations of using zebrafish as an animal model is that cardiac septation do not take place, so it is rather complicated to establish the potential relationship between the zebrafish crispants generated in this thesis and patients of this condition. In addition to this, Holt-Oram syndrome is typically produced by dominant heterozygous mutations in the *TBX5* gene and no homozygous patients have been characterized so far, possibly due to the fact that loss of both copies of *TBX5* during human development is incompatible with embryo survival³⁹⁰. Even though it is difficult to assess the relationship between the heart phenotype observed in zebrafish and that of human patients, there are indeed similarities regarding the fin/limb phenotype, as both in zebrafish and in HOS patients lack of *TBX5* leads to defects in pectoral fin or upper limbs development respectively, although further studies would be required to study this relationship more in detail (for instance, by performing Alcian or hematoxylin/eosin staining of zebrafish fin sections).

In the near future, our goal is to study further the significance of *TBX5* acetylation *in vivo*. Most of our current data is based on zebrafish experiments. Although it is an excellent animal model for quick reverse-genetics studies it still presents some limitations in comparison to other models (e.g. mouse). In order to obtain deeper knowledge on this matter, a mouse line carrying a *Tbx5* Lys339Arg point mutation has been established at the University of Nottingham. By carrying a non-conservative substitution of the *Tbx5* acetylation residue (Lys339), this mouse model serves as an effective *Tbx5* acetyl-deficient model. This line will allow us to perform a more accurate characterization of the cardiac and limb defects produced by the lack of acetylation in the *Tbx5* protein. Also, and on top of classic biochemical and cell biology experiments, we are planning to perform transcriptomic and proteomic analysis of embryonic heart samples from this mouse model in order to generate a global picture of all *Tbx5* target genes regulated via acetylation. At the time of writing of this thesis, the line is being expanded in preparation for future experiments.

To sum up, both the *in vitro* and *in vivo* data obtained in this project strongly support the fact that KAT2A and KAT2B acetyltransferases potentiate *TBX5* transcriptional activity via acetylation, and that the lack of any of them during

development lead to cardiac and fin defects due to the downregulation of *tbx5a* activity. A new regulatory mechanism of TBX5 activity is also proposed, in which acetylation at lysine 339 is required for TBX5 retention in the nucleus, thus leading to an increased pool of this factor and, subsequently, to an increased transcriptional activity (Fig. 3.26). As of today, and even though a large number of TBX5 mutations have been found in Holt-Oram syndrome patients^{291,403}, no mutations at lysine 339 have been described. However, based on the relevance that this residue seem to have in TBX5 activity, we do not discard the possibility that mutations in this lysine might contribute in some way to the onset of cardiac defects during development.

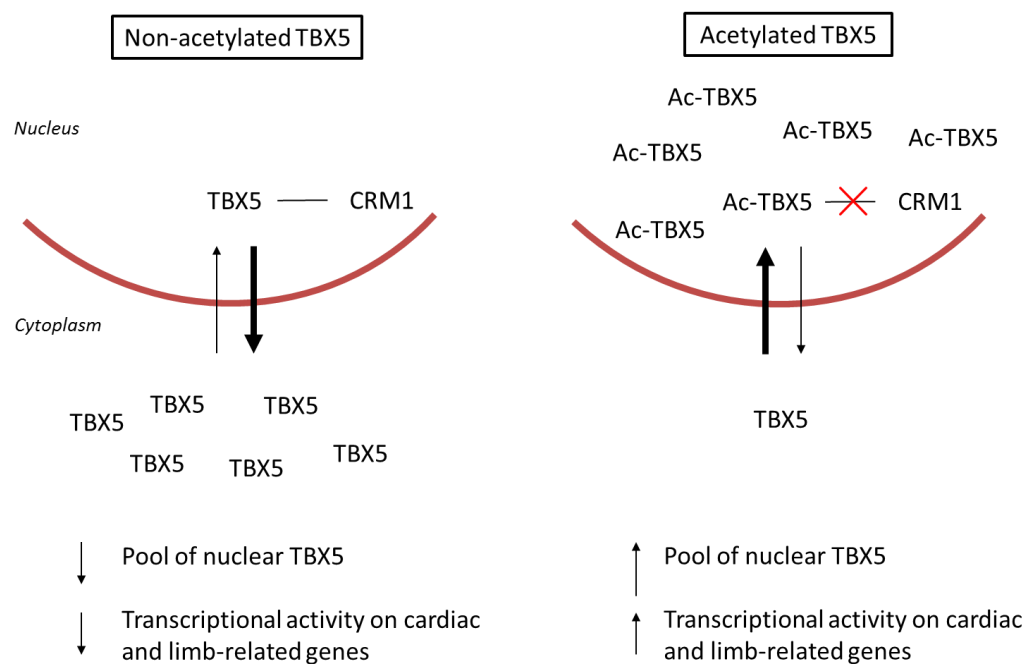


Figure 3.26. TBX5 activity regulation model. When TBX5 is acetylated by KAT2A or KAT2B in the nucleus, its affinity to the exportin CRM1 decreases, thus leading to its nuclear retention. This increases the pool of nuclear TBX5, which results in an increased transcriptional activity on TBX5 target genes, including genes involved in cardiac and limb development such as *NPPA*, *FGF10* or *BMP4*.

4. ANALYSIS OF CHD4 MUTATIONS FOUND IN SYNDROMIC-CHD PATIENTS

4.1. Introduction

4.1.1. The NuRD complex and its role in cardiac development

One of the histone-modifying complexes that has been subject to recent interest in the field of cardiac development is the NuRD complex (Nucleosome Remodelling Deacetylase complex). The NuRD complex is a group of associated proteins with both ATP-dependent chromatin remodelling and histone deacetylase activities, being comprised of several subunits: the histone deacetylase core proteins HDAC1 and HDAC2, the chromodomain/helicase/DNA-binding proteins CHD3 and CHD4, the retinoblastoma-binding proteins (or histone-binding proteins) RBBP4 and RBBP7, the methyl-CpG-binding domain proteins MBD2 and MBD3, and the metastasis-associated protein MTA1, MTA2 and MTA3^{404,405} (Fig. 4.1). The NuRD complex is required at various stages of hematopoietic differentiation, and in the context of cancer, it has been usually associated with processes such as metastasis and epithelial-to-mesenchymal transition (EMT)^{404,405}. Although the NuRD complex is comprised of several subunits, in this chapter I will mainly focus on CHD4. Information about HDAC1 and HDAC2, two of the core components of the complex, can be found in section 3.1.2.

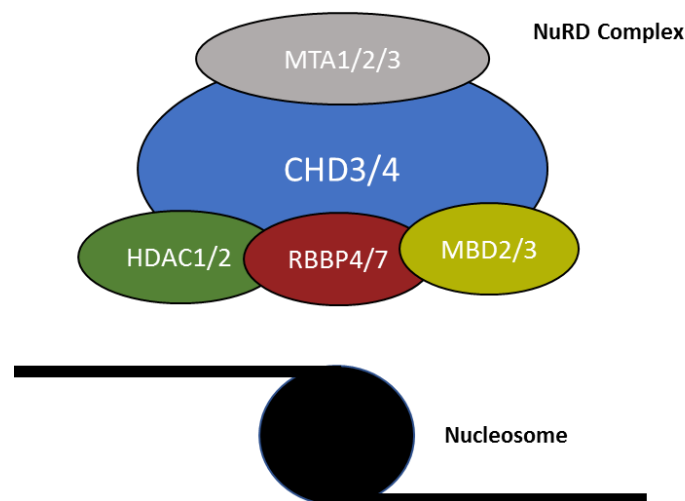


Figure 4.1. Subunits of the NuRD complex. The nucleosome/remodelling complex is comprised of several proteins, including CHD3/4 (blue), MTA1/2/3 (grey), HDAC1/2 (green), RBBP4/7 (red) and MBD2/3. Subunits of the complex belonging to the same family (such as CHD3/4 or HDAC1/2) are interchangeable, and the composition of the complex varies according to cellular conditions and signals. All subunits work together in order to perform nucleosome remodelling roles (a nucleosome is shown in black).

4.1.1.1. Chromodomain/helicase/DNA-binding protein 4 (CHD4)

The chromodomain/helicase/DNA-binding protein 4 (CHD4) is a helicase/ATPase that uses the energy from ATP to perform chromatin remodelling functions via histone sliding⁴⁰⁶. The CHD4 protein belongs to the CHD subfamily II, alongside CHD3 and CHD5⁴⁰⁶. This family is characterized by the presence of specific domains in their structure: two N-terminal plant homeodomain (PHD) zinc-finger-like domains, two chromodomains (CHD) and a centrally located ATPase/helicase domain⁴⁰⁷ (Fig. 4.2). The PHD and CHD domains are thought to mediate the binding to the nucleosomes⁴⁰⁸ and regulate CHD4 remodelling activity⁴⁰⁹, whereas the ATPase/helicase domain acts as the catalytic domain of the protein, providing the energy necessary to exert its function. No functional domains have been identified in the C-terminal region of the protein; however, it has been reported that this region is susceptible to be phosphorylated at the residue Ser1349 by ATM kinase in response to DNA damage⁴¹⁰.

CHD4 protein – 1912 aa

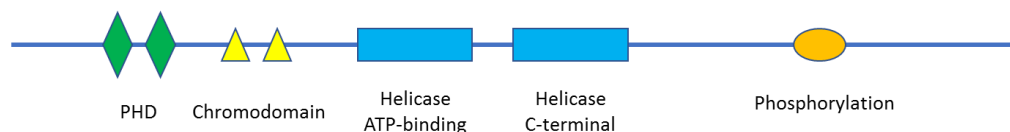


Figure 4.2. Secondary structure of the CHD4 protein. The CHD4 protein is comprised of several domains: two N-terminal plant homeodomain (PHD) zinc-finger-like domains (green), two chromodomains (CHD, yellow) and a centrally located ATPase/helicase domain (blue). Phosphorylation of CHD4 takes place in the C-terminal domain (orange) of the protein following DNA damage.

According to the Mouse Gene Expression database, *Chd4* is expressed in a wide range of different organs and tissues in the mouse embryo, including brain, heart, genitourinary system and the central nervous system among others⁴¹¹. Several *Chd4* knockout mouse models have also been generated. The International Mouse Phenotyping Consortium (IMPC) provides information on a *Chd4*-null mouse model which lacks the critical exons 11 and 12 of the *Chd4* gene (chromodomains)⁴¹². Homozygous null embryos for this model present embryonic lethality prior to organogenesis, whereas heterozygous embryos are viable but they show signs of hypoactivity⁴¹². No further information regarding this model is available as of today, as according to the search engine PubMed (<https://pubmed.ncbi.nlm.nih.gov/>), no studies using this model have been published yet. A different *Chd4*-null mouse model revealed that embryos lacking both alleles of *Chd4* are unable to successfully complete the first lineage

decision and form functional trophectoderm at the blastocyst stage, thus dying really early in development⁴¹³. Chd4 has also been conditionally ablated in specific murine organs or tissues including heart⁴¹⁴, brain⁴¹⁵, kidney⁴¹⁶ and endothelium⁴¹⁷ among others. In the brain, conditional knockout of Chd4 in the granule neurons of the cerebellar cortex impairs the establishment of granule neuron parallel fiber/Purkinje cell synapses in this tissue⁴¹⁵, thus demonstrating its role in synaptic connectivity. In the kidneys, deletion of Chd4 in cap mesenchyme cells leads to renal hypoplasia and a marked reduction in nephrons, as well as misregulation of renal progenitor cell markers⁴¹⁶. Chd4 also plays a key role in vasculogenesis, as mice lacking this protein in endothelial cells die at midgestation from vascular rupture⁴¹⁷. Cardiac abnormalities found in conditional knockouts of Chd4 in the heart are discussed in more detail in the next section.

4.1.1.2. The NuRD complex, CHD4 and cardiac development

Although the NuRD complex and its subunits have been typically linked to cancer-related processes, in the last couple of years several studies have suggested a key role for this complex in cardiac development and CHD. Waldron *et al.* showed recently that Tbx5, one of the master regulators of cardiogenesis, interacts biochemically and genetically with Chd4 in order to repress incompatible gene programs in the developing heart⁴¹⁸. Furthermore, CHD-causative mutations in TBX5 disrupt the TBX5-NuRD interaction, leading to the activation of genes that would otherwise be repressed and correlating with septal defects⁴¹⁸. A different study performed by Gómez del Arco *et al.* showed that loss of Chd4 in developing cardiomyocytes leads to a thinner ventricular compact myocardium, less intricate trabeculae, defects in ventricular septum formation and finally, embryonic lethality at E13.5⁴¹⁴. Analysis of cardiomyocyte proliferation markers also revealed the presence of reduced cardiomyocyte proliferation in the myocardium of the ventricular compact wall of these embryos, suggesting that this was the main cause of the myocardial-related defects found in this model⁴¹⁴. Deletion of Chd4 in already differentiated cardiomyocytes (adult mice) causes severe cardiomyopathy and sudden death⁴¹⁴. These adult mutant mice showed atrial dilatation and accumulation of fibrotic collagen in the left ventricle and in the ventricular septum, which suggested progression of these hearts to heart failure⁴¹⁴. Right after the publication of this study, Wilczewski *et al.* also reported the generation of a second *Chd4*-null mouse model in cardiomyocytes⁴¹⁹. Heterozygous mouse for this model are completely viable and phenotypically normal; however, lack of both copies of Chd4 in cardiomyocytes led to embryonic lethality at E12.5⁴¹⁹. At E11.5, these embryos displayed pericardial oedema and haemorrhage, as well as stunted growth. On top of this, the hearts

of these mutant embryos showed an enlarged atria, a reduced right ventricle and an enlarged left ventricle, and histological analysis revealed a reduced complexity of the trabecular myocardium in the ventricles and a decrease in the thickness of the compact layer⁴¹⁹. This study also revealed that Chd4-mediated repression of skeletal and smooth muscle myofibril isoforms is required for normal cardiac sarcomere formation and function⁴¹⁹. Regarding other NuRD subunits and cardiac development, it has also been shown that the NuRD complex interacts with Fog2, a multi-zinc finger protein that binds the transcriptional activator Gata4 and modulates Gata4-mediated regulation of target genes during heart development⁴²⁰. Mice carrying a mutation in Fog2 that disrupts NuRD binding exhibit perinatal lethality and have several cardiac malformations including ventricular and atrial septal defects and a thin ventricular myocardium, as well as presenting reduced cardiomyocyte proliferation⁴²⁰.

4.1.2. CHD4 mutations found in syndromic CHD4 patients

In the last decade, several studies have also reported the presence of Chd4 mutations in human patients. Le Gallo *et al.* identified a high frequency of somatic mutations in the *CHD4* gene in primary serous endometrial tumors⁴²¹, and *in vitro* studies using human cancer models have also showed a role for Chd4 in cancer growth and proliferation^{422,423}. Our lab also collaborated recently in a whole-exome sequencing on syndromic- and non-syndromic-CHD patients⁴²⁴. This study revealed the presence of five different *de novo* mutations in the *CHD4* gene in syndromic-CHD patients, which are the mutations I based my work on during this chapter. In the same line of work, Weiss *et al.* recently reported five individual with *de novo* mutations in the *CHD4* gene⁴²⁵ (Fig. 4.3A and B). These individuals have overlapping phenotypes including developmental delay, intellectual disability, macrocephaly and congenital heart defects such as ASD, VSD or bicuspid aortic valve, among others. A third study performed by Homsy *et al.* also revealed the presence of *de novo* mutations in the *CHD4* gene in patients with syndromic-CHD⁴²⁶.

As mentioned in the previous section, our lab recently collaborated with the Sanger institute in a whole-exome sequencing study in which three new S-CHD causative genes were discovered: *CHD4*, *CDK13* and *PRKD1*⁴²⁴. In the case of *CHD4*, five *de novo* mutations were identified in this gene, which serve as the basis of the work presented in this chapter. Four of these five mutations are missense variants (p.Cys467Tyr, p.Ser851Tyr, p.Arg1068His and p.Val1608Ile), whereas the fifth is an in-frame deletion of three amino acids (p.Leu1009_Val1011del) (Fig. 4.3B). Of these five mutations, three of them (p.Ser851Tyr, p.Leu1009_Val1011del and p.Arg1068His) are localized in the

helicase/ATPase domain of the CHD4 protein. One of the missense variants, p.Cys467Tyr, is found in one of the N-terminal plant homeodomains, whereas the variant p.Val1608Ile is situated in the C-terminal domain of the protein, whose function is unknown (Fig. 4.3B). Three of these individuals had tetralogy of Fallot or Fallot-like features, whereas the other two presented septal defects (ASD and/or VSD) and aortic coarctation. All of them manifested neurodevelopment delay, two had Chiari malformations and three of the four males presented cryptorchidism (Fig. 4.3A). These phenotypic features overlap with those of the CHARGE syndrome, a genetic condition produced by heterozygous loss-of-function mutations in *CHD7*, a paralog gene of *CHD4*^{427,428}. Mutations in *GATAD2B*, another subunit of the NuRD complex, cause a recognizable intellectual-disability syndrome⁴²⁹; however, no CHD-features have been reported in these individuals. In addition to this, several subunits of other chromatin-remodelling complexes have been reported to be mutated in patients of different developmental syndromes, including CHD in some of them^{430,431}.

Recently, a study carried out by Weiss *et al.*⁴³² performed a comprehensive investigation of the clinical spectrum of 32 individuals carrying mutations in the *CHD4* gene, including those included in the Sifrim⁴²⁴ and Weiss (2016)⁴²⁵ studies among others. This study demonstrate that some of the main clinical findings found in these patients are developmental delay (97%), brain anomalies (96%), intellectual disability (86%) and heart defects (65%). Some of the phenotypes found in murine Chd4-knockout models strongly fit with these features found in human patients. For example, as it has been mentioned before, mouse embryos lacking Chd4 in the neurons of the cerebellar cortex present defective synaptic connectivity and brain anomalies⁴¹⁵, evidence that could explain the intellectual disability and brain anomalies found in these patients. In terms of heart anomalies, patients exhibit an extensive range of different cardiac defects, such as ASD and VSD (the most common ones), pulmonary stenosis, patent ductus arteriosus or tetralogy of Fallot⁴³². Some of these defects are related to structures in which the second heart field have a major contribution during development (for example, atria and outflow tract), so CHD4 might play a key role in this cell lineage. With regards to the relationship these cardiac phenotypes and animal models, this point is still a question of debate. As mentioned before, conditional knockout models of Chd4 in the murine heart lead to ventricular septal defects, a thinner myocardium and less intricate trabeculae due to reduced cardiomyocytes proliferation^{414,419,420}. These defects fit with those present in patients who exhibit VSD and tetralogy of Fallot (as VSD is one of the main features of this condition); however, it is still

unclear how this murine phenotype correlates with other defects such as ASD, pulmonary stenosis or the outflow tract-related defects found in ToF patients.

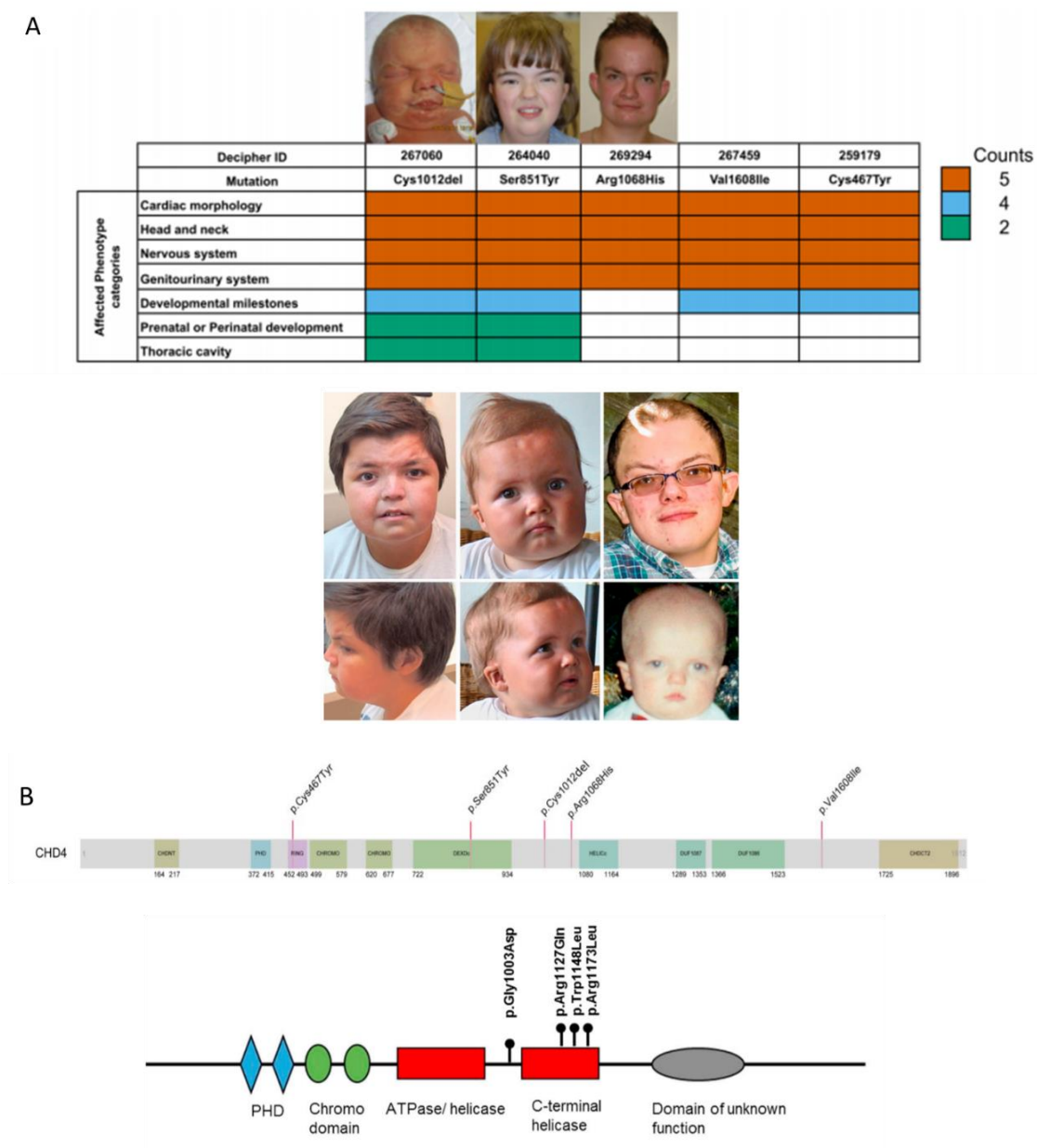


Figure 4.3. Individuals with *de novo* CHD4 mutations. Information from two different publications are included. (A) Clinical manifestations observed across patients carrying *CHD4* mutations. In the upper table, individuals are represented in each column, whereas colour of the cells indicate the number of individuals sharing that specific phenotype. Detailed information about the patients shown in the second set of pictures can be found in the next page table. (B) Illustration of the secondary structure of the CHD4 protein and the distribution of CHD4 *de novo* mutations described in two different publications. Adapted from Sifrim et al., 2016⁴²⁴ and Weiss et al., 2016⁴²⁵

Table 1. Clinical Findings in Five Subjects with De Novo Missense Variants in CHD4

	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5
CHD4 variant	c.3380G>A (p.Arg1127Gln)	c.3518G>T (p.Arg1173Leu)	c.3380G>A (p.Arg1127Gln)	c.3443G>T (p.Trp1148Leu)	c.3008G>A (p.Gly1003Asp)
Gender, age at last exam	male, 10 years	female, 16 years	male, 10 years	female, 5 years	male, 18 years
Birth weight, OFC	4 kg, 38 cm	2.8 kg, ND	3.7 kg, 37cm	2.99 kg, 35 cm	3.06 kg, ND
Height, OFC at last exam ^a	143 cm (75 th), 56 cm (>98 th)	161 cm (40 th), 62 cm (>98 th)	140 cm ^b (50 th), 56 cm (>98 th)	89.5 cm (<3 rd ; Z score -5), 49 cm (20 th)	167.5 cm (10 th), 52.5 cm at 4 years (90 th)
Developmental delay	+	+	+	+ (severe)	+
Intellectual disability	+	+	+ (mild)	+	+ (mild)
Hearing loss ^c	+	+	+	-	+
Undescended testis, micropenis	+, +	NA	+, +	NA	-, +
Macrocephaly ^d	+	+	+	relative to length	+ ^o
Widely spaced eyes ^d	+	+	+	+	+
Dysmorphic ears ^e	+	+	+	+	+
Palatal anomalies	+ ^h	-	+ ⁱ	+ ⁱ	+ ⁱ
Hypogonadotropic hypogonadism	+	-	+	NT	+
Skeletal survey	advanced bone age by 2-3 years	tarsal coalition, cervical vertebrae fusion	falx calcification	scoliosis, platybasia, fusion of C2-C3, bilateral coxa valga, fusion of the cuboid and the 3 rd cuneiforms bilaterally, brachymesophalangia	diffusely osteopenic bones
Brain MRI	enlarged lateral ventricles, congenital stroke with moyamoya disease	enlarged lateral ventricles, chiant 1 malformation	enlarged lateral ventricles	enlarged ventricles (mild), basilar, invagination and narrow foramen magnum	enlarged lateral and third ventricles
Heart	-	-	-	congenital heart defect (PDA s/p ligation, PFO, ASD, and VSD)	ASD, PDA s/p repair, VSD, bicuspid aortic valve, mild dilatation of aortic root

4.1.3. Hypothesis and aims of this project

Whole-exome sequencing studies of CHD patients is one of the most powerful tools available for the discovery of new CHD-causative genes. As indicated during the introduction of this chapter, our lab recently collaborated with the Sanger Institute in a whole-exome sequencing study on patients of non-syndromic- and syndromic-CHD, which led to the discovery of novel CHD-causative mutations in the *CHD4*, *CDK13* and *PRKD1* genes. These three genes serve as the starting point for the next steps of this project, which involve the *in vitro* characterization of these mutations and *in vivo* studies on mouse models carrying null alleles or point mutations for each one of these genes. The characterization of these mutations is a joint effort between many people in Prof. David Brook's lab. I was in charge of the *CHD4* gene, the one this chapter will be focused on.

The presence of *CHD4* mutations in syndromic-CHD patients indicates the importance of this helicase in cardiac development, a fact that has been recently supported by several studies focusing on the role of the NuRD complex and *CHD4* during cardiac development. However, we are still far from understanding the molecular mechanisms by which *CHD4* exert its function in the developing heart. In order to shed some light on this question, the main goal of this part of my thesis will be to perform an *in vitro* characterization of the five *de novo* *CHD4* mutations found in syndromic-CHD patients. To do this, I aim to answer whether these mutations have an impact on different *CHD4*-related features such as:

1. Its ability to interact with the NuRD complex
2. Its cellular localization
3. Its ATPase activity
4. Its ability to bind DNA or nucleosomes
5. Its ability to perform nucleosome remodelling

4.2. Results

4.2.1. Analysis of *CHD4* mutations found in syndromic CHD patients

The five *CHD4* mutations used in this project are described in Fig. 4.4A, including type of mutation, whether the mutation is damaging or not according to PolyPhen-2 (<http://genetics.bwh.harvard.edu/pph2/>), their domain location inside the *CHD4* protein and the cardiac defects found in patients harbouring each one of these *CHD4* variants. Fig. 4.4B shows a representation of the secondary structure of the *CHD4* protein including the location of the mutant variants. Analysis of amino acid conservation across species was also

performed by using ClustalW2 (<https://www.ebi.ac.uk/Tools/msa/clustalw2/>). As shown in Fig. 4.4C, amino acids affected in mutations M2, M3, M4 and M5 are well-conserved in CHD4 orthologs across several species (mouse, rat, chimpanzee, dog, chicken, xenopus and zebrafish); however, the valine residue affected in the mutation M1 is not well-conserved, as a leucine or proline can be found instead in other species such as mouse, rat or dog. In spite of this, the three aminoacids found in this position in different species (valine, proline and leucine) are all non-polar and hydrophobic, thus confirming that aminoacids in this position share similar structure and chemical properties.

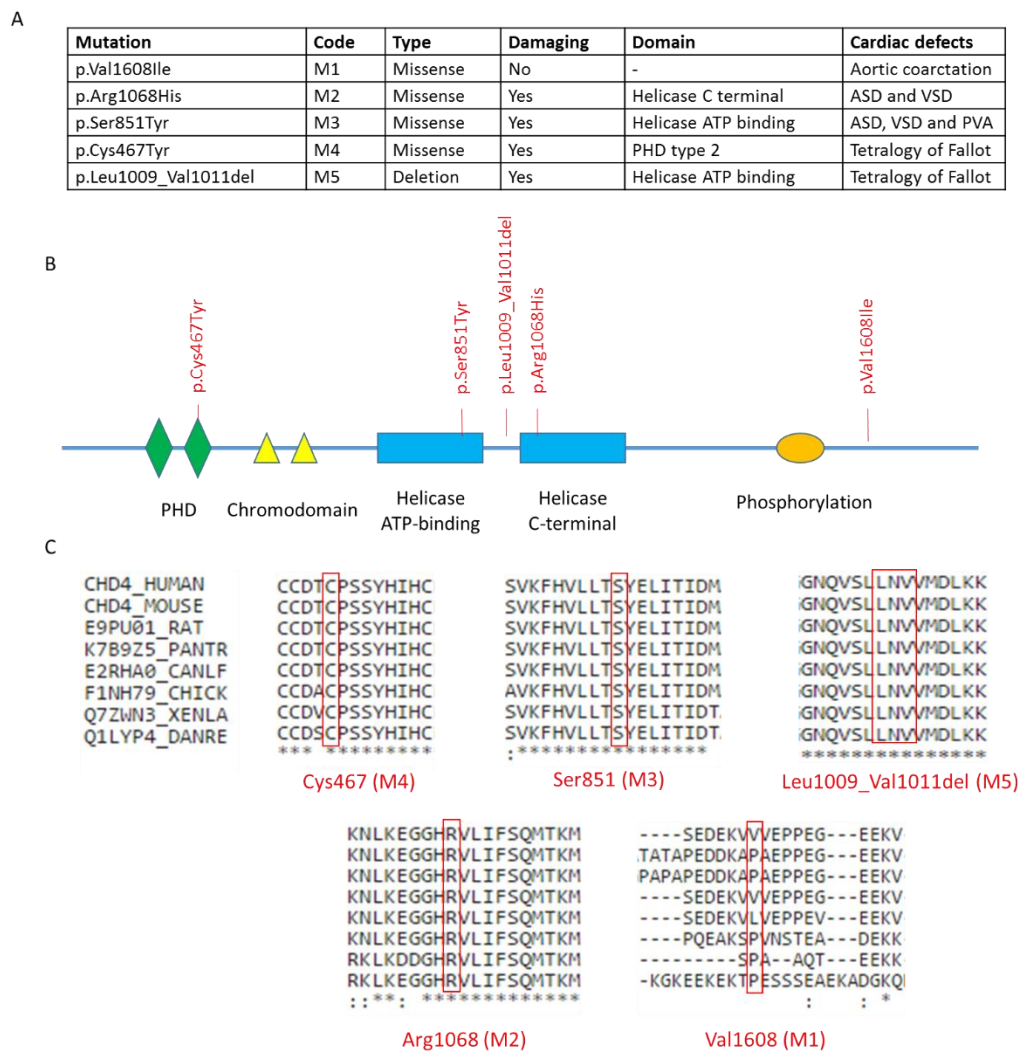


Figure 4.4. CHD4 mutations found in syndromic-CHD patients. (A) Summary of the mutations, including amino acid affected, name code, type of mutation, whether the mutation is damaging or not (based on Polyphen-2 prediction), domain in which is localized and cardiac defects found in patients harbouring each one of them. (B) Secondary structure of CHD4 and localization of each of the mutant variants within the protein. (C) Protein alignment of CHD4 orthologs across several species (human, mouse, rat, chimpanzee, dog, chicken, xenopus and zebrafish; respectively).

4.2.2. Generation and expression of CHD4 mutant variants in mammalian cells

In order to perform *in vitro* studies, I first generated a human expression vector harbouring individual CHD4 mutant variants as indicated in 2.3.1. Following site-directed mutagenesis, the generation of each CHD4 mutant variant was confirmed by Sanger sequencing, as shown in Fig. 4.5.

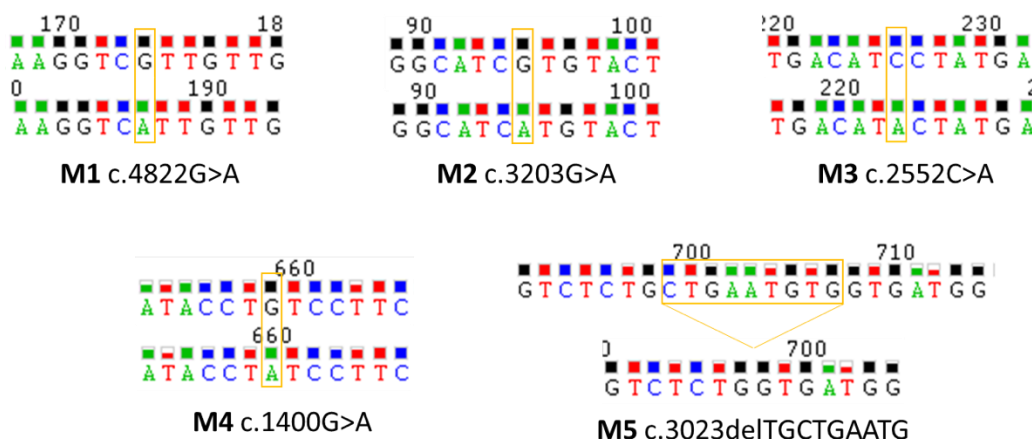


Figure 4.5. Verification of the generation of all five CHD4 mutant variants by Sanger sequencing.

In order to verify the correct protein expression of each one of the CHD4 mutant variants, I transfected them into COS7 cells and hCHD4-FLAG expression was checked via western blot 48 hours post-transfection (n=3 independent experiments). As shown in Fig. 4.6, western blot analysis of hCHD4-FLAG led to the detection of two bands: one corresponding to a molecular weight >220 KDa (~250 KDa), and another one just below the 220 KDa mark (~210 KDa). In order to ascertain which one correspond to the full-length version of CHD4, I looked in the literature for publications which report the detection of human CHD4 via western blot. As shown by Weiss⁴²⁵ and Waldron⁴¹⁸, human CHD4 is typically detected as a band of ~250 KDa, thus indicating that in my western blots, the band corresponding to the full-length version of CHD4 was the upper one. After confirming this, I wanted to gain further insight into the origin of the lower ~210 KDa band. According to the literature⁴³³ and to the western blot troubleshooting tips provided by several life science companies such as Abcam⁴³⁴ or Bio-Rad⁴³⁵, the presence of unexpected bands of a lower molecular weight can be due to different reasons, such as:

- Post-translational modification of the original protein

- Presence of an in-frame second start translation site downstream of the original one which leads to the production of a shorter version of the protein
- Proteolysis or cleavage of the original protein
- Presence of splice variants
- Another protein bearing a similar epitope is detected by the antibody

Regarding post-translational modifications of CHD4, to date, the only post-translational modification reported is phosphorylation by ATM kinase at Ser1349⁴¹⁰, which has been demonstrated to take place following DNA damage. However, in this same publication, the authors also showed that phosphorylation of CHD4 by ATM can also take place in non-DNA damaging conditions, albeit at a lower level than when DNA damage is induced⁴¹⁰. It is unlikely that CHD4 phosphorylation was taking place in my experiments as no DNA damaging reagents were used; however, based on the possibility that CHD4 might still be phosphorylated by ATM in basal conditions, and in order to completely rule out this potential explanation for the presence of the lower band, I tried to detect the phosphorylation status of CHD4 following transfection of hCHD4 in COS7 cells (n=1 independent experiment). Following immunoprecipitation of the CHD4-FLAG variants using FLAG beads, their potential phosphorylated status was checked using a generic p-Serine antibody. As shown in Fig 4.6, I was not able to detect phosphorylated CHD4 in any of the variants, thus suggesting that this post-translational modification is not the cause of the presence of the lower band.

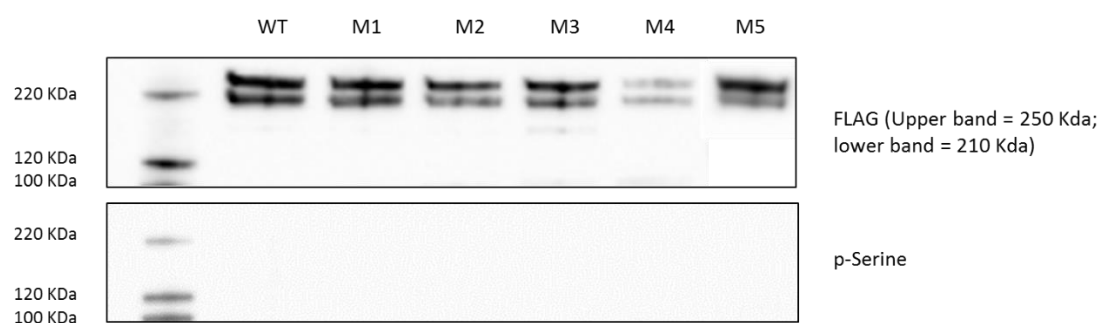


Figure 4.6. Expression of CHD4-FLAG variants in COS7 cells. Following transfection into COS7 cells, CHD4 variants were detected using an anti-FLAG antibody (n=3 independent experiments). Phosphorylated CHD4 was not detected following incubation with a p-Serine antibody (n=1 independent experiment).

Another possible explanation for the presence of the lower band is the generation of a shorter CHD4 isoform from a second translation start site downstream of the original one. In eukaryotes, the Kozak consensus sequence is a nucleic acid motif that functions as the protein translation initiation site in

most mRNA transcripts⁴³⁶. The original Kozak consensus sequence was defined as gcc(A/G)cc**ATG**G, where the start translation codon (ATG) is in bold and underlined, upper-case letters are highly conserved bases and lower-case letters indicate the most common base at a position where the base can nevertheless vary. Taking this into account, translation start sequences are characterized by the presence of an ATG codon, a guanine (G) in position +1 relative to the ATG codon, and an adenine (A) or guanine in position -3. With this in mind, I searched in the human CHD4 cDNA sequence for motifs which fulfilled these conditions and in which the ATG codon was situated in the open reading frame of the cDNA. By analysing the human CHD4 cDNA sequence retrieved from the Ensembl Genome Browser (transcript ID ENST00000544040.6), three potential translation start sites were identified downstream of the original one (Fig. 4.7A). The sequence of these additional start sites fulfil the required conditions: the ATG codon is in the open reading frame, there is a guanine in the position +1, and an adenosine or a guanine can be found in the position -3 (Fig. 4.7B). Additionally, translation start from these sequences would lead to the production of truncated CHD4 versions of 1710, 1673 and 1512 aa in length, which in a western blot would appear as 224, 219 and 198 KDa bands respectively (this is an estimation calculated by assuming that the original protein (1912 aa) appears as a 250 KDa band). The lower band displayed in my western blot corresponds to a product of ~210 KDa, so in view of the results obtained from this analysis, it is possible that this band corresponds to a shorter CHD4 variants produced from any of these additional start sites.

Overall, these results indicate that all five CHD4 mutant variants were successfully produced by using site-directed mutagenesis, as they all lead to the production of CHD4 protein in COS7 cells. The preliminary analysis performed to identify the origin of the second, lower band suggest that it might have its origin on a second translation start site downstream of the original one; however, further experiments are required to confirm this, and other potential explanations cannot be completely ruled out (a more detailed discussion on this topic can be found in section 4.3.)

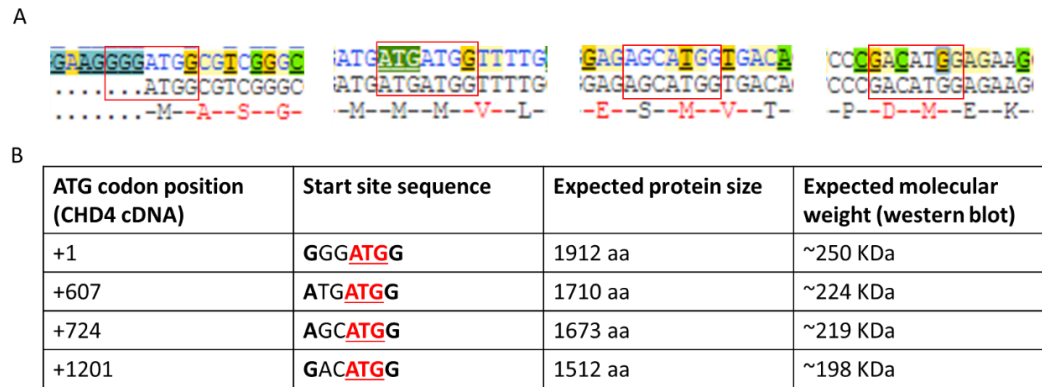


Figure 4.7. Additional translation start sequences found in the human CHD4 cDNA sequence. (A) cDNA and amino acid sequences of the original translation start site in human CHD4 (top left) and of three potential translation start sites downstream of it (middle left, middle right and top right). Translation start sequences are highlighted in red. (B) Information regarding the sequence, position, expected protein size and expected molecular weight detected by western blot based on translation from each start site. In the sequences, the ATG codon is red underlined, whereas the guanine in the position +1 and the adenine/guanine in the position -3 are bolded.

4.2.3. Physical interaction between CHD4 variants and the NuRD complex subunits

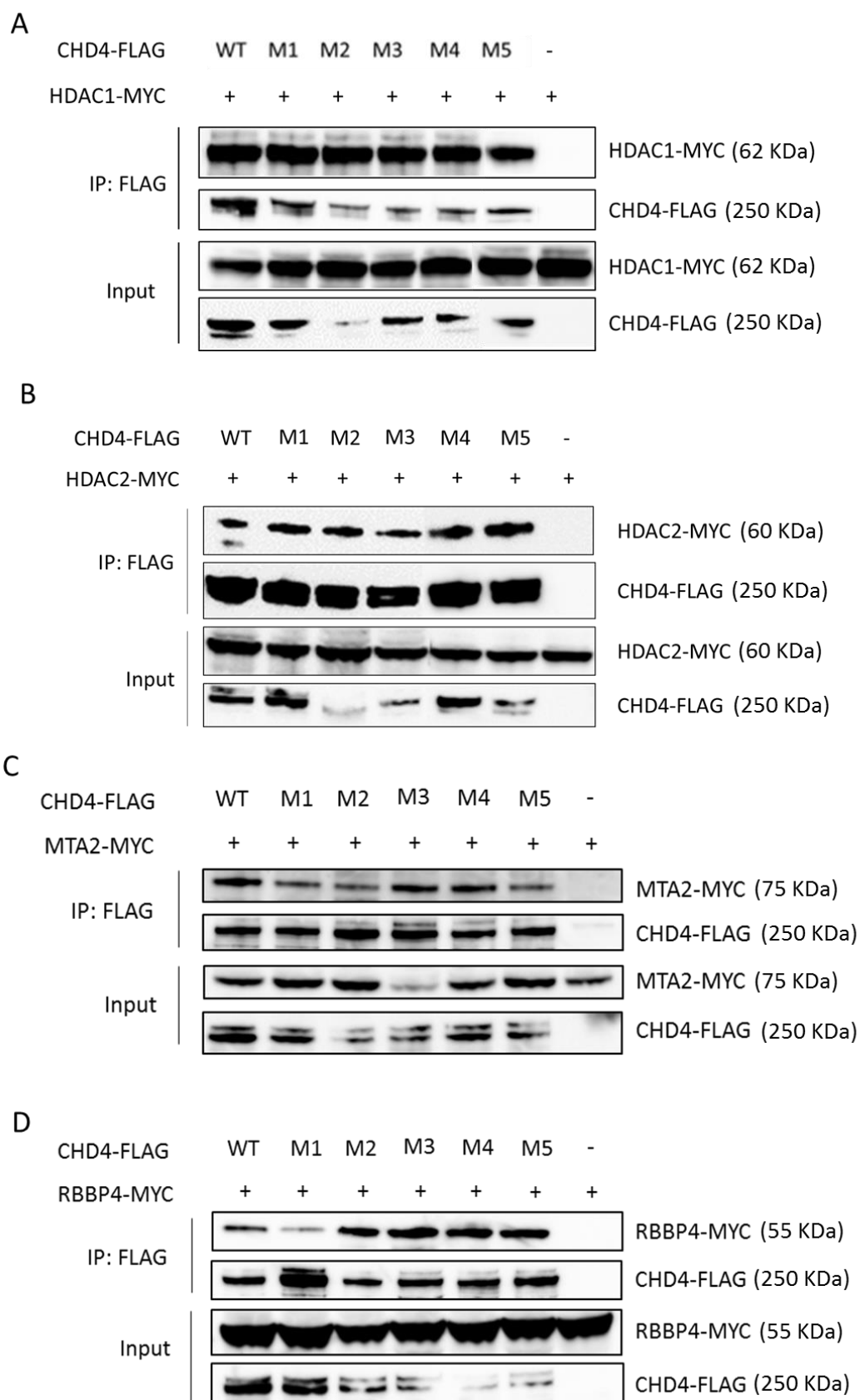
Once all CHD4 variants were generated, the first aim of this project was to check whether the interaction between CHD4 and the NuRD complex is disrupted in any of the mutant variants. The NuRD complex is comprised of several core subunits, including CHD4, HDAC1, MTA2 and RBB4 among others, so in order to address this question I performed pull-down assays by co-transfecting the different CHD4 mutant variants with expression vectors for each one of these subunits in COS7 cells (n=3 independent experiments were performed in each case).

Two of the main components of the NuRD complex are the histone deacetylases HDAC1 and HDAC2, so I initially performed pull-down assays following co-transfection of CHD4-FLAG and HDAC1-MYC or HDAC2-MYC in COS7 cells. As shown in Fig. 4.8A and 4.8B, a strong interaction between HDAC1 or HDAC2 and all CHD4 mutant variants was detected, with no significant differences between the mutant variants and CHD4 WT. Next, the physical interaction between CHD4 mutant variants and two other subunits of the NuRD complex was checked: MTA2 (metastasis-associated protein 2) and RBBP4 (retinoblastoma-binding protein 4) (Fig. 4.1). Fig. 4.8C shows that the interaction with MTA2 is well-conserved amongst all mutant variants. In the case of RBBP4, interaction between this subunit and variants M2, M3, M4 and M5 was conserved; however, the interaction between CHD4 M1 and RBBP4 was reduced in comparison to its WT counterpart (Fig. 4.8D).

In order to quantify the strength of the interaction between CHD4 M1-RBBP4 in comparison to CHD4 WT-RBBP4, I next aimed to perform densitometry on the IP RBBP4 and CHD4 bands (which are first normalized against their respective input bands) in order to calculate the relative strength of the interaction (relative strength of the interaction = intensity of the IP RBBP4 band divided by the intensity of the CHD4 band). However, as can be seen in Fig 4.8, many of the blots shown here are overexposed, thus making impossible to perform an accurate quantification of how much protein is present in each band (more details on this can be found in section 4.3). In spite of this, I still carried out this quantification: CHD4-RBBP4 interaction bands showed that the interaction between CHD4 M1 and RBBP4 was significantly reduced in comparison to the CHD4 WT-RBBP4 interaction (75% weaker, $p < 0.05$). Interaction between M2, M3, M4 or M5 and RBBP4 was generally increased in comparison to CHD4 WT, but this difference was not statistically significant (Fig 4.8E).

Again, as it was mentioned before, this quantification is not accurate as the bands used for the analysis were overexposed in some cases, so these results should be confirmed on a new set of experiments using non-overexposed bands. Nevertheless, the potentially disrupted interaction between CHD4 M1 and RBBP4 is a promising finding, so I wanted to gain further insight into this by mapping the RBBP4 interactional domain in the CHD4 protein. In order to achieve this, I planned on generating truncated versions of CHD4 and then perform truncated CHD4-RBBP4 pulldown assays in COS7 cells. Expression vectors for four different CHD4 truncated variants were successfully generated; however, due to time constraints, the pulldown assays could not be performed.

Taken together, these results indicate that the interaction between CHD4 and the NuRD complex is not disrupted in most of our CHD4 variants; however, CHD4 M1 interaction with the NuRD subunit RBBP4 seems to be reduced in comparison to its WT counterpart, thus suggesting that the interaction between this variant and the NuRD complex could be compromised in some way.



E

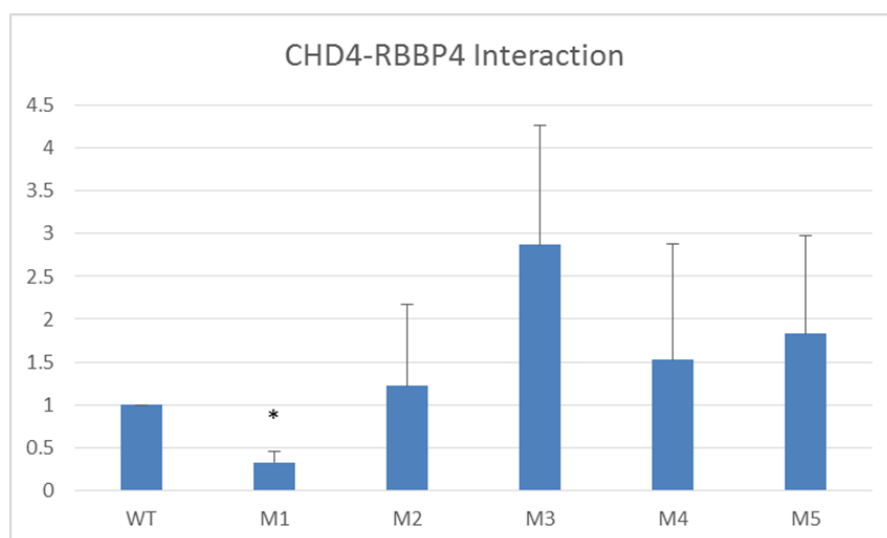


Figure 4.8. Interaction between CHD4 variants and different NuRD complex subunits. (A) and (B) Interaction between CHD4 and HDAC1 or HDAC2 is not disrupted in any of the mutant variants. (C) Interaction between CHD4 and MTA2 is not disrupted in any of the mutant variants. (D) and (E) Interaction between CHD4 M1 and RBBP4 is significantly reduced in comparison to the CHD4 WT-RBBP4 interaction. No significant difference is found for any of the other variants. Densitometry was performed on overexposed bands, so the results shown might not be accurate and should be further confirmed. Results are mean $\pm$ SD (n=3 independent experiments). *P < 0.05, Student's t-test

After studying the interaction between CHD4 and the NuRD complex, my next aim was to check the interaction between CHD4 variants and TBX5, already described elsewhere⁴¹⁸. In a similar fashion to the CHD4-HDAC/MTA2/RBBP4 interactional studies, CHD4-FLAG and TBX5-MYC expression vectors were co-transfected in COS7 cells to perform pull-down assays (n=3 independent experiments). As shown in Fig. 4.9, no interaction was detected between CHD4 WT, or any of its mutant variants, and TBX5. Different conditions were tried for the pulldown assays (e.g. buffers composition, transfection time, incubation times); however, no interaction between these two proteins was detected in any case.

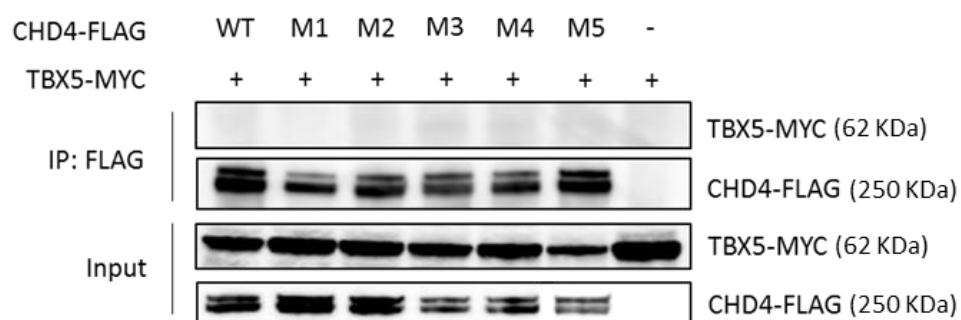


Figure 4.9. Interaction between CHD4 variants and TBX5. Following transfection in COS7 cells, pull-down assay was performed to check the interaction between CHD4 and TBX5. After trying different conditions, no interaction was detected between CHD4 WT, or any of the mutant variants, and TBX5 (n=3 independent experiments).

4.2.4. Cellular localization of CHD4 mutant variants

Following the interactional studies, I next aimed to study the cellular localization of CHD4 mutant variants in order to check whether its typical nuclear localization is compromised. To do this, I carried out an immunofluorescence staining following co-transfection of CHD4 variants and HDAC1, thus allowing me to check whether they co-localize inside the nucleus as well (a total of 50 cells were analysed per construct). Representative images are shown in Fig. 4.10. As expected, CHD4 WT presented a nuclear localization, and it co-localized with HDAC1, which is also expressed across the nucleus. All CHD4 mutant variants were also expressed ubiquitously across the nucleus, and they co-localized with HDAC1 as well. No significant differences in the cellular localization pattern between CHD4 WT and the mutant variants were found.

4.2.5. ATPase activity of CHD4 mutant variants (produced in COS7 cells)

CHD4 is an ATP-dependent helicase which catalyses the conversion of ATP into ADP in order to unwind the DNA in chromatin remodelling processes. In order to check whether the ATPase activity of CHD4 is affected in the mutant variants, I performed a luminescence-based ATPase assay (Promega) using purified CHD4 protein produced in COS7 cells, ATP and *in vitro* assembled nucleosomes, for which CHD4 presents a strong affinity (n=3 independent experiments; each independent experiment included 3 technical replicates). After normalizing the ATPase activity of the mutant variants in relation to the activity of CHD4 WT, it was revealed that mutant variants M2, M3, M4 and M5 presented a significantly reduced ATPase activity (48%, 78%, 40% and 39% of activity respectively; $p < 0.05$ for M2, M4 and M5, $p = 0.08$ for M3) in the presence of nucleosome, whereas CHD4 M1 presented a higher activity than its WT

counterpart (150%, $p < 0.05$). In the absence of nucleosomes, a really low ATPase activity was observed for all mutant variants, and no significant difference was found between them (Fig. 4.11).

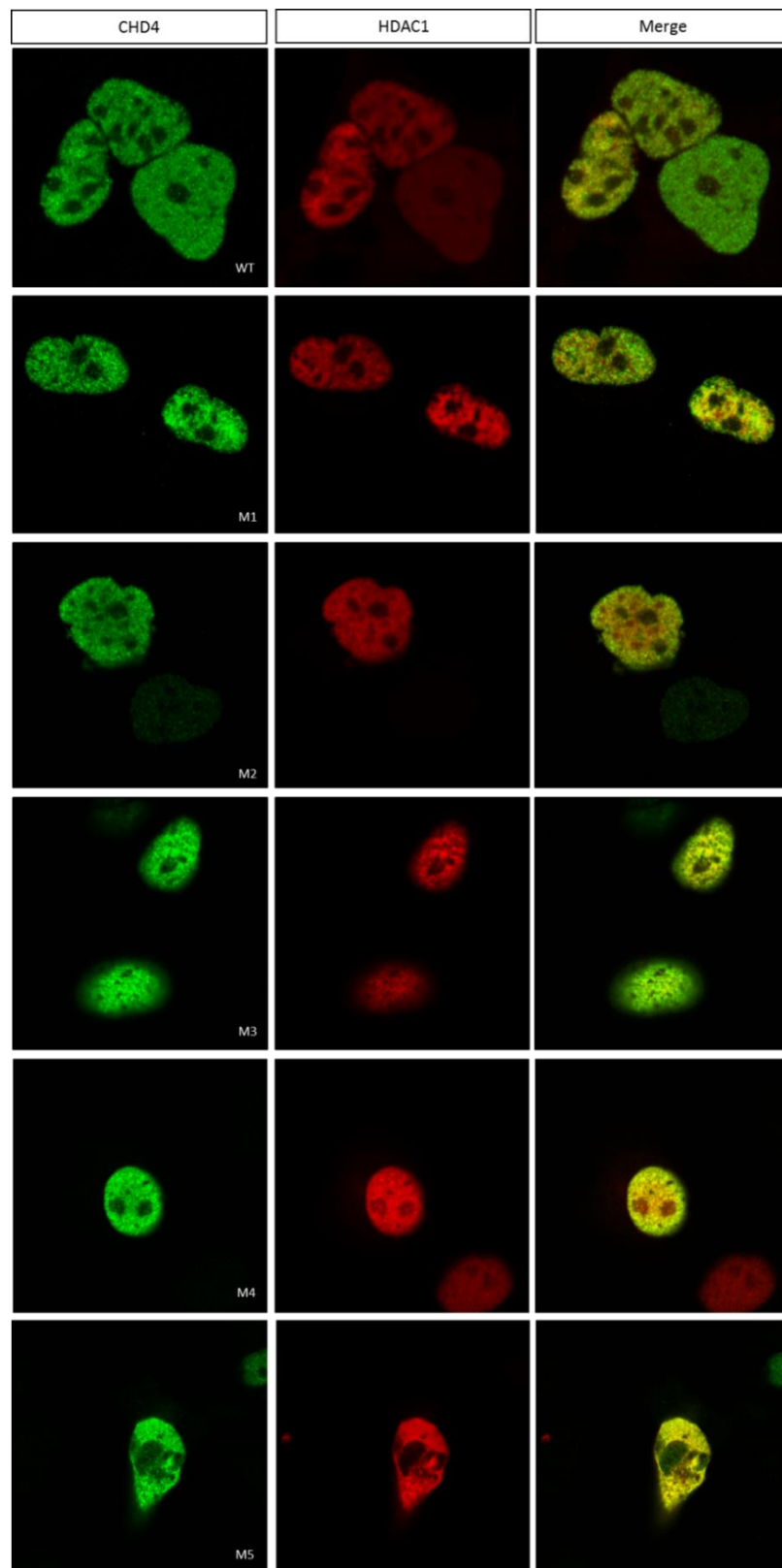


Figure 4.10. Cellular localization of CHD4 mutant variants. Immunofluorescence staining was performed on COS7 cells co-transfected with individual CHD4 variants and HDAC1. All CHD4 variants localize to the nucleus, and they co-localize with HDAC1 (a total of 50 cells were analysed per construct).

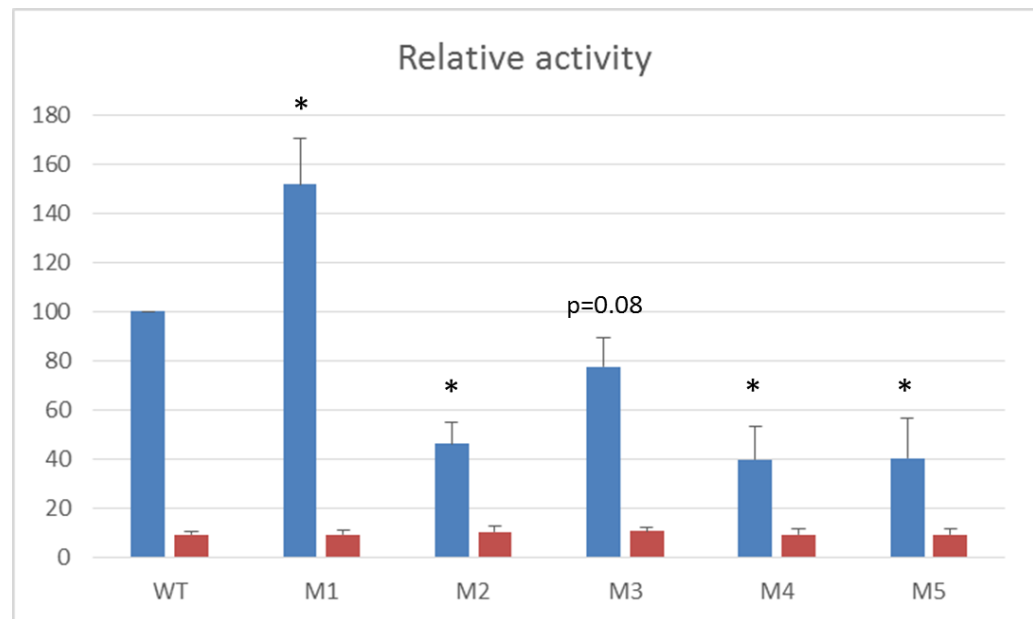


Figure 4.11. ATPase activity of CHD4 variants produced in COS7 cells. 600 nM of purified CHD4 variants were used in an ATPase assay along with 1 mM ATP and in the presence (blue bars) or absence (red bars) of 300 nM of purified nucleosomes. In the presence of nucleosomes, CHD4 M2, M3, M4 and M5 showed a reduced ATPase activity in comparison to their WT counterpart, whereas CHD4 M1 activity was increased. Lack of nucleosome in the reaction lead to a low activity of all CHD4 variants. Results are mean $\pm$ SD, *P < 0.05, Student's t-test (n=3 independent experiments; 3 technical replicates were included in each experiment).

4.2.6. CHD4 protein production in Sf9 insect cells

After analysing the results from our initial set of experiments using CHD4 protein, it would appear that the CHD4 mutations found in S-CHD patient mainly affect its ATPase activity by reducing it. This is of no surprise, as four of these five mutations are localized in the functional domain of the protein (M2, M3 and M5) or in domains which have been described to have an impact on the protein activity such as the PHD-type domains (M4). Mutation M1 seems to behave in a slightly different way, as it is localized in a domain of unknown function, and our results suggest that its ATPase activity is increased in comparison to its WT counterpart. Taking this into account, I decided to categorize these CHD4 variants into two groups according to their behaviour: on one group, mutations M2, M3, M4 and M5, which present a reduced ATPase activity and are localized in functional domains of the CHD4 protein (loss-of-function); and on the other group mutation M1, which has an increased ATP activity (gain-of-function) and is localized in a domain of unknown function (C-terminal domain of the CHD4 protein).

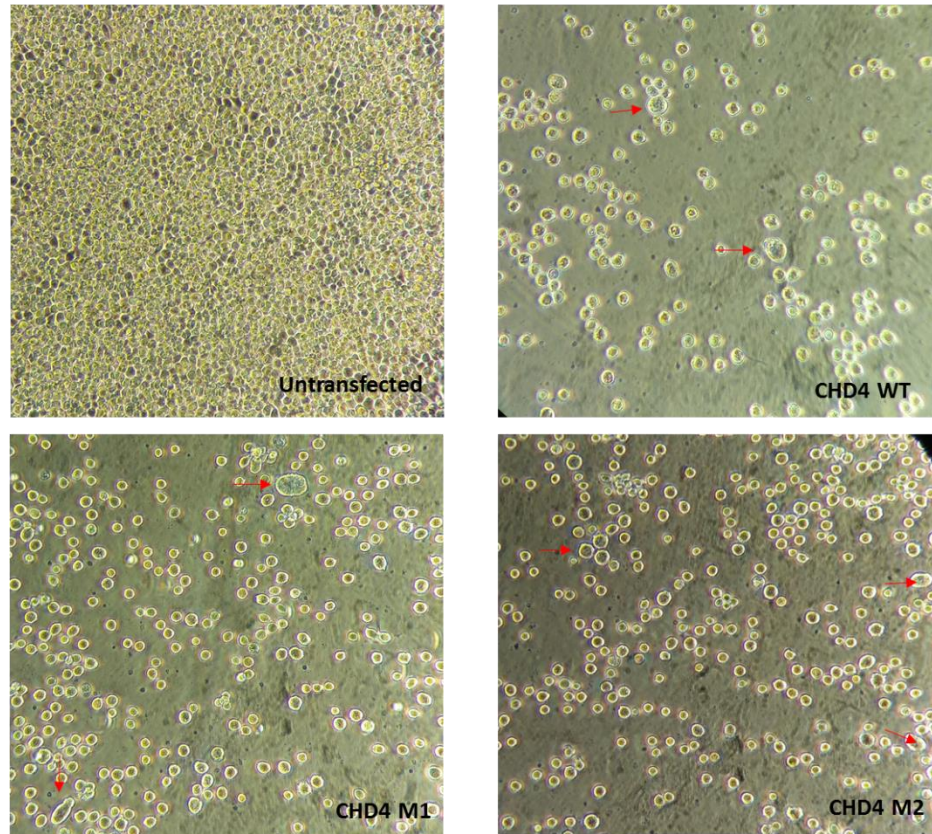
One of the main disadvantages of producing recombinant proteins in mammalian systems (e.g. COS7 cells) is that the yield of protein is relatively low in comparison to other protein production systems, a limitation for performing further *in vitro* assays with our CHD4 mutant variants. In order to overcome this problem and further characterize these two groups of variants, I moved onto the production and purification of CHD4 WT, M1 (second group of mutations) and M2 (first group of mutations) in Sf9 insect cells. These cells are easy to scale-up and grow in large volumes, and they provide high levels of protein expression with mammalian posttranslational modifications. The main disadvantage of this system is that it is quite time consuming, so instead of producing all four loss-of-function variants (all of which behave in a similar manner) only the M2 variant was chosen for production from this group.

To produce these three variants in Sf9 cells, I first transfected the cells with recombinant bacmid DNA harbouring the CHD4 WT, M1 or M2 cDNA sequence. For each construct, two wells of a 6-well plate were transfected (n=2 transfection experiments, one of which did not work as cells presented no signs of transfection). Five days post-transfection, cells were screened for signs of successful transfection and infection, which include increased cell diameter, cessation of cell-growth, loss or spherical morphology and detachment from the surface among others. As shown in Fig. 4.12A and 4.12B, cells transfected with CHD4 WT, M1 or M2 did present these signs of infection. After five days, untransfected cells generated a monolayer of cells which covered the whole surface of the flask, whereas transfected cells showed clear signs of cessation of cell growth in comparison to untransfected cells. In addition to this, many of the transfected cells showed a loss of the completely spherical Sf9 cells morphology and they presented a significantly increased cell diameter in comparison to untransfected cells (37%, 36% and 32% increase in WT, M1 and M2-transfected cells respectively, $p < 0.05$ for all three of them when compared to untransfected cells; n=10 cells analysed per construct).

Following Sf9 cells transfection, the supernatant containing the P1 virus was collected and used for posterior B1C stocks generations, and ultimately for large-scale protein production (see Materials and methods for further details). Recombinant CHD4 WT, M1 and M2 proteins from the successful transfection experiment were purified using Ni-NTA agarose beads and aliquots were run on 10% Bis-Tris gels to check for proper protein expression and purification. As shown in Fig. 4.13, all three WT, M1 and M2 variants were successfully expressed and purified as indicated by the presence of a strong ~250 KDa band in all six elution aliquots (out of a total of 10) for each one of them. In a similar fashion to what was seen in COS7 cells, an additional band of a lower molecular weight (~210 KDa) was also detected. In elutions E3 and E4, the presence of

some extra bands of lower molecular weight can also be seen, probable due to the presence of shorter translation products or unspecific proteins, but their signal is much fainter and weaker than that of the CHD4 protein.

A



B

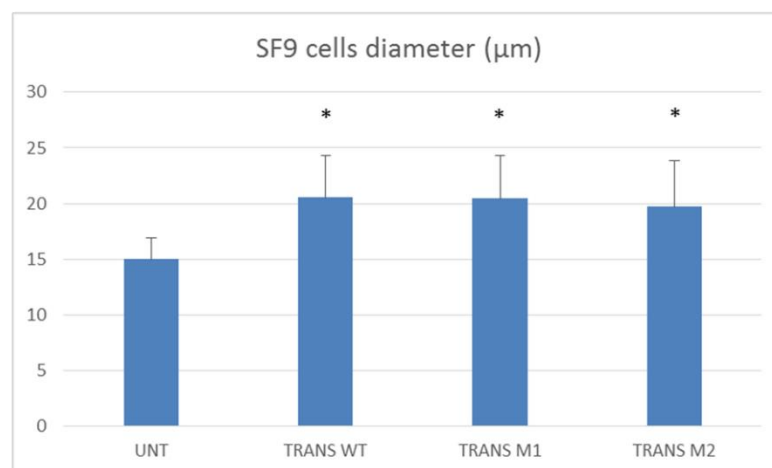


Figure 4.12. Transfection of Sf9 insect cells with CHD4 WT, M1 or M2 bacmids. (A) Representative images of Sf9 cells five days post-transfection. CHD4 WT, M1 and M2-transfected cells show signs of successful transfection and virus expression. Red arrows point to cells showing clear signs of successful transfection (n=2 transfection experiments, one of which did not work as no signs of transfection were found). (B) Mean cell diameter of untransfected and transfected Sf9 cells. The diameter of transfected cells is ~35% larger than

that of untransfected cells. Results are mean $\pm$ SD. *P < 0.05, Student's t-test (n=10 cells analysed per construct)

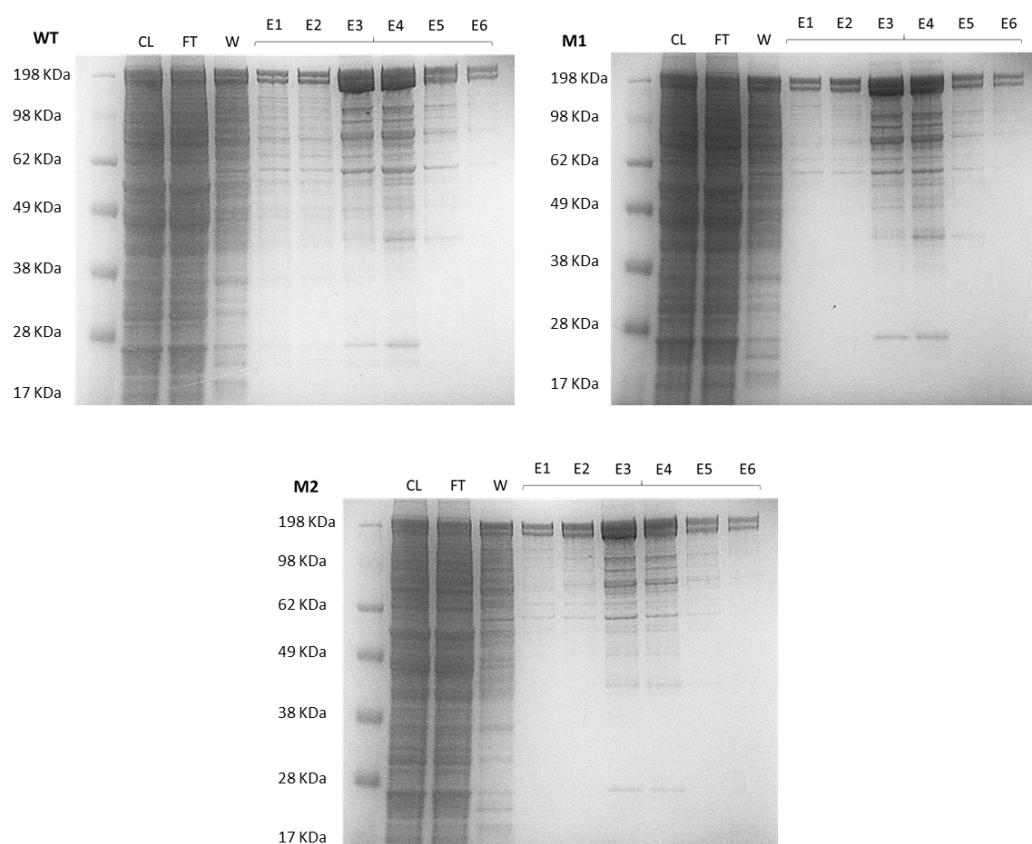


Figure 4.13. Large-scale expression and purification of CHD4 WT, M1 and M2 in Sf9 cells. For each construct, a sample of the whole cell lysate (CL), flow-through (FT) and wash (W) is shown, as well as six elution aliquots. Samples were ran on a 10% Bis-Tris gel and stained with Coomassie Blue (n=1 independent experiment).

4.2.7. ATPase assay using recombinant CHD4 produced in Sf9 cells

As with the CHD4 variants produced in COS7 cells, I first measured the ATPase activity of the recombinant CHD4 WT, M1 and M2 produced in Sf9 cells. To do this, recombinant CHD4 protein was incubated along with *in vitro* assembled nucleosomes in the presence of ATP, and the percentage of ATP hydrolysis was subsequently measured using the ADP-Glo kit (Promega) (n=3 independent experiments, each independent experiment included 3 technical replicates). As shown in Fig. 4.14, recombinant CHD4 M1 produced in Sf9 cells presented a higher ATPase activity than its WT counterpart (141% vs 100%, p<0.05), whereas the mutant variant M2 showed a lower ATPase activity than CHD4 WT (67% vs 100%, p<0.05). In both cases, the difference was statistically significant. These results were in line with the ATPase assay results we obtained using COS7 cells for CHD4 production.

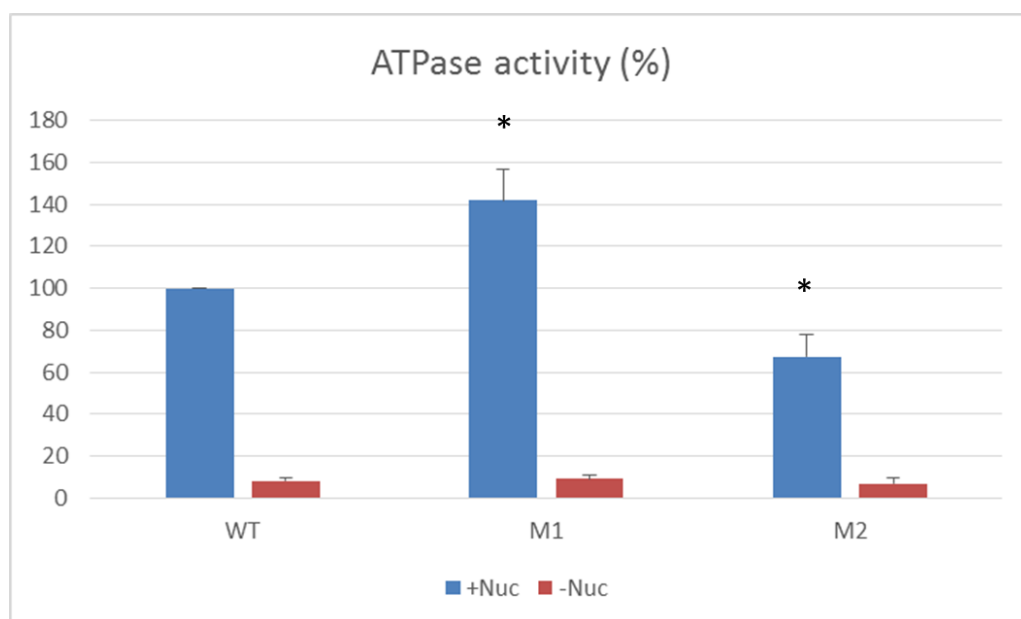


Figure 4.14. ATPase activity of CHD4 variants produced in Sf9 cells. 600 nM of purified CHD4 variants were used in an ATPase assay along with 1 mM ATP in the presence (blue bars) or absence (red bars) of 300 nM of purified nucleosomes. In the presence of nucleosomes, CHD4 M2 showed a reduced ATPase activity in comparison with their WT counterpart, whereas CHD4 M1 activity was increased. Lack of nucleosomes in the reaction led to a low activity of all three CHD4 variants. Results are mean $\pm$ SD ($n=3$ independent experiments, each independent experiment included 3 technical replicates). * $P < 0.05$, Student's t-test

4.2.8. DNA and nucleosome binding assays

In order to further characterize the CHD4 mutant variants, my next goal was to study their ability to bind free DNA as well as nucleosomes by performing a DNA/nucleosome-binding assay using 0-601-47 naked-DNA and 0-601-47 nucleosomes respectively (for a more detailed explanation on the structure of the 0-601-47 nucleosomes, and on how the assay is performed, see section 2.3.7 and 2.3.8). In these assays, increasing concentrations of CHD4 WT, M2 and M2 (50, 100, 300 and 600 nM) were incubated alongside 200 ng of 0-601-47 free-DNA (Fig. 4.15A) or 300 nM of 0-601-47 purified nucleosomes (Fig. 4.15B). These assays revealed that all three CHD4 variants (WT, M1 and M2) have the ability to bind both free DNA and nucleosomes in the same fashion and in a concentration-dependent manner, indicated by a shift of the free DNA and nucleosomes band on the gel. No significant differences were observed between the three CHD4 variants (quantification was performed by measuring via densitometry the band corresponding to the free-DNA in each case). A negative control including BSA did not produce a shift in the free DNA/nucleosomes band, thus suggesting that the CHD4 binding is indeed specific (Fig. 4.15A and B).

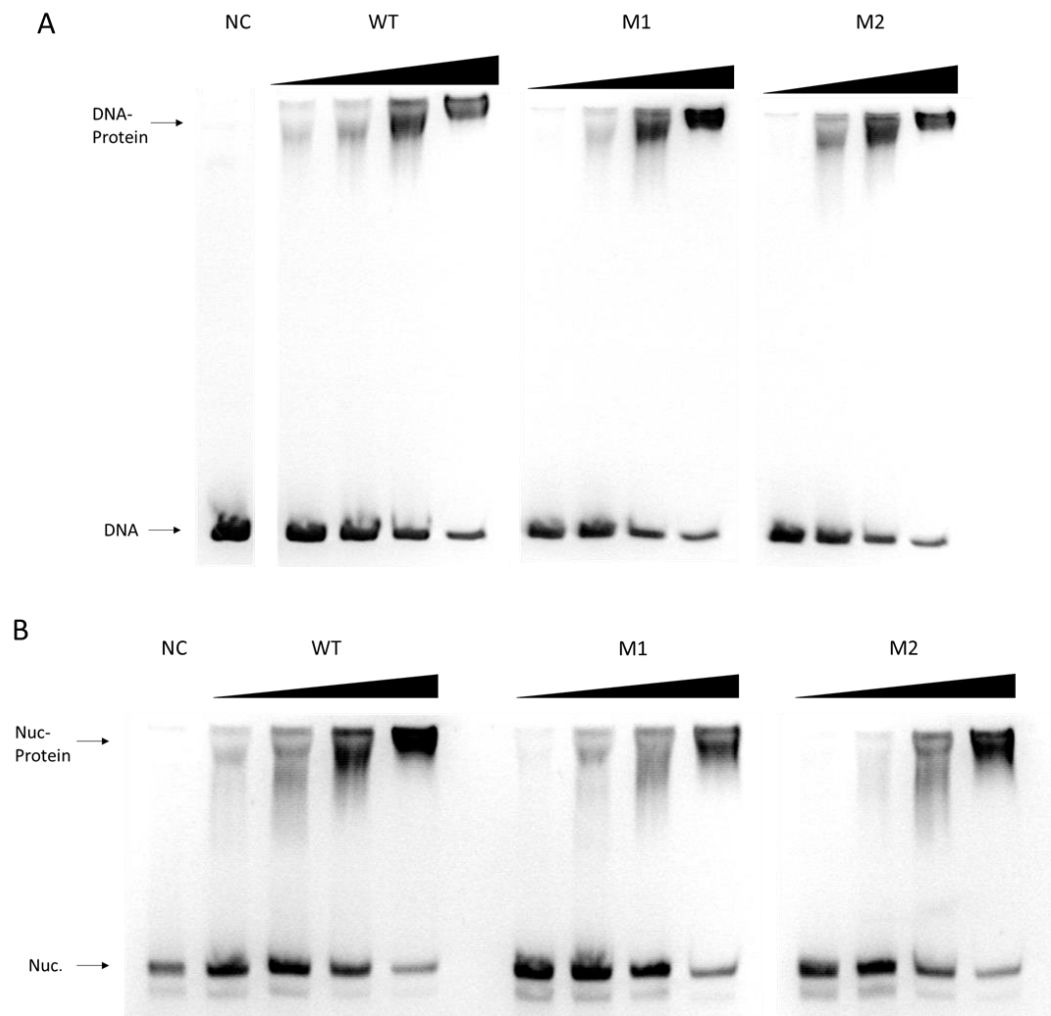


Figure 4.15. The ability of CHD4 variants to bind DNA and nucleosomes. (A) Binding of CHD4 WT, M1 and M2 to 0-601-47 free DNA fragments using increasing protein concentrations: 50, 100, 300 and 600 nM respectively. No significant differences were observed between CHD4 variants. (B) Binding of CHD4 WT, M1 and M2 to *in vitro* assembled 0-601-47 nucleosomes using increasing protein concentrations: 50, 100, 300 and 600 nM respectively. No significant differences were observed between CHD4 variants.

4.2.9. Nucleosome sliding assay

CHD4 is an ATP-dependent helicase which is able to perform nucleosome remodelling roles by itself. After confirming that CHD4 M1 and M2 are able to bind DNA and nucleosomes with the same efficiency as their WT counterpart does, my next aim was to check whether there is any difference in the nucleosome remodelling ability of these three CHD4 variants. To address this question, I carried out a nucleosome sliding assay by incubating the three CHD4 variants with 0-601-47 nucleosomes and ATP to promote nucleosome remodelling (further details on how this assay was performed can be found in 2.3.9). In a successful assay, the remodelled nucleosome is represented by a gel band of a slightly higher molecular weight than its non-remodelled

counterpart, thus representing a different nucleosome conformation (Fig. 4.16A, taken from ⁴³⁷). Unfortunately, in my assays, I was not able to find any remodelling activity in any of the three CHD4 variants (Fig. 4.16B). Different conditions were tried (e.g. different buffer composition, incubation temperature and time and nucleosome concentrations among others), but none of them triggered the nucleosome remodelling, not even in the WT case (n=3 independent experiments. Different conditions were tried on each one of them). Even though no remodelling activity was detected, I did find a reduction in the intensity of the band corresponding to the free nucleosomes at the highest CHD4 concentration. This suggests that there is indeed binding between CHD4 and nucleosome, but no remodelling is taking place or at least, not to a level detectable by this assay.

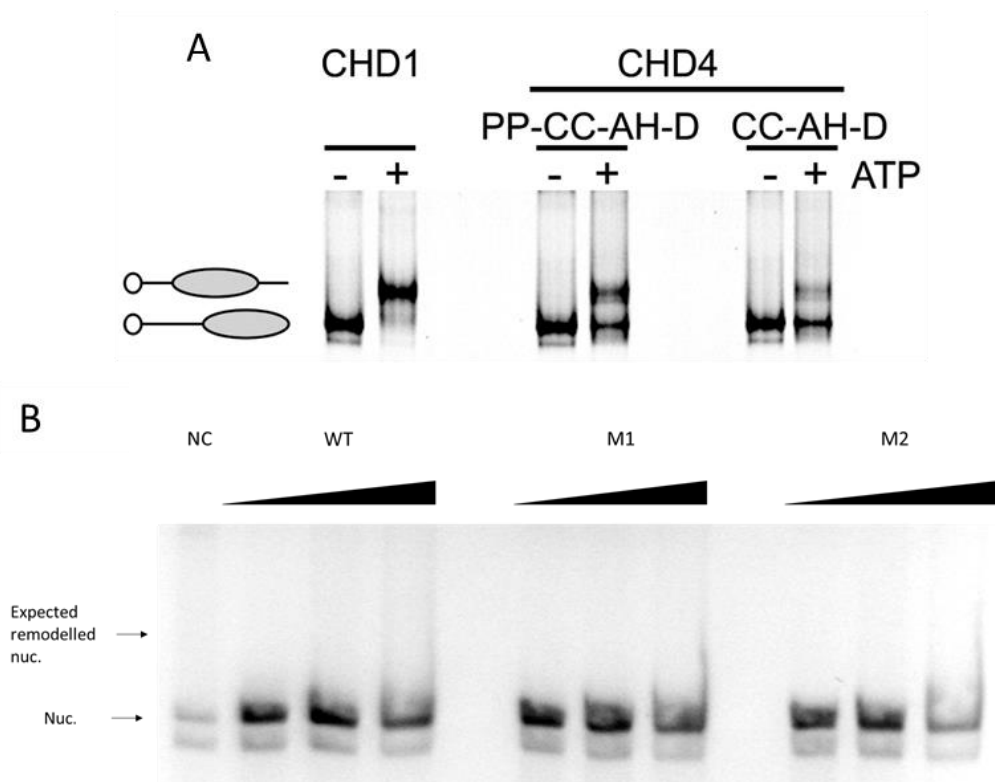


Figure 4.16. The ability of CHD4 variants to remodel nucleosomes. (A) Example of what a successful nucleosome remodelling assay looks like, taken from ⁴³⁷. The lower band represent the original nucleosome structure, with the histones situated on the right far end of the DNA fragment. When the nucleosome is remodelled by the remodelling enzyme, histones are moved to a different position in the DNA fragment, thus leading to a different nucleosome conformation and a shift in the original gel band. (B) Nucleosome sliding assay using CHD4 WT, M1 and M2. No remodelling activity was found in any of the CHD4 variants (n=3 independent experiments. Different conditions were tried on each one of them)

4.2.10. CHD4 protein production in *Escherichia coli*

In parallel to the production of CHD4 recombinant protein in Sf9 insect cells, I also carried out the production and purification of this helicase in *Escherichia*

coli as an alternative method. *E. coli* is also capable of producing high yields of recombinant protein in a quick, easy and cheap way, although it also has its limitations (e.g. lack of post-translational modifications, different molecular folding)⁴³⁸. It's been reported that producing full-length CHD4 in bacteria is a difficult task due to its size (around 200 KDa) and low solubility⁴³⁹. In order to increase the solubility of our bacterially-produced CHD4, I decided to produce a truncated, shorter version of CHD4 which lacks the C-terminal domain of the protein (Fig. 4.17A), which has no reported functions. Due to this, I was not able to generate the mutant variant M1, as this mutation is localized in the C-terminal domain of the protein. After testing different conditions (n=3 independent experiments trying different conditions), it was found that performing a slow induction (20°C for 16 hours) with 0.7 mM IPTG led to a strong induction of our recombinant proteins expression in the soluble fraction (Fig. 4.17B). Following induction and soluble protein expression, the five CHD4 variants (WT, M2, M3, M4 and M5) were purified using a polyhistidine affinity column (Fig. 4.17C).

4.2.11. ATPase assay using recombinant CHD4 produced in *E.coli*

Following the production and purification of CHD4 recombinant variants in *E.coli*, their ATPase activity was tested using the same protocol that was employed for the CHD4 recombinant proteins produced in COS7 and Sf9 cells, which involves using *in vitro* assembled nucleosomes as substrates (n=3 independent experiments). As shown in Fig. 4.18., the results obtained were quite similar to those obtained with the proteins produced in the other two systems: mutant variants M2, M3, M4 and M5 presented a significantly reduced ATPase activity than their WT counterpart in the presence on nucleosomes (60%, 65%, 41% and 56% respectively; $p < 0.05$ in all cases). In the absence of nucleosomes, the activity of all CHD4 variants was really low, and no significant differences between them were found.

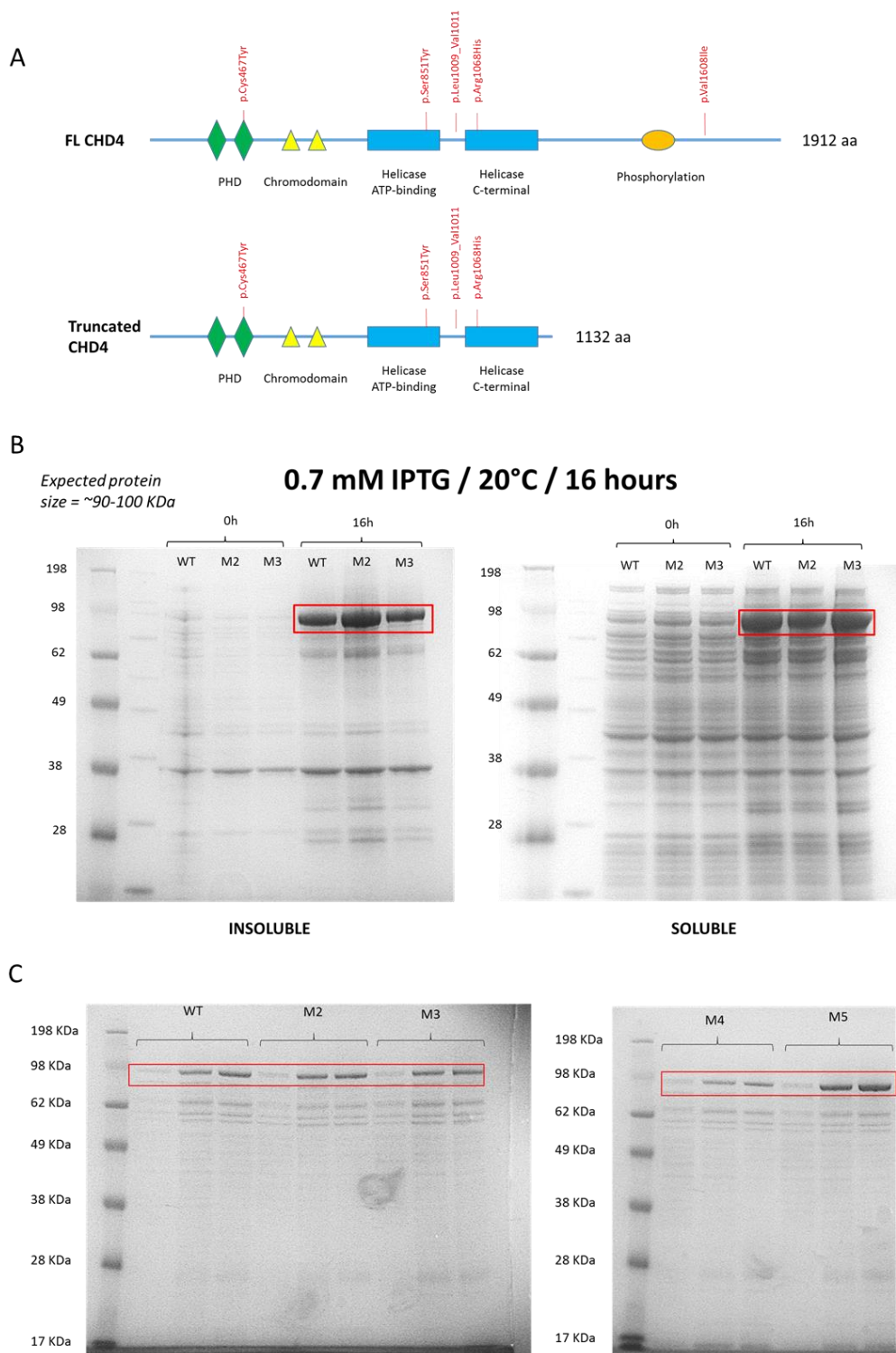


Figure 4.17. Production and purification of CHD4-His variants in *E.coli*. (A) Diagram showing the structure of full-length CHD4 (1912 aa) and that of the truncant variant used for expression in *E.coli* (1132 aa). (B) Induction of CHD4 WT, M2 and M3 variants using a slow-induction method. A strong induction of the protein is found on both insoluble (left) and soluble (right) fractions. (C) Purification of CHD4 variants using a polyhistidine affinity column. The first 3 elutions (total of 10 per variant) are shown on the gel. Truncated CHD4 bands are highlighted with a red box. (n=3 independent experiments, different conditions were assessed in each one of them)

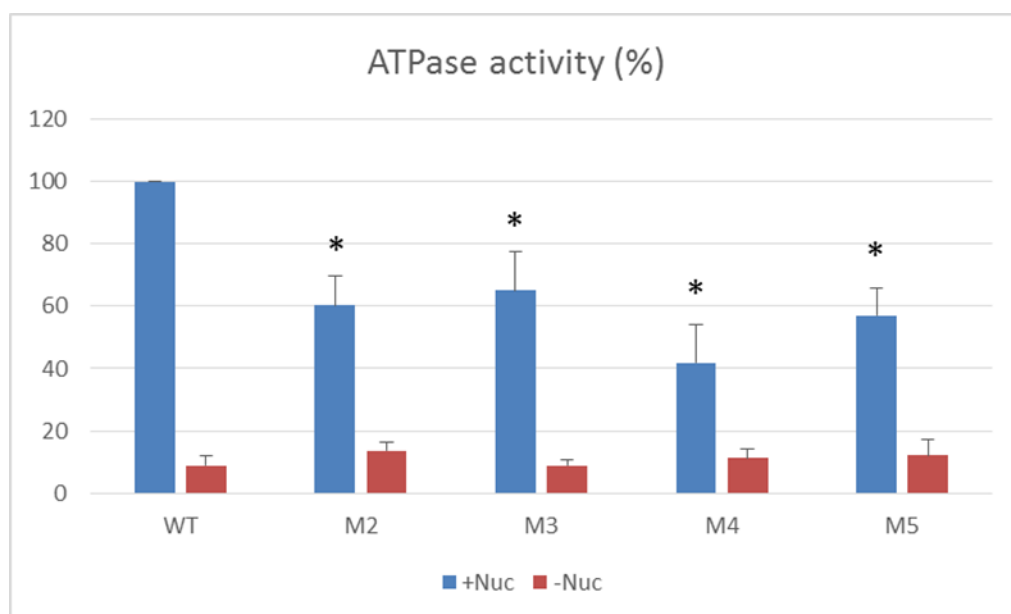


Figure 4.18. ATPase activity of CHD4 variants produced in *E.coli*. 600 nM of purified CHD4 variants were used in an ATPase assay along with 1 mM ATP in the presence (blue bars) or absence (red bars) of 300 nM of purified nucleosomes. In the presence of nucleosomes, CHD4 M2, M3, M4 and M5 present a significantly reduced ATPase activity in comparison to their WT counterpart. Lack of nucleosomes in the reactions led to a low activity of all three CHD4 variants, and no significant differences between them were observed. Results are mean $\pm$ SD (n=3 independent experiments). *P < 0.05, Student's t-test

4.3. Conclusions, discussion and perspectives for future work

The NuRD complex in general, and CHD4 in particular, has been subject to significant investigation in recent years to assess its role in cardiac development. Several publications in the field have pointed out this nucleosome remodelling complex as having a crucial role during cardiogenesis (see section 4.1. for more details). In fact, our lab recently collaborated with the Sanger Institute in a whole exome-sequencing study, which revealed five novel *de novo* mutations in the *CHD4* gene in patients of syndromic-CHD: one of them situated in one of the PHD domains (p.Cys467Tyr, M4), three in the ATPase/helicase domain (p.Ser851Tyr, M3; p.Arg1068His, M2 and p.Leu1009_Val1011del, M5), and one in a domain of unknown function situated at the C-terminal end of the protein (p.Val1608Ile, M1). CHD4 orthologs alignment revealed that the residues affected in mutations M2 to M5 are conserved across species, and alteration in any of them are predicted to be damaging according to Polyphen. However, the valine residue affected in the mutation M1 is not so well conserved, as two different amino acids (proline and leucine) can be found instead in other species, being this the reason why

it is not predicted to be damaging according to Polyphen. However, in close scrutiny, these three amino acids found in the residue 1608 share similar characteristics, as they all are non-polar and hydrophobic. This suggests that the presence of a residue with this chemical properties in this position is necessary for the correct structure or functionality of the protein.

In order to understand how these mutations lead to the onset of S-CHD, I aimed to generate those five mutant variants *in vitro* and perform several molecular assays to gain insight into their interaction with the NuRD complex, cellular localization and ATPase function, among others. Following the generation of the five CHD4 variants, my first goal was to study whether any of these mutations has an impact on the ability of CHD4 to bind other NuRD complex subunits such as HDAC1, HDAC2, RBBP4 and MTA2. Several years ago it was reported that CHD4 and HDAC1 interact through one of the PHD-type domain present in the C-terminal of the CHD4 protein⁴⁴⁰, a domain affected in one of the CHD4 mutant variants (M4). We can also hypothesize that due to the high percentage of homology between HDAC1 and HDAC2 (86% for the human protein), the latter also interacts with CHD4 through this PHD-type domain; however, apart from this study, really little is known about the CHD4 domains or regions which mediate its binding to other subunits. Pull-down assays revealed that the interaction between all of our mutant variants and HDAC1, HDAC2 and MTA2 was not affected in comparison to their WT counterpart. An effect in the interaction between CHD4 M4 (which carries a mutation in the PHD-type domain) and HDAC1 or HDAC2 was expected to be found; however, the interactions between them were completely fine. When checking the interaction between CHD4 and RBBP4, no significant differences in the interaction levels of CHD4 WT M2, M3, M4 or M5 and RBBP4 were detected; however, densitometry analysis initially revealed a significant decrease in the strength of the CHD4 M1-RBPP4 interaction in comparison to the WT protein. However, these results are not totally accurate, as the densitometry analysis was performed using overexposed western blots. In a simple way, densitometry measures the amount of light captured by the bound antibodies to the membrane: areas with more protein and bound antibodies will produce more light, thus producing a stronger and bigger signal in the blot. However, after a certain point, the bound antibodies cannot absorb any more light as they have reached their absorption limit, so further exposition to light will not produce an increase in signal (overexposition), thus leading to the loss of data and the generation of inaccurate results. Due to this, overexposed western blots should not be used for protein quantification and for the comparison of relative amounts of protein. Considering this, the quantification of the relative interaction between CHD4 variants and RBBP4 should be

repeated using non-overexposed western blots in order to confirm these preliminary results and to unravel the potential reduced interaction between CHD4 M1 (whose mutation lie at the C-terminal end of the protein) and RBBP4.

Recently it was shown that CHD4 also interacts with TBX5, one of the master regulators of cardiac development, in order to repress the expression of inappropriate genes in the heart during cardiac development⁴¹⁸. Taking this into account, my next aim was to check whether this interaction is disrupted in any of our CHD4 mutant variants. After performing pull-down assays I was not able to see any interaction between CHD4 WT (or any of the variants) and TBX5, in contrast to the positive interaction reported by Waldron *et al.* It is possible that the interaction between CHD4 and TBX5 is weak and thus, difficult to detect after performing the pull-down assay, which requires several stringent washes steps where the interaction could be disrupted. I tried performing the assay using milder conditions (e.g. reducing detergent amount in wash buffer), however, no interaction was detected. Another difference between the protocol used in this thesis and that of Waldron *et al.* lies in the tags used for immunoprecipitation. Here, FLAG-tagged CHD4 and MYC-tagged TBX5 are used, whereas Waldron *et al.* used GFP-tagged CHD4 and V5-tagged TBX5. Even though these tags are only 8-10 aminoacids in length, they still can have an impact in CHD4 or TBX5 conformation and ultimately, in their interaction with other proteins. Interestingly, Waldron *et al.* also managed to map the protein interaction domains of TBX5 and CHD4. In the CHD4 protein, this domain is localized in the N-terminal region of the protein. Both the PHD-type domains and chromodomains (situated in the N-terminal part of the protein) are not required for the CHD4-TBX5 interaction, implying that TBX5 interacts with an unidentified set of aminoacids situated in the N-terminal region of CHD4. Besides the mutation M4, localized in one of these PHD-type domains, none of the CHD4 missense mutations are localized in this N-terminal region of the protein, so even though our pull-down assays were unsuccessful, we can hypothesize that the interaction between them and TBX5 is potentially not affected. Last but not least, and in order to confirm whether the lack of interaction between CHD4 and TBX5 seen in my experiments is due to the experiments not working as intended, positive control pull-down assays for both TBX5 and CHD4 should have been run in parallel to the actual TBX5-CHD4 samples. For example, it has been shown that TBX5 is able to strongly interact *in vitro* with other cardiac transcription factors such as MEF2C³⁷⁹ or with acetyltransferases KAT2A and KAT2B⁴⁴¹, and as it has been shown in this thesis, CHD4 interacts *in vitro* with other NuRD complex subunits such as HDAC1 or HDAC2. By including a positive pull-down control for both CHD4 and TBX5 in the CHD4-TBX5 interaction experiment, it could have been elucidated whether

TBX5 and CHD4 do not interact, or whether this lack of interaction had its origin on the design and/or protocol used.

When performing pull-down assays and checking for the protein expression of CHD4, two bands that could correspond to CHD4 were detected in the western blots of all mutant variants: one of ~250 KDa and a second one of ~210 KDa. Previous reports^{418,425} have indicated that human CHD4 is detected in western blot gels as a protein of ~250 KDa, so based on this, I assumed that the upper band was the correct one. The presence of the lower band could be due to several reasons, including the presence of post-translational modifications, a second translation start site, proteolysis of the original protein or the binding of the antibody used to another epitope. As explained in 4.2.2., the only post-translational modification reported so far for CHD4 is phosphorylation by ATM kinase⁴⁴², which is greatly enhanced following DNA damage; however, no phosphorylation was detected, thus confirming that this was not the origin of the lower band. A study in close scrutiny of the human CHD4 cDNA sequence revealed the presence of three potential translation start sites downstream of the original one which would lead to the generation of protein products with a similar molecular weight to that of the lower band (~210 KDa), thus strengthening the hypothesis that this might be the reason for the appearance of this band. In spite of this, other explanations for the presence of this second band cannot be completely ruled out, such as the possibility that the anti-FLAG antibody used in these western blots binds to a different protein of a similar size or that the original CHD4 protein is being cleaved by proteases (although this is unlikely as protein lysates are always treated with proteinase cocktail inhibitor and kept at -80° C). Further experiments would have been required to fully elucidate whether the upper band was indeed the correct one, as well as the origin of the lower band. For example, for the former, a positive control (such as commercially available CHD4 protein) could have been used in western blots to compare its molecular weight with those of the bands produced by pcDNA-CHD4-FLAG; or a different primary antibody could have been tried to clarify whether the problem lies on this. For the latter, in order to confirm whether this second protein product is coming from one of the proposed alternative translation start site, conservative mutations of these sites (followed by western blot) could have been performed in order to study whether their abolishment lead to the generation of just a single protein product.

After checking the interaction between our CHD4 variants and the NuRD complex subunits, my next goal was to study whether any of these mutations affect CHD4 cellular localization. CHD4 and the NuRD complex reside within the nucleus to exert its nucleosome remodelling activity⁴⁰⁴. It has been proposed

that CHD4 nuclear localization signal (NLS) is localized in the N-terminal end of the protein^{410,443}, a region which is not affected by any of our CHD4 mutations. Indeed, immunofluorescence analysis revealed that CHD4 WT as well as the variants M1, M2, M3, M4 and M5 present an exclusively nuclear localization, being expressed ubiquitously across the nucleus and co-localizing with HDAC1, further supporting the results obtained in pull-down assays.

CHD4 is a helicase which hydrolyses ATP in order to perform its nucleosome remodelling activity. Following cellular localization studies, I aimed to study this intrinsic ATPase activity of CHD4, and see whether it is affected in the CHD4 mutant variants. An ATPase assay using *in vitro* assembled nucleosomes revealed that the ATPase activity of M2, M3, M4 and M5 variants was reduced in comparison to the CHD4 WT activity, whereas the activity of CHD4 M1 was significantly increased in contrast. These results were successfully replicated when the ATPase assay was performed using recombinant CHD4 proteins produced in Sf9 insect cells and *E.coli*. Mutant variants M2, M3 and M5 carry mutations in the ATPase/helicase domain of the CHD4 protein, so their ATPase activity could be compromised as our results indicate. CHD4 M4 present a mutation in one of the PHD-type domains of the CHD4 protein, a domain that has been shown to be required for its interaction with HDAC1⁴⁴⁰ and that is also required for the correct ATPase activity of CHD4⁴³⁷. Regarding CHD4 M1, these results point to a possible gain-of-function mutation. This is the only mutation localized in the C-terminal domain of the CHD4 protein, a poorly conserved domain of unknown function, so how this mutation can lead to an increase in CHD4 ATPase activity requires further investigation.

Recently, Weiss *et al.* also performed ATPase assay activities on several CHD4 variants found in syndromic-CHD patients, including one of the variants analysed in this thesis: p.Cys467Tyr (mutation M4). In these ATPase assays, it was demonstrated that in the presence of nucleosomes, the ATPase activity of the variant p.Cys467Tyr was reduced by 40% in comparison to its WT counterpart, thus corroborating the results obtained in this thesis (a reduction of 60% in activity in the presence on nucleosomes). In addition to testing in the presence of nucleosomes, Weiss *et al.* also measured the ATPase activity in the presence of naked-nucleosomal DNA, which also led to a similar reduction in activity of the p.Cys467Tyr variant. Unfortunately, none of the other variants analysed in this publications are the ones studied in this thesis; however, they concluded that missense substitutions in different protein domains alter CHD4 function in a variant-specific manner but result in a similar phenotype in humans, thus suggesting that there is not a strong genotype-phenotype correlation in these patients.

Even though the ATPase assay revealed extremely valuable information about how CHD4 activity is affected in these mutations, I wanted to gain further insight into this matter by performing DNA/nucleosome binding assays and nucleosome sliding assays. By performing these two experiments, we should be able to discern whether the reduced or increased ATPase activity in our mutant variants is due to a defect in substrate binding or due to a problem in ATP hydrolysis. I initially tried to perform these assays using recombinant CHD4 protein produced in COS7 cells; however, due to the low yield of recombinant protein obtained using this method, I decided to use recombinant protein produced in Sf9 cells (using CHD4 WT, M1 and M2 variants) instead. For the binding assays, either a 194 bp long DNA fragment which contains the '601' nucleosome position sequence, or *in vitro* assembled nucleosomes using this same piece of DNA and histones was used. This fragment was chosen as substrate as several publications have indicated that CHD4 presents a really high affinity for it^{407,437,444}. Both using naked DNA and nucleosomes, no significant differences were found in the DNA/nucleosome binding ability of CHD4 M1 and M2 when compared to their WT counterpart. Recently, an *in vitro* characterization of CHD4 mutations identified in endometrial carcinoma showed that mutations localized in the ATPase domain (like CHD4 M2) did not affect their nucleosome binding ability⁴⁴⁴, in line with our results. In the case of CHD4 M1, no other mutations in the same region of the protein have been studied so far; however, in the study mentioned above it was reported that nucleosome binding is mainly mediated by CHD4 chromodomains, so a normal DNA/nucleosome binding ability was also expected from this variant.

Last but not least, the nucleosome sliding capacity of CHD4 WT, M1 and M2 was also checked using a gel-based assay. Unfortunately, after many attempts using different buffers and reaction conditions, I was not able to detect any nucleosome remodelling activity, even for the WT variant of the protein, which has been shown to have a high remodelling activity *in vitro*^{407,437,444}. Taking into account that CHD4 recombinant proteins are active in other assays such as the ATPase one, a few hypotheses have been considered as the cause of this problem: one of the main reasons could be the fact that CHD4 is able to bind the nucleosome structure but then is not able either to perform the remodelling or to release the nucleosome after the remodelling is done. In fact, a slight reduction of the free-nucleosome signal was detected on most of our attempts, thus indicating that CHD4 is actually attached to the nucleosomes. An excess concentration of ATP (10 mM) was used in all reactions, so it is unlikely for this to be one of the causes of the problem. In order to release the remodelled nucleosome, competitor DNA (this is, a plasmid containing the 0-601-47 sequence) was used in excess as well, but it could be possible that this

piece of DNA is not an effective candidate to out-compete nucleosomes for CHD4 binding. Another explanation for this problem is that the remodelling activity is not high enough to allow for visualization in a gel-based assay. To stain the gels we used ethidium bromide, a typical nucleic acid stain with a detection limit of 4-5 ng/lane, so it is possible that the sensitivity of this compound is not high enough in order to detect the remodelled nucleosome. Also, in order to identify whether the problem had its origin on the purified CHD4 protein variants I was using, a positive control using commercially available CHD4 protein could have been used both for this assay and for the DNA binding assay, thus helping with the validation of both of them. Several life sciences companies such as Abcam (USA) offer in their catalogue purified CHD4 recombinant protein, and by performing the assays using this protein it could have been determined whether there was a problem with the recombinant proteins I produced or not. Due to time constraints, I could not continue carrying out further experiments in order to try and solve these issues, but it is a point that will be addressed in the near future.

To sum up, our results indicate that the main cause of dysfunction in our CHD4 mutant variants is an abnormal ATPase/helicase activity, something expected taking into account that four out of the five mutations are localized in domains involved in the regulation of this activity. To date, most of the CHD4 mutations found in patients of syndromic conditions or cancer are localized in the ATPase/helicase domain of the protein^{421,425,426}. *In vitro* characterization of some of these variants by other groups have revealed that these mutations exclusively affect the nucleosome remodelling activity of CHD4, leading to a partial or complete loss of function of the protein, whereas other features such as NuRD complex binding, localization or nucleosome binding seems to be unaffected, in line with the results obtained in this project. Interestingly, many of the phenotypic features found in the patients carrying these mutations suggest an overlap with CHARGE syndrome, which is also caused by heterozygous loss-of-function mutations in the paralogous gene CHD7⁴⁴⁵. An especially interesting case is the mutation M1, which is completely different from most of the reported mutations so far as it lies at the C-terminal end of the CHD4 protein. The results presented here indicate that in contrast to the other four CHD4 variants, this one seems to be a gain-of-function variant. Due to the poor characterization of this domain of the CHD4 protein, it is unclear at the moment how this mutation can affect the activity of CHD4. Kovac *et al.* identified a mutation situated outside of the ATPase/helicase domain (next to its C-terminal end), which presented an increased ATPase and remodelling activity, similar to CHD4 M1. The mechanistic basis of this difference in behaviour is still unclear; however, a possible explanation is that these

mutations might lead to a change in CHD4 conformation, which promotes a more efficient coupling of the ATP hydrolysis and nucleosome remodelling processes. Another question that also remains to be answered is how these mutations correlate with the cardiac phenotype of the five CHD4 mutation patients. Of these five patients, four of them present some kind of septal defect (ASD, VSD or Tetralogy of Fallot), and one of them present aortic coarctation. Interestingly, the patients presenting septal defects are those which harbour loss-of-function mutations in CHD4, whereas the aortic coarctation patient carries the mutation M1, which is suggested to be a gain-of-function variant according to our results. Even though we are still far from understanding the link between these mutations and their cardiac phenotypes, the fact that only mutation M1 (gain-of-function) shows a different cardiac phenotype than the others (loss-of-function) support our hypothesis that this mutation behaves in a completely different way than the rest.

Even though now we have a clearer picture of how these mutations affect CHD4 and subsequently lead to the onset of syndromic-CHD, further work for the characterization of these mutations is already planned and ongoing in our lab. First and foremost, the main priority at the moment is to continue with the CHD4 *in vitro* assays, and especially, with the nucleosome remodelling assay as it could provide extremely useful insight about how the mutations found in patients impact CHD4 remodelling activity. Apart from this, we will also continue with the characterization of the CHD4-RBBP4 physical interaction. Our pulldown assay results suggest that interaction between CHD4 M1 and RBBP4 is weaker than that between CHD4 WT and RBBP4. In order to gain further insight into this, we are planning on mapping the CHD4 interaction domain with RBBP4 to see whether the M1 mutation is indeed localized within it. As mentioned above, *in vitro* characterization of other CHD4 mutation present in cancer and S-CHD patients have been performed by others besides our group; however, to our understanding there are still no animal models carrying any of these patient mutations. To extend our knowledge on how these mutations affect cardiac development *in vivo*, our lab aims to produce a mouse line carrying a *Chd4* R1068H allele (M2 variant, which was found in a patient with ASD and VSD). This line will be mainly used for detailed characterization of heart morphology via high-resolution episcopic imaging, a novel 3D imaging technique that allows for easy and quick high-resolution 3D reconstruction of embryos and organs (more information about this technique can be found in chapter 5). By carrying out these experiments, our ultimate goal is to understand the precise roles of CHD4 in cardiac development and in the onset of CHD.

5. IN VIVO CHARACTERIZATION OF THE ROLES OF CHD4 AND CDK13 IN CARDIAC DEVELOPMENT

5.1. Introduction

5.1.1. Cyclin-dependent kinase 13 (CDK13)

The cyclin-dependent kinase 13 (CDK13) is a member of the cyclin-dependent serine/threonine protein kinase family, a group of 20 different ATP-dependent serine-threonine kinases regulating gene expression and cell-cycle progression⁴⁴⁶. More specifically, CDK13 is typically found forming a complex with cyclin K, which phosphorylates RNA polymerase II and is necessary for alternative splicing of RNA⁴⁴⁷. The structure of CDK13 is relatively simple, as it only contains a highly conserved serine–threonine protein kinase domain. CDK13 shares a high degree of similarity with CDK12, another member of this protein kinase family⁴⁴⁸. Since the discovery of CDK12 and CDK13 in the early 2000s, most of the published studies have focused on the former, so really little is known about the exact roles and functions of the cyclin K/CDK13 complex. The importance of the cyclin K/CDK12 complex in genome maintenance and stability has been demonstrated^{449–451}, but we can only hypothesize about a similar role for the cyclin K/CDK13 complex. A role for these two proteins in axonal elongation has also been proposed⁴⁵².

Only one study has examined the expression pattern of CDK13. Via *in situ* hybridization and western blot, Chen *et al.* showed that Cdk13 is ubiquitously expressed but not in the heart at different developmental stages (E6.5 to E9.5)⁴⁵². At later developmental stages, only expression in the brain was studied. In terms of animal models, the International Mouse Phenotyping Consortium (IMPC) provides information for a *Cdk13*-null mouse model (*Cdk13^{tm1b}*)⁴⁵³. Homozygous embryos from this line present pre-weaning lethality, whereas heterozygous mice are viable but they show increased startle reflex; however, no further information about these models, or any other Cdk13 models, is available. Two different mouse models have been reported for Cdk12: the first one is again reported in the IMPC (*Cdk12^{em1}*)⁴⁵⁴. Homozygous embryos from this line present embryonic lethality prior to organogenesis, whereas heterozygous embryos have an abnormal size as well as decreased levels of circulating HDL cholesterol and serum albumin. A second *Cdk12*-null model was generated by Juan *et al.*, and they reported that *Cdk12*-null homozygous embryos present embryonic lethality after implantation (E5.5), suggesting a strong development effect⁴⁵¹. No detailed exploration of heterozygous mice

was performed, but they were viable and indistinguishable from their WT counterparts.

Even though we are still far from fully understanding the role of CDK13 in development, recent whole-exome sequencing studies have revealed the presence of different *CDK13* mutations in individuals with syndromic-CHD. The first publication reporting mutations in the *CDK13* gene was by Sifrim *et al.*, in a study in which our lab was also involved (the *CHD4* mutations that were analysed in Chapter 4 were originally described in this publication)⁴²⁴. After performing a whole-exome sequencing study in 1891 probands, seven S-CHD individuals with clustered missense variants in the *CDK13* gene were identified: six *de novo* variants and one of unknown inheritance. All these mutations were localized in the highly conserved serine-threonine protein kinase domain, with four of the probands carrying the same missense mutation (p.Asn842Ser). These seven individuals were characterized by the presence of septal defects, microcephaly and significant developmental delay, among others. Shortly after this study, Bostwick *et al.* reported the discovery of nine new S-CHD individuals with mutations in the *CDK13* gene⁴⁵⁵. Out of these nine probands, seven carried the missense mutation p.Asn842Ser, which was previously published by Sifrim *et al.*, whereas the remaining two were novel *CDK13* variants. Congenital heart defects were found in six out of nine individuals, as well as other physical and developmental anomalies such as developmental delay, central nervous system abnormalities and gross motor delay. Last but not least, a third exome sequencing study reporting novel *CDK13* variants in S-CHD patients was published last year⁴⁵⁶. In this study, Hamilton *et al.* found nine patients carrying heterozygous mutations in the protein kinase domain of the CDK13 protein. Aside from one patient carrying a splice site mutation, all mutations were missense variants localized in the protein kinase domain of CDK13. Among them, one is the variant p.Asn842Ser, which was already found in several individuals in previous studies. Many different structural and development abnormalities were found in these individuals, including congenital cardiac defects, brain abnormalities and autism spectrum disorder, among others. As indicated, the p.Asn842Ser mutations if the most common variant found in these patients. Considering this, it would be of great interest to know what the frequency of this SNP is in the normal population, but no information regarding this has been published so far.

5.1.2. High-resolution episcopic microscopy

The study of developmental processes and morphological changes in mammalian embryos is a difficult task due to their size and lack of opacity. Analysis and imaging of histological sections using traditional methods provide

high-resolution two-dimensional (2D) images; however, this approach is far from ideal when it comes to analysing three-dimensional (3D) structures, as acquiring a comprehensive number of sections to properly study a 3D structure is quite time-consuming and technically challenging. In addition to this, the data and information is restricted to the plane that is being analysed, and technical limitations make accurate alignment of the sections an extremely difficult task. Nowadays, there are a few 3D imaging techniques available, such as ultrasound, magnetic resonance imaging (MRI) and optical projection tomography (OPT); however, they are of limited use in morphological analysis as they still do not provide high-enough resolution levels.

However, a new 3D imaging technique developed by Dr. WJ Weninger has overcome many of these problems. High-resolution episcopic microscopy (HREM) is a novel imaging technology that can generate volume data and three-dimensional computer models of whole histological samples or embryos in a simple and effective way⁴⁵⁷. In this technique, samples are embedded in a methacrylate resin called JB4, which is fluorescent as it contains eosin and acridine orange. Due to the ability of the tissue to suppress the fluorescence of the embedding plastic, the tissue can be visualized and this results in high-resolution images of the block surface. Once the sample has been embedded, the JB4 resin block is mounted in the HREM apparatus for sectioning. After every episcopic section (between 1-3 μM /section), a perfectly aligned, high-resolution image is captured, thus allowing for posterior data processing and 3D reconstruction using commonly available software⁴⁵⁸.

Although HREM is a relatively young imaging technique, it has already been successfully used in a wide variety of research fields, including cardiovascular development. In mice, it has been used in order to visualize single steps during cardiovascular development and remodelling^{459–461} and for the visualization of structural abnormalities in several knockout mouse models of human diseases^{462–464}. HREM has also been used for the visualization of cardiovascular structures in other animals models such as zebrafish⁴⁵⁸, frog⁴⁵⁷ and chick⁴⁶⁵. Due to its many advantages in comparison to other 3D imaging methods, HREM was also chosen as the main visualization technique for the phenotypic screening of E14.5 knockout mouse embryos produced in the Deciphering the Mechanisms of Developmental Disorders (DMDD) project⁴⁶⁶. In this project, nearly 700 mouse embryos from 87 different knockout lines which produce embryonically lethal offspring were screened.

5.1.3. Hypothesis and aims of this project

One of our main goals in our effort to characterize genes that contribute to congenital heart defects is to examine the diversity of heart morphology in

mouse lines carrying mutations in those genes. Traditionally, heart morphology has been extensively studied using high-resolution 2D images due to the limited resolution of classic 3D imaging techniques. However, with the development of HREM, we are now able to quickly and easily perform 3D reconstructions of potentially any embryonic organs and tissues.

In the final part of this thesis, the main goal is to take advantage of this new technology in order to study in greater detail the cardiac defects derived from mutations in either the *CHD4* or *CDK13* genes. In order to achieve this, our lab recently acquired three mouse lines carrying loss-of-function alleles for each one of these genes, as well as three lines carrying specific humanized mutations found in S-CHD patients (one per gene). Heterozygous and homozygous null mice from each one of these lines will be collected at E15.5, once the heart enters its final stages of development, and they will be subjected to HREM analysis and 3D reconstruction. Again, and as with the *in vitro* studies, this project was carried out by many people in Prof. David Brook's lab; in my case, I was in charge of the *CHD4* and *CDK13* loss-of-function models (*CHD4^{tm1b}* and *CDK13^{tm1b}* respectively).

As this part of the project was started just a few months before the end of my PhD, I only had enough time to perform an initial characterization of these lines, as well as to expand the colonies in preparation for our future HREM experiments. For these preliminary studies, I aimed to perform a lethality assessment on WT x HET and HET x HET crosses from both the *CHD4^{tm1b}* and *CDK13^{tm1b}* lines, a quick phenotype assessment of heterozygous and homozygous null embryos at different time points during development, and cardiac expression studies for both CHD4 and CDK13.

5.2. Results

5.2.1. *Chd4^{tm1b}* and *Cdk13^{tm1b}* genotyping

My first goal in the mouse project was to expand both the *Chd4^{tm1b}* and *Cdk13^{tm1b}* colonies for future experiments and to perform an initial genotype distribution assessment on WT x HET crosses that would allow us to check whether heterozygous null mice for any of these alleles are as viable as their WT counterparts. Fig. 5.1A shows the typical results routinely obtained for the genotyping of the *Chd4^{tm1b}* and *Cdk13^{tm1b}* transgenic lines. For each sample, we performed 3 different PCR reactions: a WT reaction to detect the WT allele (first lane on each sample), a MUT reaction to detect the knock-in allele and a LACZ reaction to detect the presence of the LacZ operon. As expected, in WT samples just a WT band is seen, whereas in *Chd4^{tm1b/+}* samples the WT, MUT and LACZ bands are all detected. For the *Chd4^{tm1b}* line, 52 pups were bred and genotyped,

of which 23 had a *Chd4*^{tm1b/+} genotype (44%) and 29 a *Chd4*^{+/+} genotype (56%). In the case of the *Cdk13*^{tm1b}, 134 pups were bred and genotyped (genotyping of this line was performed by Anna Wilsdon, Sophie Rochette and I): 67 presented a *Cdk13*^{tm1b/+} genotype whereas the other 67 had a *Cdk13*^{+/+} genotype (Fig 5.1B). Chi-square test on these genotype frequencies revealed that no significant differences exist between the observed and the expected genotypes distributions (p=0.405 for *Chd4*^{tm1b} and p=1 for *Cdk13*^{tm1b}) thus indicating that both *Chd4*^{tm1b/+} and *Cdk13*^{tm1b/+} heterozygous pups are viable and fertile.

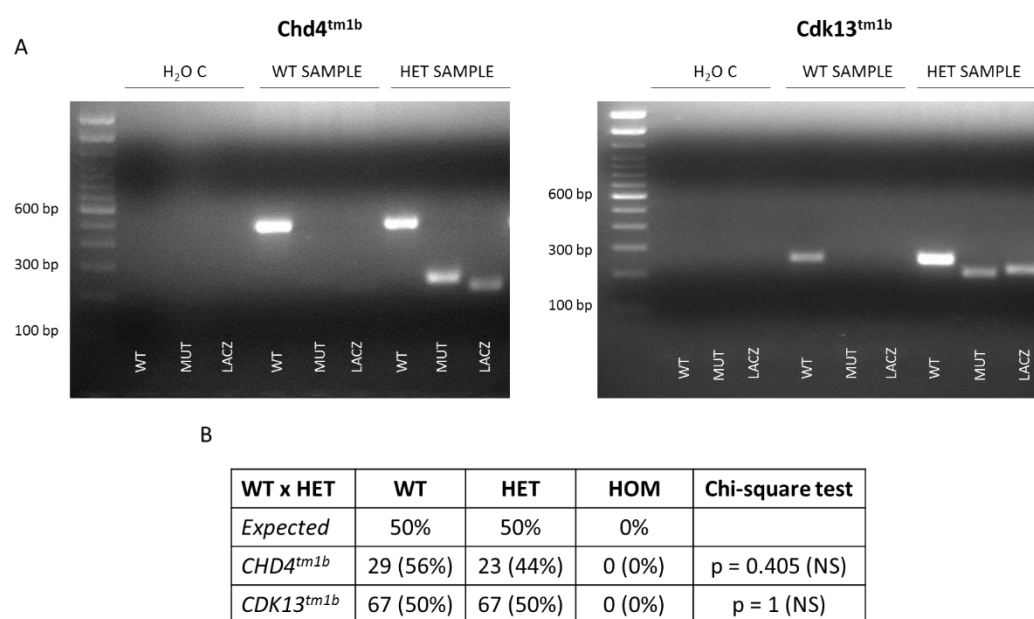


Figure 5.1. *Chd4*^{tm1b} and *Cdk13*^{tm1b} lines genotyping. (A) Agarose gel pictures showing the genotyping results for a WT and heterozygous samples from each of the transgenic lines. A water-only control is also included for each transgenic line. (B) Genotypes distribution of all the pups bred from WT x HET crosses for each of the lines. Chi-square test revealed no significant differences (NS) between the ratios observed and the Mendelian ratio. The genotyping of the *Cdk13*^{tm1b} line was performed together with Anna Wilsdon (David Brook's lab) and Sophie Rochette (Siobhan Loughna's lab)

5.2.2. Lethality assessment

According to the information provided by the Sanger Institute and the International Mouse Phenotyping Consortium, both *Chd4*^{tm1b/tm1b} and *Cdk13*^{tm1b/tm1b} homozygous mice are embryonically lethal. In order to determine at which point during development these embryos die, a preliminary lethality assessment was performed by genotyping embryos from HET x HET crosses at different time points during development. Fig. 5.2A shows the genotypes distribution obtained at different developmental stages from *Chd4*^{tm1b} HET x HET crosses. Due to time constraints I was only able to obtain 1 or 2 litters per time point (from E10.5 to E13.5), thus analysing 5 to 12 embryos,

but no homozygous null mice were found. Chi-square analysis on these distributions indicated that no significant differences were found in any stage analysed ($p \geq 0.05$ in all stages analysed): however, when the distribution of the total number of embryos analysed during development is considered, there is a significant difference in the observed distribution in comparison to the expected one ($p = 0.0004$).

The distribution of *Cdk13^{tm1b}* embryos obtained from HET x HET crosses is shown in Fig. 5.2B (genotyping of this line was performed by Anna Wilsdon, Sophie Rochette and I). A total of 6 to 29 embryos were analysed per time point, ranging from E9.5 to E15.5 (the number of litters obtained in each developmental stage is shown as well). In this case, we were able to find a total of four homozygous null embryos: one at E11.5, one at E12.5 (which had died shortly before collection) and two at E15.5. Chi-square analysis on these distributions showed that only at E9.5 and E15.5 a significant difference was found when the observed and expected distributions were compared ($p = 0.011$ and $p = 0.031$ respectively). As in the *Chd4^{tm1b}* case, when considered the total number of embryos obtained per genotype during development, statistical analysis revealed that a non-mendelian distribution was followed ($p = 0.000004$), thus indicating that homozygous *Cdk13^{-/-}* embryos die during development.

A

<i>Chd4^{tm1b}</i> HET x HET	WT	HET	HOM	REABSORBED	Chi-square test
Expected	25%	50%	25%		
E10.5 (1 litter)	2 (28%)	5 (72%)	0 (0%)	0	$p = 0.297$ (NS)
E11.5 (2 litters)	2 (17%)	10 (83%)	0 (0%)	0	$p = 0.05$ (NS)
E12.5 (2 litters)	2 (18%)	9 (82%)	0 (0%)	0	$p = 0.075$ (NS)
E13.5 (1 litter)	1 (20%)	4 (80%)	0 (0%)	2	$p = 0.333$ (NS)
Total	7 (20%)	28 (80%)	0 (0%)	2	$p = 0.0004^*$

B

<i>Cdk13^{tm1b}</i> HET x HET	WT	HET	HOM	REABSORBED	Chi-square test
Expected	25%	50%	25%		
E9.5 (1 litter)	0 (0%)	9 (100%)	0 (0%)	0	$p = 0.011^*$
E10.5 (1 litter)	2 (66%)	1 (33%)	0 (0%)	3	$p = 0.223$ (NS)
E11.5 (2 litters)	1 (11%)	7 (78%)	1 (11%)	0	$p = 0.249$ (NS)
E12.5 (3 litters)	4 (29%)	9 (64%)	1 (7%) [#]	1	$p = 0.297$ (NS)
E13.5 (2 litters)	2 (17%)	10 (83%)	0 (0%)	0	$p = 0.05$ (NS)
E14.5 (1 litter)	1 (17%)	5 (83%)	0 (0%)	0	$p = 0.223$ (NS)
E15.5 (7 litters)	6 (21%)	21 (72%)	2 (7%)	12	$p = 0.031^*$
Total	16 (19%)	62 (76%)	4 (5%)	16	$p = 0.000004^*$

Figure 5.2. *Chd4^{tm1b}* and *Cdk13^{tm1b}* lethality assessment. (A) Genotype distribution of alive embryos obtained from *Chd4^{tm1b}* HET x HET crosses from E10.5 to E13.5. (B) Genotype distribution of embryos obtained from *Cdk13^{tm1b}* HET x HET crosses from E10.5 to E15.5. The genotyping of the *Cdk13^{tm1b}* line was performed together with Anna Wilsdon (David Brook's

lab) and Sophie Rochette (Siobhan Loughna's lab). # Recent death. Chi-square test was performed on all genotype distributions (NS = not significant, *p<0.05)

A quick phenotypic analysis of each one of the embryos collected during the lethality assessment was also performed, although the numbers are probably too small to state anything conclusive from these results. In the case of the *Chd4^{tm1b}* line, heterozygous embryos for this allele were in general indistinguishable from their WT counterparts; however, a few of them were smaller in size or delayed in development in comparison to WT embryos (3/12 embryos at E11.5, 2/11 at E12.5 and 2/7 at E13.5) (Fig. 5.3 and 5.4A). Hearts were also extracted from these embryos; however, no obvious differences were found between them and WT hearts (Fig. 5.4B). Two reabsorbed embryos were also found at E13.5: these were completely necrotic when identified, so tissue could be collected for genotyping.

Gestation stage	# of embryos	Phenotype
E10.5	7	7 normal
E11.5	12	9 normal, 3 small/delayed
E12.5	11	9 normal, 2 small/delayed
E13.5	7	3 normal, 2 small/delayed, 2 necrotic

Figure 5.3. Preliminary phenotypic analysis of analysed *Chd4^{tm1b}* embryos. In general, collected embryos presented a normal phenotype (normal), a small size or delayed development (small/delayed) or they were completely necrotic (necrotic)

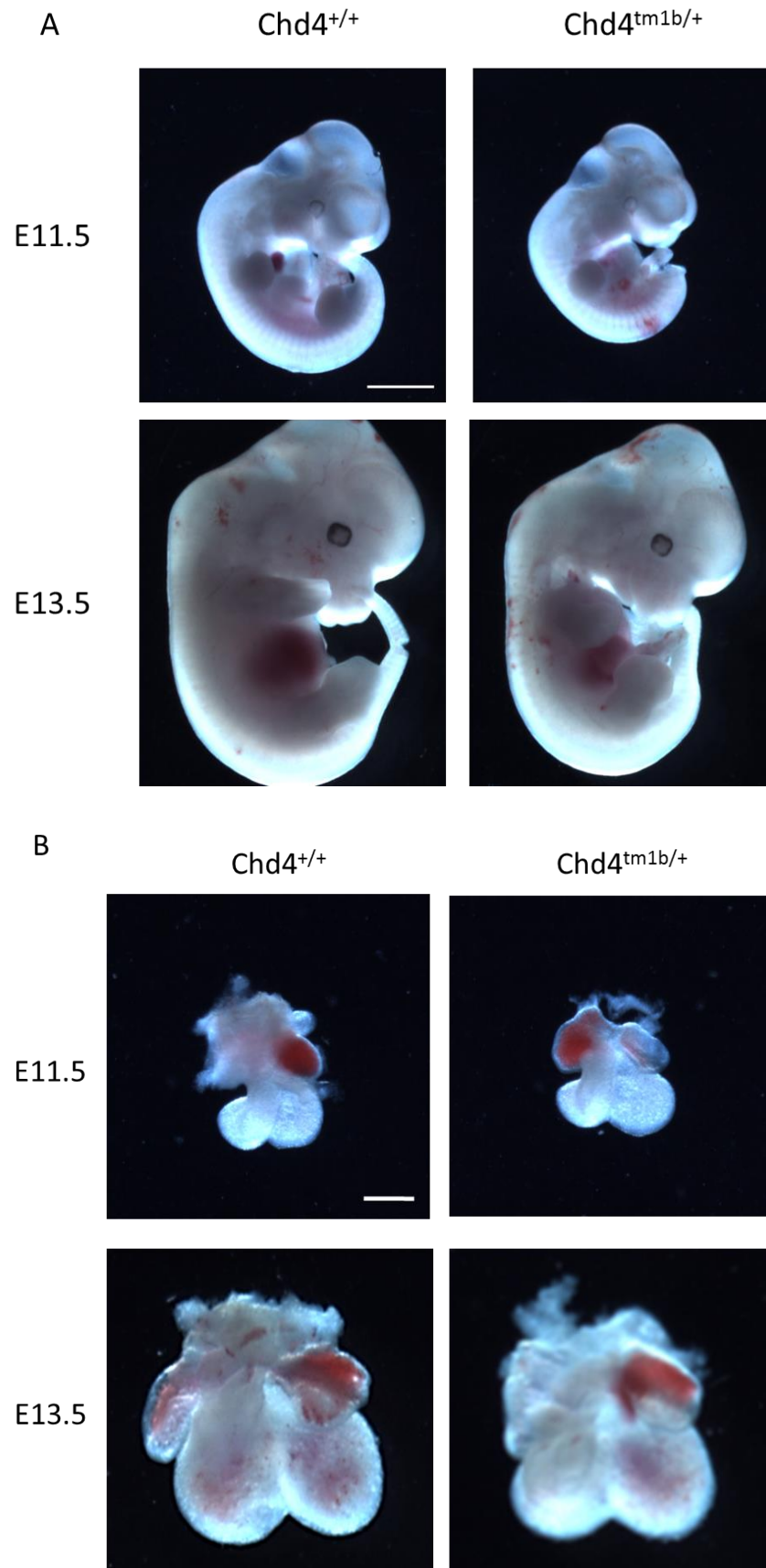


Figure 5.4. Gross morphological appearance of *Chd4*^{+/+} (left panels) and representative *Chd4*^{tm1b/+} (right panels) embryos and hearts at E11.5 and E13.5. (A) Representative pictures

of whole embryos from the *Chd4^{tm1b}* line. A slight reduction in size and developmental delay was found in some *Chd4^{tm1b/+}* embryos. Scale bar = 2mm. (B) Representative pictures of WT and heterozygous hearts. No significant differences were observed. Scale bar = 350 μ m

Regarding the *Cdk13^{tm1b}* line, we were able to analyse embryos ranging from E9.5 to E15.5 (Fig. 5.5) (phenotypic analysis of this line was carried out by Anna Wilsdon and I). In general, all or most embryos collected at each gestational stage presented a completely normal phenotype and were indistinguishable from WT embryos; however, embryos presenting a small size or delayed development, as well as necrotic embryos were also found at most stages, albeit in a lower proportion (2/9 small/delayed at E9.5, 3/16 necrotic at E10.5, 4/16 small/delayed and 1/16 necrotic at E12.5, 2/13 small/delayed and 1/13 necrotic at E13.5, and 2/41 small/delayed and 14/41 necrotic at E15.5). Representative pictures of a WT and a small/delayed heterozygous *Cdk13* embryo are shown in Fig. 5.6A. The heart of these embryos was also extracted and examined, and as expected, a reduction in the size of the ventricles and atria was found in the heterozygous embryo (Fig. 5.6B).

Gestation stage	# of embryos	Phenotype
E9.5	9	7 normal, 2 small/delayed development
E10.5	6	3 normal, 3 necrotic
E11.5	9	9 normal
E12.5	16	11 normal, 4 small/delayed, 1 necrotic
E13.5	13	10 normal, 2 small/delayed, 1 necrotic
E14.5	6	6 normal
E15.5	41	25 normal, 2 small/delayed 14 necrotic

Figure 5.5. Preliminary phenotypic analysis of analysed *Cdk13^{tm1b}* embryos. In general, collected embryos presented a normal phenotype (normal), a small size or delayed development (small/delayed) or they were completely necrotic (necrotic). Phenotypic analysis was carried out together with Anna Wilsdon (David Brook's lab).

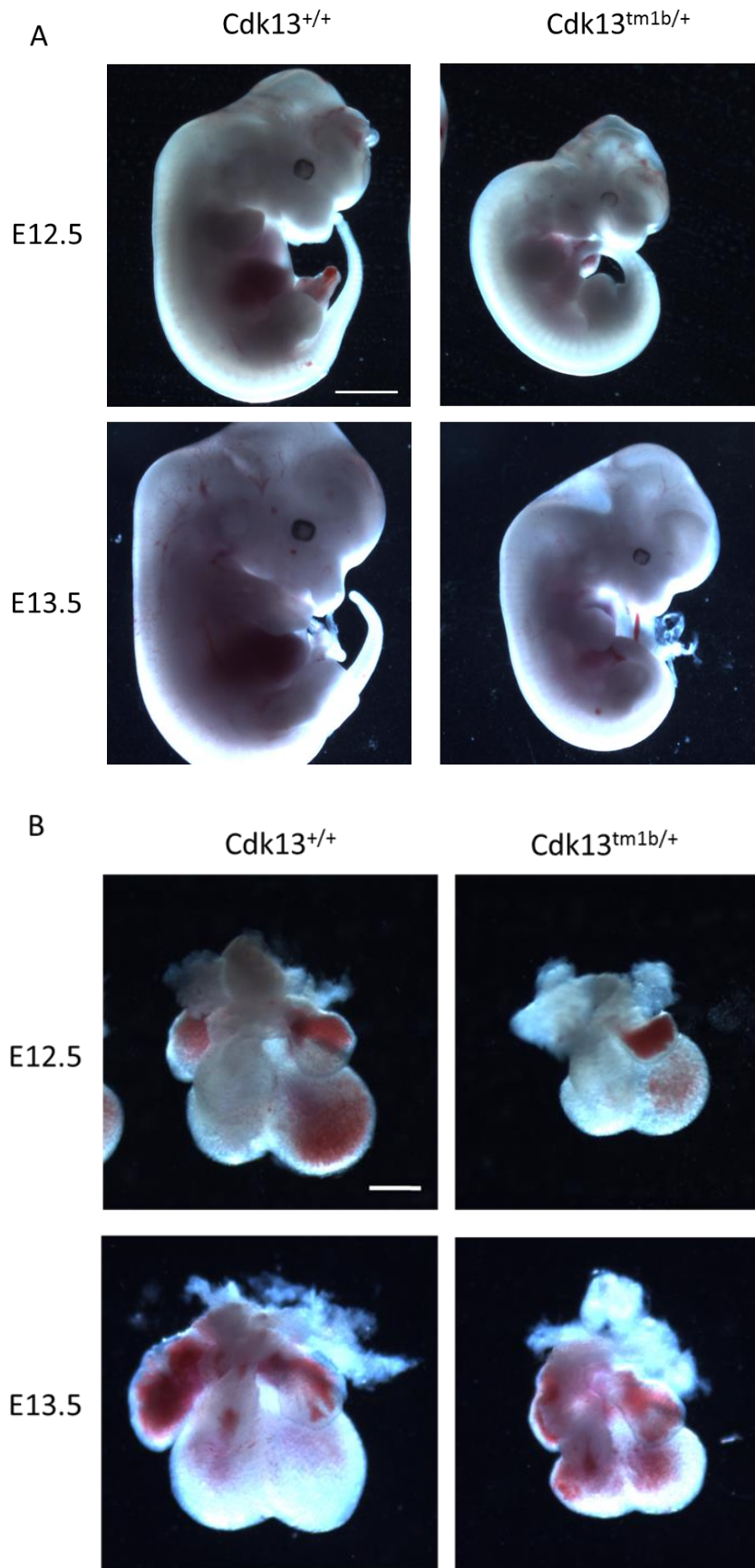


Figure 5.6. Gross morphological appearance of $Cdk13^{+/+}$ (left panels) and representative $Cdk13^{tm1b/+}$ (right panels) embryos and hearts at E12.5 and E13.5. (A) Representative pictures

of whole embryos from the *Cdk13^{tm1b}* line. A reduction in size and developmental delay was found in some *Cdk13^{tm1b/+}* embryos. Scale bar = 2 mm. (B) Representative pictures of WT and heterozygous hearts from developmentally delayed embryos. Heterozygous hearts are smaller than their WT counterparts. Scale bar = 350 μ m

5.2.3. CHD4 and CDK13 expression during development

In order to gain information about the cardiac expression of both Chd4 and Cdk13 during development, two different approaches were followed: X-Gal staining on *Chd4^{tm1b/+}* and *Cdk13^{tm1b/+}* hearts collected at E12.5 and E13.5, and western blot analysis using E12.5, E13.5 and E14.5 cardiac protein extracts from WT embryos. Fig. 5.7 shows the protein expression of Chd4 and Cdk13 in the heart during development from E12.5 to E14.5. Due to time constraints, only one western blot experiment was performed, and these preliminary results are shown in Fig. 5.8. This experiment revealed that both Chd4 and Cdk13 are detected at E12.5, E13.5 and E14.5, and that the expression of Chd4 seems to be higher than that of Cdk13 at all time points analysed. Both for Chd4 and GAPDH (which was used as a loading control), non-specific bands were also detected in the blots, especially in the case of Chd4 in which a strong, smeared band of a higher molecular weight was also detected at all time points.

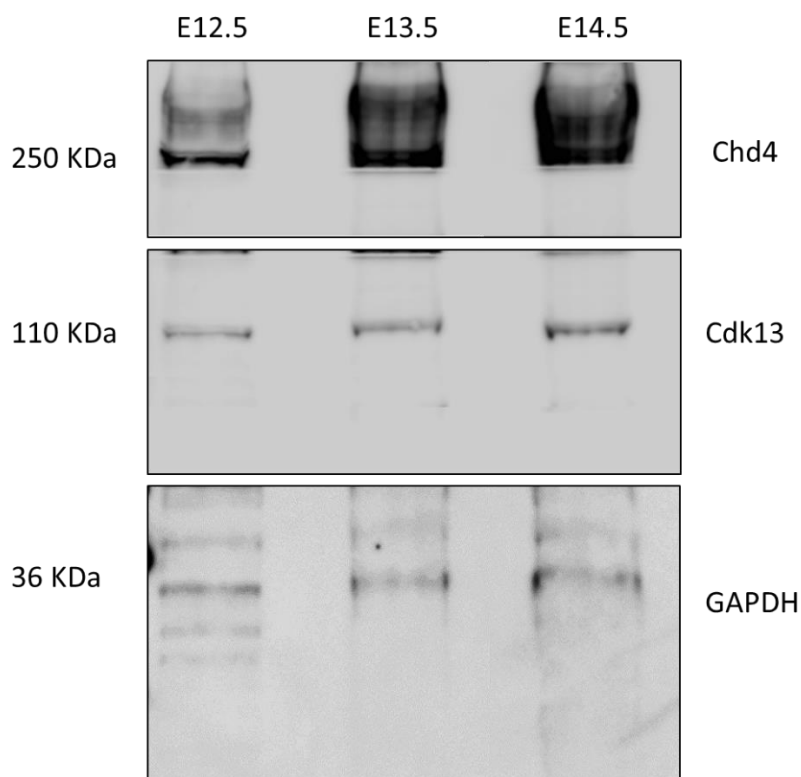


Figure 5.7. Western blot detection of Chd4 and Cdk13 in mouse heart extracts. Extracts were generated from pools of 10 WT hearts at different stages of development (E12.5, E13.5 and E14.5). Anti-GAPDH was used as loading control.

The cardiac expression of both *Chd4* and *Cdk13* was also analyzed by performing X-Gal staining on heterozygous hearts from both lines (n=5 hearts analysed per line and developmental stage) (Fig. 5.8). Staining of E13.5 *Chd4*^{tm1b/+} hearts revealed a really strong expression of *Chd4* in a ubiquitous manner throughout the heart, and more prominently in the ventricles. Expression of *Cdk13* was also detected in the heart at two different time points: E12.5 and E13.5, also in a ubiquitous manner. When comparing *Chd4* and *Cdk13* staining at E13.5, the *Chd4* signal was slightly stronger than that of *Cdk13*, thus suggesting that the expression of *Chd4* is higher. In conclusion, although further replicates and experiments are required to fully confirm it, preliminary expression analysis via western blot and X-Gal staining show the expression of *Chd4* and *Cdk13* during cardiogenesis at different time points.

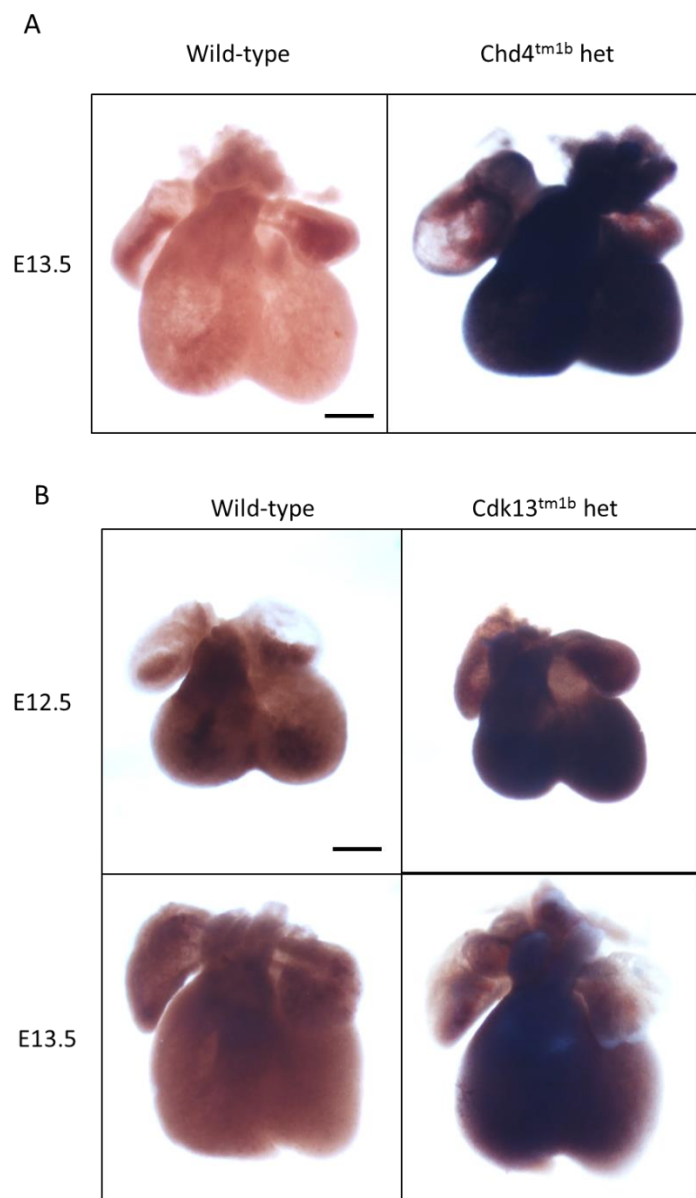


Figure 5.8. X-Gal staining of *Chd4*^{tm1b/+} and *Cdk13*^{tm1b/+} hearts. (A) Representative images of the *Chd4* expression in the heart at E13.5. Scale bar = 350 μ m (B) Representative images of the *Cdk13* expression in the heart at E12.5 and E13.5). Scale bar = 350 μ m. Wild-type hearts were used as negative controls.

5.2.4. HREM

HREM analysis on E15.5 hearts was performed in order to generate 3D models which would allow us to study, in greater detail, the cardiac morphology and anatomy of our transgenic lines. Due to time constraints, I was only able to participate in the generation of 3D reconstructions of WT hearts. Representative pictures of a WT reconstruction are shown in Fig. 5.9, showcasing different cross-sections of the heart (a short movie of this reconstruction will be shown in the viva process). In these images, several cardiac structures can be identified: right and left ventricle, right and left atrium, the ventricular septum, and in some of them, the outflow vessels (aorta and pulmonary trunk) as well as the tricuspid and mitral valves. I participated in sample collection, embedding and processing; whereas Anna Wilsdon, Siobhan Loughna and Qazi Waheed performed the software analysis and 3D reconstruction.

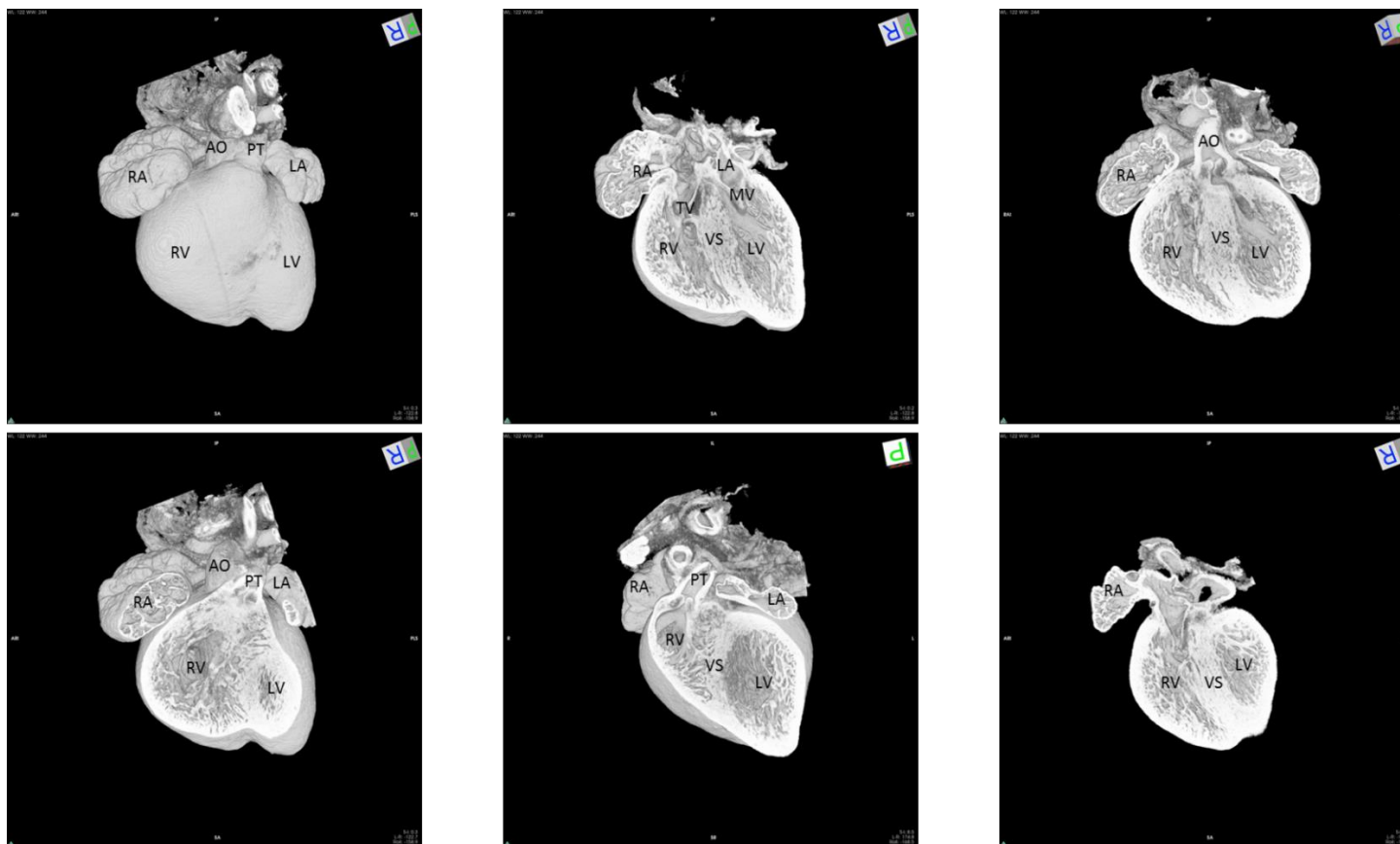


Figure 5.9. Representative HREM images of a E15.5 WT mouse heart. Several cross-sections of the heart are shown, displaying different cardiac structures such as the ventricles (RV and LV), atria (RA and LA), outflow vessels (aorta, AO, and pulmonary trunk, PT), ventricular septum (VS), and the mitral (MV) and tricuspid valves (TV). HREM sample processing was performed by myself, whereas imaging and software analysis was performed by Anna Wilsdon, Siobhan Loughna and Qazi Waheed (images provided by him and shown here with his permission).

5.3. Conclusions, discussion and perspectives for future work

The results shown in this part of my thesis constitute the preliminary experiments of a long-term project, which will be performed in our lab in collaboration with other research groups across the UK. Due to time limitations, I was not able to perform a detailed HREM analysis of the *Chd4^{tm1b}* and *Cdk13^{tm1b}* lines; however, the preliminary results obtained here give us valuable insight into them. Our main goal in this project was to establish these two lines at The University of Nottingham and to expand the colonies for future experiments. To do so, three WT x HET breeding pairs were set up for each of the lines and pups were genotyped. Genotypic distribution obtained was similar or extremely close to the Mendelian ratio, thus suggesting that mice carrying heterozygous loss-of-function mutations in the *Chd4* and *Cdk13* genes are as viable as their WT counterparts. A previously reported *Chd4* loss-of-function model also revealed that, in line with our results, heterozygous mice for a *Chd4* deletion allele were viable and fertile⁴¹³, thus suggesting that a single copy of *Chd4* is enough for viability and fertility. In the case of *Cdk13*, there are no other loss-of-function models reported in the literature; however, heterozygous mice for a *Cdk12* loss-of-function allele are also completely viable and fertile⁴⁵¹.

Even though heterozygous mice for both the *Chd4^{tm1b}* and *Cdk13^{tm1b}* lines are completely viable, according to information provided by the Sanger Institute and the International Mouse Phenotyping Consortium complete ablation of either of these genes has a strong developmental effect and they die during development (personal communication). To decipher at what stage of development these embryos die, a preliminary lethality assessment was performed using embryos collected from HET x HET crosses. Even though the number of embryos analysed was relatively low and further analysis is required, no homozygous *Chd4^{tm1b}* embryos were found as early as E10.5, and chi-square test confirmed that the genotype frequency obtained during development was significantly different to the expected one, thus confirming that homozygous *Chd4^{tm1b}* embryos die during development. These results suggest that *Chd4* play an essential role during early embryogenesis and they are in line with previous reports indicating that mice lacking *Chd4* are unable to go past the blastocyst stage⁴¹³. A quick phenotypic analysis of *Chd4^{tm1b/+}* embryos also revealed that most of them were indistinguishable from their WT counterparts; however, we found that some of these embryos were slightly smaller than *Chd4^{+/+}* embryos, even though their hearts were pretty much indistinguishable from WT hearts. Whether this smaller size is due to the lack

of *Chd4* or due to other unrelated causes could not be discerned due to the small sample size.

In the case of the *Cdk13^{tm1b}* line, 5 to 10% of the embryos analysed at E11.5, E12.5 and E15.5 were confirmed homozygotes. However, as with *Chd4^{tm1b}*, chi-square analysis of the observed genotype frequencies revealed a significant difference (in comparison to the expected ones) when all embryos analysed during development are considered, thus confirming that homozygous *Cdk13^{tm1b}* embryos also die during development and are not as viable as their heterozygous counterparts. We also found a few heterozygous embryos for both lines that presented a smaller size and/or slightly delayed development than their WT counterparts. This reduced size and developmental delay was much more noticeable in *Cdk13^{tm1b/+}* embryos than in *Chd4^{tm1b/+}* embryos, something that could indicate that the loss of one copy of *Cdk13* is more detrimental to normal development, although further sample analysis is required in order to confirm this. Also, a higher number of embryos is necessary in order to decipher at which precise stage during development homozygous embryos from both lines die.

My preliminary spatiotemporal expression analysis of these two genes showed their expression in the heart at different developmental stages, ranging from E12.5 to E14.5. During this time window, key events in murine heart morphogenesis are taking place: septation of the atria, ventricles and outflow tract have initiated by E11.5, and all are complete by E14.5. In addition to this, the valvular structures significantly mature during this period as well¹¹. In order to perform this analysis, I carried out two different techniques: X-Gal staining and western blotting. X-Gal staining was chosen because both of our transgenic lines (*Chd4^{tm1b}* and *Cdk13^{tm1b}*) carry a LacZ knock-in allele which can be used to study the spatial and temporal expression pattern of these two genes at a mRNA level. In the case of western blotting, I decided to carry out this technique as it is one of the most straightforward techniques in order to study protein expression levels, it allows the study of expression at a protein level and our lab have a great expertise on it. In spite of this, there are other techniques which would have been easier to perform and/or provide more useful information about the expression pattern of *Chd4* and *Cdk13*. For example, a quick and easy-to-perform technique to assess this is quantitative PCR, which would have provided information about the mRNA expression pattern in an easy way. In order to gain more detailed information about the spatial expression pattern of these two genes, *in situ* hybridization on cardiac tissue sections could have been performed as well. Albeit more technically-challenging, this technique would have provided information about the cardiac structures or cell populations in which these two genes are mainly expressed

during cardiogenesis. When normalized and compared to each other, these preliminary results suggest that the expression of *Chd4* is higher than that of *Cdk13* at the same developmental stages. In spite of this, further western blot replicates need to be performed to confirm these results and to obtain “cleaner” images, as the blots showed in this thesis indicate the presence of unspecific bands in *Chd4* and *GAPDH* blots.

In order to study in greater detail the cardiac phenotypes of the *Chd4*^{tm1b} and *Cdk13*^{tm1b} during development, we aimed to set up in our laboratory the HREM technique, which allows the generation of high-resolution, three-dimensional computer models of embryos and histological samples. After several months of training and equipment preparation, we successfully generated a 3D reconstruction of a E15.5 WT mouse heart, which serves as a proof-of-concept for the use of this new imaging technique in the visualization of cardiac development and characterization of heart morphology. Currently, HREM analysis of the *Chd4*^{tm1b} and *Cdk13*^{tm1b} transgenic lines is taking place in our lab and we expect to have the first results in the near future.

As mentioned above, the results presented in this chapter constitute the basis of a long-term project that will be carried out over the next years. In the near future, the main goal is to continue our HREM analysis in order to examine the diversity of heart morphology produced by the depletion of the *Chd4* and *Cdk13* genes in mice. In our lab we are currently establishing transgenic lines carrying point, humanized mutations found in syndromic-CHD patients in those genes. After characterizing the cardiac morphology of these lines, we will also study the effect of maternal diabetes and obesity on heart development by using these lines, something that will allow us to develop a predictive model for CHD. Last but not least, one of our goals is to understand how abnormal cardiac development is related to gene expression changes. To do this, we will conduct spatial transcriptomics analysis on cardiac tissue sections coming from our transgenic mice lines. All these studies will allow us to build a detailed CHD road-map and provide a CHD prediction model, which takes into account both genetic and environmental risk factors.

6. SUMMARY OF FINDINGS AND IMPLICATIONS OF THIS RESEARCH

Congenital heart defects represent the most common form of congenital abnormality. These developmental disorders affect at least 4000 new-borns each year in the UK, and present an incidence of approximately one in every 100 babies born^{46,64}. Some of the main causes of CHD are genetic mutations that arise *de novo* during development or are passed down from parents to the offspring⁹². Over the years, our lab and many others around the globe have contributed significantly to our understanding of the genes involved in cardiac development and in the onset of CHD; however, we are still far from understanding the whole picture, and an even greater, global effort is needed in order to build a precise and detailed road-map for heart development and CHD. The main goal of this PhD thesis was to contribute to this global effort by discovering and characterizing new genes and factors involved in cardiac development and CHD, and the findings and results presented here do expand our knowledge on this topic.

In the first part of this thesis, two new factors involved in cardiac and limb development were discovered: lysine acetyltransferases KAT2A and KAT2B. Our results indicate that these two acetylases play a key role in these developmental processes by potentiating the transcriptional activity of TBX5, one of the master regulators of cardiogenesis. By performing *in vitro* assays, we discovered a novel TBX5 activity-regulation-mechanism mediated by KAT2A and KAT2B, which potentiate TBX5 transcriptional activity by promoting its nuclear retention. The fact that TBX5 is susceptible to acetylation/deacetylation has already been described previously³⁴⁹; however, to our knowledge this is the first study establishing a link between KAT2A/KAT2B and any cardiac transcription factor, as well as the first one providing mechanistic details about how these acetylases regulate cardiac transcription factor activity. In addition to the *in vitro* studies, we also studied the *in vivo* link between these factors by abolishing the *kat2a* and *kat2b* activity in zebrafish using three different approaches: CRISPR/Cas9-mediated KO, morpholino-mediated KD and chemical inhibition. All these approaches revealed that zebrafish embryos lacking *kat2a* or *kat2b* show cardiac and limb defects similar to those found in *tbx5a* mutants, including relaxation of cardiac looping, bradycardia or presence of stumpy fins among others. Our experiments strongly suggest that these developmental defects are indeed due to the lack of *tbx5a* acetylation. On top of these findings, the generation of these *kat2a* and *kat2b* crispants and morphants share insight into the role of *kat2a* and *kat2b* in zebrafish, as previous studies have only addressed their function in

mouse^{355,359}. All in all, these studies unveil two new factors which play a key role in cardiac and limb development through TBX5 activity regulation, and open up a range of possibilities for future research in upstream regulators of cardiac transcription factors.

In the second part of this thesis, I carried out an *in vitro* characterization of five CHD4 mutant variants found in S-CHD patients. These *in vitro* assays revealed that four of these mutant variants have a decreased ATPase activity, whereas one of them present a significantly increased activity. No defects were found in their cellular localization and DNA-binding ability, and whereas most of them showed a normal interaction with several NuRD complex subunits, one of them (M1 variant) presented a significantly weaker interaction with RBBP4, one of the main NuRD complex subunits. These results strongly suggest that the defects found in the patients carrying these mutations are due to an impaired CHD4 helicase activity, which potentially lead to incorrect chromatin remodelling. To our knowledge, this is one of the first *in vitro* characterization studies of patient-derived CHD4 mutations, alongside with the one recently published by Weiss *et al*⁴²⁵, which also suggest an impairment in helicase activity as the main cause of an intellectual disability syndrome found in four individuals. These findings provide insight into the molecular role of CHD4 during development, help us understand the molecular basis of the defects found in S-CHD patients and pave the way for even further characterization of these and related mutations.

In the final part of this thesis we established two new transgenic mouse lines at the University of Nottingham: *Chd4*^{tm1b} and *Cdk13*^{tm1b}. These two lines carry loss-of-function alleles for the *Chd4* and *Cdk13* genes respectively, two of the genes whose mutations have been found in S-CHD patients. These preliminary studies revealed that heterozygous embryos for both of these lines are completely viable and fertile, and show no apparent defects in comparison to WT counterparts. *Chd4*^{tm1b} and *Cdk13*^{tm1b} homozygous embryos are embryonically lethal, and our lethality assessments suggest that they die at early stages during development. We also confirmed the expression of both *Chd4* and *Cdk13* in the heart at different stages of development.

Even though I have worked with apparently unrelated genes in this thesis, pathway analysis using STRING (<https://string-db.org/>) shows that indeed, most of these genes are part of a cardiac development gene network, which include many of the key regulators of this process, such as TBX5, GATA4 and MEF2C among others. Fig. 6.1 shows the interaction maps generated when just the genes studied in this project are included in the input (Fig. 6.1A), and that generated when some of the key regulators of cardiac development are

included as well (Fig. 6.1B). As shown in the first map, at first instance no direct relationship is found between most of the genes studied in this project, as interactions are found just between TBX5 and KAT2B, and KAT2B and KAT2A. However, when several master regulators of cardiac development are included in the analysis, most of them (except CDK13) are shown to be directly or indirectly linked to any of the cardiac regulators. For instance, KAT2B, besides interacting with TBX5, also interacts with SALL4, MEF2C and NKX2-5 among others, which are transcription factors involved in cardiogenesis. KAT2B, alongside KAT2A, also interact with CHD7, a paralog gene to CHD4 whose mutations lead to CHARGE syndrome. HDAC4 and HDAC5, two deacetylases which act in an opposite way to KAT2A and KAT2B and which deacetylate TBX5 *in vitro* (Dr. Tushar Ghosh, unpublished data) were also included in the analysis. The map shows that these two deacetylases are linked to KAT2A and KAT2B, as expected, as well as PRKD1, CHD4 and several cardiac transcription factors such as SALL4, GATA4 and MEF2C. Last but not least, CDK13 seems to be not connected or linked to any of the other genes included in the analysis, something that could be explained by the fact that this gene is still poorly characterized. This pathway analysis suggest that most of the genes which have been studied in this thesis are indeed part, in a direct or indirect way, of the genetic network that controls cardiac development.

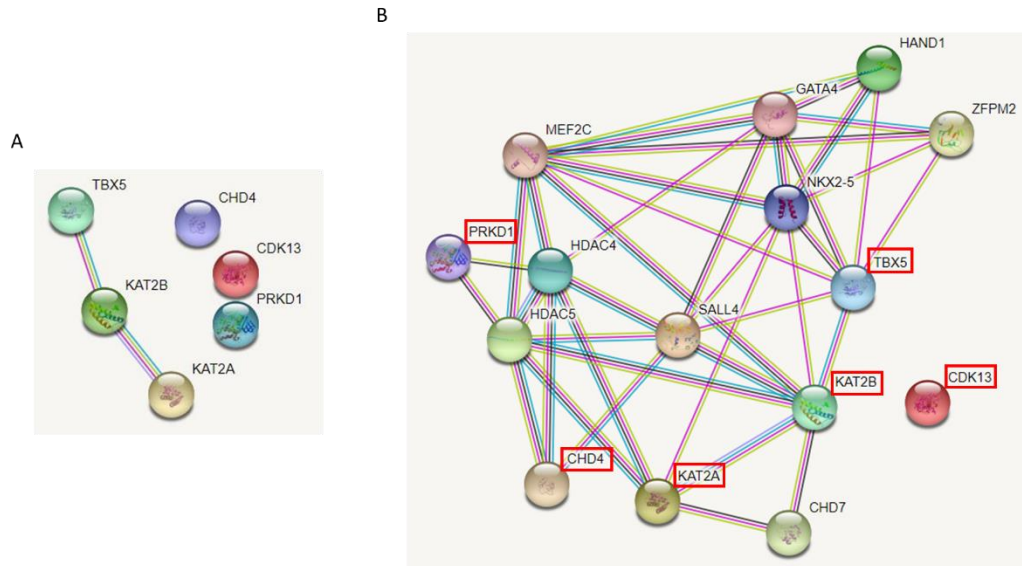


Figure 6.1. Interaction maps for TBX5, KAT2A, KAT2B, CHD4, CDK13, PRKD1 and several cardiac transcription factor factors.

Even though this thesis is comprised of three different and, at first glance, unrelated projects, they all greatly contribute to our understanding of the genetic causes of congenital heart disorders and provide further insight into the molecular mechanisms controlling cardiac development. Perhaps, the most relevant finding of the KAT2A/2B project is that post-translational modifications of a classic gene involved in cardiogenesis and in the onset of CHD (such as TBX5) do indeed play a crucial role in the correct regulation of these processes, thus providing new insights and potential new mechanisms to explain the molecular and cellular basis of congenital heart disorders. During the last few years, and thanks to the advances in the field of proteomics, several publications have provided further insight into the relationship between post-translational modifications, cardiogenesis and heart disease^{467–469}, thus supporting the fact that these PTMs play a key role during cardiogenesis.

A different field that is gaining notoriety in the context of congenital heart disorders and cardiogenesis is chromatin remodelling and epigenetics. Over the last few years, many different publications have postulated that the dynamic changes on chromatin modifications and structure might have a strong impact on cell fate during cardiogenesis, thus having a role in the occurrence of CHD^{470–472}. In addition to this, mutations in genes involved in chromatin remodelling (such as CHD3⁴⁷³, CHD4⁴⁷⁴ and CHD7⁴⁷⁵) have been described in patients presenting syndromic CHD conditions. In this same context, the work presented here provides further evidence that chromatin remodelling do play a key role in the control of cardiac development, as CHD4 mutations found in syndromic-CHD patients lead to an impaired ATPase/helicase activity of the protein that would have an impact of its chromatin remodelling ability.

All in all, these results support that not only mutations in “classic” cardiac transcription factors (such as TBX5, MEF2C or NKX2.5) seem to be the main genetic cause for congenital heart disorders, but also mutations in genes involved in chromatin remodelling and post-translational modifications (two processes which are also closely related, as demonstrated by the fact that histone tails modifications control chromatin structure and function). The increasing relevance and importance of these two processes provide a new outlook for future studies on cardiogenesis and CHDs, and broaden our scope on this field: for example, it would be of great interest to perform proteomic studies on samples coming from CHD patients in order to unravel whether the PTM-signature of these patients differs from that of healthy patients. Also, we need to be on the lookout for new chromatin remodelling genes involved in congenital heart disorders, as well as to perform an exhaustive molecular

characterization of them in order to better understand how they can lead to the onset of these defects.

To sum up, by discovering and characterizing these novel factors implicated in these processes, we have managed to expand the network of CHD and cardiogenesis-related genes, opening up new possibilities for future research in cardiac transcriptional networks.

7. REFERENCES

1. Wittig, J. & Münsterberg, A. The Early Stages of Heart Development: Insights from Chicken Embryos. *J. Cardiovasc. Dev. Dis.* (2016). doi:10.3390/jcdd3020012
2. Bruneau, B. G. *et al.* Signaling and Transcriptional Networks in Heart Development and Regeneration. *Cold Spring Harb. Lab. Perspect. Biol.* **5**, a008292 (2013).
3. Chaudhry, B., Ramsbottom, S. & Henderson, D. J. in *Progress in Molecular Biology and Translational Science* (2014). doi:10.1016/B978-0-12-386930-2.00002-1
4. Bruneau, B. G. The developmental genetics of congenital heart disease. *Nature* (2008). doi:10.1038/nature06801
5. Buckberg, G., Nanda, N., Nguyen, C. & Kocica, M. What Is the Heart? Anatomy, Function, Pathophysiology, and Misconceptions. *J. Cardiovasc. Dev. Dis.* (2018). doi:10.3390/jcdd5020033
6. Stainier, D. Y. Zebrafish genetics and vertebrate heart formation. *Nat. Rev. Genet.* **2**, 39–48 (2001).
7. Bruneau, B. G. The developmental genetics of congenital heart disease. *Nature* **451**, 943–948 (2008).
8. Desgrange, A., Garrec, J. F. Le & Meilhac, S. M. Left-right asymmetry in heart development and disease: Forming the right loop. *Development (Cambridge)* (2018). doi:10.1242/dev.162776
9. Lin, C. J., Lin, C. Y., Chen, C. H., Zhou, B. & Chang, C. P. Partitioning the heart: Mechanisms of cardiac septation and valve development. *Development (Cambridge)* (2012). doi:10.1242/dev.063495
10. Sylva, M., Van den Hoff, M. J. B. & Moorman, A. F. M. Development of the human heart. *Am. J. Med. Genet. Part A* (2014). doi:10.1002/ajmg.a.35896
11. Savolainen, S. M., Foley, J. F. & Elmore, S. A. Histology atlas of the developing mouse heart with emphasis on E11.5 to E18.5. *Toxicologic Pathology* (2009). doi:10.1177/0192623309335060
12. Al Naieb, S., Happel, C. M. & Yelbuz, T. M. A detailed atlas of chick heart development in vivo. *Ann. Anat.* (2013). doi:10.1016/j.aanat.2012.10.011
13. Kuhn, E. N. & Wu, S. M. Origin of cardiac progenitor cells in the developing and postnatal heart. *Journal of Cellular Physiology* (2010). doi:10.1002/jcp.22281

14. Solnica-Krezel, L. & Sepich, D. S. Gastrulation: Making and Shaping Germ Layers. *Annu. Rev. Cell Dev. Biol.* (2012). doi:10.1146/annurev-cellbio-092910-154043
15. Williams, M. L. & Solnica-Krezel, L. Regulation of gastrulation movements by emergent cell and tissue interactions. *Current Opinion in Cell Biology* (2017). doi:10.1016/j.ceb.2017.04.006
16. Sakabe, M. *et al.* Understanding heart development and congenital heart defects through developmental biology: A segmental approach. *Congenital Anomalies* (2005). doi:10.1111/j.1741-4520.2005.00079.x
17. Kelly, R. G., Buckingham, M. E. & Moorman, A. F. Heart fields and cardiac morphogenesis. *Cold Spring Harb. Perspect. Med.* (2014). doi:10.1101/cshperspect.a015750
18. Tam, P. P. & Zhou, S. X. The allocation of epiblast cells to ectodermal and germ-line lineages is influenced by the position of the cells in the gastrulating mouse embryo. *Dev. Biol.* **178**, 124–132 (1996).
19. Klaus, a. *et al.* Wnt/ β -catenin and Bmp signals control distinct sets of transcription factors in cardiac progenitor cells. *Proc. Natl. Acad. Sci.* **109**, 10921–10926 (2012).
20. Meilhac, S. M. & Buckingham, M. E. The deployment of cell lineages that form the mammalian heart. *Nature Reviews Cardiology* (2018). doi:10.1038/s41569-018-0086-9
21. Harvey, R. P. Patterning the vertebrate heart. *Nat. Rev. Genet.* **3**, 544–556 (2002).
22. Meilhac, S. M., Esner, M., Kelly, R. G., Nicolas, J. F. & Buckingham, M. E. The clonal origin of myocardial cells in different regions of the embryonic mouse heart. *Dev. Cell* (2004). doi:10.1016/S1534-5807(04)00133-9
23. Moorman et al. Development of the heart (1) formation of the cardiac chambers and arterial trunks. *Heart* **89**, 806–14 (2003).
24. Xin, M., Olson, E. N. & Bassel-Duby, R. Mending broken hearts: cardiac development as a basis for adult heart regeneration and repair. *Nat. Publ. Gr.* **14**, 529–541 (2013).
25. Capdevila, J., Vogan, K. J., Tabin, C. J. & Izpisua Belmonte, J. C. Mechanisms of left-right determination in vertebrates. *Cell* (2000). doi:10.1016/S0092-8674(00)80619-4
26. Nonaka, S., Shiratori, H., Saijoh, Y. & Hamada, H. Determination of left-right patterning of the mouse embryo by artificial nodal flow. *Nature* **418**, 96–99 (2002).

27. Kelly, R. G. in *Current Topics in Developmental Biology* (2012).
doi:10.1016/B978-0-12-387786-4.00002-6
28. Anderson, R. H. Development of the heart: (2) Septation of the atriums and ventricles. *Heart* (2003). doi:10.1136/heart.89.8.949
29. Armstrong, E. J. & Bischoff, J. Heart valve development: Endothelial cell signaling and differentiation. *Circulation Research* (2004).
doi:10.1161/01.RES.0000141146.95728.da
30. Markwald, R. R., Fitzharris, T. P. & Manasek, F. J. Structural development of endocardial cushions. *Am. J. Anat.* (1977).
doi:10.1002/aja.1001480108
31. Rivera-Feliciano, J. & Tabin, C. J. Bmp2 instructs cardiac progenitors to form the heart-valve-inducing field. *Dev. Biol.* (2006).
doi:10.1016/j.ydbio.2006.03.043
32. Sugi, Y., Yamamura, H., Okagawa, H. & Markwald, R. R. Bone morphogenetic protein-2 can mediate myocardial regulation of atrioventricular cushion mesenchymal cell formation in mice. *Dev. Biol.* (2004). doi:10.1016/j.ydbio.2004.01.045
33. Miquerol, L. *et al.* Embryonic development is disrupted by modest increases in vascular endothelial growth factor gene expression. *Development* (2000). doi:10.1109/PVSC.2010.5614624
34. Timmerman, L. A. *et al.* Notch promotes epithelial-mesenchymal transition during cardiac development and oncogenic transformation. *Genes Dev.* (2004). doi:10.1101/gad.276304
35. Liebner, S. *et al.* β -catenin is required for endothelial-mesenchymal transformation during heart cushion development in the mouse. *J. Cell Biol.* (2004). doi:10.1083/jcb.200403050
36. Combs, M. D. & Yutzey, K. E. Heart valve development: Regulatory networks in development and disease. *Circulation Research* (2009).
doi:10.1161/CIRCRESAHA.109.201566
37. Kirby, M. L., Gale, T. F. & Stewart, D. E. Neural crest cells contribute to normal aorticopulmonary septation. *Science* (80-.). (1983).
doi:10.1126/science.6844926
38. England, J. & Loughna, S. in *eLS* 1–11 (John Wiley & Sons, Ltd, 2018).
doi:10.1002/9780470015902.a0027698
39. de Vlaming, A. *et al.* Atrioventricular valve development: New perspectives on an old theme. *Differentiation* (2012).
doi:10.1016/j.diff.2012.04.001
40. Millington-Sanders, C., Meir, A., Lawrence, L. & Stolinski, C. Structure of

chordae tendineae in the left ventricle of the human heart. *J. Anat.* (1998). doi:10.1017/S0021878298003628

41. Jiang, X., Rowitch, D. H., Soriano, P., McMahon, A. P. & Sucov, H. M. Fate of the mammalian cardiac neural crest. *Development* (2000).
42. Anderson, R. H., Webb, S., Brown, N. A., Lamers, W. & Moorman, A. Development of the heart: (3) Formation of the ventricular outflow tracts, arterial valves, and intrapericardial arterial trunks. *Heart* (2003). doi:10.1136/heart.89.9.1110
43. Jain, R. *et al.* Cardiac neural crest orchestrates remodeling and functional maturation of mouse semilunar valves. *J. Clin. Invest.* (2011). doi:10.1172/JCI44244
44. Lamers, W. H. & Moorman, A. F. M. Cardiac septation: A late contribution of the embryonic primary myocardium to heart morphogenesis. *Circulation Research* **91**, 93–103 (2002).
45. Hoffman, J. I. . & Kaplan, S. The incidence of congenital heart disease. *J. Am. Coll. Cardiol.* **39**, 1890–1900 (2002).
46. Hoffman, J. I. E. Incidence of Congenital Heart Disease: II. Prenatal Incidence. *Pediatr. Cardiol.* **16**, 155–165 (1995).
47. Meberg, a, Otterstad, J. E., Frøland, G., Lindberg, H. & Sørland, S. J. Outcome of congenital heart defects—a population-based study. *Acta Paediatr.* **89**, 1344–51 (2000).
48. Desai, K., Rabinowitz, E. J. & Epstein, S. Physiologic diagnosis of congenital heart disease in cyanotic neonates. *Curr. Opin. Pediatr.* (2019). doi:10.1097/MOP.0000000000000742
49. Apitz, C., Webb, G. D. & Redington, A. N. [Review] Tetralogy of Fallot. *Lancet* (2009). doi:10.1016/S0140-6736(09)60657-7
50. Kelly, R. G. in *Congenital Heart Diseases: The Broken Heart: Clinical Features, Human Genetics and Molecular Pathways* (2015). doi:10.1007/978-3-7091-1883-2_33
51. Rana, M. S. *et al.* Tbx1 coordinates addition of posterior second heart field progenitor cells to the arterial and venous poles of the heart. *Circ. Res.* (2014). doi:10.1161/CIRCRESAHA.115.305020
52. Di Felice, V. & Zummo, G. Tetralogy of Fallot as a Model to Study Cardiac Progenitor Cell Migration and Differentiation During Heart Development. *Trends in Cardiovascular Medicine* (2009). doi:10.1016/j.tcm.2009.07.004
53. Théveniau-Ruissy, M. *et al.* The del22q11.2 candidate gene tbx1 controls regional outflow tract identity and coronary artery patterning.

Circ. Res. (2008). doi:10.1161/CIRCRESAHA.108.172189

54. Zhang, Z. & Baldini, A. In vivo response to high-resolution variation of Tbx1 mRNA dosage. *Hum. Mol. Genet.* (2008). doi:10.1093/hmg/ddm291
55. Targoff, K. L., Schell, T. & Yelon, D. Nkx genes regulate heart tube extension and exert differential effects on ventricular and atrial cell number. *Dev. Biol.* (2008). doi:10.1016/j.ydbio.2008.07.037
56. Prall, O. W. J. *et al.* An Nkx2-5/Bmp2/Smad1 Negative Feedback Loop Controls Heart Progenitor Specification and Proliferation. *Cell* (2007). doi:10.1016/j.cell.2007.01.042
57. Goldmuntz, E., Geiger, E. & Benson, D. W. NKX2.5 mutations in patients with tetralogy of fallot. *Circulation* (2001). doi:10.1161/hc4601.098427
58. Itoh, N., Ohta, H., Nakayama, Y. & Konishi, M. Roles of FGF signals in heart development, health, and disease. *Frontiers in Cell and Developmental Biology* (2016). doi:10.3389/fcell.2016.00110
59. Marguerie, A. *et al.* Congenital heart defects in Fgfr2-IIIb and Fgf10 mutant mice. *Cardiovasc. Res.* (2006). doi:10.1016/j.cardiores.2006.03.021
60. Vincentz, J. W., McWhirter, J. R., Murre, C., Baldini, A. & Furuta, Y. Fgf15 is required for proper morphogenesis of the mouse cardiac outflow tract. *Genesis* (2005). doi:10.1002/gene.20114
61. Donovan, J., Kordylewska, A., Jan, Y. N. & Utset, M. F. Tetralogy of fallot and other congenital heart defects in Hey2 mutant mice. *Curr. Biol.* (2002). doi:10.1016/S0960-9822(02)01149-1
62. McCright, B., Lozier, J. & Gridley, T. A mouse model of Alagille syndrome: Notch2 as a genetic modifier of Jag1 haploinsufficiency. *Development* (2002). doi:10.1093/hmg/8.5.723
63. Conway, S. J., Henderson, D. J., Kirby, M. L., Anderson, R. H. & Copp, A. J. Development of a lethal congenital heart defect in the splotch (Pax3) mutant mouse. *Cardiovasc. Res.* (1997). doi:10.1016/S0008-6363(97)00172-7
64. Hoffman, J. I. E. & Kaplan, S. The incidence of congenital heart disease. *Journal of the American College of Cardiology* (2002). doi:10.1016/S0735-1097(02)01886-7
65. Khan, R. & Jay, P. Y. in *Congenital Heart Diseases: The Broken Heart: Clinical Features, Human Genetics and Molecular Pathways* (2015). doi:10.1007/978-3-7091-1883-2_20
66. Posch, M. G., Perrot, A., Berger, F. & Özcelik, C. Molecular genetics of

congenital atrial septal defects. *Clinical Research in Cardiology* (2010). doi:10.1007/s00392-009-0095-0

67. Bruneau, B. G. *et al.* A murine model of Holt-Oram syndrome defines roles of the T-Box transcription factor Tbx5 in cardiogenesis and disease. *Cell* **106**, 709–721 (2001).
68. Rajagopal, S. K. *et al.* Spectrum of heart disease associated with murine and human GATA4 mutation. *J. Mol. Cell. Cardiol.* **43**, 677–685 (2007).
69. Hirayama-Yamada, K. *et al.* Phenotypes with GATA4 or NKX2.5 mutations in familial atrial septal defect. *Am. J. Med. Genet.* **135 A**, 47–52 (2005).
70. Olson, E. N. Gene regulatory networks in the evolution and development of the heart. *Science* **313**, 1922–7 (2006).
71. Nadeau, M. *et al.* An endocardial pathway involving tbx5, gata4, and nos3 required for atrial septum formation. *Proc. Natl. Acad. Sci. U. S. A.* (2010). doi:10.1073/pnas.0914888107
72. Nakashima, Y. *et al.* Nkx2-5 suppresses the proliferation of atrial myocytes and conduction system. *Circ. Res.* (2014). doi:10.1161/CIRCRESAHA.114.303219
73. Ching, Y. H. *et al.* Mutation in myosin heavy chain 6 causes atrial septal defect. *Nat. Genet.* (2005). doi:10.1038/ng1526
74. Matsson, H. *et al.* Alpha-cardiac actin mutations produce atrial septal defects. *Hum. Mol. Genet.* **17**, 256–265 (2008).
75. Frank, D. *et al.* Cardiac α -Actin (ACTC1) Gene Mutation Causes Atrial-Septal Defects Associated With Late-Onset Dilated Cardiomyopathy. *Circ. Genomic Precis. Med.* (2019). doi:10.1161/CIRCGEN.119.002491
76. Houyel, L. in *Congenital Heart Diseases: The Broken Heart: Clinical Features, Human Genetics and Molecular Pathways* (2015). doi:10.1007/978-3-7091-1883-2_24
77. Li, Q. Y. *et al.* Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family. *Nat. Genet.* **15**, 21–29 (1997).
78. Garg, V. *et al.* GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature* **424**, 443–7 (2003).
79. Zhou, L. *et al.* Gata4 potentiates second heart field proliferation and Hedgehog signaling for cardiac septation. *Proc. Natl. Acad. Sci. U. S. A.* (2017). doi:10.1073/pnas.1605137114
80. Panzer, A. A. *et al.* Nkx2-5 and Sarcospan genetically interact in the

development of the muscular ventricular septum of the heart. *Sci. Rep.* (2017). doi:10.1038/srep46438

81. Schott, J. J. *et al.* Congenital heart disease caused by mutations in the transcription factor NKX2-5. *Science* (80-.). (1998). doi:10.1126/science.281.5373.108
82. Franco, D. *et al.* Left and right ventricular contributions to the formation of the interventricular septum in the mouse heart. *Dev. Biol.* (2006). doi:10.1016/j.ydbio.2006.02.045
83. Togi, K. *et al.* Role of Hand1/eHAND in the Dorso-Ventral Patterning and Interventricular Septum Formation in the Embryonic Heart. *Mol. Cell. Biol.* (2004). doi:10.1128/mcb.24.11.4627-4635.2004
84. Barron, D. J. *et al.* Hypoplastic left heart syndrome. *The Lancet* (2009). doi:10.1016/S0140-6736(09)60563-8
85. Carabello, B. A. & Paulus, W. J. Aortic stenosis. *The Lancet* (2009). doi:10.1016/S0140-6736(09)60211-7
86. Chandrashekhar, Y., Westaby, S. & Narula, J. Mitral stenosis. *The Lancet* (2009). doi:10.1016/S0140-6736(09)60994-6
87. Siu, S. C. & Silversides, C. K. Bicuspid Aortic Valve Disease. *Journal of the American College of Cardiology* (2010). doi:10.1016/j.jacc.2009.12.068
88. Marino, B. S. *et al.* Quality-of-life concerns differ among patients, parents, and medical providers in children and adolescents with congenital and acquired heart disease. *Pediatrics* **123**, e708-15 (2009).
89. Oyama, K., El-Nachef, D. & Maclellan, W. R. Regeneration potential of adult cardiac myocytes. *Cell Res.* (2013). doi:10.1038/cr.2013.78
90. Miller, S. P. *et al.* Abnormal brain development in newborns with congenital heart disease. *N. Engl. J. Med.* **357**, 1928–1938 (2007).
91. Jenkins, K. J. *et al.* Noninherited risk factors and congenital cardiovascular defects: Current knowledge - A scientific statement from the American Heart Association Council on Cardiovascular Disease in the Young. *Circulation* **115**, 2995–3014 (2007).
92. Pierpont, M. E. *et al.* Genetic Basis for Congenital Heart Defects : Current Knowledge A Scientific Statement From the American Heart Association Congenital Cardiac Defects Committee , Council on Cardiovascular. *Circulation* **115**, 3015–3038 (2007).
93. Edwards, J. J. & Gelb, B. D. Genetics of congenital heart disease. *Current Opinion in Cardiology* (2016). doi:10.1097/HCO.0000000000000274

94. Basson, C. T. *et al.* Mutations in human TBX5 [corrected] cause limb and cardiac malformation in Holt-Oram syndrome. *Nat. Genet.* **15**, 30–5 (1997).
95. Schott, J. Congenital Heart Disease Caused by Mutations in the Transcription Factor NKX2-5. *Science (80-.)*. **281**, 108–111 (1998).
96. Kodo, K. *et al.* Genetic Analysis of Essential Cardiac Transcription Factors in 256 Patients With Non-Syndromic Congenital Heart Defects. *Circ. J.* **76**, (2012).
97. Ching, Y.-H. *et al.* Mutation in myosin heavy chain 6 causes atrial septal defect. *Nat. Genet.* **37**, 423–428 (2005).
98. Phillips, M. D., Mukhopadhyay, M., Poscablo, C. & Westphal, H. Dkk1 and Dkk2 regulate epicardial specification during mouse heart development. *Int. J. Cardiol.* **150**, 186–192 (2011).
99. Fouillade, C., Monet-Leprêtre, M., Baron-Menguy, C. & Joutel, A. Notch signalling in smooth muscle cells during development and disease. *Cardiovasc. Res.* **95**, 138–146 (2012).
100. Huang, Z. P. *et al.* Loss of MicroRNAs in neural crest leads to cardiovascular syndromes resembling human congenital heart defects. *Arterioscler. Thromb. Vasc. Biol.* **30**, 2575–2586 (2010).
101. Lázár, E., Sadek, H. A. & Bergmann, O. Cardiomyocyte renewal in the human heart: Insights from the fall-out. *Eur. Heart J.* (2017). doi:10.1093/eurheartj/ehx343
102. Warnes, C. A. The adult with congenital heart disease: Born to be bad? *Journal of the American College of Cardiology* **46**, 1–8 (2005).
103. Segers, V. F. & Lee, R. T. Stem-cell therapy for cardiac disease. *Nature* **451**, 937–942 (2008).
104. Wollert, K. C. & Drexler, H. Cell therapy for the treatment of coronary heart disease: a critical appraisal. *Nat. Rev. Cardiol.* **7**, 204–15 (2010).
105. Passier, R., van Laake, L. W. & Mummery, C. L. Stem-cell-based therapy and lessons from the heart. *Nature* **453**, 322–329 (2008).
106. Ptaszek, L. M., Mansour, M., Ruskin, J. N. & Chien, K. R. Towards regenerative therapy for cardiac disease. *Lancet* **379**, 933–42 (2012).
107. Ieda, M. *et al.* Direct Reprogramming of Fibroblasts into Functional Cardiomyocytes by Defined Factors. *Cell* **142**, 375–386 (2010).
108. Song, K. *et al.* Heart repair by reprogramming non-myocytes with cardiac transcription factors. *Nature* **485**, 599–604 (2012).

109. Greulich, F., Rudat, C. & Kispert, A. Mechanisms of T-box gene function in the developing heart. *Cardiovascular Research* **91**, 212–222 (2011).
110. Peterkin, T., Gibson, A., Loose, M. & Patient, R. The roles of GATA-4, -5 and -6 in vertebrate heart development. *Seminars in Cell and Developmental Biology* **16**, 83–94 (2005).
111. Black, B. L. & Cripps, R. M. in *Heart Development and Regeneration* 673–699 (2010). doi:10.1016/B978-0-12-381332-9.00030-X
112. Akazawa, H. & Komuro, I. Cardiac transcription factor Csx/Nkx2-5: Its role in cardiac development and diseases. *Pharmacology and Therapeutics* **107**, 252–268 (2005).
113. Vincentz, J. W., Barnes, R. M. & Firulli, A. B. Hand factors as regulators of cardiac morphogenesis and implications for congenital heart defects. *Birth Defects Research Part A - Clinical and Molecular Teratology* (2011). doi:10.1002/bdra.20796
114. Stennard, F. a & Harvey, R. P. T-box transcription factors and their roles in regulatory hierarchies in the developing heart. *Development* **132**, 4897–4910 (2005).
115. Scambler, P. J. 22q11 Deletion syndrome: A role for TBX1 in pharyngeal and cardiovascular development. in *Pediatric Cardiology* **31**, 378–390 (2010).
116. Christoffels, V. M. *et al.* Formation of the venous pole of the heart from an Nkx2-5-negative precursor population requires Tbx18. *Circ. Res.* **98**, 1555–1563 (2006).
117. Stennard, F. a *et al.* Murine T-box transcription factor Tbx20 acts as a repressor during heart development, and is essential for adult heart integrity, function and adaptation. *Development* **132**, 2451–2462 (2005).
118. Yagi, H. *et al.* Role of TBX1 in human del22q11.2 syndrome. *Lancet* **362**, 1366–1373 (2003).
119. Kirk, E. P. *et al.* Mutations in cardiac T-box factor gene TBX20 are associated with diverse cardiac pathologies, including defects of septation and valvulogenesis and cardiomyopathy. *Am. J. Hum. Genet.* **81**, 280–91 (2007).
120. Collavoli, A. *et al.* TBX5 nuclear localization is mediated by dual cooperative intramolecular signals. *J. Mol. Cell. Cardiol.* (2003). doi:10.1016/S0022-2828(03)00231-1
121. Kulisz, A. & Simon, H.-G. An Evolutionarily Conserved Nuclear Export Signal Facilitates Cytoplasmic Localization of the Tbx5 Transcription Factor. *Mol. Cell. Biol.* (2008). doi:10.1128/MCB.00935-07

122. Bruneau, B. G. *et al.* Chamber-specific cardiac expression of Tbx5 and heart defects in Holt- Oram syndrome. *Dev. Biol.* (1999). doi:10.1006/dbio.1999.9298
123. Chapman, D. L. *et al.* Expression of the T-box family genes, Tbx1-Tbx5, during early mouse development. *Dev. Dyn.* **206**, 379–390 (1996).
124. Gibson-Brown, J. J. *et al.* Evidence of a role for T-box genes in the evolution of limb morphogenesis and the specification of forelimb/hindlimb identity. *Mech. Dev.* (1996). doi:10.1016/0925-4773(96)00514-X
125. Begemann, G. & Ingham, P. W. Developmental regulation of Tbx5 in zebrafish embryogenesis. *Mech. Dev.* (2000). doi:10.1016/S0925-4773(99)00246-4
126. Garrity, D. M., Childs, S. & Fishman, M. C. The heartstrings mutation in zebrafish causes heart/fin Tbx5 deficiency syndrome. *Development* **129**, 4635–4645 (2002).
127. Vanlerberghe, C. *et al.* Holt-Oram syndrome: clinical and molecular description of 78 patients with TBX5 variants. *European Journal of Human Genetics* (2019). doi:10.1038/s41431-018-0303-3
128. Ghosh, T. *et al.* Studies of Genes Involved in Congenital Heart Disease. *J. Cardiovasc. Dev. Dis.* **1**, 134–145 (2014).
129. Ogura, T. in *Cardiovascular Development and Congenital Malformations: Molecular & Genetic Mechanisms* 75–77 (2007). doi:10.1002/9780470988664.ch18
130. Hiroi, Y. *et al.* Tbx5 associates with Nkx2-5 and synergistically promotes cardiomyocyte differentiation. *Nat. Genet.* **28**, 276–80 (2001).
131. Fan, C., Liu, M. & Wang, Q. Functional analysis of TBX5 missense mutations associated with Holt-Oram syndrome. *J. Biol. Chem.* **278**, 8780–8785 (2003).
132. Bodmer, R. The gene tinman is required for specification of the heart and visceral muscles in *Drosophila*. *Development* **118**, 719–729 (1993).
133. Lints, T. J., Parsons, L. M., Hartley, L., Lyons, I. & Harvey, R. P. Nkx-2.5: a novel murine homeobox gene expressed in early heart progenitor cells and their myogenic descendants. *Development* **119**, 419–431 (1993).
134. Kasahara, H., Bartunkova, S., Schinke, M., Tanaka, M. & Izumo, S. Cardiac and extracardiac expression of Csx/Nkx2.5 homeodomain protein. *Circ. Res.* **82**, 936–46 (1998).
135. Shiojima, I. *et al.* Context-dependent transcriptional cooperation mediated by cardiac transcription factors Csx/Nkx-2.5 and GATA-4. *J.*

Biol. Chem. **274**, 8231–8239 (1999).

136. Yamagishi, H. *et al.* The combinatorial activities of Nkx2.5 and dHAND are essential for cardiac ventricle formation. *Dev. Biol.* **239**, 190–203 (2001).
137. Barth, J. L. *et al.* Jarid2 is among a set of genes differentially regulated by Nkx2.5 during outflow tract morphogenesis. *Dev. Dyn.* **239**, 2024–2033 (2010).
138. Cleaver, O. B., Patterson, K. D. & Krieg, P. a. Overexpression of the tinman-related genes XNkx-2.5 and XNkx-2.3 in *Xenopus* embryos results in myocardial hyperplasia. *Development* **122**, 3549–56 (1996).
139. Chen, J. N. & Fishman, M. C. Zebrafish tinman homolog demarcates the heart field and initiates myocardial differentiation. *Development* **122**, 3809–3816 (1996).
140. Lyons, I. *et al.* Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5. *Genes Dev.* **9**, 1654–1666 (1995).
141. Jay, P. Y. *et al.* Nkx2-5 mutation causes anatomic hypoplasia of the cardiac conduction system. *J. Clin. Invest.* **113**, 1130–1137 (2004).
142. Lentjes, M. H. F. M. *et al.* The emerging role of GATA transcription factors in development and disease. *Expert Reviews in Molecular Medicine* (2016). doi:10.1017/erm.2016.2
143. Tremblay, M., Sanchez-Ferras, O. & Bouchard, M. Gata transcription factors in development and disease. *Dev.* (2018). doi:10.1242/dev.164384
144. Leonard, M. W., Lim, K. C. & Engel, J. D. Expression of the chicken GATA factor family during early erythroid development and differentiation. *Development* (1993).
145. Fujiwara, Y., Browne, C. P., Cunniff, K., Goff, S. C. & Orkin, S. H. Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc. Natl. Acad. Sci. U. S. A.* (1996). doi:10.1073/pnas.93.22.12355
146. Tsai, F. Y. *et al.* An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature* (1994). doi:10.1038/371221a0
147. Pandolfi, P. P. *et al.* Targeted disruption of the GATA3 gene causes severe abnormalities in the nervous system and in fetal liver haematopoiesis. *Nat. Genet.* (1995). doi:10.1038/ng0995-40
148. Brewer, A. & Pizzey, J. GATA factors in vertebrate heart development and disease. *Expert Rev. Mol. Med.* **8**, 1–20 (2006).

149. Pikkarainen, S., Tokola, H., Kerkelä, R. & Ruskoaho, H. GATA transcription factors in the developing and adult heart. *Cardiovascular Research* **63**, 196–207 (2004).
150. Pu, W. T., Ishiwata, T., Juraszek, A. L., Ma, Q. & Izumo, S. GATA4 is a dosage-sensitive regulator of cardiac morphogenesis. *Dev. Biol.* **275**, 235–244 (2004).
151. Rojas, A. *et al.* GATA4 is a direct transcriptional activator of cyclin D2 and Cdk4 and is required for cardiomyocyte proliferation in anterior heart field-derived myocardium. *Mol. Cell. Biol.* **28**, 5420–5431 (2008).
152. Searcy, R. D., Vincent, E. B., Liberatore, C. M. & Yutzey, K. E. A GATA-dependent nkx-2.5 regulatory element activates early cardiac gene expression in transgenic mice. *Development* **125**, 4461–4470 (1998).
153. Maitra, M. *et al.* Interaction of Gata4 and Gata6 with Tbx5 is critical for normal cardiac development. *Dev. Biol.* **326**, 368–377 (2009).
154. Morin, S., Charron, F., Robitaille, L. & Nemer, M. GATA-dependent recruitment of MEF2 proteins to target promoters. *EMBO J.* **19**, 2046–55 (2000).
155. Kuo, C. T. *et al.* GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. *Genes Dev.* **11**, 1048–1060 (1997).
156. Molkenstin, J. D., Lin, Q., Duncan, S. A. & Olson, E. N. Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. *Genes Dev.* **11**, 1061–1072 (1997).
157. Chen, Y. *et al.* A novel mutation in GATA4 gene associated with dominant inherited familial atrial septal defect. *J. Thorac. Cardiovasc. Surg.* **140**, 684–687 (2010).
158. Morrissey, E. E., Ip, H. S., Tang, Z., Lu, M. M. & Parmacek, M. S. GATA-5: A transcriptional activator expressed in a novel temporally and spatially-restricted pattern during embryonic development. *Dev. Biol.* (1997). doi:10.1006/dbio.1996.8485
159. Laforest, B., Andelfinger, G. & Nemer, M. Loss of Gata5 in mice leads to bicuspid aortic valve. *J. Clin. Invest.* (2011). doi:10.1172/JCI44555
160. Singh, M. K. *et al.* Gata4 and Gata5 cooperatively regulate cardiac myocyte proliferation in mice. *J. Biol. Chem.* (2010). doi:10.1074/jbc.M109.038539
161. Wei, D. *et al.* GATA5 loss-of-function mutation responsible for the congenital ventriculoseptal defect. *Pediatr. Cardiol.* (2013). doi:10.1007/s00246-012-0482-6

162. Shi, L. M. *et al.* GATA5 loss-of-function mutations associated with congenital bicuspid aortic valve. *Int. J. Mol. Med.* (2014). doi:10.3892/ijmm.2014.1700
163. Morrissey, E. E., Ip, H. S., Lu, M. M. & Parmacek, M. S. GATA-6: A zinc finger transcription factor that is expressed in multiple cell lineages derived from lateral mesoderm. *Dev. Biol.* (1996). doi:10.1006/dbio.1996.0165
164. Morrissey, E. E. *et al.* GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. *Genes Dev.* (1998). doi:10.1101/gad.12.22.3579
165. Lepore, J. J. *et al.* GATA-6 regulates semaphorin 3C and is required in cardiac neural crest for cardiovascular morphogenesis. *J. Clin. Invest.* (2006). doi:10.1172/JCI27363
166. Tian, Y. *et al.* Characterization and In Vivo Pharmacological Rescue of a Wnt2-Gata6 Pathway Required for Cardiac Inflow Tract Development. *Dev. Cell* (2010). doi:10.1016/j.devcel.2010.01.008
167. Kodo, K. *et al.* GATA6 mutations cause human cardiac outflow tract defects by disrupting semaphorin-plexin signaling. *Proc. Natl. Acad. Sci.* (2009). doi:10.1073/pnas.0904744106
168. Maitra, M., Koenig, S. N., Srivastava, D. & Garg, V. Identification of GATA6 sequence variants in patients with congenital heart defects. *Pediatr. Res.* (2010). doi:10.1203/PDR.0b013e3181ed17e4
169. Potthoff, M. J. & Olson, E. N. MEF2: a central regulator of diverse developmental programs. *Development* (2007). doi:10.1242/dev.008367
170. Edmondson, D. G., Lyons, G. E., Martin, J. F. & Olson, E. N. Mef2 gene expression marks the cardiac and skeletal muscle lineages during mouse embryogenesis. *Development* **120**, 1251–1263 (1994).
171. Black, B. L. & Olson, E. N. TRANSCRIPTIONAL CONTROL OF MUSCLE DEVELOPMENT BY MYOCYTE ENHANCER FACTOR-2 (MEF2) PROTEINS. *Annu. Rev. Cell Dev. Biol.* (1998). doi:10.1146/annurev.cellbio.14.1.167
172. Ghosh, T. K. *et al.* Physical Interaction between TBX5 and MEF2C Is Required for Early Heart Development. *Mol. Cell. Biol.* (2009). doi:10.1128/mcb.01923-08
173. Black, B. L., Molkentin, J. D. & Olson, E. N. Multiple roles for the MyoD basic region in transmission of transcriptional activation signals and interaction with MEF2. *Mol. Cell. Biol.* (1998). doi:10.1128/MCB.18.1.69
174. Desjardins, C. & Naya, F. The Function of the MEF2 Family of Transcription Factors in Cardiac Development, Cardiogenomics, and

Direct Reprogramming. *J. Cardiovasc. Dev. Dis.* (2016).
doi:10.3390/jcdd3030026

175. Naya, F. J. *et al.* Mitochondrial deficiency and cardiac sudden death in mice lacking the MEF2A transcription factor. *Nat. Med.* (2002).
doi:10.1038/nm789
176. Lin, Q., Schwarz, J., Bucana, C. & Olson, E. N. Control of mouse cardiac morphogenesis and myogenesis by transcription factor MEF2C. *Science* (80-). (1997). doi:10.1126/science.276.5317.1404
177. Bi, W., Drake, C. J. & Schwarz, J. J. The transcription factor MEF2C-null mouse exhibits complex vascular malformations and reduced cardiac expression of angiopoietin 1 and VEGF. *Dev. Biol.* (1999).
doi:10.1006/dbio.1999.9307
178. Xu, J. *et al.* Myocyte enhancer factors 2A and 2C induce dilated cardiomyopathy in transgenic mice. *J. Biol. Chem.* (2006).
doi:10.1074/jbc.M510217200
179. Kim, Y. *et al.* The MEF2D transcription factor mediates stress-dependent cardiac remodeling in mice. *J. Clin. Invest.* (2008).
doi:10.1172/JCI33255
180. Dai, Y. S., Cserjesi, P., Markham, B. E. & Molkentin, J. D. The transcription factors GATA4 and dHAND physically interact to synergistically activate cardiac gene expression through a p300-dependent mechanism. *J. Biol. Chem.* (2002).
doi:10.1074/jbc.M202490200
181. Thomas, T., Yamagishi, H., Overbeek, P. A., Olson, E. N. & Srivastava, D. The bHLH factors, dHAND and eHAND, specify pulmonary and systemic cardiac ventricles independent of left-right sidedness. *Dev. Biol.* (1998).
doi:10.1006/dbio.1998.8849
182. Srivastava, D., Thomas, T., Kirby, M. L., Brown, D. & Olson, E. N. Regulation of cardiac mesodermal and neural crest development by the bHLH transcription factor, dHAND. *Nat. Genet.* (1997).
doi:10.1038/ng0697-154
183. Firulli, A. B., McFadden, D. G., Lin, Q., Srivastava, D. & Olson, E. N. Heart and extra-embryonic mesodermal defects in mouse embryos lacking the bHLH transcription factor Hand1. *Nat. Genet.* (1998).
doi:10.1038/ng0398-266
184. Li, L. *et al.* HAND1 loss-of-function mutation contributes to congenital double outlet right ventricle. *Int. J. Mol. Med.* (2017).
doi:10.3892/ijmm.2017.2865
185. Cheng, Z. *et al.* Two novel HAND1 mutations in Chinese patients with

ventricular septal defect. *Clin. Chim. Acta* (2012).
doi:10.1016/j.cca.2011.10.014

186. Sun, Y.-M. *et al.* A HAND2 Loss-of-Function Mutation Causes Familial Ventricular Septal Defect and Pulmonary Stenosis. *G3: Genes/Genomes/Genetics* (2016). doi:10.1534/g3.115.026518
187. Lu, C. X. *et al.* A novel HAND2 loss-of-function mutation responsible for tetralogy of Fallot. *Int. J. Mol. Med.* (2016).
doi:10.3892/ijmm.2015.2436
188. Koshiba-Takeuchi, K. *et al.* Cooperative and antagonistic interactions between Sall4 and Tbx5 pattern the mouse limb and heart. *Nat. Genet.* **38**, 175–183 (2006).
189. Smagulova, F. O., Manuylov, N. L., Leach, L. L. & Tevosian, S. G. GATA4/FOG2 transcriptional complex regulates Lhx9 gene expression in murine heart development. *BMC Dev. Biol.* **8**, 67 (2008).
190. Granados-Riveron, J. T. *et al.* Alpha-cardiac myosin heavy chain (MYH6) mutations affecting myofibril formation are associated with congenital heart defects. *Hum. Mol. Genet.* **19**, 4007–16 (2010).
191. Khan, S. H. Genome-Editing Technologies: Concept, Pros, and Cons of Various Genome-Editing Techniques and Bioethical Concerns for Clinical Application. *Molecular Therapy - Nucleic Acids* (2019).
doi:10.1016/j.omtn.2019.02.027
192. Chapman, J. R., Taylor, M. R. G. & Boulton, S. J. Playing the End Game: DNA Double-Strand Break Repair Pathway Choice. *Molecular Cell* **47**, 497–510 (2012).
193. Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* **32**, 347–55 (2014).
194. Silva, G. *et al.* Meganucleases and Other Tools for Targeted Genome Engineering: Perspectives and Challenges for Gene Therapy. *Curr. Gene Ther.* (2011). doi:10.2174/156652311794520111
195. Guha, T. K. & Edgell, D. R. Applications of alternative nucleases in the age of CRISPR/Cas9. *International Journal of Molecular Sciences* (2017).
doi:10.3390/ijms18122565
196. Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S. & Gregory, P. D. Genome editing with engineered zinc finger nucleases. *Nature Reviews Genetics* (2010). doi:10.1038/nrg2842
197. Moehle, E. A. *et al.* Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc. Natl. Acad. Sci.* (2007). doi:10.1073/pnas.0611478104

198. Ramirez, C. L. *et al.* Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat. Methods* (2008). doi:10.1038/nmeth0508-374
199. Christian, M. *et al.* Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* (2010). doi:10.1534/genetics.110.120717
200. Moscou, M. J. & Bogdanove, A. J. A simple cipher governs DNA recognition by TAL effectors. *Science* (80-.). (2009). doi:10.1126/science.1178817
201. Bogdanove, A. J. & Voytas, D. F. TAL effectors: Customizable proteins for DNA targeting. *Science* (2011). doi:10.1126/science.1204094
202. Reyon, D. *et al.* FLASH assembly of TALENs for high-throughput genome editing. *Nat. Biotechnol.* (2012). doi:10.1038/nbt.2170
203. Tsai, S. Q. *et al.* Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat. Biotechnol.* (2014). doi:10.1038/nbt.2908
204. Guha, T. K., Wai, A. & Hausner, G. Programmable Genome Editing Tools and their Regulation for Efficient Genome Engineering. *Computational and Structural Biotechnology Journal* (2017). doi:10.1016/j.csbj.2016.12.006
205. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* (2005). doi:10.1007/s00239-004-0046-3
206. Horvath, P. & Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**, 167–170 (2010).
207. Terns, M. P. & Terns, R. M. CRISPR-based adaptive immune systems. *Current Opinion in Microbiology* **14**, 321–327 (2011).
208. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* (80-.). **337**, 816–821 (2012).
209. Li, X. L. *et al.* Highly efficient genome editing via CRISPR-Cas9 in human pluripotent stem cells is achieved by transient BCL-XL overexpression. *Nucleic Acids Res.* (2018). doi:10.1093/nar/gky804
210. Hruscha, A. *et al.* Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish. *Development* **140**, 4982–7 (2013).
211. Karvelis, T. *et al.* Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. *Genome Biol.* (2015). doi:10.1186/s13059-015-0818-7
212. Ran, F. A. *et al.* Double nicking by RNA-guided CRISPR cas9 for

enhanced genome editing specificity. *Cell* (2013).
doi:10.1016/j.cell.2013.08.021

213. Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* (2013). doi:10.1038/nbt.2675
214. Malik, N. & Rao, M. S. A review of the methods for human iPSC derivation. in *Methods in Molecular Biology* (2013). doi:10.1007/978-1-62703-348-0_3
215. Takahashi, K. *et al.* Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell* (2007).
doi:10.1016/j.cell.2007.11.019
216. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* (80-.). (2007). doi:10.1126/science.1151526
217. Gaj, T., Sirk, S. J., Shui, S. L. & Liu, J. Genome-editing technologies: Principles and applications. *Cold Spring Harb. Perspect. Biol.* (2016).
doi:10.1101/cshperspect.a023754
218. Kim, K. *et al.* Epigenetic memory in induced pluripotent stem cells. *Nature* (2010). doi:10.1038/nature09342
219. Rouhani, F. *et al.* Genetic Background Drives Transcriptional Variation in Human Induced Pluripotent Stem Cells. *PLoS Genet.* (2014).
doi:10.1371/journal.pgen.1004432
220. Wang, G. *et al.* Modeling the mitochondrial cardiomyopathy of Barth syndrome with induced pluripotent stem cell and heart-on-chip technologies. *Nat. Med.* (2014). doi:10.1038/nm.3545
221. Hinson, J. T. *et al.* Titin mutations in iPS cells define sarcomere insufficiency as a cause of dilated cardiomyopathy. *Science* (80-.). (2015). doi:10.1126/science.aaa5458
222. Theodoris, C. V. *et al.* Human disease modeling reveals integrated transcriptional and epigenetic mechanisms of NOTCH1 haploinsufficiency. *Cell* (2015). doi:10.1016/j.cell.2015.02.035
223. Capecchi, M. R., Evans, M. J. & Smithies, O. Generating mice with targeted mutations. *Nat. Med.* (2001). doi:10.1038/nm1001-1086
224. Carroll, K. J. *et al.* A mouse model for adult cardiac-specific gene deletion with CRISPR/Cas9. *Proc. Natl. Acad. Sci.* (2016).
doi:10.1073/pnas.1523918113
225. Endres, B. T. *et al.* Mutation of *Plekha7* attenuates salt-sensitive hypertension in the rat. *Proc. Natl. Acad. Sci.* (2014).
doi:10.1073/pnas.1410745111

226. Niimi, M. *et al.* ApoE knockout rabbits: A novel model for the study of human hyperlipidemia. *Atherosclerosis* (2016). doi:10.1016/j.atherosclerosis.2015.12.002
227. Kotani, H., Taimatsu, K., Ohga, R., Ota, S. & Kawahara, A. Efficient multiple genome modifications induced by the crRNAs, tracrRNA and Cas9 Protein complex in zebrafish. *PLoS One* **10**, e0128319 (2015).
228. German, D. M., Mitalipov, S., Mishra, A. & Kaul, S. Therapeutic Genome Editing in Cardiovascular Diseases. *JACC: Basic to Translational Science* (2019). doi:10.1016/j.jacbts.2018.11.004
229. Ding, Q. *et al.* Permanent alteration of PCSK9 with in vivo CRISPR-Cas9 genome editing. *Circ. Res.* (2014). doi:10.1161/CIRCRESAHA.115.304351
230. Wang, X. *et al.* CRISPR-Cas9 targeting of PCSK9 in human hepatocytes in vivo - Brief report. *Arterioscler. Thromb. Vasc. Biol.* (2016). doi:10.1161/ATVBAHA.116.307227
231. Yin, H. *et al.* Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nat. Biotechnol.* (2014). doi:10.1038/nbt.2884
232. Long, C. *et al.* Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy. *Science* (80-.). (2016). doi:10.1126/science.aad5725
233. Krishnan, A. *et al.* A detailed comparison of mouse and human cardiac development. *Pediatr. Res.* (2014). doi:10.1038/pr.2014.128
234. Webb, S., Brown, N. A., Wessels, A. & Anderson, R. H. Development of the murine pulmonary vein and its relationship to the embryonic venous sinus. *Anat. Rec.* (1998). doi:10.1002/(SICI)1097-0185(199803)250:3<325::AID-AR7>3.0.CO;2-Z
235. Wessels, A. *et al.* Atrial development in the human heart: An immunohistochemical study with emphasis on the role of mesenchymal tissues. *Anat. Rec.* (2000). doi:10.1002/1097-0185(20000701)259:3<288::AID-AR60>3.0.CO;2-D
236. Wessels, A. & Sedmera, D. Developmental anatomy of the heart: a tale of mice and man. *Physiol. Genomics* (2003). doi:10.1152/physiolgenomics.00033.2003
237. England, J. & Loughna, S. Vertebrate Embryo : An Overview of Heart Development and the Utilisation of Multiple. *eLS. John Wiley Sons Ltd, Chichester* **Mar**, 1–11 (2018).
238. Guénet, J. L. The mouse genome. *Genome Research* (2005). doi:10.1101/gr.3728305

239. King, A. J. F. The use of animal models in diabetes research. *British Journal of Pharmacology* (2012). doi:10.1111/j.1476-5381.2012.01911.x
240. Cheon, D.-J. & Orsulic, S. Mouse Models of Cancer. *Annu. Rev. Pathol. Mech. Dis.* (2011). doi:10.1146/annurev.pathol.3.121806.154244
241. Dutta, S. & Sengupta, P. Men and mice: Relating their ages. *Life Sciences* (2016). doi:10.1016/j.lfs.2015.10.025
242. Gordon, J. W. & Ruddle, F. H. Integration and stable germ line transmission of genes injected into mouse pronuclei. *Science* (80-.). (1981). doi:10.1126/science.6272397
243. Camacho, P., Fan, H., Liu, Z. & He, J. Q. Small mammalian animal models of heart disease. *American Journal of Cardiovascular Disease* (2016). doi:10.3390/jcdd3040030
244. Digilio, M., Marino, B., Capolino, R. & Dallapiccola, B. Clinical manifestations of Deletion 22q11.2 syndrome (DiGeorge/Velo-Cardio-Facial syndrome). *Images Paediatr. Cardiol.* (2005).
245. Morrow, B. E., McDonald-McGinn, D. M., Emanuel, B. S., Vermeesch, J. R. & Scambler, P. J. Molecular genetics of 22q11.2 deletion syndrome. *American Journal of Medical Genetics, Part A* (2018). doi:10.1002/ajmg.a.40504
246. Jerome, L. A. & Papaioannou, V. E. DiGeorge syndrome phenotype in mice mutant for the T-box gene, Tbx1. *Nat. Genet.* (2001). doi:10.1038/85845
247. Lindsay, E. A. *et al.* Tbx1 haploinsufficiency in the DiGeorge syndrome region causes aortic arch defects in mice. *Nature* (2001). doi:10.1038/35065105
248. Pereira, L. *et al.* Targetting of the gene encoding fibrillin-1 recapitulates the vascular aspect of marfan syndrome. *Nature Genetics* (1997). doi:10.1038/ng1097-218
249. Li, D. Y. *et al.* Elastin is an essential determinant of arterial morphogenesis. *Nature* (1998). doi:10.1038/30522
250. Xue, Y. *et al.* Embryonic lethality and vascular defects in mice lacking the Notch ligand Jagged1. *Hum. Mol. Genet.* (1999). doi:10.1093/hmg/8.5.723
251. de Lange, W. J., Grimes, A. C., Hegge, L. F. & Ralphe, J. C. Ablation of cardiac myosin-binding protein-C accelerates contractile kinetics in engineered cardiac tissue. *J. Gen. Physiol.* (2013). doi:10.1085/jgp.201210837

252. Unsld, B. *et al.* Age-dependent changes in contractile function and passive elastic properties of myocardium from mice lacking muscle LIM protein (MLP). *Eur. J. Heart Fail.* (2012). doi:10.1093/eurjhf/hfs020
253. Cao, Q. *et al.* Phenotype and Functional Analyses in a Transgenic Mouse Model of Left Ventricular Noncompaction Caused by a <i>DTNA</i> Mutation. *Int. Heart J.* (2017). doi:10.1536/ihj.16-019
254. Flenniken, A. M. A Gja1 missense mutation in a mouse model of oculodentodigital dysplasia. *Development* (2005). doi:10.1242/dev.02011
255. Simon, A. M., Goodenough, D. A. & Paul, D. L. Mice lacking connexin40 have cardiac conduction abnormalities characteristic of atrioventricular block and bundle branch block. *Curr. Biol.* (1998). doi:10.1016/S0960-9822(98)70113-7
256. Asnani, A. & Peterson, R. T. The zebrafish as a tool to identify novel therapies for human cardiovascular disease. *DMM Disease Models and Mechanisms* (2014). doi:10.1242/dmm.016170
257. Stainier, D. Y., Lee, R. K. & Fishman, M. C. Cardiovascular development in the zebrafish. I. Myocardial fate map and heart tube formation. *Development* (1993).
258. Scott, I. C. *et al.* The G Protein-Coupled Receptor Agtr1b Regulates Early Development of Myocardial Progenitors. *Dev. Cell* (2007). doi:10.1016/j.devcel.2007.01.012
259. Garavito-Aguilar, Z. V. *et al.* Distinct phases of cardiomyocyte differentiation regulate growth of the zebrafish heart. *Development* (2009). doi:10.1242/dev.030924
260. Bussmann, J., Bakkers, J. & Schulte-Merker, S. Early endocardial morphogenesis requires Scl/Tal1. *PLoS Genet.* (2007). doi:10.1371/journal.pgen.0030140
261. Holtzman, N. G., Schoenebeck, J. J., Tsai, H.-J. & Yelon, D. Endocardium is necessary for cardiomyocyte movement during heart tube assembly. *Development* (2007). doi:10.1242/dev.02857
262. Smith, K. A. *et al.* Rotation and Asymmetric Development of the Zebrafish Heart Requires Directed Migration of Cardiac Progenitor Cells. *Dev. Cell* (2008). doi:10.1016/j.devcel.2007.11.015
263. Rohr, S., Otten, C. & Abdelilah-Seyfried, S. Asymmetric involution of the myocardial field drives heart tube formation in zebrafish. *Circ. Res.* (2008). doi:10.1161/CIRCRESAHA.107.165241
264. Chen, J. N. *et al.* Left-right pattern of cardiac BMP4 may drive

asymmetry of the heart in zebrafish. *Development* (1997).

265. Serluca, F. C. Development of the proepicardial organ in the zebrafish. *Dev. Biol.* (2008). doi:10.1016/j.ydbio.2007.10.007
266. Bakkers, J. Zebrafish as a model to study cardiac development and human cardiac disease. *Cardiovasc. Res.* **91**, 279–288 (2011).
267. Howe, K. *et al.* The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**, 498–503 (2013).
268. Asnani, a. & Peterson, R. T. The zebrafish as a tool to identify novel therapies for human cardiovascular disease. *Dis. Model. Mech.* **7**, 763–767 (2014).
269. Tu, S. & Chi, N. C. Zebrafish models in cardiac development and congenital heart birth defects. *Differentiation* **84**, 4–16 (2012).
270. Baker, K., Warren, K. S., Yellen, G. & Fishman, M. C. Defective ‘pacemaker’ current (I_h) in a zebrafish mutant with a slow heart rate. *Proc. Natl. Acad. Sci.* (1997). doi:10.1073/pnas.94.9.4554
271. Chi, N. C. *et al.* Genetic and physiologic dissection of the vertebrate cardiac conduction system. *PLoS Biol.* (2008). doi:10.1371/journal.pbio.0060109
272. Poss, K. D., Wilson, L. G. & Keating, M. T. Heart regeneration in zebrafish. *Science* (80-.). (2002). doi:10.1126/science.1077857
273. Bakkers, J. Zebrafish as a model to study cardiac development and human cardiac disease. *Cardiovascular Research* (2011). doi:10.1093/cvr/cvr098
274. Brown, D., Samsa, L., Qian, L. & Liu, J. Advances in the Study of Heart Development and Disease Using Zebrafish. *J. Cardiovasc. Dev. Dis.* (2016). doi:10.3390/jcdd3020013
275. England, J. & Loughna, S. in *eLS* (2018). doi:10.1002/9780470015902.a0027698
276. Bill, B. R., Petzold, A. M., Clark, K. J., Schimmenti, L. a & Ekker, S. C. A primer for morpholino use in zebrafish. *Zebrafish* **6**, 69–77 (2009).
277. Bedell, V. M. *et al.* In vivo genome editing using a high-efficiency TALEN system. *Nature* **491**, 114–8 (2012).
278. Hwang, W. Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* **31**, 227–9 (2013).
279. Summerton, J. & Weller, D. Morpholino antisense oligomers: Design, preparation, and properties. *Antisense Nucleic Acid Drug Dev.* (1997).

doi:10.1089/oli.1.1997.7.187

280. Stainier, D. Y. R. *et al.* Guidelines for morpholino use in zebrafish. *PLOS Genet.* (2017). doi:10.1371/journal.pgen.1007000
281. Karlen, S. & Rebagliati, M. A morpholino phenocopy of the cyclops mutation. *Genesis* (2001). doi:10.1002/gene.1046
282. Garrity, D. M., Childs, S. & Fishman, M. C. The heartstrings mutation in zebrafish causes heart/fin Tbx5 deficiency syndrome. *Development* (2002).
283. Kok, F. O. *et al.* Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. *Dev. Cell* **32**, 97–108 (2015).
284. Rossi, A. *et al.* Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature* (2015). doi:10.1038/nature14580
285. El-Brolosy, M. A. & Stainier, D. Y. R. Genetic compensation: A phenomenon in search of mechanisms. *PLoS Genetics* (2017). doi:10.1371/journal.pgen.1006780
286. Eisen, J. S. & Smith, J. C. Controlling morpholino experiments: don't stop making antisense. *Development* (2008). doi:10.1242/dev.001115
287. Foley, J. E. *et al.* Targeted mutagenesis in zebrafish using customized zinc-finger nucleases. *Nat. Protoc.* (2009). doi:10.1038/nprot.2009.209
288. Zu, Y. *et al.* TALEN-mediated precise genome modification by homologous recombination in zebrafish. *Nat. Methods* (2013). doi:10.1038/nmeth.2374
289. Irion, U., Krauss, J. & Nusslein-Volhard, C. Precise and efficient genome editing in zebrafish using the CRISPR/Cas9 system. *Development* **141**, 4827–30 (2014).
290. Burger, A. *et al.* Maximizing mutagenesis with solubilized CRISPR-Cas9 ribonucleoprotein complexes. *Development* **143**, 2025–2037 (2016).
291. Mori, A. D. & Bruneau, B. G. TBX5 mutations and congenital heart disease: Holt-Oram syndrome revealed. *Current Opinion in Cardiology* (2004). doi:10.1097/00001573-200405000-00004
292. Tessadori, F. *et al.* Effective CRISPR/Cas9-based nucleotide editing in zebrafish to model human genetic cardiovascular disorders. *Dis. Model. Mech.* (2018). doi:10.1242/dmm.035469
293. Green, M. R. & Sambrook, J. Precipitation of DNA with ethanol. *Cold Spring Harb. Protoc.* (2016). doi:10.1101/pdb.prot093377

294. Froger, A. & Hall, J. E. Transformation of Plasmid DNA into E. Coli using the heat shock method. *J. Vis. Exp.* (2007). doi:10.3791/253
295. Rio, D. C., Ares, M., Hannon, G. J. & Nilsen, T. W. Purification of RNA using TRIzol (TRI Reagent). *Cold Spring Harb. Protoc.* (2010). doi:10.1101/pdb.prot5439
296. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* (2001). doi:10.1006/meth.2001.1262
297. Wang, K. K. W., Posner, A. & Hajimohammadreza, I. Total protein extraction from cultured cells for use in electrophoresis and Western blotting. *Biotechniques* (1996). doi:10.2144/19962004662
298. LOWRY, O. H., ROSEBROUGH, N. J., FARR, A. L. & RANDALL, R. J. Protein measurement with the Folin phenol reagent. *J. Biol. Chem.* (1951). doi:10.1016/0304-3894(92)87011-4
299. Eslami, A. & Lujan, J. Western blotting: Sample preparation to detection. *J. Vis. Exp.* (2010). doi:10.3791/2359
300. Gluzman, Y. SV40-transformed simian cells support the replication of early SV40 mutants. *Cell* (1981). doi:10.1016/0092-8674(81)90282-8
301. Yaffe, D. & Saxel, O. Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. *Nature* (1977). doi:10.1038/270725a0
302. Vaughn, J. L., Goodwin, R. H., Tompkins, G. J. & McCawley, P. The establishment of two cell lines from the insect *spodoptera frugiperda* (lepidoptera; noctuidae). *In Vitro* (1977). doi:10.1007/BF02615077
303. Van Meerloo, J., Kaspers, G. J. L. & Cloos, J. Cell sensitivity assays: The MTT assay. *Methods Mol. Biol.* (2011). doi:10.1007/978-1-61779-80-5_20
304. Ghosh, T. K. *et al.* Characterization of the TBX5 binding site and analysis of mutations that cause Holt-Oram syndrome. *Hum. Mol. Genet.* (2001). doi:10.1093/hmg/10.18.1983
305. Kudo, N. *et al.* Leptomycin B inactivates CRM1/exportin 1 by covalent modification at a cysteine residue in the central conserved region. *Proc. Natl. Acad. Sci. U. S. A.* (1999). doi:10.1073/pnas.96.16.9112
306. Balasubramanyam, K. *et al.* Polyisoprenylated benzophenone, garcinol, a natural histone acetyltransferase inhibitor, represses chromatin transcription and alters global gene expression. *J. Biol. Chem.* (2004). doi:10.1074/jbc.M402839200
307. Biswas, A. *et al.* Opposing action of casein kinase 1 and calcineurin in

nucleo-cytoplasmic shuttling of mammalian translation initiation factor eIF6. *J. Biol. Chem.* (2011). doi:10.1074/jbc.M110.188565

308. Domazetovska, A. *et al.* Mechanisms underlying intranuclear rod formation. *Brain* (2007). doi:10.1093/brain/awm247
309. Labun, K., Montague, T. G., Gagnon, J. A., Thyme, S. B. & Valen, E. CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Res.* (2016). doi:10.1093/nar/gkw398
310. Bae, S., Park, J. & Kim, J. S. Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* (2014). doi:10.1093/bioinformatics/btu048
311. Moreno-Mateos, M. A. *et al.* CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat. Methods* **12**, 982–8 (2015).
312. Gagnon, J. A. *et al.* Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PLoS One* **9**, 5–12 (2014).
313. Jao, L.-E., Wente, S. R. & Chen, W. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 13904–9 (2013).
314. Grainger, S. *et al.* CRISPR Guide RNA Validation in Vitro. *Zebrafish* (2017). doi:10.1089/zeb.2016.1358
315. Guschin, D. Y. *et al.* A rapid and general assay for monitoring endogenous gene modification. *Methods Mol. Biol.* **649**, 247–256 (2010).
316. Thisse, C. & Thisse, B. High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nat. Protoc.* **3**, 59–69 (2008).
317. Chernyavskaya, Y., Ebert, A. M., Milligan, E. & Garrity, D. M. Voltage-gated calcium channel CACNB2 (??2.1) protein is required in the heart for control of cell proliferation and heart tube integrity. *Dev. Dyn.* **241**, 648–662 (2012).
318. Parrie, L. E., Renfrew, E. M., Wal, A. Vander, Mueller, R. L. & Garrity, D. M. Zebrafish *tbx5* paralogs demonstrate independent essential requirements in cardiac and pectoral fin development. *Dev. Dyn.* **242**, 485–402 (2013).
319. Zhao, X. *et al.* Zebrafish *cul4a*, but not *cul4b*, modulates cardiac and forelimb development by upregulating *tbx5a* expression. *Hum Mol Genet* **24**, 853–864 (2015).

320. Sasore, T. & Kennedy, B. Deciphering combinations of PI3K/AKT/mTOR pathway drugs augmenting anti-angiogenic efficacy in vivo. *PLoS One* (2014). doi:10.1371/journal.pone.0105280
321. Li, J., Wu, P., Liu, Y., Wang, D. & Cheng, C. H. K. Temporal and spatial expression of the four Igf ligands and two Igf type 1 receptors in zebrafish during early embryonic development. *Gene Expr. Patterns* (2014). doi:10.1016/j.gep.2014.05.006
322. Tang, R., Dodd, A., Lai, D., McNabb, W. C. & Love, D. R. Validation of zebrafish (*Danio rerio*) reference genes for quantitative real-time RT-PCR normalization. *Acta Biochim. Biophys. Sin. (Shanghai)*. (2007). doi:10.1111/j.1745-7270.2007.00283.x
323. Hu, Y., Xie, S. & Yao, J. Identification of novel reference genes suitable for qRT-PCR normalization with respect to the zebrafish developmental stage. *PLoS One* (2016). doi:10.1371/journal.pone.0149277
324. Wasilko, D. J. *et al.* The titerless infected-cells preservation and scale-up (TIPS) method for large-scale production of NO-sensitive human soluble guanylate cyclase (sGC) from insect cells infected with recombinant baculovirus. *Protein Expr. Purif.* (2009). doi:10.1016/j.pep.2009.01.002
325. Lee, K.-M. & Narlikar, G. in *Current Protocols in Molecular Biology* (2001). doi:10.1002/0471142727.mb2106s54
326. Loughna, S. & Henderson, D. Methodologies for staining and visualisation of β -galactosidase in mouse embryos and tissues. *Methods Mol. Biol.* (2007). doi:10.1385/1-59745-549-0:1
327. Bruneau, B. G. Signaling and transcriptional networks in heart development and regeneration. *Cold Spring Harb. Perspect. Biol.* (2013). doi:10.1101/cshperspect.a008292
328. Waardenberg, A. J., Ramialison, M., Bouvere, R. & Harvey, R. P. Genetic networks governing heart development. *Cold Spring Harb. Perspect. Med.* (2014). doi:10.1101/cshperspect.a013839
329. Hübner, M. R., Eckersley-Maslin, M. A. & Spector, D. L. Chromatin organization and transcriptional regulation. *Current Opinion in Genetics and Development* (2013). doi:10.1016/j.gde.2012.11.006
330. Verdone, L., Caserta, M. & Di Mauro, E. Role of histone acetylation in the control of gene expression. *Biochemistry and Cell Biology* (2005). doi:10.1139/o05-041
331. Mariño-Ramírez, L., Kann, M. G., Shoemaker, B. A. & Landsman, D. Histone structure and nucleosome stability. *Expert Review of Proteomics* (2005). doi:10.1586/14789450.2.5.719

332. Tessarz, P. & Kouzarides, T. Histone core modifications regulating nucleosome structure and dynamics. *Nat. Rev. Mol. Cell Biol.* **15**, 703–708 (2014).
333. Grunstein, M. Histone acetylation in chromatin structure and transcription. *Nature* **389**, 349–352 (1997).
334. Cheung, W. L., Briggs, S. D. & Allis, C. D. Acetylation and chromosomal functions. *Current Opinion in Cell Biology* **12**, 326–333 (2000).
335. McKinsey, T. a, Zhang, C. L. & Olson, E. N. Control of muscle development by dueling HATs and HDACs. *Curr. Opin. Genet. Dev.* **11**, 497–504 (2001).
336. Backs, J. & Olson, E. N. Control of cardiac growth by histone acetylation/deacetylation. *Circulation Research* **98**, 15–24 (2006).
337. Haberland, M., Montgomery, R. L. & Olson, E. N. The many roles of histone deacetylases in development and physiology: implications for disease and therapy. *Nat. Rev. Genet.* **10**, 32–42 (2009).
338. Han, P., Hang, C. T., Yang, J. & Chang, C.-P. Chromatin Remodeling in Cardiovascular Development and Physiology. *Circ. Res.* **108**, 378–396 (2011).
339. Yang, X. J. & Seto, E. Collaborative spirit of histone deacetylases in regulating chromatin structure and gene expression. *Current Opinion in Genetics and Development* **13**, 143–153 (2003).
340. Guenther, M. G., Barak, O. & Lazar, M. A. The SMRT and N-CoR corepressors are activating cofactors for histone deacetylase 3. *Mol. Cell. Biol.* **21**, 6091–101 (2001).
341. Passier, R. *et al.* CaM kinase signaling induces cardiac hypertrophy and activates the MEF2 transcription factor in vivo. *J. Clin. Invest.* **105**, 1395–1406 (2000).
342. Vega, R. B. *et al.* Protein kinases C and D mediate agonist-dependent cardiac hypertrophy through nuclear export of histone deacetylase 5. *Mol. Cell. Biol.* **24**, 8374–85 (2004).
343. Zhang, C. L. *et al.* Class II histone deacetylases act as signal-responsive repressors of cardiac hypertrophy. *Cell* **110**, 479–488 (2002).
344. Chang, S. *et al.* Histone deacetylases 5 and 9 govern responsiveness of the heart to a subset of stress signals and play redundant roles in heart development. *Mol. Cell. Biol.* **24**, 8467–8476 (2004).
345. Vega, R. B. *et al.* Histone deacetylase 4 controls chondrocyte hypertrophy during skeletogenesis. *Cell* **119**, 555–566 (2004).

346. Chang, S. *et al.* Histone Deacetylase 7 Maintains Vascular Integrity by Repressing Matrix Metalloproteinase 10. *Cell* **126**, 321–334 (2006).
347. Montgomery, R. L. *et al.* Histone deacetylases 1 and 2 redundantly regulate cardiac morphogenesis, growth, and contractility. *Genes Dev.* **21**, 1790–1802 (2007).
348. Montgomery, R. L. *et al.* Maintenance of cardiac energy metabolism by histone deacetylase 3 in mice. *J. Clin. Invest.* **118**, 3588–3597 (2008).
349. Lewandowski, S. L. *et al.* Histone deacetylase 3 modulates Tbx5 activity to regulate early cardiogenesis. *Hum. Mol. Genet.* **23**, 3801–3809 (2014).
350. Trivedi, C. M., Lu, M. M., Wang, Q. & Epstein, J. a. Transgenic overexpression of Hdac3 in the heart produces increased postnatal cardiac myocyte proliferation but does not induce hypertrophy. *J. Biol. Chem.* **283**, 26484–9 (2008).
351. Carrozza, M. J., Utley, R. T., Workman, J. L. & C??t??, J. The diverse functions of histone acetyltransferase complexes. *Trends in Genetics* **19**, 321–329 (2003).
352. Nagy, Z. & Tora, L. Distinct GCN5/PCAF-containing complexes function as co-activators and are involved in transcription factor and global histone acetylation. *Oncogene* (2007). doi:10.1038/sj.onc.1210604
353. Koutsogiannouli, E. A. *et al.* Differential effects of histone acetyltransferase GCN5 or PCAF knockdown on urothelial carcinoma cells. *Int. J. Mol. Sci.* (2017). doi:10.3390/ijms18071449
354. Zhao, M. *et al.* PCAF/GCN5-Mediated Acetylation of RPA1 Promotes Nucleotide Excision Repair. *Cell Rep.* (2017). doi:10.1016/j.celrep.2017.08.015
355. Yamauchi, T. *et al.* Distinct but overlapping roles of histone acetylase PCAF and of the closely related PCAF-B/GCN5 in mouse embryogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 11303–11306 (2000).
356. Jin, Q. *et al.* Gcn5 and PCAF Regulate PPAR and Prdm16 Expression To Facilitate Brown Adipogenesis. *Mol. Cell. Biol.* (2014). doi:10.1128/MCB.00622-14
357. Kurooka, H. & Honjo, T. Functional interaction between the mouse Notch1 intracellular region and histone acetyltransferases PCAF and GCN5. *J. Biol. Chem.* (2000). doi:10.1074/jbc.M000909200
358. Wilde, J. J., Siegenthaler, J. A., Dent, S. Y. R. & Niswander, L. A. Diencephalic Size Is Restricted by a Novel Interplay Between GCN5 Acetyltransferase Activity and Retinoic Acid Signaling. *J. Neurosci.* (2017). doi:10.1523/JNEUROSCI.2121-16.2017

359. Xu, W. *et al.* Loss of Gcn5l2 leads to increased apoptosis and mesodermal defects during mouse development. *Nat. Genet.* **26**, 229–232 (2000).
360. Bu, P., Evrard, Y. A., Lozano, G. & Dent, S. Y. R. Loss of Gcn5 Acetyltransferase Activity Leads to Neural Tube Closure Defects and Exencephaly in Mouse Embryos. *Mol. Cell. Biol.* (2007). doi:10.1128/MCB.00066-07
361. Sapountzi, V. & C??t??, J. MYST-family histone acetyltransferases: Beyond chromatin. *Cellular and Molecular Life Sciences* (2011). doi:10.1007/s00018-010-0599-9
362. Squatrito, M., Gorrini, C. & Amati, B. Tip60 in DNA damage response and growth control: many tricks in one HAT. *Trends in Cell Biology* (2006). doi:10.1016/j.tcb.2006.07.007
363. Voss, A. K. & Thomas, T. MYST family histone acetyltransferases take center stage in stem cells and development. *BioEssays* (2009). doi:10.1002/bies.200900051
364. Avvakumov, N. & Côté, J. The MYST family of histone acetyltransferases and their intimate links to cancer. *Oncogene* (2007). doi:10.1038/sj.onc.1210608
365. Kim, M. S., Merlo, X., Wilson, C. & Lough, J. Co-activation of atrial natriuretic factor promoter by Tip60 and serum response factor. *J. Biol. Chem.* (2006). doi:10.1074/jbc.M513593200
366. Fisher, J. B. *et al.* Depletion of Tip60 from in vivo cardiomyocytes increases myocyte density, followed by cardiac dysfunction, myocyte fallout and lethality. *PLoS One* (2016). doi:10.1371/journal.pone.0164855
367. Voss, A. K. *et al.* MOZ Regulates the Tbx1 Locus, and Moz Mutation Partially Phenocopies DiGeorge Syndrome. *Dev. Cell* (2012). doi:10.1016/j.devcel.2012.07.010
368. Vanyai, H. K., Thomas, T. & Voss, A. K. Mesodermal expression of Moz is necessary for cardiac septum development. *Dev. Biol.* (2015). doi:10.1016/j.ydbio.2015.04.011
369. Chen, G. *et al.* Spatiotemporal expression of histone acetyltransferases, p300 and CBP, in developing embryonic hearts. *J. Biomed. Sci.* (2009). doi:10.1186/1423-0127-16-24
370. Yao, T. P. *et al.* Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300. *Cell* **93**, 361–372 (1998).
371. Yanazume, T. *et al.* Cardiac p300 is involved in myocyte growth with

- decompensated heart failure. *Mol. Cell. Biol.* **23**, 3593–3606 (2003).
372. Dai, Y. S. & Markham, B. E. p300 Functions as a Coactivator of Transcription Factor GATA-4. *J. Biol. Chem.* (2001). doi:10.1074/jbc.M103731200
 373. Ma, K., Chan, J. K. L., Zhu, G. & Wu, Z. Myocyte Enhancer Factor 2 Acetylation by p300 Enhances Its DNA Binding Activity, Transcriptional Activity, and Myogenic Differentiation. *Mol. Cell. Biol.* (2005). doi:10.1128/mcb.25.9.3575-3582.2005
 374. Roth, S. Y., Denu, J. M. & Allis, D. Histone Acetyltransferases. *Annu. Rev. Biochem* **70**, 81–120 (2001).
 375. Takaya, T. *et al.* Identification of p300-targeted acetylated residues in GATA4 during hypertrophic responses in cardiac myocytes. *J. Biol. Chem.* **283**, 9828–9835 (2008).
 376. Ma, K., Chan, J. K. L., Zhu, G. & Wu, Z. Myocyte enhancer factor 2 acetylation by p300 enhances its DNA binding activity, transcriptional activity, and myogenic differentiation. *Mol. Cell. Biol.* **25**, 3575–3582 (2005).
 377. Kawamura, T. *et al.* Acetylation of GATA-4 is involved in the differentiation of embryonic stem cells into cardiac myocytes. *J. Biol. Chem.* **280**, 19682–8 (2005).
 378. Lewandowski, S. L., Janardhan, H. P. & Trivedi, C. M. Histone deacetylase 3 coordinates deacetylase-independent epigenetic silencing of TGF β 1 to orchestrate second heart field development. *J. Biol. Chem.* jbc.M115.684753 (2015). doi:10.1074/jbc.M115.684753
 379. Ghosh, T. K. *et al.* Physical Interaction between TBX5 and MEF2C Is Required for Early Heart Development. *Mol. Cell. Biol.* (2009). doi:10.1128/MCB.01923-08
 380. Gu, W. & Roeder, R. G. Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. *Cell* (1997). doi:10.1016/S0092-8674(00)80521-8
 381. Rojas, J. R. *et al.* Structure of Tetrahymena GCN5 bound to coenzyme A and a histone H3 peptide. *Nature* (1999). doi:10.1038/43487
 382. Camarata, T. *et al.* LMP4 regulates Tbx5 protein subcellular localization and activity. *J. Cell Biol.* (2006). doi:10.1083/jcb.200511109
 383. Bimber, B., Dettman, R. W. & Simon, H. G. Differential regulation of Tbx5 protein expression and sub-cellular localization during heart development. *Dev. Biol.* (2007). doi:10.1016/j.ydbio.2006.09.023
 384. Park, S. J. *et al.* Analysis of the molecular and regulatory properties of

active porcine endogenous retrovirus gamma-1 long terminal repeats in kidney tissues of the NIH-Miniature pig. *Mol. Cells* (2010). doi:10.1007/s10059-010-0121-0

385. Doench, J. G. *et al.* Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* **32**, 1262–7 (2014).
386. Chari, R., Mali, P., Moosburner, M. & Church, G. M. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nat. Methods* **12**, 1–7 (2015).
387. Kim, K. H. *et al.* *Lrrc10* is required for early heart development and function in zebrafish. *Dev. Biol.* (2007). doi:10.1016/j.ydbio.2007.06.005
388. Lin, C. Y. *et al.* Low Temperature Mitigates Cardiac Bifida in Zebrafish Embryos. *PLoS One* (2013). doi:10.1371/journal.pone.0069788
389. Lu, J. K. *et al.* Pectoral fin anomalies in *tbx5a* knockdown zebrafish embryos related to the cascade effect of n-cadherin and extracellular matrix formation. *J. Dev. Biol.* (2019). doi:10.3390/jdb7030015
390. Steimle, J. D. & Moskowitz, I. P. in *Current Topics in Developmental Biology* (2017). doi:10.1016/bs.ctdb.2016.08.008
391. Beketaev, I. *et al.* Potentiation of Tbx5-mediated transactivation by SUMO conjugation and protein inhibitor of activated STAT 1 (PIAS1). *Int. J. Biochem. Cell Biol.* (2014). doi:10.1016/j.biocel.2014.02.007
392. Barron, M. E. *et al.* Serum response factor, an enriched cardiac mesoderm obligatory factor, is a downstream gene target for Tbx genes. *J. Biol. Chem.* (2005). doi:10.1074/jbc.M412408200
393. Soutoglou, E., Katrakili, N. & Talianidis, I. Acetylation regulates transcription factor activity at multiple levels. *Mol. Cell* (2000). doi:10.1016/S1097-2765(00)80253-1
394. Gay, F. *et al.* Acetylation regulates subcellular localization of the Wnt signaling nuclear effector POP-1. *Genes Dev.* (2003). doi:10.1101/gad.1042403
395. Zhao, L. J., Subramanian, T., Zhou, Y. & Chinnadurai, G. Acetylation by p300 regulates nuclear localization and function of the transcriptional corepressor CtBP2. *J. Biol. Chem.* (2006). doi:10.1074/jbc.M509051200
396. Sung, Y., Kim, J., Kim, H., Lee, J. & Jeon, J. Highly efficient gene knockout in mice and zebrafish with RNA-guided endonucleases. *Genome Res.* **24**, 125–131 (2014).
397. Jin, Q. *et al.* Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation.

EMBO J. **30**, 249–62 (2011).

- 398. Jin, Q. *et al.* Gcn5 and PCAF negatively regulate interferon- β production through HAT-independent inhibition of TBK1. *EMBO Rep.* **15**, 1192–201 (2014).
- 399. Puskaric, S. *et al.* Shox2 mediates Tbx5 activity by regulating Bmp4 in the pacemaker region of the developing heart. *Hum. Mol. Genet.* **19**, 4625–4633 (2010).
- 400. Li, M., Luo, J., Brooks, C. L. & Gu, W. Acetylation of p53 inhibits its ubiquitination by Mdm2. *J. Biol. Chem.* **277**, 50607–50611 (2002).
- 401. Wang, X. & Hayes, J. J. Acetylation mimics within individual core histone tail domains indicate distinct roles in regulating the stability of higher-order chromatin structure. *Mol Cell Biol* **28**, 227–236 (2008).
- 402. Pi-Roig, A., Martin-Blanco, E. & Minguillon, C. Distinct tissue-specific requirements for the zebrafish tbx5 genes during heart, retina and pectoral fin development. *Open Biol.* (2014). doi:10.1098/rsob.140014
- 403. Debeer, P., Race, V., Gewillig, M., Devriendt, K. & Frijns, J. P. Novel TBX5 mutations in patients with Holt-Oram syndrome. *Clin. Orthop. Relat. Res.* (2007). doi:10.1097/BLO.0b013e3181123ffe
- 404. Denslow, S. a & Wade, P. a. The human Mi-2/NuRD complex and gene regulation. *Oncogene* **26**, 5433–8 (2007).
- 405. Lai, A. Y. & Wade, P. A. Cancer biology and NuRD: a multifaceted chromatin remodelling complex. *Nat Rev Cancer* **11**, 588–596 (2011).
- 406. Marfella, C. G. A. & Imbalzano, A. N. The Chd family of chromatin remodelers. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* (2007). doi:10.1016/j.mrfmmm.2006.07.012
- 407. Low, J. K. K. *et al.* CHD4 is a peripheral component of the nucleosome remodeling and deacetylase complex. *J. Biol. Chem.* (2016). doi:10.1074/jbc.M115.707018
- 408. Mansfield, R. E. *et al.* Plant homeodomain (PHD) fingers of CHD4 are histone H3-binding modules with preference for unmodified H3K4 and methylated H3K9. *J. Biol. Chem.* (2011). doi:10.1074/jbc.M110.208207
- 409. Morra, R., Lee, B. M., Shaw, H., Tuma, R. & Mancini, E. J. Concerted action of the PHD, chromo and motor domains regulates the human chromatin remodelling ATPase CHD4. *FEBS Lett.* (2012). doi:10.1016/j.febslet.2012.06.017
- 410. Urquhart, A. J., Gatei, M., Richard, D. J. & Khanna, K. K. ATM mediated phosphorylation of CHD4 contributes to genome maintenance. *Genome Integr.* **2**, 1 (2011).

411. CHD4 Gene Expression. Gene Expression Database (GXD), Mouse Genome Informatics Web Site. World Wide Web (URL: <http://www.informatics.jax.org>). [Feb, 2018]. (2018).
412. CHD4 gene (MGI:1344380). International Mouse Phenotyping Consortium Web Site. World Wide Web (URL: <https://www.mousephenotype.org/data/genes/MGI:1344380>) [Feb, 2018].
413. O'Shaughnessy-Kirwan, A., Signolet, J., Costello, I., Gharbi, S. & Hendrich, B. Constraint of gene expression by the chromatin remodelling protein CHD4 facilitates lineage specification. *Development* (2015). doi:10.1242/dev.125450
414. Gomez-Del Arco, P. *et al.* The Chromatin Remodeling Complex Chd4/NuRD Controls Striated Muscle Identity and Metabolic Homeostasis. *Cell Metab.* **23**, 881–892 (2016).
415. Yamada, T. *et al.* Promoter decommissioning by the NuRD chromatin remodeling complex triggers synaptic connectivity in the mammalian brain. *Neuron* (2014). doi:10.1016/j.neuron.2014.05.039
416. Denner, D. R. & Rauchman, M. Mi-2/NuRD is required in renal progenitor cells during embryonic kidney development. *Dev. Biol.* (2013). doi:10.1016/j.ydbio.2012.11.018
417. Ingram, K. G., Curtis, C. D., Silasi-Mansat, R., Lupu, F. & Griffin, C. T. The NuRD Chromatin-Remodeling Enzyme CHD4 Promotes Embryonic Vascular Integrity by Transcriptionally Regulating Extracellular Matrix Proteolysis. *PLoS Genet.* (2013). doi:10.1371/journal.pgen.1004031
418. Waldron, L. *et al.* The Cardiac TBX5 Interactome Reveals a Chromatin Remodeling Network Essential for Cardiac Septation. *Dev. Cell* **36**, 262–275 (2016).
419. Wilczewski, C. M. *et al.* CHD4 and the NuRD complex directly control cardiac sarcomere formation. *Proc. Natl. Acad. Sci.* (2018). doi:10.1073/pnas.1722219115
420. Garnatz, A. S. *et al.* FOG-2 mediated recruitment of the NuRD complex regulates cardiomyocyte proliferation during heart development. *Dev. Biol.* **395**, 50–61 (2014).
421. Le Gallo, M. *et al.* Exome sequencing of serous endometrial tumors identifies recurrent somatic mutations in chromatin-remodeling and ubiquitin ligase complex genes. *Nat. Genet.* (2012). doi:10.1038/ng.2455
422. Beronja, S. *et al.* RNAi screens in mice identify physiological regulators of oncogenic growth. *Nature* (2013). doi:10.1038/nature12464

423. Bhm, M. *et al.* Helicase CHD4 is an epigenetic coregulator of PAX3-FOXO1 in alveolar rhabdomyosarcoma. *J. Clin. Invest.* (2016). doi:10.1172/JCI85057
424. Sifrim, A. *et al.* Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. *Nat. Genet.* (2016). doi:10.1038/ng.3627
425. Weiss, K. *et al.* De Novo Mutations in CHD4, an ATP-Dependent Chromatin Remodeler Gene, Cause an Intellectual Disability Syndrome with Distinctive Dysmorphisms. *Am. J. Hum. Genet.* (2016). doi:10.1016/j.ajhg.2016.08.001
426. Homsy, J. *et al.* De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. *Science* (80-.). (2015). doi:10.1126/science.aac9396
427. Lalani, S. R. *et al.* Spectrum of CHD7 Mutations in 110 Individuals with CHARGE Syndrome and Genotype-Phenotype Correlation. *Am. J. Hum. Genet.* (2006). doi:10.1086/500273
428. Vissers, L. E. L. M. *et al.* Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nat. Genet.* (2004). doi:10.1038/ng1407
429. De Ligt, J. *et al.* Diagnostic exome sequencing in persons with severe intellectual disability. *Obstetrical and Gynecological Survey* (2013). doi:10.1097/01.ogx.0000428160.59063.a6
430. Fitzgerald, T. W. *et al.* Large-scale discovery of novel genetic causes of developmental disorders. *Nature* (2015). doi:10.1038/nature14135
431. Zaidi, S. *et al.* De novo mutations in histone-modifying genes in congenital heart disease. *Nature* **498**, 220–223 (2013).
432. Weiss, K. *et al.* The CHD4-related syndrome: a comprehensive investigation of the clinical spectrum, genotype–phenotype correlations, and molecular basis. *Genet. Med.* (2019). doi:10.1038/s41436-019-0612-0
433. Mahmood, T. & Yang, P. C. Western blot: Technique, theory, and trouble shooting. *N. Am. J. Med. Sci.* (2012). doi:10.4103/1947-2714.100998
434. Abcam - Western blot troubleshooting tips (URL: <https://www.abcam.com/protocols/western-blot-troubleshooting-tips>) [December 2019].
435. Bio-Rad - Western Blot Troubleshooting: Unusual or Unexpected Bands (URL: <https://www.bio-rad-antibodies.com/western-blot-unusual-unexpected-bands-western-blotting.html>) [December 2019].

436. Kozak, M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger rNAS. *Nucleic Acids Res.* (1987).
doi:10.1093/nar/15.20.8125
437. Watson, A. A. *et al.* The PHD and chromo domains regulate the atpase activity of the human chromatin remodeler CHD4. *J. Mol. Biol.* **422**, 3–17 (2012).
438. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Frontiers in Microbiology* (2014). doi:10.3389/fmicb.2014.00172
439. Morra, R., Fessl, T., Wang, Y., Mancini, E. J. & Tuma, R. in *Methods in Molecular Biology* (2016). doi:10.1007/978-1-4939-3631-1_14
440. Zhang, Y., LeRoy, G., Seelig, H. P., Lane, W. S. & Reinberg, D. The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. *Cell* **95**, 279–289 (1998).
441. Ghosh, T. K. *et al.* Acetylation of TBX5 by KAT2B and KAT2A regulates heart and limb development. *J. Mol. Cell. Cardiol.* **114**, 185–198 (2018).
442. Urquhart, A. J., Gatei, M., Richard, D. J. & Khanna, K. K. ATM mediated phosphorylation of CHD4 contributes to genome maintenance. *Genome Integr.* (2011). doi:10.1186/2041-9414-2-1
443. Polo, S. E., Kaidi, A., Baskcomb, L., Galanty, Y. & Jackson, S. P. Regulation of DNA-damage responses and cell-cycle progression by the chromatin remodelling factor CHD4. *EMBO J.* (2010).
doi:10.1038/emboj.2010.188
444. Kovač, K. *et al.* Tumour-associated missense mutations in the dMi-2 ATPase alters nucleosome remodelling properties in a mutation-specific manner. *Nat. Commun.* (2018). doi:10.1038/s41467-018-04503-2
445. Bouazoune, K. & Kingston, R. E. Chromatin remodeling by the CHD7 protein is impaired by mutations that cause human developmental disorders. *Proc. Natl. Acad. Sci.* **109**, 19238–19243 (2012).
446. Malumbres, M. Cyclin-dependent kinases. *Genome Biol.* (2014).
doi:10.1186/gb4184
447. Greifenberg, A. K. *et al.* Structural and Functional Analysis of the Cdk13/Cyclin K Complex. *Cell Rep.* (2016).
doi:10.1016/j.celrep.2015.12.025
448. Greenleaf, A. L. Human CDK12 and CDK13, multi-tasking CTD kinases for the new millenium. *Transcription* (2019).
doi:10.1080/21541264.2018.1535211

449. Blazek, D. *et al.* The cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. *Genes Dev.* (2011). doi:10.1101/gad.16962311
450. Blazek, D. The cyclin K/Cdk12 complex: An emerging new player in the maintenance of genome stability. *Cell Cycle* (2012). doi:10.4161/cc.11.6.19678
451. Juan, H. C., Lin, Y., Chen, H. R. & Fann, M. J. Cdk12 is essential for embryonic development and the maintenance of genomic stability. *Cell Death Differ.* (2016). doi:10.1038/cdd.2015.157
452. Chen, H. R., Lin, G. T., Huang, C. K. & Fann, M. J. Cdk12 and Cdk13 regulate axonal elongation through a common signaling pathway that modulates Cdk5 expression. *Exp. Neurol.* (2014). doi:10.1016/j.expneurol.2014.06.024
453. CDK13 gene (MGI:1916812). International Mouse Phenotyping Consortium Web Site. World Wide Web (URL: <https://www.mousephenotype.org/data/genes/MGI:1916812>) [Feb, 2018].
454. CDK12 gene (MGI:1098802). International Mouse Phenotyping Consortium Web Site. World Wide Web (URL: <https://www.mousephenotype.org/data/genes/MGI:1098802>) [Feb, 2018].
455. Bostwick, B. L. *et al.* Phenotypic and molecular characterisation of CDK13-related congenital heart defects, dysmorphic facial features and intellectual developmental disorders. *Genome Med.* (2017). doi:10.1186/s13073-017-0463-8
456. Hamilton, M. J. *et al.* Heterozygous mutations affecting the protein kinase domain of CDK13 cause a syndromic form of developmental delay and intellectual disability. *J. Med. Genet.* (2018). doi:10.1136/jmedgenet-2017-104620
457. Weninger, W. J. *et al.* High-resolution episcopic microscopy: A rapid technique for high detailed 3D analysis of gene activity in the context of tissue architecture and morphology. *Anat. Embryol. (Berl)*. (2006). doi:10.1007/s00429-005-0073-x
458. Geyer, S. H., Maurer-Gesek, B., Reissig, L. F. & Weninger, W. J. High-resolution Episcopic Microscopy (HREM) - Simple and Robust Protocols for Processing and Visualizing Organic Materials. *J. Vis. Exp.* (2017). doi:10.3791/56071
459. Weninger, W., Maurer, B., Zendron, B., Dorfmeister, K. & Geyer, S. Measurements of the diameters of the great arteries and semi-lunar valves of chick and mouse embryos. *J. Microsc.* (2009).

doi:10.1111/j.1365-2818.2009.03159.x

- 460. Geyer, S. H. & Weninger, W. J. Some mice feature 5th pharyngeal arch arteries and double-lumen aortic arch malformations. *Cells Tissues Organs* (2012). doi:10.1159/000330789
- 461. Paun, B., Bijmens, B., Cook, A. C., Mohun, T. J. & Butakoff, C. Quantification of the detailed cardiac left ventricular trabecular morphogenesis in the mouse embryo. *Med. Image Anal.* (2018). doi:10.1016/j.media.2018.08.001
- 462. Lana-Elola, E. *et al.* Genetic dissection of Down syndrome- associated congenital heart defects using a new mouse mapping panel. *Elife* (2016). doi:10.7554/eLife.11614.001
- 463. Le Garrec, J. F. *et al.* A predictive model of asymmetric morphogenesis from 3D reconstructions of mouse heart looping dynamics. *Elife* (2017). doi:10.7554/eLife.28951
- 464. Captur, G. *et al.* Morphogenesis of myocardial trabeculae in the mouse embryo. *J. Anat.* (2016). doi:10.1111/joa.12465
- 465. Maurer, B., Geyer, S. H. & Weninger, W. J. A Chick Embryo with a yet unclassified type of cephalothoracopagus malformation and a hypothesis for explaining its genesis. *J. Vet. Med. Ser. C Anat. Histol. Embryol.* (2013). doi:10.1111/ahe.12002
- 466. Mohun, T. *et al.* Deciphering the Mechanisms of Developmental Disorders (DMDD): a new programme for phenotyping embryonic lethal mice. *Dis. Model. Mech.* (2013). doi:10.1167/iavs.08-2025
- 467. Liddy, K. A., White, M. Y. & Cordwell, S. J. Functional decorations: Post-translational modifications and heart disease delineated by targeted proteomics. *Genome Medicine* (2013). doi:10.1186/gm424
- 468. Yan, K., Wang, K. & Li, P. The role of post-translational modifications in cardiac hypertrophy. *Journal of Cellular and Molecular Medicine* (2019). doi:10.1111/jcmm.14330
- 469. Fert-Bober, J., Murray, C. I., Parker, S. J. & Van Eyk, J. E. Precision profiling of the cardiovascular post-translationally modified proteome where there is a will, there is a way. *Circ. Res.* (2018). doi:10.1161/CIRCRESAHA.118.310966
- 470. Moore-Morris, T., van Vliet, P. P., Andelfinger, G. & Puceat, M. Role of epigenetics in cardiac development and congenital diseases. *Physiological Reviews* (2018). doi:10.1152/physrev.00048.2017
- 471. Vecoli, C., Pulignani, S., Foffa, I. & Andreassi, M. Congenital Heart Disease: The Crossroads of Genetics, Epigenetics and Environment. *Curr. Genomics* (2014). doi:10.2174/1389202915666140716175634

- 472. Zaidi, S. & Brueckner, M. Genetics and Genomics of Congenital Heart Disease. *Circulation Research* (2017). doi:10.1161/CIRCRESAHA.116.309140
- 473. Snijders Blok, L. *et al.* CHD3 helicase domain mutations cause a neurodevelopmental syndrome with macrocephaly and impaired speech and language. *Nat. Commun.* (2018). doi:10.1038/s41467-018-06014-6
- 474. Sifrim, A. *et al.* Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. *Nat. Genet.* **48**, 1–9 (2016).
- 475. Corsten-Janssen, N. *et al.* The cardiac phenotype in patients with a CHD7 mutation. *Circ. Cardiovasc. Genet.* (2013). doi:10.1161/CIRCGENETICS.113.000054

